Random Quantum Circuits and Their Simulation Complexity

An Analysis With Statistical Mechanics

Thesis by
Alexander M. Dalzell

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended August 31, 2021

Alexander M. Dalzell
ORCID: 0000-0002-3756-8500

ACKNOWLEDGEMENTS

Thinking back on how this thesis came to be, I can feel nothing except immensely fortunate for the circumstances and for the people that led me to this point. Firstly, I owe special gratitude to my advisors Fernando Brandão and John Preskill. It can be difficult to maintain motivation during the inevitable challenges that surface during a PhD. Fernando's guidance and timely assistance in crucial moments have been essential for keeping my journey in motion, and John's encouragement and everlasting excitement for this subject have inspired me to always look ahead with enthusiasm, even in moments of frustration.

I would also like to thank my collaborators: Rolando La Placa, John Napp, Nick Hunter-Jones, Fernando, and Aram Harrow, without whom the work in this thesis would not have existed. I particularly want to thank Rolando for guiding me into the field of quantum information in the first place, now many years ago, and for being a great friend along the way. I also acknowledge my previous undergraduate mentors for their indispensable role in setting me on this path: Aram, as well as Ike Chuang, Ted Yoder, Guang Hao Low, and Dax Koh.

I cannot overstate my gratitude for the Caltech quantum information community, from which I've learned so much, scientifically and otherwise. I greatly appreciate all of my great conversations with the group's many postdocs and fellow grad students, especially Victor Albert, Thom Bohdanowicz, Alex Buser, Bailey Gu, Robert Huang, Tomas Jochym-O'Connor, Kohtaro Kato, Richard Kueng, Angelo Lucia, Burak Şahinoğlu, Grant Salton, and Eugene Tang. Thanks also to the other members of my thesis committee, Thomas Vidick and Oskar Painter, for their feedback on my work.

I acknowledge financial support from the National Science Foundation Graduate Research Fellowship Program (DGE-1745301) as well as the Dominic Orr Fellowship at Caltech.

Most importantly, I could not have gotten here without the undying support of my close friends and family. To my best friend and partner, Paola, I cannot express how lucky I am to have you by my side, through both the difficult moments and the triumphant ones. Finally to my sister Maddie, and especially to my parents, Meredith and Kevin: none of this could have been possible without you.

ABSTRACT

Random circuit simulation, the task of replicating the output of a randomly chosen noiseless quantum computation, has been proposed as a problem that should be easy for quantum devices but hard for classical ones. Establishing the existence of such tasks and accomplishing them on actual quantum hardware is important for benchmarking progress in an era where quantum devices are hampered by small sizes and high noise rates. Additionally, at a fundamental level, the assertion that *random* quantum circuits are hard to classically simulate is a statement that quantum advantage is not only possible, but ubiquitous. In this thesis, we scrutinize the random circuit simulation dilemma from both sides. On the one hand, we investigate whether the task is classically hard—we find that, in certain non-trivial cases, it can actually be easy, complicating a potential general proof of hardness. On the other hand, we investigate whether the task can be easily accomplished on realistic quantum devices, which are subject to substantial noise rates—we find that, indeed, a version of the circuit simulation task can be salvaged even on a noisy quantum device performing the computation with low fidelity, as long as the noise meets certain conditions. Thus, this thesis emphasizes that, to construct a strong argument of quantum advantage via random circuit simulation on noisy quantum hardware, the core theoretical challenge remains proving lower bounds on the classical complexity of the task; doing so will require new ideas to circumvent the barriers presented by our work.

On the classical simulation side, we propose a classical algorithm for approximate random circuit simulation of constant-depth 2D circuits. We prove that the algorithm is efficient for one specific family of 2D circuits, and we give evidence that it is efficient more generally as long as the circuit depth is sufficiently shallow. This is surprising because, under plausible conjectures from complexity theory, it is known that no efficient simulation algorithm exists that *exactly* computes the probabilities for most instances or exactly samples from the distribution in every instance. Thus, our algorithm demonstrates that allowing error (as is necessary when comparing with *noisy* quantum computers) can greatly reduce the classical complexity of the simulation problem. We also give evidence that the algorithm becomes inefficient when the circuit depth exceeds some constant threshold value by connecting the complexity of the simulation problem to phase transitions in statistical mechanical systems.

Next, we study the output probability distributions of noiseless random quantum circuits; a classical or quantum device that simulates the random circuits should be able to sample from these distributions, to some degree of precision. We prove that these distributions achieve the anti-concentration property—meaning that the probability mass has spread out roughly evenly over all possible outcomes—at a much shallower circuit depth than previously believed. We consider the case where gates are nearest-neighbor when the

qubits are arranged in 1D, as well as the case where gates are completely non-local; in both, we show that, for systems with n qubits, $\Theta(n \log(n))$ random two-qubit gates are necessary and sufficient for anti-concentration, and we give evidence that this fact is true in the general case as well. Having the anti-concentration property is evidence that the simulation task is classically hard, and it is advantageous for this to occur after the fewest possible number of gates, as noise in near-term devices accumulates with the size of the circuit.

Finally, we examine the impact of noise on a quantum device running a random quantum circuit; we show that random quantum circuits quickly scramble local noise, allowing it to be treated as global white noise. Specifically, as long as the local noise is incoherent and its strength ϵ satisfies $\epsilon^{-1} \gg \tilde{\Omega}(n)$, the output distribution p_{noisy} of a noisy random quantum circuit with s gates is approximately $F p_{\text{ideal}} + (1 - F) p_{\text{unif}}$, where p_{ideal} is the output distribution for the noiseless circuit, p_{unif} is the uniform distribution, and $F = e^{-\Theta(\epsilon s)}$ is the circuit fidelity. We show that the error in the approximation, as measured by the total variation distance, is bounded by $O(F\epsilon\sqrt{s})$. Thus, when $\epsilon^2 s \ll 1$, the output p_{noisy} is well described by a combination of signal from the ideal noiseless computation (weighted by F) and white noise (weighted by $1 - F$); this allows the signal to be extracted from the noisy outputs simply by repetition of the experiment. One implication of this is that low-fidelity random circuit experiments are essentially just as hard to classically replicate as high-fidelity random circuit experiments, bolstering claims of quantum advantage on devices even at realistic noise rates.

The main analytical technique we utilize for each of these results is the statistical mechanics method for random quantum circuits, which maps random quantum circuits made from local Haar-random gates to partition functions of classical statistical mechanical systems. This thesis demonstrates the utility of this method by applying it in several new ways. In some cases, we use it for heuristic reasoning about the behavior of random quantum circuits; in others, we go further and perform rigorous calculations of the resulting partition function, leading to precise technical statements about random quantum circuits. For example, our anti-concentration analysis produces sharp upper and lower bounds that match even up to the constant prefactor of the leading $\Theta(n \log(n))$ term.

PUBLISHED CONTENT AND CONTRIBUTIONS

Chapters 3 and 4 include content adapted from previously written standalone research articles. The content of Chapter 5 was adapted into a research article that was publicly released shortly after the defense of this thesis.

Chapter 3 J. Napp, R. L. La Placa, A. M. Dalzell, F. G. S. L. Brandão, and A. W. Harrow, "Efficient classical simulation of random shallow 2D quantum circuits," arXiv:2001.00021. AMD made intellectual contributions to all aspects of the paper and was the main author for sections about the statistical mechanics method.

Chapter 4 A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, "Random quantum circuits anti-concentrate in log depth," arXiv:2011.12277. AMD was the main contributor.

Chapter 5 A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, "Random quantum circuits transform local noise into global white noise," arXiv:2111.14907. AMD was the main contributor.

While at Caltech, the author also contributed to the following publications, which are not included in the content of this thesis.

A. M. Dalzell and F. G. S. L. Brandão, "Locally accurate MPS approximations for ground states of one-dimensional gapped local Hamiltonians," *Quantum* **3** (2019) 187, arXiv:1903.10241. AMD was the main contributor.

A. M. Dalzell, A. W. Harrow, D. E. Koh, and R. L. La Placa, "How many qubits are needed for quantum computational supremacy?" *Quantum* **4** (2020) 264, arXiv:1805.05224. AMD was the main contributor.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION: THE COMPLICATIONS OF A NOISY QUANTUM WORLD

The desire to build quantum computers is driven by the belief that it will not be possible to replicate their behavior on the computers we already have. Indeed, with both public [1] and private (e.g., [2]) yearly investment in quantum computing now measured in hundreds of millions of dollars, there is building anticipation of a transformed world where quantum computers solve important and previously insurmountable computational problems.

In this thesis, we shall not speculate on if or when such a future might become reality, except to say that it is not imminent. With only dozens of qubits, state-of-the-art quantum computers are still quite small and, crucially, they are also error prone. Nevertheless, they are complex enough that the task of simulating them is arguably impossible on modern classical computers. We are now entering the Noisy Intermediate-Scale Quantum (NISQ) era [3], where quantum devices are good enough to potentially carry out interesting computations, but too small and too noisy to implement the quantum error-correction schemes that will be required to realize the full potential of quantum computing.

Benchmarking progress in the NISQ era requires careful examination of the tasks that quantum computers can accomplish, as well as the reasons to believe that the same tasks are difficult for classical computers. One way to argue that a task is classically intractable is to determine the quantity of resources (e.g., time, computer memory, money) required to solve the task using the best known classical method for doing so, and observe that this quantity is overwhelmingly large. The problem here is that better algorithms might be discovered that dramatically reduce the resource cost of classically performing the task. This is a risk especially when the task has been chosen not because its classical complexity has been extensively studied, but rather for the simple reason that NISQ computers can actually do it.

A stronger claim is that the task would be intractable even using classical methods we might have yet to discover. After all, the intuition that quantum computers should be superior to classical computers amounts to more than a lack of classical ingenuity. Rather, fundamental features of quantum mechanics—in particular, the fact that a system with n particles lives in a Hilbert space with dimension exponentially large in n—should make quantum computation inherently more powerful. It is notoriously hard to definitively prove this stronger notion of intractability, but the field of complexity theory provides a framework for giving concrete evidence in favor of such claims.

For example, it can be shown that noiseless quantum computations cannot generally be efficiently classically simulated if certain complexity-theoretic assumptions are imposed; these assumptions are widely believed to be true for reasons entirely independent from quantum computing. This conclusion forms the bedrock of a research program that attempts to classify certain kinds of quantum computations by whether they are hard to classically simulate in this strong complexity-theoretic sense. Knowing that certain kinds of computations are classically hard is, of course, practically important for measuring the computational value added by quantum computing and informing design choices toward that end. But it is also of purely theoretical interest, as it allows us to determine which features are essential for powerful quantum computation and which are extraneous, leading to a better understanding of the fundamental source of quantum advantage embedded in the laws of physics.

On this front, the limitations of the NISQ era lead to interesting questions. In particular, the fact that the hardware is noisy means that NISQ computers are only capable of simulating a noiseless version of themselves in an approximate sense. Is *approximate* simulation of quantum computations still hard for classical computers? Moreover, the impact of noise accumulates as computations get longer, eventually to a point where nothing interesting can be accomplished. Thus, special attention must be given to computations of relatively shallow depths, and in this case restrictions in the layout of the qubits (e.g., 2D lattice) and implementable gate set can have a more significant impact on the tasks that can be performed and the strength of the evidence that classically simulating the quantum computation is hard.

The content of this thesis concerns a particular computational task called Random Circuit Sampling (RCS) [4, 5] , which is amenable to implementation in the NISQ era and, in fact, has already been attempted on quantum devices [6, 7]. Our discussion so far has not drawn a hard distinction between classical simulation of a quantum computation that achieves a certain task and classically accomplishing the task in some other way. That's because for RCS, simulation itself *is* the task. The idea is to choose a quantum circuit at random, subject only to certain restrictions in depth and layout, and let the task be to generate a sample from the output of a noiseless implementation of the circuit. The fact that a noiseless quantum computer can perform the RCS task becomes essentially tautological, but since NISQ computers are not noiseless, the real-world situation is considerably more murky. Beyond the possibility of NISQ implementation, the RCS task is of fundamental theoretical interest, as random quantum circuits embody generic quantum evolution and allow us to probe which features of quantum computing are only observed in special situations, and which are ubiquitous.

The contributions contained here clarify certain aspects of the RCS landscape. The results are technical in nature, but their significance relates back to the following complementary questions:

(1) Are there efficient classical algorithms for RCS?

(2) How well can noisy quantum computers solve RCS despite the presence of errors and the limitation of shallow depth?

In other words, we are interested in both the classical complexity and the quantum complexity of simulation of ideal random quantum circuits. The technical statements and the methods for achieving our results have other potential applications, some of which we point out in later chapters, but the lens of classical simulability provides a unifying framework to understand how they fit together.

Another unifying feature of our results is the underlying techniques used in the analysis. Specifically, we employ a method that associates random quantum circuits with partition functions of classical statistical mechanical systems. In some cases, we can glean intuition from the stat mech system that yields insights for the random quantum circuits. In other cases, we go further and perform rigorous calculations on the partition functions, producing precise upper and lower bounds on important random-quantum-circuit quantities. Overall, one takeaway from this thesis is simply the utility of the stat mech method, and we speak more on this method in Chapter 2.

In the remainder of this chapter, we dive deeper into the central concepts needed to understand the significance of the contributions of this thesis. We introduce technical definitions as necessary to explain our results, but full details and further commentary on motivation and meaning appears individually in each chapter.

1.1 When is simulating quantum computations hard, and how do we know?

Completely describing a quantum state on n qubits requires specifying 2^n complex numbers, the amplitudes for each of the basis states in the exponentially large Hilbert space. This is the foundation for the belief that quantum computers should have an exponential computational advantage over classical computers. Indeed, the exponential scaling of Hilbert space was Richard Feynman's initial reason for proposing the concept of a quantum computer in 1981 [8].

What this fact illustrates is that a classical algorithm that simulates a quantum computation by storing a complete description of the n-qubit quantum state and updating it after each gate will not be efficient[1]. However, this is by no means the only way to attempt to simulate a quantum computation. In certain special cases, there are known efficient simulation algorithms, often because in these cases there is a much more efficient way to represent the

[1]Here and throughout, we follow the convention of calling an algorithm "efficient" when its runtime scales like some polynomial in its input size. When the input is a description of a quantum circuit on n-qubits with poly(n) gates, the input size is itself polynomial in n.

quantum state. For instance, quantum computations that generate a small amount of entanglement can be simulated with matrix product states or other tensor network methods [9–12], and quantum computations consisting entirely of Clifford operations lead only to states that can be efficiently represented and updated within the stabilizer formalism [13, 14]. How can one be sure there does not exist a more efficient way to represent general quantum states that would allow for an efficient general simulation algorithm?

Showing hardness of simulation by embedding a hard computational problem into a quantum circuit

Complexity theory allows this question to be explored in a more precise manner. In particular, one can encode the answer to a very hard classical computational problem into the output of a simple quantum computation, such that an efficient simulation of the quantum computation would lead to an efficient solution to the hard computational problem, a highly dubious conclusion. This can be done as follows. Let $f : \{0,1\}^n \to \{0,1\}$ be an arbitrary efficiently computable Boolean function on n-bit inputs, and let the quantity $\#f$ be the number of inputs x for which $f(x) = 1$. The quantity $\#f$ is called a $\#\mathsf{P}$ function in complexity theory, and it is widely believed that, at least for some functions f, there is no efficient way to compute $\#f$. After all, the naive way of computing $\#f$ by enumerating all 2^n inputs x and counting how many yield $f(x) = 1$ has exponential run time. Yet, we can design an efficient quantum algorithm[2] whose output is related to $\#f$. We do so by computing f on all 2^n inputs in superposition and engineering interference between the 2^n outcomes. Specifically, using Hadamard gates, we first prepare the initial superposition state

$$\frac{1}{2^{n/2}} \sum_{x\in\{0,1\}^n} |x\rangle \otimes \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right) . \tag{1.1}$$

Then, we compute the Boolean function $f(x)$ (recall it is efficiently computable) and record the answer into the last qubit, that is, we perform the unitary transformation U_f for which $U_f|x, b\rangle = |x, b \oplus f(x)\rangle$. This flips the sign of terms for which $f(x) = 1$, yielding

$$\frac{1}{2^{n/2}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)}|x\rangle \otimes \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right) . \tag{1.2}$$

Finally, we throw away the last qubit and apply another round of Hadamard gates on the remaining n qubits to arrive at

$$\frac{1}{2^n} \sum_{x\in\{0,1\}^n} \left(\sum_{y\in\{0,1\}^n} (-1)^{f(y)+x\cdot y} \right) |x\rangle , \tag{1.3}$$

[2]The algorithm described here is identical to the Deutsch-Jozsa algorithm [15, 16]. In that context, there is the additional promise that either $\#f \in \{0, 2^n\}$, or $\#f = 2^{n-1}$, and the problem is to determine which is the case.

where $x \cdot y$ is the normal dot product between n-bit Boolean vectors. We now note that the coefficient of $|0^n\rangle$ in the final expression is precisely $1 - \#f/2^{n-1}$. Thus, if there were a classical algorithm that could efficiently compute the amplitudes in the output for a given quantum computation, then there would also be a classical algorithm that could efficiently compute any $\#\mathsf{P}$ function, a conclusion complexity theorists consider highly unlikely. This provides a formal answer to the question posed above: assuming that $\#\mathsf{P}$ functions cannot be efficiently computed, there cannot be a classical description of general quantum states that allows for an efficient simulation of quantum computation, or at least no description that allows us to compute output amplitudes.

However, computing output amplitudes (or the output probabilities, which are the square of the magnitude of the amplitudes) is not something even a quantum computer is capable of doing efficiently. Output states of quantum computations ultimately must be measured, producing a random output. In our example above, a measurement on all n output qubits would yield the outcome $|0^n\rangle$ with probability $(1 - \#f/2^{n-1})^2$. The quantum computation would in general need to be repeated an exponential number of times to produce a good estimate for $\#f$. This emphasizes the distinction between a *strong simulation* of a quantum computation, which is an algorithm for computing the output probabilities, and a *weak simulation*, which is an algorithm that produces random outputs according to the same distribution as the quantum computation. In this language, we could say that a quantum computer naturally performs the task of weakly simulating itself, but not the task of strongly simulating itself. The argument above essentially showed that classical computers cannot efficiently perform the task of strongly simulating a quantum computation without surprising complexity-theoretic consequences; the strong simulation task is $\#\mathsf{P}$-hard.

It turns out that the argument can be extended to rule out efficient weak classical simulations as well, although showing this requires a deeper detour into complexity theory. To summarize this, first note that the amplitude for $|0^n\rangle$ is 0 if and only if $\#f = 2^{n-1}$, that is, if exactly half the inputs x have $f(x) = 1$. Thus, the quantum computation produces output $|0^n\rangle$ with non-zero probability only when $\#f \neq 2^{n-1}$. This was only possible due to destructive interference between positive and negative amplitudes that exactly cancel only when half the inputs have $f(x) = 1$, an inherently quantum phenomenon, and it would be surprising if a classical computer could efficiently replicate this feat. If a classical computer could efficiently weakly simulate the quantum computation, then there would also be an efficient randomized classical algorithm that outputs 0^n with non-zero probability only when $\#f \neq 2^{n-1}$. In complexity-theoretic language, this reads[3] $\mathsf{NP} \supset \mathsf{coC_{=}P}$. However, this would be a surprising conclusion, since it is known to imply that a set of complexity

[3]The class NP can be defined as containing problems for which there is an efficient randomized algorithm that outputs YES with non-zero probability if and only if the correct answer is YES. The class $\mathsf{coC_{=}P}$ can be defined as containing problems where there is an

classes called the polynomial hierarchy (PH) "collapses" [17, 18]. It is not vital to understand the complexity-theoretic details here, except to know that it is widely believed that the polynomial hierarchy does not in fact collapse, for reasons that have nothing to do with quantum computing. In fact, the assumption that the PH does not collapse is a slightly stronger version of the well-known $\mathsf{P} \neq \mathsf{NP}$ conjecture, which is virtually unquestioned but famously hard to prove. In any case, the implication is that under plausible complexity-theoretic assumptions, even weak classical simulation of quantum computation cannot be efficient; the weak simulation task is PH-hard. This conclusion is a key piece of justification in our pursuit of quantum computation: quantum computers are doing things that genuinely cannot be efficiently replicated on classical computers, even considering the possibility that better classical simulation algorithms might be developed in the future.

One interesting theoretical direction building from this insight has been to examine whether these kinds of arguments still work as certain restrictions to the quantum computation are imposed. In some cases, hardness can still be shown despite the restrictions. In other cases, restrictions make the computation efficient to simulate. By looking at both sides, one can begin to assess the necessary and sufficient ingredients for quantum computation to transcend the capabilities of efficient classical computation.

Noise and its connection to random quantum computations

Noise is one example of a restriction that puts the hardness-of-simulation argument in jeopardy. Without employing error-correcting schemes, noisy quantum devices are incapable of encoding a hard classical problem into their output amplitude to the precision required for the hardness argument above to apply. To see this, let p_x be the probability the circuit above outputs $|x\rangle$, and let $\tilde{p}_x$ be the probability a noisy implementation of it outputs $|x\rangle$. If the noisy output distribution is δ-far from the ideal distribution in total variation distance (that is, $\frac{1}{2}\sum_x |\tilde{p}_x - p_x| = \delta$) the entirety of that δ error can be allocated to the $|0^n\rangle$ output probability. In other words, $|\tilde{p}_{0^n} - p_{0^n}| = \delta$ is possible in the worst case. Note that the quantity $\#f$ containing the answer to the hard classical problem is related to this output probability p_{0^n} by the relation $\#f = 2^{n-1}(1 \pm \sqrt{p_{0^n}})$, which means it is exponentially sensitive to deviations in p_{0^n} due to error; computing $\#f$ is no longer a #P-hard problem if additive errors on p_{0^n} of size $\delta = O(1)$[4] are allowed.

efficient randomized algorithm that outputs YES with probability exactly 1/2 if and only if the correct answer is NO. The containment $\mathsf{NP} \subset \mathsf{coC_{=}P}$ can be easily shown: given an efficient randomized algorithm M that produces output $M(x,r) \in \{\text{NO}, \text{YES}\}$ on input x and random bits r, one can define a new algorithm M' with one additional random bit b, such that $M'(x,r,b) = \text{YES}$ if $b = 1$ or if $M(x,r) = \text{YES}$, and otherwise $M'(x,r,b) = \text{NO}$. M' outputs YES for exactly half of the choices of (r,b) if and only if M outputs NO for all choices of r.

[4] Throughout this thesis, we use big-O notation, where $g(x) = O(f(x))$ indicates there is a constant c such that $g(x) \leq cf(x)$ for x sufficiently large, $g(x) = \Omega(f(x))$ indicates there

One way forward is to "hide" the answer to the hard computational problem by introducing randomness into the computational problem instance [19]. For each of the 2^n outputs $|x\rangle$, we may define a Boolean function f_x by $f_x(y) = f(y) + x \cdot y$. Under this notation, $f_{0^n} = f$. In the quantum circuit described above, the output probability for outcome $|x\rangle$ is precisely $p_x = (1 - \#f_x/2^{n-1})^2$, so all 2^n outputs are associated with some #P function. We may now argue that, *when x is chosen uniformly at random*, less than $O(\delta/2^n)$ error will be allocated to the $|x\rangle$ output with high probability—that is, $|\tilde{p}_x - p_x| \leq O(\delta/2^n)$ for most x. The idea is, even if the noise acts adversarially, the x we chose has been hidden, and the noise cannot know which of the 2^n outputs it should attack; the best it can do is to allocate the δ error roughly evenly over all 2^n outputs. Moreover, the mean of the quantity p_x when x is chosen uniformly at random is $1/2^n$ (since the 2^n output probabilities must sum to 1), and as long as the output distribution has a property called *anti-concentration*[5], the random fluctuations away from the $1/2^n$ mean are not too large. When this is the case, the $O(\delta/2^n)$ error on the value of p_x will usually be a small fraction of the value of p_x itself; that is, $\frac{|\tilde{p}_x - p_x|}{p_x} \leq O(\delta)$ for most x. In other words, $O(\delta)$ additive error in the worst case corresponds to $O(\delta)$ *relative* error in the average case. Importantly, the task of estimating $p_x = (1 - \#f_x/2^{n-1})^2$ up to $O(1)$ relative error for *every* #P function $\#f_x$ can be shown to be essentially #P-hard, and the task of sampling from a distribution $\tilde{p}$ for which $\tilde{p}_x$ is within $O(1)$ relative error of p_x for *every* x is PH-hard (see, e.g., Refs. [20, 21]).

Thus, using the average case instead of the worst case partially recovers the computational hardness of the strong simulation problem, even in the presence of noise, but completing the argument now requires that these relative-error tasks are hard not only in the worst case, but also in the average case. In other words, to show that *noisy* quantum computations are difficult to simulate classically using this argument, one must show that *random* quantum computations are difficult to simulate. While the exact ensemble of random quantum computations that are chosen can be malleable (e.g., Ref. [19] examines Haar-random linear optical networks, and Ref. [22] examines random "IQP" circuits), in all cases the question becomes some form of: are quantum computations just as hard to simulate in the average-case as they are in the worst-case? Beyond its connection to noise, this question is fundamentally interesting: it asks whether classical hardness of simulation is a generic feature of quantum evolution, or if it only appears in specific, contrived settings.

---

is a constant c such that $g(x) \geq cf(x)$ for x sufficiently large, and $g(x) = \Theta(f(x))$ means $g(x) = O(f(x))$ and $g(x) = \Omega(f(x))$ simultaneously.

[5] Anti-concentration in random quantum circuits is the subject of Chapter 4. In particular, refer to Section 4.4.2 for a discussion on the role of anti-concentration in arguments for hardness of simulation.

1.2 Random quantum circuits and Random Circuit Sampling

Perhaps the simplest ensemble of random quantum computations are random quantum circuits made of two-qubit gates arranged in some predetermined layout. This setting has been studied in a variety of contexts. For example, physicists have used random quantum circuits to understand the onset of chaos in strongly interacting systems and as a model for the dynamics inside black holes. In a sense, random quantum circuits capture generic unitary evolution where the sole constraint is locality (the gates act only on pairs of qubits, arranged in some fashion). However, random quantum circuits also have specific technical benefits when it comes to questions of classical simulation complexity in the NISQ era, including the hiding property outlined above. In this context, the task of random circuit sampling (RCS) was proposed [4, 5, 23] not because it is useful for anything in particular, but rather because it is experimentally feasible while providing a fertile testing ground for studying the extent to which noisy quantum computers are difficult to simulate on classical computers, both in theory and in practice.

Definition of RCS

The input to an instance of the RCS task is a description of a quantum circuit with n qubits and s gates. For concreteness, assume that the initial state of the circuit is always $|0^n\rangle$, and that each of the s gates acts on only a pair of qubits, chosen according to the particular *architecture* of interest; for instance, in the 1D architecture, the gates are chosen to act on nearest-neighbor pairs when the qubits are arranged in a ring. Each gate implements a 4×4 unitary transformation, and the sequence of 4×4 unitaries (along with the sequence of qubit pairs) taken together implements some $2^n \times 2^n$ unitary U. When we choose an instance, we always choose the s underlying 4×4 unitaries independently at random, usually according to the Haar measure[6] over unitaries of dimension 4, which for a fixed architecture induces some ensemble over $2^n \times 2^n$ unitaries U. We would like to make statements that we can prove must hold in expectation over choice of U from this ensemble.

If a computational basis measurement is performed on all n qubits of the quantum circuit after the application of the s two-qubit gates, some bit string $x \in \{0,1\}^n$ is obtained randomly. For a certain fixed instance U, let $p_{\text{ideal}}(x)$ denote the probability that the outcome x is obtained after measurement of that instance, assuming no error (i.e. the output is ideal).

$$p_{\text{ideal}}(x) = |\langle x|U|0^n\rangle|^2 \,. \tag{1.4}$$

[6] We use the Haar measure because it has certain properties that make the analysis easier, and any conclusions we make should carry over to most other measures. In practice, we would likely choose the gates at random from some discrete gate set because compiling Haar-random unitaries into the discrete set implementable on our device would incur significant overhead.

The circuit sampling task (weak simulation) is to generate outputs x according to the probability distribution p_{ideal}. We can also consider the task of computing $p_{\text{ideal}}(x)$ for some specific choice of x (strong simulation). If a device completes this task successfully for every possible U, we say it is solving the task in the *worst case*. For RCS, we relax this by requiring success only for a large fraction of the instances randomly chosen according to the random quantum circuit ensemble. We can refer to this as the *average case*. By default, RCS refers to the weak simulation task, but, in a slight abuse of language, we will say "strong RCS" to refer to the task of computing output probabilities in the average case.

RCS with noise

We now relax the task to account for the possibility of error due to noise. We measure error by the total variation distance (TVD) between the noisy distribution p_{noisy} and the ideal distribution p_{ideal}.

$$\text{TVD}(p_{\text{ideal}}, p_{\text{noisy}}) = \frac{1}{2}\|p_{\text{ideal}} - p_{\text{noisy}}\|_1 = \frac{1}{2}\sum_{x\in\{0,1\}^n} |p_{\text{ideal}}(x) - p_{\text{noisy}}(x)| \quad (1.5)$$

If the sampled distribution has total variation distance from the ideal distribution that is exponentially small in n, then we call the task *near-exact* simulation, and we treat this as essentially equivalent to exact simulation. If the total variation distance is a small constant, we call it *approximate* simulation. In the context of RCS, these TVD bounds must hold for a large fraction of randomly chosen instances. If a quantum computer is noisy but the noise rate is sufficiently weak and localized, quantum error correction can be employed to implement a random quantum circuit fault-tolerantly, and the distribution p_{ideal} can be near-exactly sampled with only polynomial overhead in number of gates and number of qubits [24]. (However, this polynomial overhead is beyond the capabilities of NISQ-era devices.) Without error correction, the quantum computer can only sample p_{ideal} approximately.

In practice, even approximate sampling can be difficult for a noisy quantum device since the effect of noise accumulates quickly. Consider a computation with s gates, where each two-qubit gate is followed by a depolarizing channel with error probability ϵ on each qubit involved of the gate. In this case, the noisy distribution can differ from the ideal distribution in TVD by a quantity $O(\epsilon s)$. Doing interesting computations on dozens or hundreds of qubits will require s to be at least in the hundreds or thousands, and error rates must be proportionally small to keep the TVD beneath a small constant. In current superconducting qubit systems, error rates are on the order of 10^{-2} to 10^{-3}, which is not good enough to approximately perform interesting computations.

A weaker form of noisy sampling is to generate samples from the *white-noise* distribution

$$p_{\text{wn}}(x) = F p_{\text{ideal}}(x) + (1 - F)2^{-n} \quad (1.6)$$

for some small number F. The white-noise distribution is a mixture of the ideal distribution p_{ideal} and the uniform distribution p_{unif}, for which each outcome x has an equal probability 2^{-n} of being obtained (complete white noise). Note that for typical random circuits, $\text{TVD}(p_{\text{unif}}, p_{\text{ideal}}) = \Theta(1)$, and hence $\text{TVD}(p_{\text{unif}}, p_{\text{wn}}) = \Theta(F)$. We define the task *white-noise* sampling as sampling from a distribution p_{noisy} such that $\text{TVD}(p_{\text{noisy}}, p_{\text{wn}}) \leq \delta F$ for some small constant δ (thus, the sampled distribution p_{noisy} is much closer to p_{wn} than p_{unif} is to p_{wn}).

The upshot of the white-noise distribution is that, while it is far in total variation distance from the ideal distribution p_{ideal}, it retains a weak signal of the ideal distribution, which can be extracted by repeating the experiment many times. For example, suppose we are interested in some quantity $Q(x) \in [-1, 1]$ that has mean μ when x is drawn from p_{ideal}, but mean 0 when x is drawn from the uniform distribution. Then Q has expectation value μF when x is drawn from the white-noise distribution. The standard deviation of Q is bounded by a constant; hence, the error on our estimate of the mean of Q decreases with the number of samples T like $O(1/\sqrt{T})$ and we may estimate μ up to precision η using $O(\eta^{-2}F^{-2})$ samples from p_{wn}. As expected, smaller F means more repetitions are required to extract the signal.

But why should we expect noisy quantum devices to sample from the white-noise distribution? First of all, we expect noise in actual NISQ devices to be fairly localized. In Chapter 5, we will model localized noise by inserting single-qubit noise channels that act on each qubit involved in a gate immediately after the gate. This is an imperfect approximation since noise in actual devices can be correlated from qubit to qubit, but experimentalists can successfully suppress this kind of noise [6]. Note that quantum error correction also requires an assumption of localized noise; the idea behind quantum error correcting codes is to encode logical information into non-local degrees of freedom that are unlikely to be corrupted by an environment that produces local physical errors. Under our local error model, we can generally think of each noise channel as doing nothing with probability $1 - \epsilon$ and applying some error operator with probability ϵ, for some parameter ϵ. Hence, the chance F of performing s two-qubit gates without any errors decays like $F = (1 - \epsilon)^{2s}$. Meanwhile, there is a $1 - F$ probability that at least one of the gates experiences an error. The *white-noise assumption* is essentially the assertion that the output distribution conditioned on the occurrence of at least one error is very close to the uniform distribution, and hence $p_{\text{noisy}} \approx p_{\text{wn}}$. A priori it is not clear if and when this assumption would hold, but it is plausible that it does in the case of random quantum circuits, which are expected to quickly scramble local errors such that they contribute to the output probability in a way that is random and uncorrelated with the ideal output. Numerical evidence in favor of the white-noise assumption for random quantum circuits under a local noise model was provided in Ref. [5].

Task	Condition	NISQ feasibility
Near-exact RCS	$\frac{1}{2}\|p_{\text{ideal}} - p_{\text{noisy}}\|_1 \leq e^{-\Omega(n)}$	Not feasible; requires error correction
Approximate RCS	$\frac{1}{2}\|p_{\text{ideal}} - p_{\text{noisy}}\|_1 \leq \delta$	Feasible only if noise is very weak
White-noise RCS	$\frac{1}{2}\|p_{\text{wn}} - p_{\text{noisy}}\|_1 \leq \delta F$	Feasible with many repetitions if white-noise assumption holds

Table 1.1: Summary of different versions of the RCS task. The error tolerance for each task is stated in terms of the total variation distance between distributions, where p_{ideal} is the ideal (noiseless) distribution, p_{noisy} is the noisy distribution sampled by the device, and p_{wn} is the white-noise distribution (mixture of ideal with weight F and uniform with weight $1-F$), and $\delta = O(1)$ is a constant much smaller than 1. The condition must hold for large constant fraction of random quantum circuit instances.

1.3 Progress on hardness of simulation for random circuits in prior literature

It is generally believed that RCS is a hard classical task regardless of which version of it we choose. A main reason for this is simply that we have no good ideas on how we might simulate random quantum circuits efficiently. In situations where we do know how to efficiently simulate quantum computations on classical computers, the features that make them simulable also make them not generic. Choosing a computation completely at random would be expected to avoid these special cases with high probability.

Yet, relatively little progress has been made on formally connecting this intuition with results from complexity theory. Unlike worst-case weak simulation, which is PH-hard, average-case weak simulation (i.e., RCS) could be easy for classical computers and there would be no surprising consequences. This is true even for exact RCS; showing that approximate RCS or white-noise RCS is a hard classical task would be an even more formidable challenge. In Ref. [4], the hardness of approximate RCS was formally conjectured, building from similar conjectures for other kinds of random computations in Refs. [19, 22], which considered random linear-optical networks and random "IQP" circuits.

The one point of progress on the complexity of average-case simulation has been for strong simulation, not weak simulation. In an intriguing line of work [4, 25–27], it was shown that computing the output probabilities of random quantum circuits is a hard task. The method for doing this was a reduction from the worst-case (which is known to be #P-hard, as previously discussed),

to the average case. Essentially, it was shown that for any worst-case instance, one can generate a set of random instances and then infer the output probability of the worst-case instance from the output probabilities of the random instances. Thus, if one could compute the output probability of most random instances, then one could compute the output probability of any instance with high probability, implying that the average-case task is also #P-hard. Moreover, it has been shown that this inference process is robust to a small amount of error: as long as the average-case output probabilities are computed to additive precision at most $e^{-\Omega(n \log n)}$ [26, 27], the worst-case output probability can be inferred precisely enough for the argument to go through. Deviations of size $e^{-\Theta(n \log n)}$ on each of the 2^n possible outputs corresponds to $2^n e^{-\Theta(n \log n)}$ total variation distance, which is exponentially small and therefore qualifies as "near-exact." Interestingly, if the robustness in the argument could be improved to allow $\Omega(2^{-n})$ additive precision, this would allow the argument to extend from strong simulation to weak simulation! However, there are significant barriers that make such improvements unlikely (see the discussion in Chapter 3 and Ref. [26]).

The status of these different tasks and their known hardness is listed in Table 1.2. The most realistic tasks that can be performed on a noisy quantum device without error-correction—approximate RCS (if noise is very weak) or white-noise RCS (if noise is reasonably weak and the white-noise assumption holds)—have unknown complexity, and are only conjectured to be classically hard. In fact, these conjectures are in a sense two steps away from any concrete statement that can be proved, since they deal with approximate average-case weak simulation, but the only concrete results deal with (near-)exact average-case strong simulation, or (near-)exact worst-case weak simulation: in each case two qualifiers must be changed for something to be known.

1.4 Quantum computational supremacy on noisy devices

Random quantum circuits have also featured prominently in recent experiments aimed at achieving "quantum computational supremacy" [28]. Indeed, it was the proposals for these experiments that originally motivated much of the theoretical attention on the RCS task. In quantum computational supremacy experiments, the explicit goal is simply to perform a well-defined computational task on a quantum device that would be hard to perform on any existing classical device, whether or not that task is useful. In 2019, a team at Google declared they had achieved quantum computational supremacy after performing a version of the RCS task on their noisy device with 53 superconducting qubits arranged in a 2D grid [6]. In 2021, a collaboration at the University of Science and Technology of China (USTC) performed a very similar experiment on 56 superconducting qubits [7].

When n qubits are arranged in a 2D grid, the diameter of the grid is roughly $\sqrt{n}$. These experiments implemented circuits of depth exceeding the circuit diameter, which is necessary for the circuit to be capable of spreading

	Exact worst-case simulation	**Near-exact RCS**	**Approximate RCS**	**White-noise RCS**
Weak simulation	PH-hard	conjectured PH-hard	conjectured PH-hard*	
Strong simulation	#P-hard	#P-hard [25–27]	conjectured #P-hard	conjectured #P-hard

Table 1.2: Summary of known complexity of various simulation tasks. The first row (weak simulation) is the more realistic sampling task that is naturally performed by quantum devices. Weak approximate RCS and weak white-noise RCS (*) are the only tasks a NISQ device might be capable of accomplishing. In Chapter 5, we show they are equivalent in the sense that one is PH-hard if and only if the other is (assuming that F is greater than inverse polynomial in n). In the second row, by strong simulation of RCS, we mean computation of output probabilities for most instances. The classification #P-hard means that an efficient classical algorithm for the task would imply that #P functions can be computed in randomized polynomial time. The classification PH-hard roughly means that an efficient algorithm would imply the collapse of the polynomial hierarchy. Both implications are widely believed to be unlikely (the first even moreso than the second).

local information over the entire system. As n grows, this requires $\Theta(n^{3/2})$ total two-qubit gates. In Google's 53-qubit experiment, there were 430 two-qubit gates (as well as more than a thousand single-qubit gates). The two-qubit gates each had small error rates of less than 1%. Nevertheless, the overall fidelity of their experiment, which is roughly speaking the chance that no errors occur during the computation, was a very small 0.2%.

Since at least one error occurs 998 times out of 1000, the output distribution of the quantum device is not very close (in total variation distance) to the ideal distribution, and the quantum device is not capable of performing the approximate RCS task. Accomplishing the approximate RCS task with TVD on the order of 0.1 would require error rates to be improved by multiple orders of magnitude, and better still as n increases (error rate must scale as $n^{-3/2}$). This is decidedly out-of-reach in the near-term. The only version of the RCS task that Google's device can claim to have performed is white-noise RCS, with the parameter $F = 0.002$. Indeed, in the supplementary material of their paper, they made a complexity-theoretic argument that sampling exactly from p_{wn} can only be a factor of F easier for a classical computer than sampling exactly from p_{ideal}. In reality, even if the white-noise distribution is a good approximation for the output of the device, we do expect there to be some small total variation distance error between the noisy distribution sampled by

the device and the white-noise distribution; it remains conjecture that p_{wn} is hard to sample even when we tolerate a small amount of error.

Verification of the experiment with linear cross-entropy benchmarking

Given the high error rate of Google's experiment, it is of particular importance that they verify that their device actually performs the white-noise RCS task with substantial fidelity. Their solution to the verification problem was to use the linear cross-entropy metric, which is defined for a set of samples $\{x_1, \ldots, x_T\}$ to be

$$\mathcal{F} = \frac{1}{T} \sum_{t=1}^{T} 2^n p_{\text{ideal}}(x_t) - 1 \,. \tag{1.7}$$

In practice, the quantities $p_{\text{ideal}}(x_t)$ are computed by running exponential-time (strong) classical simulation algorithms; thus, they are intractable to calculate in the regime where "quantum computational supremacy" is being declared. However, Google calculated $\mathcal{F}$ on smaller versions of its experiment, and for a variation that omitted a relatively small number of the gates specifically to make calculating $p_{\text{ideal}}(x_t)$ classically easier.

If there is no noise and the samples x_t are drawn from the ideal distribution for sufficiently deep random quantum circuits, then the mean of $\mathcal{F}$ can be shown to be roughly equal to 1. If noise causes the samples x_t to be drawn from the white-noise distribution with parameter F, then the mean of $\mathcal{F}$ is F. Thus, the empirical quantity $\mathcal{F}$ can be used to benchmark the overall fidelity of the experiment. However, the standard deviation of the quantity $\mathcal{F}$ is $O(1/\sqrt{T})$. This means $T = O(1/F^2)$ samples must be taken to differentiate the parameter F from zero. This illustrates how the white-noise assumption is important for justifying usage of the linear cross-entropy to benchmark noise in the experiment. If the white-noise assumption fails and the noisy portion of the output distribution is non-uniform in a way that is correlated with the ideal output distribution, then the quantity $\mathcal{F}$ will not necessarily be a measure of the underlying noise in the device.

However, calculating $\mathcal{F}$ is not alone sufficient for certifying that the output distribution is close to the white-noise distribution, since other distributions can lead to the same value. In fact, there is no hope of definitively verifying that the output distribution is the white-noise distribution without taking an exponential number of samples. We could circumvent this issue simply by re-defining the quantum computational supremacy task to be producing samples with a non-negligible linear cross entropy score $\mathcal{F}$. This weaker task is related to the Heavy Output Generation (HOG) task proposed in Ref. [23]. The issue with these tasks is that it is more difficult to give complexity-theoretic evidence that the task is classically hard. In fact, there is some evidence that scoring well could be classically easy, or at least much easier than a genuine simulation

of random quantum circuits: in Ref. [29], a classical algorithm for "spoofing" the benchmark, i.e. scoring well despite not performing a full simulation of the quantum computation, was developed for shallow circuits. While the algorithm breaks down for deep circuits, or for 2D circuits, it is some indication that the linear cross-entropy benchmark does not fully capture the classical hardness of the simulation task.

Conjectures behind quantum computational supremacy

This clarifies the theoretical challenges associated with Google's claim of quantum computational supremacy. There are essentially two unproven assumptions.

(1) (White-noise assumption) The output distribution of noisy random quantum circuits implemented on Google's device is sufficiently close to the white-noise distribution for most instances.

(2) (Hardness of white-noise RCS) Sampling from a distribution close to the white-noise distribution on most instances is a hard classical task.

1.5 Overview of results

The technical results in this thesis contribute to clarifying the situation regarding the classical simulability of RCS and the theoretical backing of quantum computational supremacy demonstrations. Here we summarize these results, which appear in Chapters 3, 4, and 5.

Efficient classical simulation of shallow 2D random circuits

In Chapter 3, we propose two classical algorithms for the approximate RCS task in any setting where the random circuits act on qubits arranged in a 2D grid and have only a constant number of layers of gates. This setup is identical to recent quantum computational supremacy experiments by Google [6] and USTC [7], except that the depth of the circuits in their experiment was a deeper $\Theta(\sqrt{n})$ instead of $\Theta(1)$. In one specific setting, we can rigorously prove that one of our algorithms efficiently performs the approximate RCS task, even though worst-case simulation and near-exact RCS are known to be hard in the same setting. We also give numerical and analytical evidence that our algorithms are efficient in a much wider range of settings where the circuit depth is sufficiently small. Crucially, our algorithms exploit both the average-case and approximate nature of the task, and cannot successfully perform near-exact RCS, nor can they perform any form of worst-case simulation. An important takeaway, therefore, is that moving from the worst case to the average case or from near-exact to approximate simulation can make the simulation task much easier. This might be regarded as a setback in the journey to give evidence for the claim that NISQ devices performing RCS go beyond the capabilities of classical computers (quantum computational supremacy),

since the main source of formal evidence that the task is hard has been the results on hardness of near-exact strong simulation discussed above.

On the other hand, our algorithms fail to be efficient once the circuits become too deep, and cannot efficiently simulate the deep circuits Google implemented. In some sense, the failure of our algorithm at large depth could be regarded as positive evidence that the approximate RCS task can indeed be hard in that setting. We also show that this transition from efficient to inefficient as the depth increases is related to order-disorder thermal phase transitions in certain classical statistical mechanical systems. Thus, our work exposes the richness of the RCS landscape, where classical simulation complexity has interesting dependencies on the depth and layout of the underlying random circuits and has deep connections to statistical mechanics.

Anti-concentration depth of random quantum circuits

Our algorithm for approximate shallow RCS in 2D leaves open the possibility that circuits with more layers or different layouts can be hard to approximately simulate. While this has not been proven for any RCS setting, the arguments that get closest to a proof require an ingredient called anti-concentration. Anti-concentration roughly means that none of the 2^n possible measurement outcomes are exponentially more likely to be obtained than the others. Anti-concentration is important for avoiding the situation where most output probabilities are very close to zero, and even very tiny errors lead to deviations that are much larger than the probabilities themselves.

In Chapter 4, we rigorously prove that the anti-concentration property is achieved by random quantum circuits at a shallower depth than previously believed. Google designed its quantum computational supremacy experiment to have depth exceeding the $\Theta(\sqrt{n})$ diameter of the 2D qubit array in part because it was believed that this was necessary to gain important properties like anti-concentration. Our work suggests that anti-concentration can be achieved after only $\Theta(\log(n))$ depth. We only manage to prove this statement in the case that the qubits are arranged in 1D and in the case of a "complete-graph" architecture where there is no spatial arrangement whatsoever. However, our framework gives strong heuristic evidence that a similar statement should hold in nearly any natural random quantum circuit layout. This conclusion indicates that fewer gates are needed to reach the regime where classical simulation should be hard. Moreover, we provide lower bounds on the depth needed for anti-concentration. In general, $\Omega(\log(n))$ depth is required, potentially providing an alternate explanation for why constant-depth 2D circuits are easy to approximately simulate. In the 1D and complete-graph case, we prove a tight lower bound that matches the upper bound even up to the constant prefactor of the leading term.

Although anti-concentration is generally thought to imply that simulation is hard, there are some situations where it can indicate that simulation is easy;

it has been a necessary ingredient for certain classical simulation algorithms to be efficient [29–31]. In any case, it is clear that knowing precisely when anti-concentration holds is generally important for assessing the difficulty of simulation tasks. This is not surprising, as anti-concentration is a basic property of the output distribution of a quantum circuit, and ultimately sampling the output distribution is the only way to actually access the quantum information processed during a quantum algorithm.

A proof of the white-noise assumption for random quantum circuits

Finally, in Chapter 5 we prove a version of the white-noise assumption for noisy random quantum circuits, assuming that the noise is local and that the circuit is deep enough to be anti-concentrated. Recall that if a quantum device has s gates each with local depolarizing noise with strength ϵ, approximate RCS can be performed only if $\epsilon s \ll 1$, or in other words, the chance of any errors happening is small. We show that, if each two-qubit gate is followed by single-qubit noise channels with noise strength ϵ on each qubit involved in the gate, then when we choose the fidelity parameter $F = \exp(-2\epsilon s + O(s\epsilon^2)))$, the distance between the output distribution and the white-noise distribution scales as $O(F\epsilon\sqrt{s})$. This means one can perform the white-noise RCS task, as previously defined, on a noisy device so long as the condition $\epsilon^2 s \ll 1$ holds. That is, even when gate errors are common overall, the white-noise assumption is true as long it is rare for there to be a *pair* of gate errors at nearby locations in the circuit. One caveat is that the error rate must also satisfy $1/\epsilon \geq \tilde{\Omega}(n)$, where the tilde suppresses $\log(n)$ factors in the lower bound.

Our results also hold for more general single-qubit noise channels, not just depolarizing noise: our theorems depend only on the average infidelity and the unitarity of the noise channel. En route to proving the white-noise assumption, we also give tight bounds on the decay of the linear cross-entropy benchmark (which depends only on the average infidelity) and the rate of convergence of the output distribution toward the uniform distribution (which depends only on the unitarity). We find that the white-noise assumption only holds when the noise channel is mostly incoherent.

This adds important justification for the use of the white-noise assumption to ground the hardness of quantum computational supremacy experiments based on RCS and communicates how various circuit parameters must scale with n for the white-noise assumption to remain true as n grows. It also justifies the use of the linear cross-entropy benchmark as a method of measuring the overall noise in the experiment.

Beyond the narrow question of whether RCS-based quantum computational supremacy experiments actually perform a classically hard computation, the fact that random quantum circuits (approximately) transform local gate-level errors into white-noise is a valuable insight for the NISQ era, suggesting that it may be possible to salvage computations performed with low

fidelity on a noisy quantum computer by repeating the experiment many times to extract the signal from the white noise. Our bounds inform when to expect that such a strategy could be possible. For example, random quantum circuits are believed to be a good model for evolution by chaotic local Hamiltonians, so simulation of these dynamics on noisy quantum devices are likely to benefit from some version of the white-noise approximation.

Ch.	Result	Significance for simulation complexity
3	Efficient classical algorithms for approximate RCS of shallow 2D circuits	• Proves that it is possible for approximate RCS to be easy even when near-exact RCS and exact worst-case simulation are hard • Presents barrier to proving classical hardness for NISQ-implementable task
4	$\Theta(n \log(n))$ random two-qubit gates are necessary and sufficient to achieve the anti-concentration property	• Clarifies how many gates are needed to attain key ingredient in justification of RCS hardness conjecture
5	Noisy random quantum circuits approximately obey white-noise approximation	• Proves one of two key assumptions (under an idealized noise model) behind low-fidelity quantum computational supremacy experiments • Justifies using linear cross-entropy metric to benchmark NISQ implementation of RCS

Table 1.3: Summary of results in this thesis and their implications for the classical complexity of RCS.

1.6 Outlook for simulation complexity of random quantum circuits

The contributions of this thesis cut both ways on the two complementary questions of (1) whether RCS is hard for a classical computer and (2) whether a well-defined version of RCS can be achieved on a NISQ device.

Regarding (1), our classical algorithm for approximate constant-depth 2D RCS is a significant reason to discount previous evidence that approximate RCS is generally a hard classical task. However, our observation that constant-depth random quantum circuits are not anti-concentrated illustrates one key sense in which constant-depth 2D random quantum circuits are different from deeper 2D random quantum circuits, something that would need to be leveraged if the latter are to be shown to be hard to approximately simulate.

Regarding (2), our proof of the white-noise assumption for devices with local noise—and generally our contribution to understanding the way that random quantum circuits transform local noise in NISQ devices—fills an important gap in connecting what a realistic noisy quantum computer can accomplish to the formal RCS task. If we trust that our device is well characterized by local, mostly incoherent, and sufficiently weak noise, then we can be confident that the white-noise RCS task is indeed being performed. While we would ideally have some independent verification of the task using the output samples alone, a theoretical guarantee under an idealized noise model is perhaps the best that we can hope for in this case. Moreover, the exponential decay of the fidelity F with the number of gates s highlights the importance of minimizing the number of gates in the circuit. On this front, our proof that the anti-concentration property is gained after fewer gates than previously believed is good news for the hardness of NISQ-implementable tasks.

The core theoretical conjecture in this research program, and the essence of the complexity of RCS, is whether classical (weak) simulation of *generic* quantum computations is essentially just as difficult as the worst case. Our contributions demonstrate that proving this remains the central barrier to a stronger claim of quantum advantage in the NISQ era, and that despite significant attention and plenty of intuitive reasons to believe it should be true—we elaborate on some of these reasons in Section 3.2.1 of Chapter 3—formal evidence in favor of the conjecture is lacking. This perspective should not be mistaken for significant doubt in the conjecture; in an empirical sense, the longer the RCS task is studied without the discovery of an efficient classical algorithm to solve it, the more confident we can be that the conjecture is true. Rather, it is our perspective that, in our journey to utilize the laws of quantum mechanics to transcend the capabilities of classical computation, we should demand a more rigorous kind of confidence. Ideally, any contradiction of our expectations that quantum computers are more powerful than classical computers should bring about unexpected and exciting consequences of its own. Reaching that point for tasks that can be implemented on NISQ devices is a worthwhile goal, and it is our hope that the contributions of this thesis have clarified the remaining obstacles toward that end.

1.7 Outline of this thesis

In Chapter 2, we give an introduction to the common technique used in each of the following chapters, namely, the statistical mechanics method for random quantum circuits. Chapters 3, 4, and 5 contain the technical results mentioned above. Each of these chapters follows a roughly similar structure. First, an independent introduction and motivation for the contents of the chapter are provided; these introduction sections are broader than and complementary to the motivation presented in this chapter. Then, an overview of the technical results are stated, along with context on how they fit in with prior literature. Next, the methods are discussed in more detail, but some of

the most technical details are deferred to appendices. Each chapter has its own appendix; the appendix sections are labeled with letters as opposed to numbers.

Chapter 2

THE STATISTICAL MECHANICS METHOD FOR RANDOM QUANTUM CIRCUITS

The previous chapter illustrated how our results tie back to the question of random circuit simulation on classical computers and on noisy quantum devices. However, the story of this thesis could just as easily have been told from the perspective of the underlying techniques. In each of our three technical chapters, we rely on variations of a common approach: the statistical mechanics ("stat mech") method for random quantum circuits. In this chapter, we give an overview of the method and comment on its origins.

The stat mech method creates a map from random quantum circuits to classical stat mech systems. In particular, it maps expectation values of quantities that depend on the circuit instance to partition functions of the stat mech system. It is useful to briefly define and review these concepts individually before discussing the map itself.

2.1 Random quantum circuits and their moments

Random quantum circuits appear in many settings in physics and quantum information science. As a theoretical model, random circuits are broadly useful because they capture a notion of *generic* evolution of quantum systems, free of any built-in symmetries or bias, where the only constraint is the locality of the underlying gates, a constraint that mirrors the local nature of physical laws: interactions in physics are fundamentally few-body and typically require those bodies to be spatially near one another. In the context of quantum computation, random quantum circuits are appealing because they are more or less straightforward to implement once you have a functional (programmable) quantum computer.

To be more precise about notation, a random quantum circuit is a length-s sequence of local (usually two-qudit) randomly chosen unitary gates acting on an n-qudit system. Each qudit has local Hilbert space dimension q; for qubits, $q = 2$. Denote the unitaries enacted by the s gates by $(U^{(1)}, U^{(2)}, \ldots, U^{(s)})$, which can be organized into a quantum circuit diagram, as in Figure 2.1. Each $U^{(t)}$ is chosen independently at random according to some measure on the unitary group. In our analysis, we use the Haar measure, the only measure invariant under multiplication by any unitary. Together these gates determine a global $q^n \times q^n$ unitary U acting on the n-qudit system.[1]

[1] In Chapter 3 and Chapter 5, we consider the alternate scenario where weak measurements and single-qudit noise channels, respectively, occur in between gates. In this case,

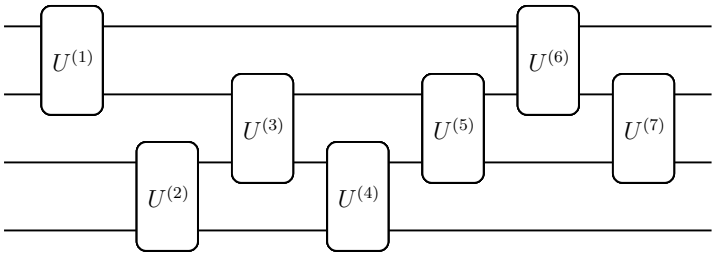


Figure 2.1: Example of a quantum circuit diagram on $n = 4$ qudits with $s = 7$ two-qudit gates. For random quantum circuits, each unitary $U^{(t)}$ is chosen at random, typically from the Haar measure.

The obvious first issue with analyzing random quantum circuits is that the induced measure over U is difficult to grasp. There is an expectation that if we perform enough random two-qudit gates, the global measure should look very similar to the Haar measure on the entire q^n-dimensional Hilbert space, but at finite circuit size it is unclear exactly how to make precise characterizations of the measure. The stat mech method is a solution to this problem. Specifically, the stat mech method is a prescription for calculating expectation values $\mathbb{E}_U[f(U)]$ of some U-dependent quantity $f(U)$ over choice of U from the random quantum circuit ensemble. However, the method only works when f is a linear function in $U^{\otimes k} \otimes U^{*\otimes k}$ for some integer k, where X^* is the complex conjugate of X; that is, the stat mech method describes how to compute kth moment information for the random quantum circuit ensemble. An example of one such function is $f(U) = \langle 0^n|^{\otimes 2k} U^{\otimes k} \otimes U^{*\otimes k} |0^n\rangle^{\otimes 2k} = p_{\text{ideal}}(0^n)^k$, where p_{ideal} is given in Eq. (1.4).

For probability distributions over a single real variable x, the moments $\mathbb{E}[x^k]$ are known to uniquely determine the entire distribution, as long as they are reasonably well behaved [32]. Similarly, we expect the moments of the random quantum circuit measure to collectively contain everything we might want to learn. As we will see, the issue in practice is that the stat mech method becomes increasingly complicated as k increases. In fact, for $k \geq 3$, there has been a scarcity of concrete results using the stat mech method (except in the limit of $q \to \infty$) and the first moment, $k = 1$, is often trivial. Indeed, as the results in this thesis illustrate, it is primarily the second moment, $k = 2$, where the stat mech method truly shines.

2.2 Statistical mechanics, partition functions, and the Ising model

Classical statistical mechanics connects macroscopic properties of physical systems like temperature, energy, and entropy to their microscopic descriptions. The core idea is that for a certain macroscopic state of the system, there

the transformation enacted by the circuit is not strictly unitary, but the stat mech method is still useful.

is some ensemble of possible corresponding microstates. After all, knowing the temperature, pressure, and volume of gas in a room hardly tells you the exact location of all the particles, but it does tell you something.

The Ising model is an illustrative example of classical statistical mechanics in action, and one that we will come back to in the context of random quantum circuits. In the Ising model, we have a system of m spin-1/2 particles, each with two internal states, so the system microstates are labeled by a choice of $\sigma_i = \pm 1$ for each $i = 1, \ldots, m$, collectively denoted by σ. The energy of a microstate is given by

$$H(\sigma) = -\sum_{i=1}^{m} \sum_{j=i+1}^{m} J_{ij} \sigma_i \sigma_j \,, \tag{2.1}$$

where the matrix J_{ij} encodes the interaction strengths between particle i and particle j. The pairs of particles $\langle ij \rangle$ (with $i < j$) for which $J_{ij} \neq 0$ form the edges of the *interaction graph* for the model; often we restrict this interaction graph to be spatially local, for example on a 1D or 2D lattice. One macroscopic quantity of interest is the magnetization $M = \sum_i \sigma_i$, which represents the total magnetic moment of the system as a whole.

If the system is in thermal equilibrium at some temperature T, then introductory statistical mechanics dictates that the system is in the *canonical ensemble* and the probability that the system is in microstate σ is

$$\Pr[\sigma] = \frac{1}{Z} \exp\left(-\frac{H(\sigma)}{k_B T}\right) , \tag{2.2}$$

where k_B is the Boltzmann constant and Z is the partition function, given as follows:

$$Z = \sum_{\sigma} \exp\left(-\frac{H(\sigma)}{k_B T}\right) = \sum_{\sigma} \prod_{i=1}^{m} \prod_{j=i+1}^{m} \exp\left(\frac{J_{ij} \sigma_i \sigma_j}{k_B T}\right) \tag{2.3}$$

$$= \sum_{\sigma} \prod_{\langle ij \rangle} \text{weight}_{\langle ij \rangle}(\sigma) \,, \tag{2.4}$$

where the sum over $\langle ij \rangle$ denotes a sum over edges of the interaction graph. The definition

$$\text{weight}_{\langle ij \rangle}(\sigma) = \exp\left(\frac{J_{ij} \sigma_i \sigma_j}{k_B T}\right) \tag{2.5}$$

emphasizes that the partition function is simply a weighted sum over all possible microstates of the m-particle system, where all the weights are positive numbers, and furthermore each weight can be decomposed into a product of edge weights for each edge $\langle ij \rangle$ in the interaction graph; the edge weight for edge $\langle ij \rangle$ only depends on the internal states of particles i and j.

The Ising model is used to understand ferromagnetism in materials as the temperature changes. Suppose all J_{ij} are non-negative so that the minimum

energy microstates are the microstates where all of the particles have the same internal state: either $\sigma_i = 1$ for all i or $\sigma_i = -1$ for all i. In either of these microstates, the system is highly polarized, with the total magnetization $M = \pm m$. At $T = \infty$, all microstates are equally likely, but as T decreases, lower energy states become more probable. In some cases, there is a critical value of T that divides two distinct phases. The high-temperature "paramagnetic" phase is characterized by disorder and no macroscopic magnetization; that is, microstates drawn from the canonical ensemble typically have small values of $|M|$. The low-temperature "ferromagnetic" phase is characterized by long-range order and macroscopic magnetization; that is, $|M| = \Theta(m)$. This kind of order-disorder thermal phase transition happens for the 2D Ising model, but not for the 1D Ising model, where the paramagnetic phase persists for all $T > 0$. In Chapter 3, we argue that a similar phase transition occurs in the classical model associated with 2D random quantum circuits. This phase transition is driven not by temperature but rather by features of the quantum circuit—specifically, the depth and the local Hilbert space dimension of the qudits.

2.3 The map from random quantum circuits to classical partition functions

The map from quantum circuits to classical stat mech systems depends on the quantum circuit diagram (i.e., the arrangement of two-qudit gates), the local Hilbert space dimension q of the qudits, and the particular quantity of interest that relates to the kth moment of the random circuits. It is simplest to assume that $k = 2$ and then generalize to larger k.

For $k = 2$, we are interested in the expectation value of quantities $f(U)$ where

$$f(U) = L\left(U^{\otimes 2} \otimes U^{*\otimes 2}\right) \tag{2.6}$$

for some linear function L. In fact, since U is composed of the smaller unitaries $U^{(t)}$ for $t = 1, 2, \ldots, s$, the function L is linear in

$$U^{(t)\otimes 2} \otimes U^{(t)*\otimes 2} \tag{2.7}$$

for each t. Since each $U^{(t)}$ is chosen independently at random, we can perform the expectation value over choice of $U^{(t)}$ individually for each t, which is possible because second moment expectation values over the Haar measure have a closed-form expression. We express this key formula for single-qudit $q \times q$ Haar-random unitaries. To so, we first choose any basis $\{|i\rangle\}_{i=0}^{q-1}$ for the Hilbert space and define vectors $|I\rangle$ and $|S\rangle$, which live in a four-fold tensor

product of the Hilbert space.

$$|I\rangle = \sum_{i_1=0}^{q-1} \sum_{i_2=0}^{q-1} |i_1, i_2\rangle \otimes |i_1, i_2\rangle \tag{2.8}$$

$$|S\rangle = \sum_{i_1=0}^{q-1} \sum_{i_2=0}^{q-1} |i_1, i_2\rangle \otimes |i_2, i_1\rangle \,. \tag{2.9}$$

Then, with $\int dV$ denoting integration over the Haar measure, we have the following formula [33]:

$$\int dV \; V^{\otimes 2} \otimes V^{*\otimes 2}$$
$$= \frac{1}{q^2-1}|I\rangle\langle I| + \frac{1}{q^2-1}|S\rangle\langle S| - \frac{1}{q(q^2-1)}|I\rangle\langle S| - \frac{1}{q(q^2-1)}|S\rangle\langle I| \,. \tag{2.10}$$

In the case that V is a two-qudit unitary, we simply send $q \to q^2$, $|I\rangle \to |I\rangle^{\otimes 2}$ and $|S\rangle \to |S\rangle^{\otimes 2}$ in the formula above.

The appearance of the stat mech partition function can be seen directly from Eq. (2.10): each two-qudit gate $U^{(t)}$ in the circuit diagram gets replaced with a *pair* of particles, an incoming particle associated with the bras in the equation and an outgoing particle associated with the kets. These particles can be in one of two internal states $|I\rangle^{\otimes 2}$ or $|S\rangle^{\otimes 2}$, and the formula is a weighted sum over all possible internal states of those two particles. Let τ_t denote the internal state of the tth incoming particle and σ_t denote the internal state of the tth outgoing particle. The coefficients of the four terms above turn into edge weights for the partition function.

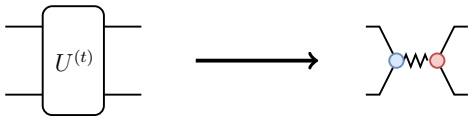


Figure 2.2: Map from a two-qudit gate in the random circuit diagram to a pair of particles, an incoming (blue) particle and an outgoing (red) particle. These particles have an interaction given by Eq. (2.11) and depicted by the zigzag line.

By applying the formula to each of the random gates that comprise U, a circuit with s gates turns into a stat mech system with $2s$ particles, each with two possible internal states, which gives a total of 2^{2s} system microstates. There is an interaction between the two particles arising from the same unitary t, which is denoted by the edge $\langle t \rangle$. The edge weight of $\langle t \rangle$ is read off from

Eq. (2.10), with q replaced by q^2 since the gates act on two qudits.

$$\text{weight}_{\langle t\rangle}(\sigma,\tau) = \begin{cases} (q^4-1)^{-1} & \text{if } \sigma_t = \tau_t \\ -q^{-2}(q^4-1)^{-1} & \text{if } \sigma_t \neq \tau_t\,. \end{cases} \tag{2.11}$$

There are also interactions between an outgoing particle from unitary u and an incoming particle from unitary v if the two unitaries act in succession on the same qudit. We denote this edge as $\langle uv\rangle$. The edge weight of $\langle uv\rangle$ is given by

$$\text{weight}_{\langle uv\rangle}(\sigma,\tau) = \begin{cases} q^2 & \text{if } \sigma_v = \tau_u \\ q & \text{if } \sigma_v \neq \tau_u\,, \end{cases} \tag{2.12}$$

owing to the fact that $\langle I|I\rangle = \langle S|S\rangle = q^2$, and $\langle I|S\rangle = \langle S|I\rangle = q$. In the case that measurements or noise channels act in between unitaries u and v, the weight formula would be modified.

The final ingredient to the correspondence is the choice of boundary conditions at the beginning and end of the quantum circuit diagram. These will depend on the quantity f that we are trying to compute. For example, in Chapter 4, we compute $\mathbb{E}_U[\langle 0^n|^{\otimes 4} U^{\otimes 2} \otimes U^{*\otimes 2} |0^n\rangle^{\otimes 4}]$ and in that case we have open boundary conditions on both sides. However, in Chapter 3, we will see an instance where more complicated boundary conditions at the end of the circuit are required. For this general discussion, we will stick to looking only at the bulk properties of the system. An example of the map from circuit diagram to interaction graph appears in Figure 2.4.

Together, these observations allow us to write

$$\mathop{\mathbb{E}}_{U}[f(U)] = \sum_{\sigma,\tau} \prod_{\langle t\rangle} \text{weight}_{\langle t\rangle}(\sigma,\tau) \prod_{\langle uv\rangle} \text{weight}_{\langle uv\rangle}(\sigma,\tau)\,, \tag{2.13}$$

mirroring the equation for the partition function in Eq. (2.4).

One looming difference between this partition function and that of the Ising model is the possibility of negative weights, as seen in Eq. (2.11). This is a manifestation of the *sign problem*, and is problematic for a few reasons. First of all, it is impossible to interpret the stat mech system as an actual physical system at a real-valued temperature, making connections to conventional statistical mechanics less direct. Second, it is possible that the positive terms and negative terms in the sum are both very large but cancel out to yield something small, complicating combinatorial methods that might try to put upper and lower bounds on the quantity by taking the absolute value of the individual terms. Later, we will show how the issue of negative weights can be circumvented in the case of $k = 2$.

Higher moments

The generalization of the method to moments for $k \geq 3$ is straightforward. In fact, the interaction graph for the stat mech system is independent of k.

However, instead of two internal states, each particle has $k!$ possible internal states, which can each be associated with an element from the symmetric group $\mathcal{S}_k$. These elements are permutations of the indices $\{1, \ldots, k\}$ and can be written in cycle notation; for example, the swap operation for $k = 2$ could be written as (12) instead of S. For $\nu \in \mathcal{S}_k$, we define

$$|\nu\rangle = \sum_{i_1,\ldots,i_k=0}^{q-1} |i_1, \ldots, i_k\rangle \otimes |i_{\nu^{-1}(1)}, \ldots, i_{\nu^{-1}(k)}\rangle \tag{2.14}$$

on $2k$ copies of the Hilbert space, generalizing Eqs. (2.8) and (2.9). This definition allows for an updated integration formula [33, 34]

$$\int dV\ V^{\otimes k} \otimes V^{*\otimes k} = \sum_{\nu,\mu\in\mathcal{S}_k} \mathcal{W}\mathrm{g}(q, \nu^{-1}\mu)|\nu\rangle\langle\mu|\,, \tag{2.15}$$

where $\mathcal{W}\mathrm{g}(q, \nu)$ is the Weingarten function [33–35], and updated weight formulas

$$\mathrm{weight}_{\langle t\rangle}(\sigma, \tau) = \mathcal{W}\mathrm{g}(q^2, \tau_t^{-1}\sigma_t) \tag{2.16}$$

$$\mathrm{weight}_{\langle uv\rangle}(\sigma, \tau) = q^{C(\sigma_u^{-1}\tau_v)}\,, \tag{2.17}$$

with $C(\nu)$ the function that returns the number of cycles in the permutation $\nu \in \mathcal{S}_k$.

Getting rid of negative weights by decimating incoming particles

We now consider decimating the incoming particles; that is, in the partition function in Eq. (2.13), we explicitly perform the sum over τ. This removes the incoming particles from the system, giving a new stat mech system with half as many particles. The remaining sum over σ can still be interpreted as a partition function on this system, but the interactions become three-body instead of two-body: we now have an interaction hypergraph consisting of hyperedges on sets of three particles. These hyperedges $\langle uvw\rangle$ exist whenever unitary u acts on a pair of qudits that were most recently acted upon by unitaries v and w. We can compute this hyperedge weight by summing over the $k!$ possible values of τ_u, as follows:

$$\mathrm{weight}_{\langle uvw\rangle}(\sigma) = \sum_{\tau_u\in\mathcal{S}_k} q^{C(\sigma_v^{-1}\tau_u)} q^{C(\sigma_w^{-1}\tau_u)}\, \mathcal{W}\mathrm{g}(q^2, \tau_u^{-1}\sigma_u)\,, \tag{2.18}$$

which for $k = 2$, yields the simple expression

$$\mathrm{weight}_{\langle uvw\rangle}(\sigma) = \begin{cases} 1 & \text{if } \sigma_u = \sigma_v = \sigma_w \\ \frac{1}{q+q^{-1}} & \text{if } \sigma_v \neq \sigma_w \\ 0 & \text{if } \sigma_u \neq \sigma_v = \sigma_w\,. \end{cases} \tag{2.19}$$

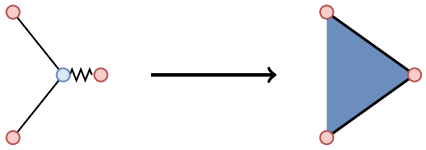


Figure 2.3: Decimation of incoming (blue) particles creates a three-body interaction between leftover outgoing particles.

The partition function for this version of the method reads

$$\mathop{\mathbb{E}}_{U}[f(U)] = \sum_{\sigma} \prod_{\langle uvw \rangle} \text{weight}_{\langle uvw \rangle}(\sigma)\,. \tag{2.20}$$

At the expense of going from two-local to three-local interactions, we have arrived at a stat mech system with only non-negative interaction weights, which makes analysis easier. We can also see that this model is similar in spirit to the ferromagnetic Ising model since the largest weights occur when all of the spins agree.

Unfortunately, for $k \geq 3$, some of the three-body weights can still be negative [36, 37]. This observation, combined with the larger number of internal states and more complex interaction weights makes analyzing systems for $k \geq 3$ much more difficult than for $k = 2$.

An equivalent picture of evolving n-bit configurations

In Chapter 4 and Chapter 5, we will essentially be analyzing the system described above with three-body weights, but we organize our analysis in a slightly different way. For any choice of t with $0 \leq t \leq s$ and each $0 \leq a < n$, let $u_{a,t}$ be the maximum integer such that $u_{a,t} \leq t$ and qudit a was one of the qudits involved in the gate at time step $u_{a,t}$. Then we can construct what we call a *configuration* $\vec{\gamma}^{(t)} \in \{I, S\}^n$ by letting $\gamma_a^{(t)} = \sigma_{u_{a,t}}$. The sequence of configurations $\gamma = (\vec{\gamma}^{(0)}, \vec{\gamma}^{(1)}, \ldots, \vec{\gamma}^{(s)})$ is called a *trajectory*, where $\vec{\gamma}^{(t)}$ is the same as $\vec{\gamma}^{(t-1)}$ except potentially at the positions that are involved in the tth gate. Each trajectory γ is associated with a microstate of the system in the three-body stat mech picture as long as $\gamma_a^{(t)} = \gamma_b^{(t)}$ whenever unitary t acts on qudits a and b. When this rule is obeyed, we can see from Eq. (2.19) that the overall weight of a trajectory γ in the partition function decreases by a factor of $q + q^{-1}$ each time gate t acts on two qudits that have different assignments in configuration $\vec{\gamma}^{(t-1)}$. This version of the stat mech method is derived in a self-contained manner in Chapter 4.

2.4 Past and future of the stat mech method

The stat mech method for random quantum circuits is very similar to a method for analyzing random tensor networks, first introduced in 2016 by Hayden et al. [38] as a model for holographic duality, and further studied in

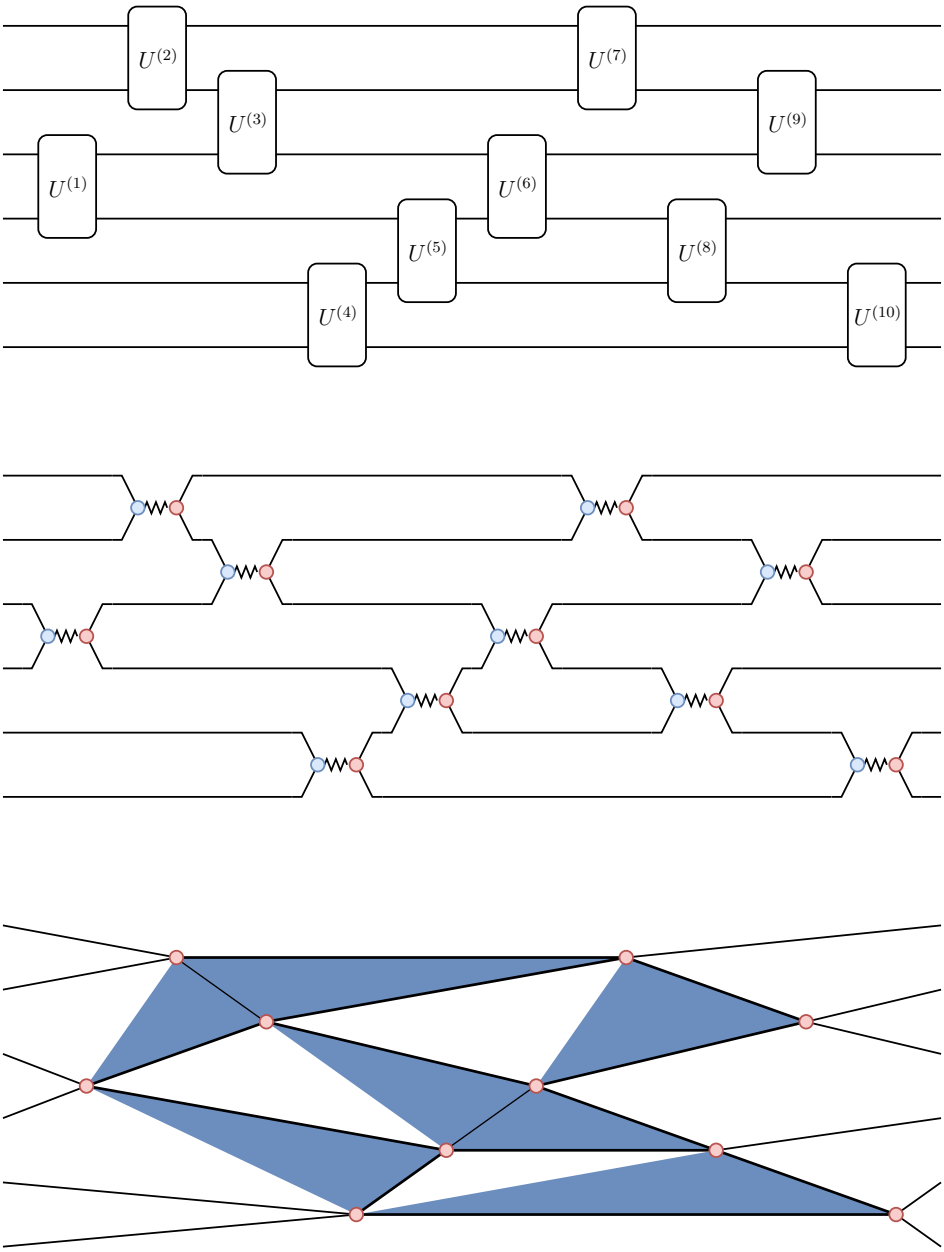


Figure 2.4: Complete example of stat mech map on a random quantum circuit with $s = 10$ gates on $n = 6$ qudits. Top: Circuit diagram. Middle: Interaction graph for corresponding stat mech system. There are two types of interactions, denoted by straight and zigzag lines between particles. Expectation values of kth moment quantities are given by partition functions for different choices of boundary conditions at the input and output of the circuit. Bottom: Interaction hypergraph with three-body interactions (blue triangles) that results from decimating all of the incoming (blue) particles from the middle diagram.

Refs. [39, 40]. The random-quantum-circuit version of the method, as described in this chapter, first appeared in 2017 in two simultaneous papers, one by Nahum, Vijay, and Haah [41] and another by von Keyserlingk, Rakovszky, Pollmann, and Sondhi [42]. Both of these papers sought to understand chaotic quantum dynamics by studying entanglement growth and the spreading of local operators in random quantum circuits. They employed the $k = 2$ version of the method on 1D random quantum circuits to compute the average out-of-time-order correlator (OTOC), a measure of operator spreading, as well as the average purity (exponential of the Rényi-2 entanglement entropy) of subregions of the chain. They observed that for $k = 2$, the three-body version of the method maps 1D circuit diagrams to an interaction graph on a square lattice, where each particle has two possible internal states. Contiguous regions of particles with the same internal state define a "domain" and "domain walls" divide these domains. The OTOC and average purity were computed by counting domain wall patterns on the 2D lattice. Further calculations of related quantities using a similar stat-mech-based approach later appeared in Ref. [43].

In 2018, the method was considerably extended by Zhou and Nahum [36], who combined it with the *replica trick* to compute the scaling of the average Rényi-k entanglement entropy for different k in 1D random quantum circuits, as a perturbative series in $1/q$. They discussed the method for general k, although explicit calculations were only performed for $k = 2$ and $k = 3$. For $k \geq 3$, there are more than two possible internal states for each particle, and thus more than one flavor of domain wall. When domain walls of different flavors come together, the combinatorial approach becomes more complicated, in part because of the possibility of negative weights.

In 2019, Hunter-Jones [37] used the method to study the convergence of 1D random quantum circuits to unitary k-designs, again by domain wall counting. They showed that previously known results for $k = 2$ could be reproduced using the stat mech method, and made progress toward extending these results to larger k. Under a conjecture that single-domain-wall configurations dominate the calculation, it was shown that approximate k-designs are achieved on n qudits in $O(nk)$ circuit depth.

Later in 2019, the method was used in parallel papers by Jian, You, Vasseur, and Ludwig [44] and Bao, Choi, and Altman [45] to study 1D random quantum circuits that include measurements of some fraction of the qubits after each layer of gates. It had been numerically observed in previous literature (e.g., Ref. [46]) that the output state of these "hybrid" circuits undergoes a phase transition from a high-entanglement volume-law phase to a low-entanglement area-law phase as the measurement fraction is increased. These two papers gave an analytical explanation for the transition using the stat mech method: the $k = 2$ version of the method allows for direct analysis of an entanglement-entropy-like quantity that is similar to the Rényi-2 entropy,

and it was seen that this quantity undergoes a disorder-order Ising-like phase transition. Additionally, higher k could be successfully analyzed in the $q \to \infty$ limit, which, using the replica trick, allowed a deduction about the average von Neumann entropy in that limit. Similar themes of an entanglement phase transition appear in Chapter 3 of this thesis, which was adapted from work [47] that first appeared publicly in early 2020.

The stat mech method has also been connected heuristically to random stabilizer quantum error-correcting codes in Refs. [48, 49] and employed to give evidence for exponential decay of fidelity in noisy random quantum circuits in Ref. [50] (more on this in Chapter 5). Besides the work in this thesis, the only other appearance of the method in relation to classical algorithms for simulating random quantum circuits was in Ref. [29], which studied anti-concentration in 1D circuits in order to argue their proposed algorithm is efficient. We discuss that work further in Chapter 4.

Taken together, and in combination with the contributions in this thesis, these results demonstrate that the stat mech method can be used in a variety of settings. However, they also illustrate how the same hurdles appear regardless of the application in mind. These hurdles are primarily (1) the difficulty of analyzing the stat mech system for $k \geq 3$ due to a larger number of internal states for each particle and the possibility of negative weights, and (2) the difficulty of going beyond 1D random quantum circuit architectures. On that point, the work in this thesis pushes the method in a new direction since in Chapter 3, we apply the method to several different families of shallow 2D circuits. (However, there is a sense in which shallow 2D circuits and deep 1D circuits are very similar, with the second spatial dimension in the former playing the role of time in the latter.) Moreover, in Chapter 4 and Chapter 5, we move beyond 1D random quantum circuits and give a precise analysis for a fully connected architecture.

Moving forward, given the success of the $k = 2$ version of the stat mech method in many disparate settings, we are confident that there are many more situations where the $k = 2$ method should be useful. It would seem that any second-moment quantity for random quantum circuits, even if it does not allow for rigorous analysis, would at least benefit from heuristic reasoning under the stat mech method. Additionally, we are optimistic that further study of the stat mech systems for $k \geq 3$ might allow the obstacles discussed to be surmounted in some circumstances.

Chapter 3

EFFICIENT SIMULATION OF SHALLOW 2D RANDOM QUANTUM CIRCUITS

This chapter has been adapted from joint work with John Napp, Rolando L. La Placa, Fernando G. S. L. Brandão, and Aram W. Harrow in Ref. [47].

3.1 Motivation

Simulating quantum systems on classical computers requires effort that scales exponentially with the size of the quantum system, at least for a general-purpose simulation. This was the original motivation of the field of quantum computing, since it shows that time evolution (natural or controlled) of quantum systems can perform tasks that are intractable for classical computers. However, many special cases are known where quantum time evolution *can* be efficiently simulated, including limited entanglement/interaction, free fermions, and Clifford circuits. Understanding the boundary between classically simulable and intractable gets at a crucial question of quantum computers: what makes quantum computing powerful?

There are two main ways to answer this question concretely. We can find more classes of easy-to-simulate quantum dynamics, or we can find evidence that other classes of quantum dynamics are hard for classical computers to simulate. The latter approach is related to the goal of *quantum computational supremacy* [28], which involves finding a well-defined computational task with evidence for classical intractability (usually based on a plausible conjecture from complexity theory), then actually performing the task on quantum hardware and verifying the result. Achieving this feat (as has been claimed by Refs. [6, 7, 51] is the computational analogue of a Bell inequality violation: theoretically it would merely confirm orthodox interpretations of quantum mechanics, but practically it would be a milestone in our ability to coherently control quantum systems.

A leading proposal for demonstrating quantum computational supremacy is Random Circuit Sampling (RCS), meaning that the quantum computer applies random unitary gates and then measures all the qubits. This was used by Google [6] and, as we will discuss below, it is a plausible candidate for an intractable class of dynamics. Indeed, the previously known examples of efficiently simulable quantum dynamics were all in some ways special: using only Clifford gates, or only non-entangling gates, for example. So it would be reasonable to assume that random gates would be the best way to avoid any known or unknown structure in the circuits that would facilitate simulation.

The main contribution of this chapter is to show that RCS becomes easy to simulate at low enough circuit depth and local dimension. We do this by developing classical algorithms for RCS that are efficient (polynomial-time) in some settings in which all previously known classical algorithms are inefficient (exponential-time). Moreover, in these regimes no efficient classical sampling algorithms are possible for arbitrary (i.e., non-random) circuits, assuming standard complexity theoretic assumptions (specifically, the "non-collapse of the Polynomial Hierarchy (PH)"). Our results thus show that for natural problems, random instances can be much easier than a worst-case analysis would suggest. On the other hand, we also find evidence that our algorithms exhibit computational phase transitions into inefficient regimes when certain parameters of the circuits are tuned.

3.2 Overview of contributions

As discussed in the previous section, a fundamental question in computer science and physics is to understand where the boundary between classically-intractable and classically-simulable quantum systems or quantum circuits lies. A more specific question within the context of quantum computational supremacy is to understand what types of quantum gate sequences are hardest to classically simulate. So far, our answers to these questions have been informal or incomplete. On the simulation side, Markov and Shi [52] showed that a quantum circuit could be classically simulated by contracting a tensor network with cost exponential in the treewidth of the graph induced by the circuit. (Treewidth is a measure of how far from a tree a graph is; it is 1 for a tree and $\sim L^{D-1}$ for a D-dimensional lattice with side length L.) When applied to n qubits in a line running a circuit with depth d, the simulation cost of this algorithm is $\exp(\Theta(\min(n, d)))$. More generally we could consider $n = L_1 L_2$ qubits arranged in an $L_1 \times L_2$ grid running for depth d, in which case the simulation cost would be

$$\exp\Big(\Theta(\min(L_1 L_2, L_1 d, L_2 d))\Big). \tag{3.1}$$

In other words, we can think of the computation as taking up a space-time volume of $L_1 \times L_2 \times d$ and the simulation cost is dominated by the size of the smallest cut bisecting this volume. An exception is for depth $d = 1$ or $d = 2$, which have simple exact simulations [53]. Some restricted classes such as stabilizer circuits [13] or one-dimensional systems that are sufficiently unentangled [9–11] may also be simulated efficiently. However, the conventional wisdom has been that in general, for 2D circuits with $d \geq 3$, the simulation cost scales as Eq. (3.1).

These considerations led IBM to propose the benchmark of "quantum volume" [54] which in our setting is $\exp(\sqrt{d \min(L_1, L_2)})$; this does not exactly coincide with Eq. (3.1) but qualitatively captures a similar phenomenon. The idea of quantum volume is to compare quantum computers with possibly different architectures by evaluating their performance on a simple benchmark.

This benchmark task is to perform n layers of random two-qubit gates on n qubits, and being able to perform this with ~ 1 expected gate errors corresponds to a quantum volume of $\exp(n)$.[1] Google's quantum computing group has also proposed random unitary circuits as a benchmark task for quantum computers [5]. While their main goal has been quantum computational supremacy [6, 56], random circuits could also be used to diagnose errors including those that go beyond single-qubit error models by more fully exploring the configuration space of the system [54].

These proposals from industry reflect a rough consensus that simulating a 2D random quantum circuit should be nearly as hard as exactly simulating an arbitrary circuit with the same architecture, or in other words that random circuit simulation is nearly as hard as the *worst case*, given our current state of knowledge.

To the contrary, we prove (assuming standard complexity-theoretic conjectures) that for a certain family of constant-depth architectures, classical simulation of typical instances with small allowed error is easy, despite worst-case simulation being hard (by which we mean, it is classically intractable to simulate an arbitrary random circuit realization with arbitrarily small error). For these architectures, we show that a certain algorithm exploiting the randomness of the gates and the allowed small simulation error can run much more quickly than the scaling in Eq. (3.1), running in time $O(L_1 L_2)$. While our proof is architecture-specific, we give numerical and analytical evidence that for sufficiently low constant values of d, the algorithm remains efficient more generally. The intuitive reason for this is that the simulation of 2D shallow random circuits can be reduced to the simulation of a form of effective 1D dynamics which includes random local unitaries and weak measurements. The measurements cause the 1D process to generate much less entanglement than it could in the worst case, making efficient simulation possible. Such dynamics consisting of random local gates with interspersed measurements has in fact recently become the subject of an intensive research focus [44–46, 48, 57–80], and our simulation algorithm can be viewed as an application of this line of work. Furthermore, the measurement-strength-driven entanglement phase transitions observed in these processes are closely related to the computational phase transition we observe for our algorithms.

3.2.1 Evidence from prior literature that simulating random circuits is hard

Before discussing our results in greater detail, we briefly review the main technical arguments for the prevailing belief that random circuit simulation

[1] Our calculation of quantum volume for 2D circuits above uses the additional fact that, assuming for simplicity that $L_1 \leq L_2$, we can simulate a fully connected layer of gates on $L_2 x$ qubits (for $x \leq L_1$) with $O(x L_2 / L_1)$ locally connected 2D layers using the methods of [55]. Then x is chosen to maximize $\min(L_2 x, d/(x L_2/L_1))$.

should be nearly as hard as the worst case.

1. Evidence from complexity theory. A long line of work has shown that it is worst-case hard to either sample from the output distributions of quantum circuits or compute their output probabilities with exponentially small error [19, 22, 28, 30, 53, 81, 82]. While the requirements of worst-case and near-exact simulation are rather strong, these results do apply to any quantum circuit family that becomes universal once post-selection [81] is allowed, thereby including noninteracting bosons [19] and 2D depth-3 circuits [53]. The hardness results are also based on the widely believed conjecture that the polynomial hierarchy (PH) is infinite, or more precisely that approximate counting is weaker than exact counting. Since these results naturally yield worst-case hardness, they do not obviously imply that *random* circuit simulation should be hard. In some cases, additional conjectures can be made to extend the hardness results to some form of average-case hardness (as well as ruling out approximate simulations) [19, 22, 23], but these conjectures have not received widespread scrutiny. Besides stronger conjectures, these hardness results usually require that the quantum circuits have an "anti-concentration" property, meaning roughly that their outputs are not too far from the uniform. This is the subject of Chapter 4. While random circuits are certainly not the only route to anti-concentration (applying a Hadamard gate to each qubit of $|0\rangle^{\otimes n}$ would do), they are a natural way to combine anti-concentration with an absence of any obvious structure (e.g., Clifford gates) that might admit a simple simulation (however, note that in Chapter 4, we show that constant-depth random quantum circuits do not have the anti-concentration property). Furthermore, a line of work beginning with Ref. [4] (see [25–27] for subsequent improvements) has established that random circuit simulation admits a worst-to-average case reduction for the computation of output probabilities. In particular, the ability to near-exactly compute the probability of some output string for a $1 - 1/\operatorname{poly}(n)$ fraction of Haar-random circuit instances on some architecture is essentially as hard as computing output probabilities for an arbitrary circuit instance with this architecture, which is known to be #P-hard even for certain 2D depth-3 architectures.

2. Near-maximal entanglement in random circuits. Haar-random states on n qudits are nearly maximally entangled across all cuts simultaneously [83, 84]. Random quantum circuits on $L \times L \times \cdots$ arrays of qudits achieve similar near-maximal entanglement across all possible cuts once the depth is at least $\sim L$ [85, 86] and before this time, the entanglement often spreads "ballistically" [87]. Random tensor networks with large bond dimension nearly obey a min-flow/max-cut-type theorem [38, 88], again meaning that they achieve nearly maximal values of an entanglement-like quantity. These results suggest that when running algorithms based on tensor contraction, random gates should be nearly the hardest possible gates to simulate.

3. Absence of algorithms taking advantage of random inputs. There are not many algorithmic techniques known that simulate random circuits more easily than worst-case circuits. There are a handful of exceptions. In the presence of any constant rate of noise, random circuits [31, 89], IQP circuits [30] and (for photon loss) boson sampling [90, 91] can be efficiently simulated. These results can also be viewed as due to the fact that fault-tolerant quantum computing is not a generic phenomenon and requires structured circuits to achieve (see [30] for discussion in the context of IQP). Permanents of random matrices whose entries have small nonzero mean can be approximated efficiently [92], while the case of boson sampling corresponds to entries with zero mean and the approach of [92] is known to fail there. A heuristic approximate simulation algorithm based on tensor network contraction [93] was recently proposed and applied to random circuits, although for this algorithm it is unclear how the approximations made are related to the overall simulation error incurred (in contrast, our algorithm based on matrix product states can bound the overall simulation error it is making, even when comparison with exact simulation is not feasible). In practice, evidence for a hardness conjecture often is no more than the absence of algorithms. Indeed, while some approximation algorithms are known for estimating output probabilities of constant-depth circuits [94], IQP circuits [95] and boson sampling [19] up to additive error δ in time $\mathrm{poly}(n, 1/\delta)$, these are not very helpful for random circuits where typical output probabilities are $\sim 2^{-n}$.

Despite the above intuitive arguments for why the simulation of uniformly random circuits should be nearly as hard as the worst case, we (1) prove that there exist architectures for which this is not the case, and (2) give evidence that this result is not architecture-specific, but is rather a general property of sufficiently shallow random circuits. To this end, we propose and implement a simulation algorithm based on a 2D-to-1D mapping in conjunction with tensor network methods. In Appendix 3.B, we introduce and study a second simulation algorithm (referred to as `Patching`) based on locally simulating spatially disconnected regions which are then "stitched" together. The performance of both algorithms is related to certain entropic quantities.

We also give evidence of computational phase transitions for our proposed simulation algorithms driven by circuit depth and qudit dimension. Previously it was known that phase transitions between classical and quantum computation existed as a function of the noise parameter in conventional quantum computation [96–102] as well as in measurement-based quantum computing (MBQC) [103, 104]. In the noiseless setting, besides the gap between depth-2 and depth-3 circuits [53], a computational phase transition as function of rate of qubit loss during the preparation of a resource state for MBQC [105] and (under additional assumptions) as a function of duration of time evolution for simulating dynamics generated by quadratic bosonic Hamiltonians [106, 107] was also known.

3.2.2 Our results

We give two classes of results, which we summarize in more detail below. The first consists of rigorous separations in complexity between worst-case simulation[2] and approximate average-case simulation (for sampling) and between near-exact average-case simulation and approximate average-case simulation (for computing output probabilities) for random circuit families defined with respect to certain circuit architectures. While these results are rigorous, they are proved with respect to a contrived architecture and therefore do not address the question of whether random shallow circuits are classically simulable more generally. To address this issue, we also give conjectures on the performance of our algorithms for more general and more natural architectures. Our second class of results consists of analytical and numerical evidence supporting these conjectures.

Provable complexity separations

We now summarize our provable results for particular circuit architectures. We first define more precisely what we mean by an "architecture."

Definition 3.1 (Architecture)**.** *An* architecture A *is an efficiently computable mapping from positive integers L to circuit layouts $A(L)$ defined on rectangular grids with sidelengths $L \times f(L)$ for some function $f(L) \leq \mathrm{poly}(L)$. A "circuit layout" is a specification of locations of gates in space and time and the number of qudits acted on by each gate. (The gate itself is not specified.) For any architecture A, we obtain the associated Haar-random circuit family acting on qudits of constant dimension q, $C_{A,q}$, by specifying every gate in A to be distributed according to the Haar measure and to act on qudits of dimension q which are initialized in a product state $|0\rangle^{\otimes(L\times f(L))}$.*

In this paper, we only consider architectures that are constant depth and spatially 2-local (that is, a gate either acts on a single site or two adjacent sites); therefore, "architecture" for our purposes always refers to a constant-depth spatially 2-local architecture. The above definition permits architectures for which the layout of the circuit itself may be different for different sizes. However, it is natural for a circuit architecture to be spatially periodic, and furthermore for the "unit cells" of the architecture to be independent of L. We formalize this as a notion of *uniformity*, which we define more precisely below.

Definition 3.2 (Uniformity)**.** *We call a constant-depth architecture A* uniform *if there exists some spatially periodic circuit layout B on an infinite square lattice such that, for all positive integers L, $A(L)$ is a restriction of B to a rectangular sub-grid with sidelengths $L \times f(L)$ for some $f(L) \leq \mathrm{poly}(L)$. A*

[2]Unless specified otherwise, we use *worst-case simulation* to refer to the problem of exactly simulating an arbitrary circuit instance.

random circuit family $C_{A,q}$ associated with a uniform architecture A is said to be a uniform random circuit family.

While uniformity is a natural property for a circuit architecture to possess, our provable separations are with respect to certain non-uniform circuit families. In particular, we prove in Section 3.4 that for any fixed $0 < c < 1$, there exists some non-uniform circuit architecture A acting on n qubits such that, if C_A is the Haar-random circuit family associated with A acting on qubits, the following are true:

1. **Exact worst-case sampling is hard**: There does not exist a poly(n)-time classical algorithm that exactly samples from the output distribution of arbitrary realizations of C_A unless the polynomial hierarchy collapses to the third level.

2. **Near-exact average-case computation of output probabilities is hard**: Given an arbitrary fixed output string $\mathbf{x}$, there does not exist a poly(n)-time classical algorithm for computing the probability of obtaining $\mathbf{x}$, $|\langle \mathbf{x}|C_A|0\rangle^{\otimes n}|^2$, up to additive error $\leq 2^{-cn\log(n)}$ for a constant $c > 0$, with probability at least $1 - 1/\operatorname{poly}(n)$ over choice of circuit instance, unless a #P-hard function can be computed in randomized polynomial time.

3. **Approximate average-case sampling is easy**: There exists a classical algorithm that runs in time $O(n)$ and, with probability at least $1 - 2^{-n^c}$ over choice of circuit instance, samples from the output distribution of C_A up to error at most 2^{-n^c} in total variation distance.

4. **Approximate average-case computation of output probabilities is easy**: There exists a classical algorithm that runs in time $O(n)$ and, for an arbitrary output string $\mathbf{x}$, with probability at least $1 - 2^{-n^c}$ over choice of circuit instance, estimates $|\langle \mathbf{x}|C_A|0\rangle^{\otimes n}|^2$ up to additive error $2^{-n}/2^{n^c}$. (This should be compared with 2^{-n}, which is the average output probability over choices of $\mathbf{x}$.)

The first two points above follow readily from prior works (respectively [53] and [26, 27]), while the latter two follow from an analysis of the behavior of one of our simulation algorithms for this architecture. These algorithms improve on the previously best known simulation time for this family of architectures of $2^{O(L)} = 2^{O(n^{c'})}$ for some constant $c'(c) < 1$ based on an exact simulation based on tensor network contraction. We refer to the architectures for which we prove the above separations as "extended brickwork architectures" (see Figure 3.3 for a specification), as they are related to the "brickwork architecture" [108] studied in the context of MBQC.

Implications for quantum computational supremacy. The worst-case to average-case reductions that imply the second item above have been widely cited as evidence for the conjectures that underpin random-circuit-based quantum computational supremacy proposals. Yet, the existence of an architecture for which the fourth item is also true indicates that the robustness of the reduction could not be sufficiently improved to actually prove those conjectures, barring the introduction of some new technique that is sensitive to the circuit depth. Thus, although our algorithms can only efficiently simulate shallow random circuits, they accentuate a fundamental weakness in the main source of formal evidence for hardness even in the case of deep circuits. (See Appendix 3.D for further discussion of the relationship to this line of work.)

Conjectures for uniform architectures

While the above results are provable, they are unfortunately proved with respect to a unnatural non-uniform architecture, and furthermore do not provide good insight into how the simulation runtime scales with simulation error and simulable circuit fraction. An obvious question is then whether efficient classical simulation remains possible for more natural random circuit families that are sufficiently shallow, and if so, how the runtime scales with system size and error parameters. We argue that it does, but that a computational phase transition occurs for our algorithms when the depth (d) or local Hilbert space dimension (q) becomes too large. Here we are studying the simulation cost as $n \to \infty$ for fixed d and q. Intuitively, there are many constant-depth random circuit families for which efficient classical simulation is possible, including many "natural" circuit architectures (it seems plausible that *any* depth-3 random circuit family on qubits is efficiently simulable). However, we expect a computational phase transition to occur for sufficiently large constant depths or qudit dimensions, at which point our algorithms become inefficient. The location of the transition point will in general depend on the details of the architecture. The conjectures stated below are formalizations of this intuition.

We now state our conjectures more precisely. Conjecture 3.1 essentially states that there are *uniform* random circuit families for which worst-case simulation (in the sense of sampling or computing output probabilities) is hard, but approximate average-case simulation can be performed efficiently. (Worst-case hardness for computing probabilities also implies a form of average-case hardness for computing probabilities, as discussed above.) This is stated in more-or-less the weakest form that seems to be true and would yield a polynomial-time simulation. However, we suspect that the scaling is somewhat more favorable. Our numerical simulations and toy models are in fact consistent with a stronger conjecture, Conjecture 3.1′, which, if true, would yield stronger run-time bounds. Conversely, Conjecture 3.2 states that if the depth or local qudit dimension of such an architecture is made to be a sufficiently large constant, our two proposed algorithms experience computational

phase transitions and become inefficient even for approximate average-case simulation.

Conjecture 3.1. *There exist uniform architectures and choices of q such that, for the associated random circuit family $C_{A,q}$, (1) worst-case simulation of $C_{A,q}$ (in terms of sampling or computing output probabilities) is classically intractable unless the polynomial hierarchy collapses, and (2) our algorithms approximately simulate $C_{A,q}$ with high probability. More precisely, given parameters ε and δ, our algorithms run in time bounded by $\mathrm{poly}(n, 1/\varepsilon, 1/\delta)$ and can, with probability $1-\delta$ over the random circuit instance, sample from the classical output distribution produced by C_q up to variational distance error ε and compute a fixed output probability up to additive error ε/q^n.*

Conjecture 3.1′. *For any uniform random circuit family $C_{A,q}$ satisfying the conditions of Conjecture 3.1, efficient simulation is possible with runtime replaced by*

$$n^{1+o(1)} \cdot \exp\Big(O(\sqrt{\log(1/\varepsilon\delta)})\Big). \tag{3.2}$$

Conjecture 3.2. *For any uniform random circuit family $C_{A,q}$ satisfying the conditions of Conjecture 3.1, there exists some constant q^* such that our algorithms become inefficient for simulating $C_{A,q'}$ for any constant $q' > q^*$, where $C_{A,q'}$ has the same architecture as as C_q but acts on qudits of dimension q'. There also exists some constant k^* such that, for any constant $k > k^*$, our algorithms become inefficient for simulating the composition of k layers of the random circuit, $C^k_{A,q} \circ \cdots \circ C^2_{A,q} \circ C^1_{A,q}$, where each $C^i_{A,q}$ is i.i.d. and distributed identically to $C_{A,q}$. In the inefficient regime, for fixed ε and δ the runtime of our algorithms is $2^{O(L)}$.*

Our evidence for these conjectures, which we elaborate upon in the following sections, consists primarily of the following elements:

1. A rigorous reduction from the 2D simulation problem to a 1D simulation problem that can be efficiently solved with high probability if certain conditions on expected entanglement in the 1D state are met (Section 3.3).

2. Convincing numerical evidence that these conditions are indeed met for a specific worst-case-hard uniform random circuit family and that in this case the algorithm is extremely successful in practice (Section 3.5).

3. Heuristic analytical evidence for both conjectures using a mapping from random unitary circuits to classical statistical mechanical models (Section 3.6), and for Conjecture 3.1′ using a toy model which can be more rigorously studied (Section 3.3.4).

The uniform random circuit family for which we collect the most evidence for classical simulability is associated with the depth-3 "brickwork architecture" [108] (see also Figure 3.3 for a specification).

In the remainder of the paper we develop the evidence for our conjectures outlined in the three items above, and also present our rigorous complexity separation in Section 3.4.

3.3 Simulation by reduction to 1D dynamics

We reduce the problem of simulating a constant-depth quantum circuit acting on a $L \times L'$ grid of qudits to the problem of simulating an associated "effective dynamics" in 1D on L qudits which is iterated for L' timesteps, or alternatively on L' qudits which is iterated for L timesteps. This mapping is rigorous and is related to previous maps from 2D quantum systems to 1D system evolving in time [109–111]. The effective 1D dynamics is then simulated using the time-evolving block decimation algorithm of Vidal [10]. By analogy, we call this algorithm space-evolving block decimation (`SEBD`). In Section 3.3.1, we specify the details of `SEBD` and rigorously bound the simulation error made by the algorithm in terms of quantities related to the entanglement spectra of the effective 1D dynamics and give conditions in which it is provably asymptotically efficient for sampling and estimating output probabilities with small error. `SEBD` is self-certifying in the sense that it can construct confidence intervals for its own simulation error and for the fraction of random circuit instances it can simulate. This makes numerically studying the algorithm's performance feasible, and is a crucial difference between `SEBD` and heuristics based on approximate tensor network contractions (e.g., [93]) in which the error incurred by truncating bonds of the tensor network cannot be directly related to operational measures such as trace-distance error.

A 1D unitary quantum circuit on L qubits iterated for L^c timesteps with $c > 0$ is generally hard to simulate classically in poly(L)-time, as the entanglement across any cut can increase linearly in time. However, the form of 1D dynamics that a shallow circuit maps to includes measurements as well as unitary gates. While the unitary gates tend to build entanglement, the measurements tend to destroy entanglement and make classical simulation more tractable. It is *a priori* unclear which effect has more influence.

Fortunately, unitary-and-measurement processes have been studied in a flurry of recent papers from the physics community [44–46, 48, 57–80]. The consensus from this work is that processes consisting of entanglement-creating unitary evolution interspersed with entanglement-destroying measurements can be in one of two phases, where the entanglement entropy equilibrates to either an area law (constant), or to a volume law (extensive). When we vary parameters like the fraction of qudits measured between each round of unitary evolution, a phase transition is observed. The existence of a phase transition appears to be robust to variations in the exact model, such as replacing projective measurements on a fraction of the qudits with weak measurements on all of the qudits [46, 60], or replacing Haar-random unitary evolution with Clifford [46, 57, 61, 62] or Floquet [57, 59] evolution. This suggests that the efficiency of the `SEBD` algorithm depends on whether the particular circuit depth and

architecture being simulated yields effective 1D dynamics that falls within the area-law or the volume-law regime. It also suggests a computational phase transition in the complexity of the SEBD algorithm. Essentially, decreasing the measurement strength or increasing the qudit dimension in these models is associated with moving toward a transition into the volume-law phase. Since increasing the 2D circuit depth is associated with decreasing the measurement strength and increasing the local dimension of the associated effective 1D dynamics, this already gives substantial evidence in favor of a computational phase transition in SEBD.

SEBD is inefficient if the effective 1D dynamics are on the volume-law side of the transition, and we expect it to be efficient on the area-law side because, in practice, dynamics obeying an area law for the von Neumann entanglement entropy are generally efficiently simulable. However, technically SEBD (like matrix-product states) requires that Rényi entropy S_0 to be bounded, not the von Neumann entropy S, and $S_0 \geq S$. Indeed, there are known contrived examples of states where $S_0 \gg S$, so that they obey an area law but cannot be efficiently simulated with matrix product states [112]. We address this concern by directly studying the entanglement spectrum of unitary-and-measurement processes in the area-law phase. To do this, we introduce a toy model for such dynamics which may be of independent interest. For this model, discussed more in Section 3.3.4, we rigorously derive an asymptotic scaling of Schmidt values across some cut as $\lambda_i \propto \exp\big(-\Theta(\log^2 i)\big)$ which is consistent with the scaling observed in our numerical simulations. Moreover, for this toy model we show that with probability at least $1-\delta$, the equilibrium state after iterating the process can be ε-approximated by a state with Schmidt rank $r \leq \exp\Big(O(\sqrt{\log(n/\varepsilon\delta)})\Big)$. Taking this toy model analysis as evidence that the bond dimension of SEBD when simulating a circuit whose effective 1D dynamics is in an area-law phase obeys this asymptotic scaling leads to Conjecture 3.1′.

3.3.1 Specification of algorithm

In this section, we assume that the reader is familiar with standard tensor network methods, particularly algorithms for manipulating matrix product states (see e.g. [12, 113] for reviews).

For concreteness, we consider a rectangular grid of qudits with local Hilbert space dimension q, although the algorithm could be similarly defined for different lattices. Assume WLOG that the grid consists of $n = L_1 \times L_2$ qudits, where L_1 is the number of rows, L_2 is the number of columns, and $L_1 \leq L_2$. For each qudit, let $|i\rangle, i \in [q] = \{0, 1, \ldots, q-1\}$ label a set of basis states which together form the computational basis. Assume that all gates act on one site or two neighboring sites, and the starting state is $|0\rangle^{\otimes n}$. Let d denote the circuit depth, which should be regarded as a constant. For a fixed circuit instance C, the goal is to sample from a distribution close to $\mathcal{D}_C$,

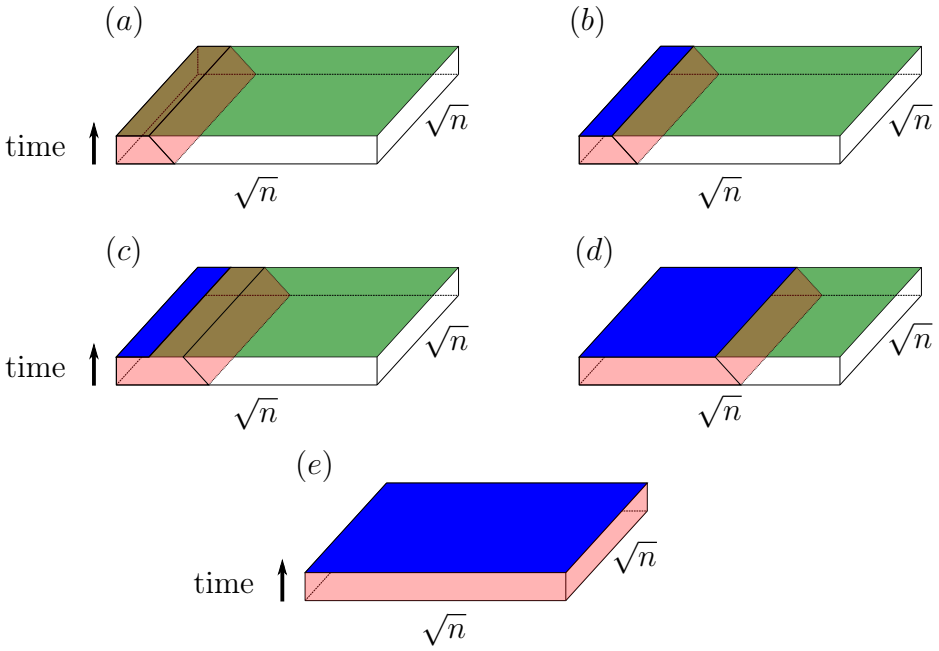


Figure 3.1: Schematic depiction of `SEBD` simulating a shallow 2D circuit. In all figures, the 2D circuit is depicted as a spacetime volume, with time flowing upwards. The blue regions correspond to sites for which measurements have been simulated, while green regions correspond to unmeasured sites. In (a), we apply all gates in the lightcone of `column 1`, namely, those gates intersecting the spacetime volume shaded red. In (b), we simulate the computational basis measurement of `column 1`. In (c), we apply all gates in the lightcone of `column 2` that were previously unperformed. Figure (d) depicts the algorithm at an intermediate stage of the simulation, after the measurements of about half of the qudits have been simulated. The algorithm stores the current state as an MPS at all times, which may be periodically compressed to improve efficiency. Figure (e) depicts the algorithm at completion: the measurements of all n of the qudits have been simulated.

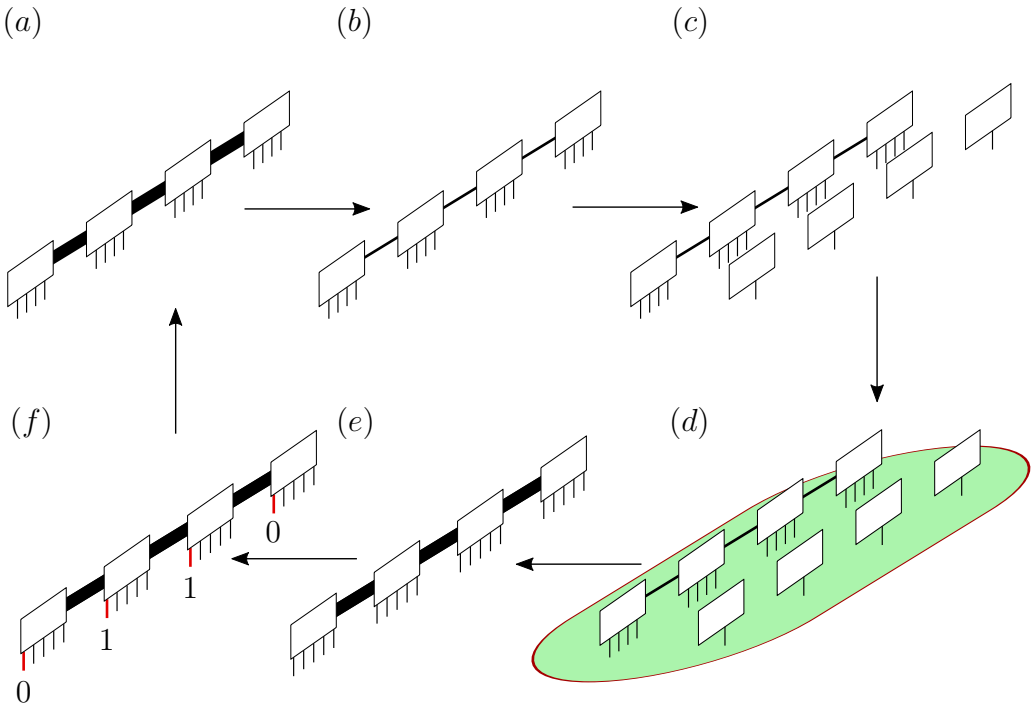


Figure 3.2: Iteration of `SEBD`. In (a), we begin with an MPS describing the current state ρ_j. In (b), the MPS is compressed via truncation of small Schmidt values. This will generally decrease the bond dimension of the MPS, depicted in the cartoon by a reduction in the thickness of the lines between tensors. In (c), qudits acted on by V_j that are not already incorporated into the current state are added to the MPS (increasing the physical bond dimension of the MPS) and initialized in $|0\rangle$ states. In (d), the unitary gates associated with V_j are applied. Figure (e) depicts the MPS after the application of V_j; the virtual bond dimension generally is increased by the application of V_j. In (f), the measurement of `column j` is performed, and the outcome `0110` is obtained. Subsequently, `column j` is projected onto the outcome `0110`, removing the physical legs associated with these sites from the MPS. The resulting state is ρ_{j+1}.

defined to be the distribution of the output of C upon measuring all qudits in the computational basis. For an output string $\mathbf{x} \in [q]^n$, we let $\mathcal{D}_C(\mathbf{x})$ denote the probability of the circuit outputting $\mathbf{x}$ after measurement. The high-level behavior of the algorithm is illustrated in Figure 3.1. Recall that C can always be *exactly* simulated in time $L_2 q^{\Theta(dL_1)}$ using standard tensor network algorithms [52].

Since all of the single-qudit measurements commute, we can measure the qudits in any order. In particular, we can first measure all of the sites in `column 1`, then those in `column 2`, and iterate until we have measured all L_2 columns. This is the measurement order we will take. Now, consider the first step in which we measure `column 1`. Instead of applying all of the gates of the circuit and then measuring, we may instead apply only the gates in the *lightcone* of `column 1`, that is, the gates that are causally connected to the measurements in `column 1`. We may ignore qudits that are outside the lightcone, by which we mean qudits that are outside the support of all gates in the lightcone.

Let $\rho_1 = |0\rangle\langle 0|^{\otimes L_1}$ denote the trivial starting state that is a tensor product of $|0\rangle$ states in `column 1`, which the algorithm represents as an MPS. Let V_1 denote the isometry corresponding to applying all gates in the lightcone of this column. The algorithm simulates the application of V_1 by adding qudits in the lightcone of `column 1` as necessary and applying the associated unitary gates, maintaining the description of the state as an MPS of length L_1 as illustrated in Figure 3.2. Since there are up to $d+1$ columns in the lightcone of `column 1`, each tensor of the MPS after the application of V_1 has up to $d+1$ dangling legs corresponding to physical indices, for a total physical dimension of at most q^{d+1}. Since in the application of V_1, there are up to $O(d^2)$ gates that act between any two neighboring rows, the (virtual) bond dimension of the updated MPS is at most $q^{O(d^2)}$.

We now simulate the computational basis measurement of `column 1`. More precisely, we measure the qudits of `column 1` one by one. We first compute the respective probabilities $p_1, p_2, \ldots, p_q$ of the q possible measurement outcomes for the first qudit. This involves contracting the MPS encoding $V_1 \rho_1 V_1^\dagger$. We now use these probabilities to classically sample an outcome $i \in [q]$, and update the MPS to condition on this outcome. That is, if (say) we obtain outcome `1` for site i, we apply the projector $|0\rangle\langle 0|$ to site i of the state and subsequently renormalize. After doing this for every qudit in the column, we have exactly sampled an output string $\mathbf{x}_1 \in [q]^{L_1}$ from the marginal distribution on `column 1`, and are left with an MPS description of the pure, normalized, post-measurement state ρ_2 proportional to $\mathrm{tr}_{\texttt{column 1}}\left(\Pi_1^{\mathbf{x}} V_1 \rho_1 V_1^\dagger \Pi_1^{\mathbf{x}}\right)$, where $\Pi_1^{\mathbf{x}}$ denotes the projection of `column 1` onto the sampled output string $\mathbf{x} = \mathbf{x}_1$. Using standard tensor network algorithms, the time complexity of these steps is $L_1 q^{O(d^2)}$.

We next consider `column 2`. At this point, we add the qudits and apply the gates that are in the lightcone of `column 2` but were not applied previously. Denote this isometry by V_2. It is straightforward to see that this step respects causality. That is, if some gate U is in the lightcone of `column 1`, then any gate W that is in the lightcone of `column 2` but not `column 1` cannot be required to be applied before U, because if it were, then it would be in the lightcone of `column 1`. Hence, when we apply gates in this step, we never apply a gate that was required to be applied before some gate that was applied in the first step. After this step, we have applied all gates in the lightcone of `columns (1, 2)`, and we have also projected `column 1` onto the measurement outcomes we observed.

By simulating the measurements of `column 2` in a similar way to those of `column 1`, we sample a string $\mathbf{x}_2$ from the marginal distribution on `column 2`, conditioned on the previously observed outcomes from `column 1`. Each time an isometry V_j is applied, the bond dimension of the MPS representation of the current state will in general increase by a multiplicative factor. In particular, if we iterate this procedure to simulate the entire lattice, we will eventually encounter a maximal bond dimension of up to $q^{O(dL_1)}$ and will obtain a sample $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{L_2}) \in [q]^n$ from the true output distribution.

To improve the efficiency at the expense of accuracy, we may compress the MPS in each iteration to one with smaller bond dimension using standard MPS compression algorithms. In particular, in each iteration j before we apply the corresponding isometry V_j, we first discard as many of the smallest singular values (i.e. Schmidt values) associated with each cut of the MPS as possible up to a total *truncation error* per bond of ϵ, defined as the sum of the squares of the discarded singular values. The bond dimension across any cut is reduced by the number of discarded values. This truncation introduces some error that we quantify below.

If the maximal bond dimension of this truncated version of the simulation algorithm is D, the total runtime of the full algorithm to obtain a sample is bounded by (taking q and d to be constants) $O(nD^3)$ using standard MPS compression algorithms.

We assume that for a specified maximal bond dimension D and truncation error per bond ϵ, if a bond dimension ever exceeds D then the algorithm terminates and outputs a failure flag FAIL. Hence, the runtime of the algorithm when simulating some circuit C with parameters ϵ and D is bounded by $O(nD^3)$, and the algorithm has some probability of failure $p_{f,C}$. We summarize the `SEBD` algorithm in Algorithm 1.

Algorithm 1 SEBD

Input: circuit instance C, truncation error ϵ, bond dimension cutoff D
Output: string $\mathbf{x} \in [q]^n$ or FAIL
Runtime: $O(nD^3)$ [q and d assumed to be constants]

initialize an MPS in the state $|0\rangle\langle 0|^{\otimes L_1}$, corresponding to `column 1`
for $t = 1 \dots L_2$ **do**
    compress MPS describing state by truncating small singular values, up to error ϵ per bond
    apply V_t, corresponding to gates in the lightcone of not yet applied
    if some bond dimension is greater than D, terminate and output FAIL
    simulate measurement of all qudits in `column` t via MPS contraction and sampling
    apply $\Pi_t^{\mathbf{x}_t}$ to condition on measurement string $\mathbf{x}_t$ observed for that column
return $(\mathbf{x}_1, \dots, \mathbf{x}_{L_2}) \in [q]^n$

The untruncated version of the algorithm presented above samples from the true distribution $\mathcal{D}_C$ of the measurement outcomes of the original 2D circuit C. However, due to the MPS compression which we perform in each iteration and the possibility of failure, the algorithm incurs some error which causes it to instead sample from some distribution $\mathcal{D}'_C$. Here, we bound the total variation distance between these distributions, given by

$$\frac{1}{2}\|\mathcal{D}'_C - \mathcal{D}_C\|_1 = \frac{1}{2}\sum_{\mathbf{x}} |\mathcal{D}'_C(\mathbf{x}) - \mathcal{D}_C(\mathbf{x})| + \frac{1}{2}p_{f,C}, \tag{3.3}$$

where the sum runs over the q^n possible output strings (not including FAIL), in terms of the truncation error made by the algorithm.

We first obtain a very general bound on the error made by `SEBD` with no bond dimension cutoff in terms of the truncation error. Note that the truncation error may depend on the (random) measurement outcomes, and is itself therefore a random variable. See Appendix 3.E for a proof.

Lemma 3.1. *Let ϵ_i denote the sum of the squares of all singular values discarded in the compression during iteration i of the simulation of a circuit C with output distribution $\mathcal{D}_C$ by `SEBD` with no bond dimension cutoff, and let Λ denote the sum of all singular values discarded over the course of the algorithm. Then the distribution $\mathcal{D}'_C$ sampled from by `SEBD` satisfies*

$$\frac{1}{2}\|\mathcal{D}'_C - \mathcal{D}_C\|_1 \leq \mathbb{E}\sum_{i=1}^{L_2}\sqrt{2\epsilon_i} \leq \sqrt{2}\,\mathbb{E}\,\Lambda, \tag{3.4}$$

where the expectations are over the random measurement outcomes.

From Lemma 3.1, we immediately obtain two corollaries. The first is useful for empirically bounding the sampling error in total variation distance made by SEBD when the algorithm also has a bond dimension cutoff. The second is a useful asymptotic statement. The corollaries follow straightforwardly from the coupling formulation of variational distance, Markov's inequality, and the triangle inequality.

Corollary 3.1. *Let $\mathcal{A}$ denote a SEBD algorithm with truncation error parameter ϵ and bond dimension cutoff D. Consider a fixed circuit C, and suppose that $\mathcal{A}$ applied to this circuit fails with probability $p_{f,C}$. Then $\mathcal{A}$ samples from the output distribution of C with total variation distance error bounded by $L_2\sqrt{2\epsilon L_1} + p_{f,C}$.*

If the failure probability of $\mathcal{A}$ averaged over random choice of circuit instance and measurement outcome is p_f, then for any δ, on at least $1-\delta$ fraction of circuit instances, $\mathcal{A}$ samples from the true output distribution with total variation distance error bounded by $L_2\sqrt{2\epsilon L_1} + p_f/\delta$.

In practice, the variational distance error of SEBD with truncation error ϵ applied to the simulation of some circuit C can be bounded by constructing a confidence interval for $p_{f,C}$ and applying the above bound.

Corollary 3.2. *Let $\mathcal{A}$ denote a SEBD algorithm with truncation error parameter ϵ and no bond dimension cutoff. Suppose that, for some random circuit family with $q = O(1)$ and $d = O(1)$, the expected bond dimension across any cut is bounded by $\mathrm{poly}(n, 1/\epsilon)$. Then, SEBD with some choice of $\epsilon = 1/\mathrm{poly}(n)$ and $D = \mathrm{poly}(n)$ runs in time $\mathrm{poly}(n, 1/\varepsilon, 1/\delta)$ and, with probability at least $1-\delta$ over the choice of circuit instance C, samples from the output distribution of C with variational distance error less than ε.*

Thus, to prove the part of Conjecture 3.1 about sampling up to total variation distance error ε for uniform random circuit families, it would suffice to show that there is a 2D constant-depth uniform random quantum circuit family with the worst-case-hard property for which the expected bond dimension across any cut while running SEBD with truncation parameter ϵ is bounded by $\mathrm{poly}(n, 1/\epsilon)$. Later, we will introduce two candidate circuit families for which we can give numerical and analytical evidence that this criterion is indeed met.

In the next subsection, we show how the other part of Conjecture 3.1, regarding computing output probabilities, would also follow from a $\mathrm{poly}(n, 1/\epsilon)$ bound on the bond dimension of states encountered by SEBD on uniform worst-case-hard circuit families.

3.3.2 Computing output probabilities with SEBD

In the previous section, we described how a SEBD algorithm with a truncation error parameter ϵ and a bond dimension cutoff D applied to a circuit

C samples from a distribution $\mathcal{D}'_C$ satisfying $\|\mathcal{D}'_C - \mathcal{D}_C\|_1 \leq 2L_2\sqrt{2\epsilon L_1} + 2p_{f,C}$ where $p_{f,C}$ is the probability that some bond dimension exceeds D and the algorithm terminates and indicates failure. Expanding the expression for the 1-norm and rearranging, we have

$$\frac{1}{q^n}\sum_{\mathbf{x}} |\mathcal{D}'_C(\mathbf{x}) - \mathcal{D}_C(\mathbf{x})| \leq \frac{2L_2\sqrt{2\epsilon L_1} + p_{f,C}}{q^n}. \tag{3.5}$$

`SEBD` with bond dimension cutoff D can be used to compute $\mathcal{D}'_C(\mathbf{x})$ for any output string $\mathbf{x}$ in time $O(nD^3)$ (taking q and d to be constants). To do this, for a fixed output string $\mathbf{x}$, `SEBD` proceeds similarly to the case in which it is being used for sampling, but rather than sampling from the output distribution of some column, it simply projects that column onto the outcome specified by the string $\mathbf{x}$, and computes the conditional probability of that outcome via contraction of the MPS. That is, at iteration t, the algorithm computes the conditional probability of measuring the string $\mathbf{x}_t \in [q]^{L_1}$ in `column` t, $\mathcal{D}'_C(\mathbf{x}_t|\mathbf{x}_1, \ldots, \mathbf{x}_{t-1})$, by projecting `column` t onto the relevant string via the projector $\Pi_t^{\mathbf{x}_t}$ and then contracting the relevant MPS. If the bond dimension ever exceeds D, then it must hold that $\mathcal{D}'_C(\mathbf{x}) = 0$, and so the algorithm outputs zero and terminates. Otherwise, the algorithm outputs $\mathcal{D}'_C(\mathbf{x}) = \prod_{t=1}^{L_2} \mathcal{D}'_C(\mathbf{x}_t|\mathbf{x}_1, \ldots, \mathbf{x}_{t-1})$. We summarize this procedure in Algorithm 2.

Algorithm 2 `SEBD` for computing output probabilities

Input: circuit instance C, truncation error ϵ, bond dimension cutoff D, string $\mathbf{x} \in [q]^n$

Output: $\mathcal{D}'_C(\mathbf{x})$

Runtime: $O(nD^3)$ [q and d assumed to be constants]

1: initialize an MPS in the state $|0\rangle\langle 0|^{\otimes L_1}$, corresponding to `column 1`
2: **for** $t = 1 \ldots L_2$ **do**
3: compress MPS describing state by truncating small singular values, up to error ϵ per bond
4: apply V_t, corresponding to gates in the lightcone of `column` t not yet applied
5: if some bond dimension is greater than D, terminate and output zero
6: apply $\Pi_t^{\mathbf{x}_t}$ to condition on string $\mathbf{x}_t$
7: compute $\mathcal{D}'_C(\mathbf{x}_t|\mathbf{x}_1, \ldots, \mathbf{x}_{t-1})$ via MPS contraction
return $\mathcal{D}'_C(\mathbf{x}) = \prod_{t=1}^{L_2} \mathcal{D}'_C(\mathbf{x}_t|\mathbf{x}_1, \ldots, \mathbf{x}_{t-1})$

We have therefore shown the following:

Lemma 3.2. *Let $p_{f,C}$ be the failure probability of `SEBD` when used to simulate a circuit instance C with truncation error parameter ϵ and bond dimension cutoff D. Suppose $\mathbf{x} \in [q]^n$ is an output string drawn uniformly at random.*

Then Algorithm 2 outputs a number $\mathcal{D}'_C(\mathbf{x})$ satisfying

$$\mathop{\mathbb{E}}_{\mathbf{x}} |\mathcal{D}'_C(\mathbf{x}) - \mathcal{D}_C(\mathbf{x})| \leq \frac{2L_2\sqrt{2\epsilon L_1} + p_{f,C}}{q^n}. \tag{3.6}$$

The above lemma bounds the expected error incurred while estimating a uniformly random output probability for a fixed circuit instance C. We may use this lemma to straightforwardly bound the expected error incurred while estimating the probability of a fixed output string over a distribution of random circuit instances. The corollary is applicable if the distribution of circuit instances has the property of being invariant under an application of a final layer of arbitrary single-qudit gates. This includes circuits in which all gates are Haar-random (as long as every qudit is acted on by some gate), but is more general. In particular, any circuit distribution in which the final gate to act on any given qudit is Haar-random satisfies this property. This fact will be relevant in subsequent sections.

Corollary 3.3. *Let p_f be the failure probability of `SEBD` when used to simulate a random circuit instance C with truncation error parameter ϵ and bond dimension cutoff D, where C is drawn from a distribution that is invariant under application of a final layer of arbitrary single-qudit gates. Then for any fixed string $\mathbf{x} \in [q]^n$, the output of Algorithm 2 satisfies*

$$\mathop{\mathbb{E}}_{C} |\mathcal{D}'_C(\mathbf{x}) - \mathcal{D}_C(\mathbf{x})| \leq \frac{2L_2\sqrt{2\epsilon L_1} + p_f}{q^n}. \tag{3.7}$$

Proof. Averaging the bound of Eq. (3.6) over random circuit instances, we have

$$\mathop{\mathbb{E}}_{\mathbf{y}} \mathop{\mathbb{E}}_{C} |\mathcal{D}'_C(\mathbf{y}) - \mathcal{D}_C(\mathbf{y})| \leq \frac{2L_2\sqrt{2\epsilon L_1} + p_f}{q^n}. \tag{3.8}$$

Let $L_{\mathbf{y}}$ denote a layer of single-qudit gates with the property that $L_{\mathbf{y}}|\mathbf{x}\rangle = |\mathbf{y}\rangle$. By assumption, C is distributed identically to the composition of C with $L_{\mathbf{y}}$, denoted $L_{\mathbf{y}} \circ C$. Together with the observation that $\mathcal{D}_{L_{\mathbf{y}}\circ C}(\mathbf{y}) = \mathcal{D}_C(\mathbf{x})$, we have

$$\mathop{\mathbb{E}}_{\mathbf{y}} \mathop{\mathbb{E}}_{C} |\mathcal{D}'_C(\mathbf{y}) - \mathcal{D}_C(\mathbf{y})| = \mathop{\mathbb{E}}_{\mathbf{y}} \mathop{\mathbb{E}}_{C} |\mathcal{D}'_{L_{\mathbf{y}}\circ C}(\mathbf{y}) - \mathcal{D}_{L_{\mathbf{y}}\circ C}(\mathbf{y})| \tag{3.9}$$

$$= \mathop{\mathbb{E}}_{C} |\mathcal{D}'_C(\mathbf{x}) - \mathcal{D}_C(\mathbf{x})|, \tag{3.10}$$

from which the result follows. □

The following asymptotic statement follows straightforwardly.

Corollary 3.4. *Let $\mathcal{A}$ denote a `SEBD` algorithm with truncation error parameter ϵ and no bond dimension cutoff. Suppose that, for some random circuit family with $q = O(1)$ and $d = O(1)$, the expected bond dimension across any*

cut is bounded by $\mathrm{poly}(n, 1/\epsilon)$*. Then,* ***SEBD*** *with some choice of* $\epsilon = 1/\mathrm{poly}(n)$ *and* $D = \mathrm{poly}(n)$ *runs in time* $\mathrm{poly}(n, 1/\varepsilon, 1/\delta)$ *and, with probability at least* $1 - \delta$ *over the choice of circuit instance* C*, estimates* $\mathcal{D}_C(\mathbf{x})$ *for some fixed* $\mathbf{x} \in [q]^n$ *up to additive error bounded by* ε/q^n*.*

Corollary 3.4 shows how the part of Conjecture 3.1 about computing arbitrary output probabilities to error ε/q^n would follow from a bound on the bond dimension across any cut when SEBD runs on a uniform worst-case-hard circuit family.

3.3.3 Example: SEBD applied to cluster state with Haar-random measurements (CHR)

To illustrate the connection between SEBD and random unitary and measurement dynamics, we now study the SEBD algorithm in more detail for a simple uniform family of 2D random circuits that possesses the worst-case-hard property required by Conjecture 3.1. The model we consider is the following: start with a 2D cluster state of n qubits arranged in a $\sqrt{n} \times \sqrt{n}$ grid, apply a single-qubit Haar-random gate to each qubit, and then measure all qubits in the computational basis. Recall that a cluster state may be created by starting with the product state $|+\rangle^{\otimes n}$ before applying CZ gates between all adjacent sites. An equivalent formulation which we will find convenient in the subsequent section is to measure each qubit of the cluster state in a Haar-random basis. We refer to this model as CHR, for "cluster state with Haar-random measurements."

Following Ref. [82], it is straightforward to show that sampling from the output distribution of CHR is classically *worst-case hard* assuming the polynomial hierarchy (PH) does not collapse to the third level. It can also be readily shown, following Refs. [26, 27], that near-exactly computing output probabilities of CHR is #P-hard in the average case. These results rule out, under standard conjectures, the existence of a classical sampling algorithm for CHR that succeeds for all instances, or a classical algorithm for efficiently computing most output probabilities of CHR near-exactly. A natural question is then whether efficient approximate average-case versions of these algorithms may exist. We formalize these questions as the problems $\mathtt{CHR}_{\pm}^{\mathtt{samp/prob}}$.

Problem 3.1 ($\mathtt{CHR}_{\pm}^{\mathtt{samp/prob}}$)**.** *Given as input a random instance* C *of* ***CHR*** *(specified by a sidelength* $\sqrt{n}$ *and a set of* n *single-qubit Haar-random gates applied to the* $\sqrt{n} \times \sqrt{n}$ *cluster state) and error parameters* ε *and* δ*, perform the following computational task in time* $\mathrm{poly}(n, 1/\varepsilon, 1/\delta)$*.*

- $\mathtt{CHR}_{\pm}^{\mathtt{samp}}$*. Sample from a distribution* $\mathcal{D}'_C$ *that is* ε*-close in total variation distance to the true output distribution* $\mathcal{D}_C$ *of circuit* C*, with probability of success at least* $1 - \delta$ *over the choice of measurement bases.*

- $\texttt{CHR}_{\pm}^{\texttt{prob}}$. *Estimate $\mathcal{D}_C(\mathbf{0})$, the probability of obtaining the all-zeros string upon measuring the output state of C in the computational basis, up to additive error at most $\varepsilon/2^n$, with probability of success at least $1-\delta$ over the choice of measurement bases.*

We now show that SEBD solves $\texttt{CHR}_{\pm}^{\texttt{samp/prob}}$ if a certain form of 1D dynamics involving local unitary gates and measurements is classically simulable. We first consider the sampling variant of SEBD. Specializing to the CHR model, the algorithm takes on a particularly simple form due to the fact that the cluster state is built by applying CZ gates between all neighboring pairs of qubits, which are initialized in $|+\rangle$ states. Due to this structure, the radius of the lightcone for this model is simply one. In particular, the only gates in the lightcone of `columns 1-j` are the Haar-random single-qubit gates acting on qubits in these columns, as well as CZ gates that act on at least one qubit within these columns. This permits a simple prescription for SEBD applied to this problem.

Initialize the simulation algorithm in the state $\rho_1 = |+\rangle\langle+|^{\otimes\sqrt{n}}$ corresponding to `column 1`. To implement the isometry V_1, initialize the qubits of `column 2` in the state $|+\rangle\langle+|^{\otimes\sqrt{n}}$ and apply CZ gates between adjacent qubits that are both in `column 1` and between adjacent qubits in separate columns. Now, measure the qubits of `column 1` in the specified Haar-random bases (equivalently, apply the specified Haar-random gates and measure in the computational basis), inducing a pure state ρ_2 with support in `column 2`. Iterating this process, we progress through a random sequence of 1D states on $\sqrt{n}$ qubits $\rho_1 \to \rho_2 \to \cdots \to \rho_{\sqrt{n}}$ which we will see can be equivalently understood as arising from a 1D dynamical process consisting of alternating layers of random unitary gates and weak measurements.

It will be helpful to introduce notation. Define $|\theta,\phi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$. In other words, let $|\theta,\phi\rangle$ denote the single-qubit pure state with polar angle θ and azimuthal angle ϕ on the Bloch sphere. Let $\theta_i^{(t)}$ and $\phi_i^{(t)}$ specify the measurement basis of the qubit in row i and column t; that is, the projective measurement on the qubit in row i and column t is $\{\Pi^0_{\theta_i^{(t)},\phi_i^{(t)}}, \Pi^1_{\theta_i^{(t)},\phi_i^{(t)}}\}$ with $\Pi^0_{\theta_i^{(t)},\phi_i^{(t)}} = \left|\theta_i^{(t)},\phi_i^{(t)}\right\rangle\!\left\langle\theta_i^{(t)},\phi_i^{(t)}\right|$ and $\Pi^1_{\theta_i^{(t)},\phi_i^{(t)}} = I - \Pi^0_{\theta_i^{(t)},\phi_i^{(t)}}$. We also define

$$M_0(\theta,\phi) = \begin{pmatrix} \cos(\theta/2) & 0 \\ 0 & e^{-i\phi}\sin(\theta/2) \end{pmatrix} \tag{3.11a}$$

$$M_1(\theta,\phi) = \begin{pmatrix} \sin(\theta/2) & 0 \\ 0 & e^{i\phi}\cos(\theta/2) \end{pmatrix}. \tag{3.11b}$$

Note that $\{M_0(\theta,\phi), M_1(\theta,\phi)\}$ defines a weak single-qubit measurement. We now describe, in Algorithm 3, a 1D process which we claim produces a sequence

of states identical to that encountered by SEBD for the same choice of measurement bases and measurement outcomes, and also has the same measurement statistics.

Algorithm 3 Effective 1D dynamics of a fixed instance of CHR

$\varphi_1 \leftarrow |+\rangle\langle+|^{\otimes\sqrt{n}}$.
for $t = 1 \ldots \sqrt{n} - 1$ **do**
    apply a CZ gate between every adjacent pair of qubits
    measure $\{M_0(\theta_i^{(t)}, \phi_i^{(t)}), M_1(\theta_i^{(t)}, \phi_i^{(t)})\}$ on qubit i, obtaining $X_i^{(t)}$, for $i \in \{1, \ldots, \sqrt{n}\}$
    apply a Hadamard transform
    $\varphi_{t+1} \leftarrow$ resulting state
measure $\{\Pi^0_{\theta_i^{(\sqrt{n})}, \phi_i^{(\sqrt{n})}}, \Pi^1_{\theta_i^{(\sqrt{n})}, \phi_i^{(\sqrt{n})}}\}$ on qubit i, obtaining $X_i^{(\sqrt{n})}$, for $i \in \{1, \ldots, \sqrt{n}\}$

Lemma 3.3. *For a fixed choice of $\{\theta_i^{(t)}, \phi_i^{(t)}\}$ parameters, the joint distribution of outcomes $\{X_i^{(t)}\}_{i,t}$ is identical to that of $\{Y_i^{(t)}\}_{i,t}$, where $\{Y_i^{(t)}\}_{i,t}$ are the measurement outcomes obtained upon measuring all qubits of a $\sqrt{n} \times \sqrt{n}$ cluster state, with the measurement on the qubit in row i and column t being $\{\Pi^0_{\theta_i^{(t)}, \phi_i^{(t)}}, \Pi^1_{\theta_i^{(t)}, \phi_i^{(t)}}\}$. Furthermore, for any fixed choice of measurement outcomes, $\varphi_j = \rho_j$ for all $j \in \{1, \ldots, \sqrt{n}\}$, where ρ_j is the state at the beginning of iteration j of the SEBD algorithm.*

Proof. The lemma follows from the above description of the behavior of SEBD applied to CHR, as well as the following identities holding for any single-qubit state $|\xi\rangle$ which may be verified by straightforward calculation:

$$(\Pi^0_{\theta,\phi} \otimes I)CZ(|\xi\rangle \otimes |+\rangle) = |\theta, \phi\rangle \otimes HM_0(\theta, \phi)|\xi\rangle \tag{3.12}$$

$$(\Pi^1_{\theta,\phi} \otimes I)CZ(|\xi\rangle \otimes |+\rangle) = |\pi - \theta, -\phi\rangle \otimes HM_1(\theta, \phi)|\xi\rangle. \tag{3.13}$$

□

We have seen that, for a fixed choice of single-qubit measurement bases $\{\theta_j^{(t)}, \phi_j^{(t)}\}_{t,j}$ associated with an instance C, we can define an associated 1D process consisting of alternating layers of single-qubit weak measurements and local unitary gates, such that simulating this 1D process is sufficient for sampling from $\mathcal{D}_C$.

Now, recall that in the context of simulating CHR, each single-qubit measurement basis is chosen randomly according to the Haar measure. That is, the Bloch sphere angles $(\theta_i^{(t)}, \phi_i^{(t)})$ are Haar-distributed. If we define $x_i^{(t)} \equiv \cos\theta_i^{(t)}$, we find that $x_i^{(t)}$ is uniformly distributed on the interval $[-1, 1]$. The parameters $\phi_i^{(t)}$ are uniformly distributed on $[0, 2\pi]$. Using these observations, as well

as the observation that the outcome probabilities of the measurement of qubit i in iteration t are independent of the azimuthal angle $\phi_i^{(t)}$ when $t < \sqrt{n}$, we may derive effective dynamics of a random instance.

Define the operators

$$N(x) = \begin{pmatrix} \sqrt{\frac{1+x}{2}} & 0 \\ 0 & \sqrt{\frac{1-x}{2}} \end{pmatrix}, \quad x \in [-1, 1].$$

Note that $\{N(x), N(-x)\}$ defines a weak measurement. Also, define the phase gate

$$P(\phi) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}, \quad \phi \in [0, 2\pi].$$

By randomizing each single-qubit measurement basis according to the Haar distribution, one finds that the dynamics of Algorithm 3 (which applies for a fixed choice of measurement bases) may be written as Algorithm 4 below, where the notation $x \in_U [-1, 1]$ means that x is a random variable uniformly distributed on $[-1, 1]$. That is, the distribution of random sequences $\varphi_1 \to \varphi_2 \to \cdots \to \varphi_{\sqrt{n}}$ and distribution of output statistics produced by Algorithm 4 is identical to that produced by SEBD applied to CHR.

Algorithm 4 Effective 1D dynamics of CHR

$\varphi_1 \leftarrow |+\rangle\langle+|^{\otimes\sqrt{n}}$.
for $t = 1 \ldots \sqrt{n} - 1$ **do**
    apply a CZ gate between every adjacent pair of qubits
    for $i = 1 \ldots \sqrt{n}$ **do**
        measure $\{N(x), N(-x)\}$ on qubit i with $x \in_U [-1, 1]$
        apply the gate $P(\phi)$ with $\phi \in_U [0, 2\pi]$ to qubit i
    apply a Hadamard transform
    $\varphi_{t+1} \leftarrow$ resulting state
perform a projective measurement on each qubit in a Haar-random basis

Hence, if TEBD can efficiently simulate the process of Algorithm 4 with high probability, then SEBD can solve $\mathtt{CHR}_{\pm}^{\mathtt{samp}}$ and $\mathtt{CHR}_{\pm}^{\mathtt{prob}}$. We formalize this in the following lemma.

Lemma 3.4. *Suppose that TEBD can efficiently simulate the process described in Algorithm 4 in the sense that the expected bond dimension across any cut is bounded by* $\mathrm{poly}(n, 1/\epsilon)$ *where* ϵ *is the truncation error parameter. Then SEBD can be used to solve* $\mathtt{CHR}_{\pm}^{\mathtt{samp}}$ *and* $\mathtt{CHR}_{\pm}^{\mathtt{prob}}$.

Proof. Follows from Corollary 3.2, Corollary 3.4, and the equivalence to Algorithm 4 discussed above. □

We have shown how SEBD applied to CHR can be reinterpreted as TEBD applied to a 1D dynamical process involving alternating layers of random unitaries and weak measurements. Up until this point, there has been little reason to expect that SEBD is efficient for the simulation of CHR. In particular, with no truncation, the bond dimension of the MPS stored by the algorithm grows exponentially as the algorithm sweeps across the lattice.

We now invoke the findings of a number of related recent works [44–46, 48, 57–80]. to motivate the possibility that TEBD can efficiently simulate the effective 1D dynamics. These works study various 1D dynamical processes involving alternating layers of measurements and random local unitaries. In some cases, the measurements are considered to be projective and only occur with some probability p. In other cases, similarly to Algorithm 4, weak measurements are applied to each site with probability one. The common finding of these papers is that such models appear to exhibit an entanglement phase transition driven by measurement probability p (in the former case), or measurement strength (in the latter case). On one side of the transition, the entanglement entropy obeys an area law, scaling as $O(1)$ with the length L. On the other side, it obeys a volume law, scaling as $O(L)$.

Based on these works, one expects the entanglement dynamics to saturate to an area-law or volume-law phase. And in fact, our numerical studies (presented in Section 3.5) suggest that these dynamics saturate to an area-law phase. The common intuition that 1D quantum systems obeying an area law for the von Neumann entropy are easy to simulate with matrix product states therefore suggests that SEBD applied to this problem is efficient. While counterexamples to this common intuition are known [112], they are contrived and do not present an obvious obstruction for our algorithm. To better understand the relationship between maximal bond dimension and truncation error when the effective dynamics is in the area-law phase as well as rule out such counterexamples, in the following section we describe a toy model for a unitary-and-measurement process in the area-law phase, which predicts a superpolynomial decay of Schmidt values across any cut and therefore predicts that a polynomial runtime is sufficient to perform the simulation to $1/\operatorname{poly}(n)$ error. Our numerical results (presented in Section 3.5) suggest that the effective dynamics of the random circuit architectures we consider are indeed in the area-law phase, with entanglement spectra consistent with those predicted by the toy model dynamics. Further analytical evidence for efficiency is given in Section 3.6.

Note that, although we explicitly derived the effective 1D dynamics for the CHR model and observed it to be a simple unitary-and-measurement process, the interpretation of the effective 1D dynamics as a unitary-and-measurement process is not specific to CHR and is in fact general. In the general case, SEBD tracks $O(r)$ columns simultaneously where r is the radius of the lightcone corresponding to the circuit. In each iteration, new qudits that have come

into the lightcone are added, unitary gates that have come into the lightcone are performed, and finally projective measurements are performed on a single column of qudits. Similarly to the case of CHR, this entire procedure can be viewed as an application of unitary gates followed by weak measurements on a 1D chain of qudits of dimension $q^{O(r)}$. Intuitively, increasing the circuit depth corresponds both to increasing the local dimension in the effective 1D dynamics and decreasing the measurement strength. The former is due to the fact that in general the lightcone radius r will increase as depth is increased, and the local dimension of the effective dynamics is $q^{O(r)}$. The latter is due to the fact that as r increases, the number of tracked columns increases but the number of measured qudits in a single round stays constant. Hence the fraction of measured qudits decreases, and intuitively we expect this to correspond to a decrease in effective measurement strength. This intuition together with the findings of prior works on unitary-and-measurement dynamics suggests that the effective dynamics experiences an entanglement phase transition from an area-law to volume-law phase as q or d is increased, and therefore SEBD experiences a computational phase transition, supporting Conjecture 3.2. While this analogy is not perfect, we provide further analytical evidence in Section 3.6 that the effective 1D dynamics indeed undergoes such a phase transition.

3.3.4 Conjectured entanglement spectrum of unitary-and-measurement dynamics in an area-law phase

Numerical (Section 3.5) and analytical (Section 3.6) evidence suggests that the effective 1D dynamics corresponding to the uniform 2D shallow random circuit families we consider are in the area-law phase, making efficient simulation via SEBD very plausible. However, it is desirable to have clear predictions for the scaling of the entanglement spectra for states of the effective 1D dynamics, as this allows us to make concrete predictions for error scaling of SEBD and rule out (contrived) examples of states [112] which cannot be efficiently represented via MPS despite obeying an area law for the von Neumann entanglement entropy.

To this end, we study a simple toy model of how entanglement might scale in the area-law phase of a unitary-and-measurement circuit. Consider a chain of n qubits where we are interested in the entanglement across the cut between $1, \ldots, n/2$ and $n/2+1, \ldots, n$ (assume that n is even). We model the dynamics as follows. In each time step, we perform the following three steps:

1. Set the state of sites $n/2$ and $n/2+1$ to be an EPR pair $|\Phi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$.
2. Perform the cyclic permutations of qubits $(n/2, n/2-1, \ldots, 1)$ and $(n/2+1, n/2+2, \ldots, n)$. That is, move each qubit one step away from the central cut, except for qubits 1 and n, which are moved to $n/2$ and $n/2+1$, respectively.

3. Perform a weak measurement on each qubit with Kraus elements $M_0(\theta) = \cos(\theta/2)\,|0\rangle\langle 0| + \sin(\theta/2)\,|1\rangle\langle 1|$ and $M_1(\theta) = \sin(\theta/2)\,|0\rangle\langle 0| + \cos(\theta/2)\,|1\rangle\langle 1|$. This is based on Eq. (3.11), but the phases will not matter here, so we have dropped them for simplicity.

Without the measurements, this would create one EPR pair in each time step until the system had $n/2$ EPR pairs across the cut after time $n/2$. However, the measurements have the effect of reducing the entanglement. For this process, we derive the functional form of the asymptotic scaling of half-chain Schmidt coefficients $\lambda_1 \geq \lambda_2 \geq \cdots$. Moreover, bounds on the scaling of the entanglement spectrum allows us to derive a relation between the truncation error (sum of squares of discarded Schmidt values) ϵ incurred upon discarding small Schmidt values, and the rank r of the post-truncation state. The bounds are given in the following lemma, which is proved in Appendix 3.E.

Lemma 3.5. *Let $\lambda_1 \geq \lambda_2 \geq \cdots$ denote the half-chain Schmidt values after at least $n/2$ iterations of the toy model process. Then, with probability at least $1-\delta$, the half-chain Schmidt values indexed by $i \geq i^* = \exp\Big(\Theta(\sqrt{\log(n/\delta)})\Big)$ obey the asymptotic scaling*

$$\lambda_i \propto \exp\big(-\Theta(\log^2(i))\big)\,. \tag{3.14}$$

Furthermore, upon truncating the smallest Schmidt coefficients up to a truncation error of ϵ, with probability at least $1-\delta$, the half-chain Schmidt rank r of the post-truncation state obeys the scaling

$$r \leq \exp\Big(\Theta\Big(\sqrt{\log(n/\epsilon\delta)}\Big)\Big)\,. \tag{3.15}$$

This is the basis for our Conjecture 3.1′. More precisely, we take this analysis as evidence that the bond dimension D, truncation error ϵ, and system size n obey the scaling $D \leq \exp\Big(\Theta\Big(\sqrt{\log(n/\epsilon\delta)}\Big)\Big)$ with probability $1-\delta$ over random circuit instance and random measurement outcomes when `SEBD` simulates a random constant-depth 2D circuit whose effective 1D dynamics lie in the area-law phase. Recalling that the runtime of `SEBD` scales like $O(nD^3)$ for a maximal bond dimension of D and using the relationship between truncation error, failure probability, variational distance error, and simulable circuit fraction given in Corollary 3.1, we conclude that `SEBD` with a maximal bond dimension cutoff scaling as $\exp\Big(\Theta\Big(\sqrt{\log(n/\epsilon\delta)}\Big)\Big)$ runs in time $n^{1+o(1)}\exp\Big(\Theta\Big(\sqrt{\log(1/\varepsilon\delta)}\Big)\Big)$ and simulates $1-\delta$ fraction of random circuit instances up to variational distance error ε.

It is important to note what this heuristic argument leaves out. While a 1D unitary-and-measurement circuit will indeed create $O(1)$ ebits across any given cut in each round, these will not remain in the form of distinct pairs

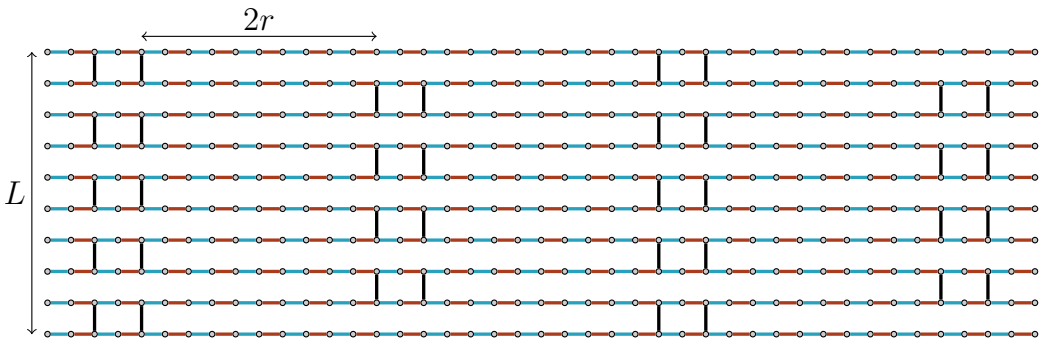


Figure 3.3: Extended brickwork architecture with n qubits. Here, circles represent qubits initialized in the state $|0\rangle^{\otimes n}$, turquoise lines represent the first layer of gates to act, red lines represent the second layer, and black lines represent the third and final layer. All gates are chosen Haar-randomly. We let $\texttt{Brickwork}(L, r, v)$ denote the corresponding random circuit with circuit layout depicted in the figure above with vertical sidelength L, "extension parameter" $2r$ (which gives the distance between vertical gates acting on adjacent pairs of rows), and number of pairs of columns of vertical gates v. In the above example, $r = 5$ and $v = 4$. The standard brickwork architecture corresponds to $r = 1$. Note that $n = \Theta(Lrv)$.

of qubits. The unitary dynamics *within* each side of the cut will have the effect of transforming the Schmidt bases into entangled ones. This will make the measurements less effective at reducing the entanglement, for reasons that can be understood in terms of quantum state merging [61, 114]. Another simplification of the toy model is that the measurement angle θ is taken to be a fixed constant rather than random. Finally, in the toy model, we assume for simplicity that the EPR pairs move cyclically. We expect that, if this effect is significant, it is more likely to make the toy model overly pessimistic compared with the real situation. Despite these simplifications, we believe this model is qualitatively accurate in the area-law phase. Indeed, the scaling of Schmidt values predicted by our toy model analysis is consistent with the scaling we find numerically in Figure 3.5.

3.4 Rigorous analysis of SEBD for the "extended brickwork architecture"

In this section, we show that `SEBD` is provably efficient for certain random circuit families that are worst-case hard. We define the circuit architecture in Figure 3.3. It follows readily from prior works that exactly sampling from the output distribution of this random circuit family for arbitrary circuit instances or near-exactly computing a specific output probability with high probability is classically hard under standard complexity theoretic assumptions. We summarize these observations in the following lemma.

Lemma 3.6. *Let $r(L)$ and $v(L)$ be any polynomially bounded functions, with $v(L) \geq L^a$ for some $a > 0$. Suppose that there exists a classical algorithm that runs in time* $\mathrm{poly}(n)$ *and samples from the output distribution of an arbitrary realization of* $\texttt{Brickwork}(L, r(L), v(L))$*, as defined in Figure 3.3. Then the polynomial hierarchy collapses to the third level. Suppose there exists a classical algorithm that runs in time* $\mathrm{poly}(n, 1/\delta)$ *and, for an arbitrary fixed output string* $\mathbf{x}$*, with probability at least* $1-\delta$ *over choice of random instance, computes the output probability of* $\mathbf{x}$ *up to additive error* $2^{-\tilde{\Theta}(n^2)}$*. Then there exists a probabilistic polynomial-time algorithm for computing a* #P*-hard function.*

Proof. We first note that $\texttt{Brickwork}(L, r(L), v(L))$ supports universal MBQC, in the sense that a specific choice of gates can create a resource state that is universal for MBQC. This is an immediate consequence of the proof of universality of the "standard" brickwork architecture (corresponding to $r = 1$) proved in [108]. Indeed, when using the extended brickwork architecture for MBQC, measurements on the long 1D stretches of length $2r$ may be chosen such that the effective state is simply teleported to the end when computing from left to right (i.e., measurements may be chosen such that the long 1D segments simply amount to applications of identity gates on the effective state). The scaling $v \geq L^a$ ensures that MBQC with an extended brickwork resource state suffices to simulate any BQP computation with polynomial overhead. Since a specific choice of gates creates a resource state for universal MBQC, an algorithm that can simulate an arbitrary circuit realization can be used to simulate arbitrary single-qubit measurements on a resource state universal for MBQC. Under post-selection, such an algorithm can therefore simulate PostBQP [109] and hence cannot be efficiently simulated classically unless the polynomial hierarchy collapses to the third level [82].

Similarly, for some subsets of instances, it is #P-hard to compute the output probability of an arbitrary string, since (by choosing gates to create a resource state for universal MBQC) this would allow one to compute output probabilities of universal polynomial-size quantum circuit families which is known to be #P-hard. The result of Refs. [26, 27] is then applicable, which implies that if the gates are chosen Haar-randomly, efficiently computing the output probability of some fixed string with probability $1 - 1/\mathrm{poly}(n)$ over the choice of instance up to additive error bounded by $2^{-\tilde{\Theta}(n \log(n))}$ implies the ability to efficiently compute a #P-hard function with high probability. □

Our goal is to prove that SEBD can efficiently approximately simulate the extended brickwork architecture in the average case for choices of extension parameters for which the above hardness results apply. To this end, we first show a technical lemma which describes how measurements destroy entanglement in 1D shallow random circuits. In particular, given a 1D state generated by a depth-2 Haar-random circuit acting on qubits, after measuring some contiguous region of spins B, the expected entanglement entropy of the resulting

post-measurement pure state across a cut going through B is exponentially small in the length of B. We defer the proof to Appendix 3.E.

Lemma 3.7. *Suppose a 1D random circuit C is applied to qubits $\{1,\dots,n\}$ consisting of a layer of 2-qubit Haar-random gates acting on qubits $(k,k+1)$ for odd $k \in \{1,\dots,n-1\}$, followed by a layer of 2-qubit Haar-random gates acting on qubits $(k,k+1)$ for even $k \in \{1,\dots,n-1\}$. Suppose the qubits of region $B = \{i,i+1,\dots,j\}$ for $j \geq i$ are measured in the computational basis, and the outcome b is obtained. Then, letting $|\psi_b\rangle$ denote the post-measurement pure state on the unmeasured qubits, and letting $A = \{1,2,\dots,i-1\}$ denote the qubits to the left of B,*

$$\mathbb{E}\, S(A)_{\psi_b} \leq c^{|B|} \tag{3.16}$$

for some universal constant $c < 1$, where the expectation is over measurement outcomes and choice of random circuit C.

We now outline the argument for why `SEBD` should be efficient for the extended brickwork architecture for sufficiently large extension parameters; full details may be found in Appendix 3.E. During the evolution of `SEBD`, as it sweeps from left to right across the lattice, it periodically encounters long stretches of length $2r$ in which no vertical gates are applied. We call these "1-local regions" since the maps applied in the corresponding effective 1D dynamics are 1-local when the algorithm is in such a region. Hence, in the effective 1D dynamics, no 2-qubit maps are applied and therefore the bond dimension of the associated MPS cannot increase during these stretches. It turns out that in 1-local regions, not only does the bond dimension needed to represent the state not increase, but it in fact rapidly decays in expectation. If r is sufficiently large, then the effective 1D state at the end of the 1-local region is very close to a product state with high probability, regardless of how entangled the state was before the region. Hence, when `SEBD` compresses the MPS describing the effective state at the end of the region, it may compress the bond dimension of the MPS to some fixed constant with very small incurred error. The two-qubit maps that are performed in-between 1-local regions only increase the bond dimension by a constant factor. Hence, with high probability, `SEBD` can use a $O(1)$ maximal bond dimension cutoff and simulate a random circuit with extended brickwork architecture with high probability. More precisely, it turns out that the scaling $r \geq \Theta(\log(n))$ is sufficient to guarantee efficient simulation with this argument. A more precise statement of the efficiency of `SEBD` for this architecture is given in the below lemma, whose proof may be found in Appendix 3.E.

Lemma 3.8. *Let C be an instance of `Brickwork`(L,r,v). Then, with probability at least $1-2^{-\Theta(r)}$ over the circuit instance, `SEBD` running with maximal bond dimension cutoff $D = \Theta(1)$ and truncation error parameter $\epsilon = 2^{-\Theta(r)}$ can be used to (1) sample from the output distribution of C up to error $n2^{-\Theta(r)}$*

in variational distance and (2) compute the output probability of an arbitrary output string up to additive error $n2^{-\Theta(r)}/2^n$ in runtime $\Theta(n)$.

With an appropriate choice of $r = \Theta(\log(L))$, the above result implies that SEBD can perform the simulation with error $1/\operatorname{poly}(n)$ for at least $1-1/\operatorname{poly}(n)$ fraction of instances. Similarly, choosing r to be a sufficiently large polynomial in L, SEBD can perform the simulation with error $2^{-n^{1-\delta}}$ for $1-2^{-n^{1-\delta}}$ fraction of instances, for any constant $\delta > 0$. We summarize these observations as the following corollary.

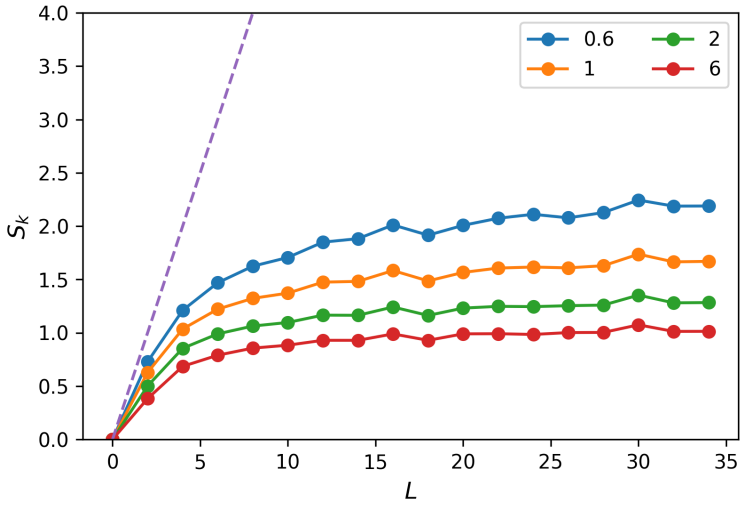
Corollary 3.5. *For any choice of polynomially bounded v, p_1, p_2, for any sufficiently large constant c SEBD can simulate $1 - 1/p_1(n)$ fraction of instances of Brickwork$(L, \lceil c\log(L)\rceil, v(L))$ up to error $\varepsilon \leq 1/p_2(n)$ in time $O(n)$. For any choice of $\delta > 0$ and $v(L) \leq \operatorname{poly}(L)$, for any sufficiently large constant c SEBD can simulate $1-2^{-n^{1-\delta}}$ fraction of instances of Brickwork$(L, \lceil L^c\rceil, v(L))$ up to error $\varepsilon \leq 2^{-n^{1-\delta}}$ in time $O(n)$. Here, "simulate with error ε" implies the ability to sample with variational distance error ε and compute the output probability of some fixed string $\mathbf{x}$ with additive error $\varepsilon/2^n$.*

3.5 Numerical results

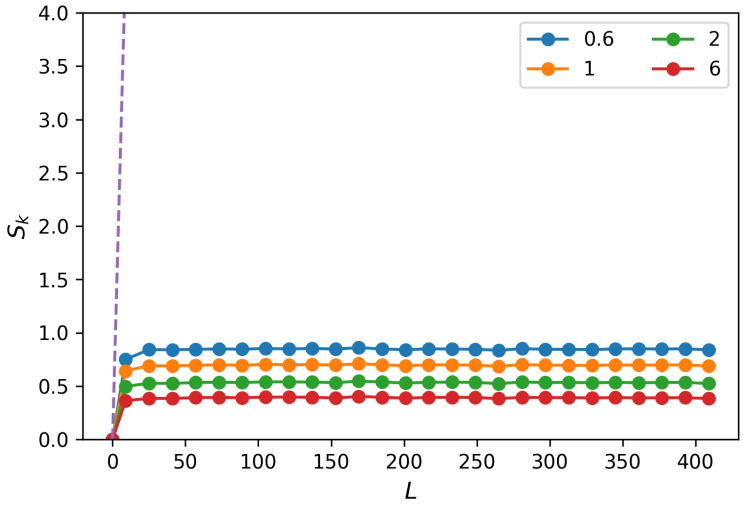
We implemented[3] SEBD on two families of random circuits: one consisting of depth-3 random circuits defined on a "brickwork architecture" consisting of three layers of two-qubit Haar-random gates (Figure 3.3 with parameter $r = 1$), and the other being the random circuit family obtained by applying single-qubit Haar-random gates to all sites of a cluster state—we referred to this problem as CHR previously. Note that the former architecture has depth three (not including the measurement layer) and the latter has depth four, and both architectures support universal measurement-based quantum computation [108], meaning they have the worst-case-hard property relevant for Conjecture 3.1. We did not implement Patching, due to its larger overhead.

Implementing SEBD on a standard laptop, we could simulate typical instances of the 409×409 brickwork model with truncation error 10^{-14} per bond with a runtime on the order of one minute per sample, and typical instances of the 34×34 CHR model with truncation error 10^{-10} per bond with a runtime on the order of five minutes per sample (these truncation error settings correspond to sampling errors of less than 0.01 in variational distance as derived previously in Section 3.3). We in fact simulated instances of CHR with grid sizes as large as 50×50, although due to the significantly longer runtime for such instances we did not perform large numbers of trials for these cases. In the case of the 409×409 brickwork model, performing over 3000 trials (consisting of generating a random circuit instance and generating a sample from its output distribution using a truncation error of 10^{-14}) and finding no

[3]The code for our implementation is available at https://github.com/random-shallow-2d/random-shallow-2d.



(a) CHR



(b) Brickwork

Figure 3.4: Rényi half-chain entanglement entropies S_k versus sidelength L in the effective 1D dynamics for the CHR and brickwork models, after 80 (resp. 550) iterations. Each point represents the entanglement entropy averaged over 50 random circuit instances, and over the final 10 (resp. 50) iterations for the CHR (resp. brickwork) model. Dashed lines depict the half-chain entanglement entropy scaling of a maximally entangled state, which can be created with a "worst-case" choice of gates for both architectures. The maximal truncation error per bond ϵ was 10^{-10} for CHR and 10^{-14} for the brickwork model.

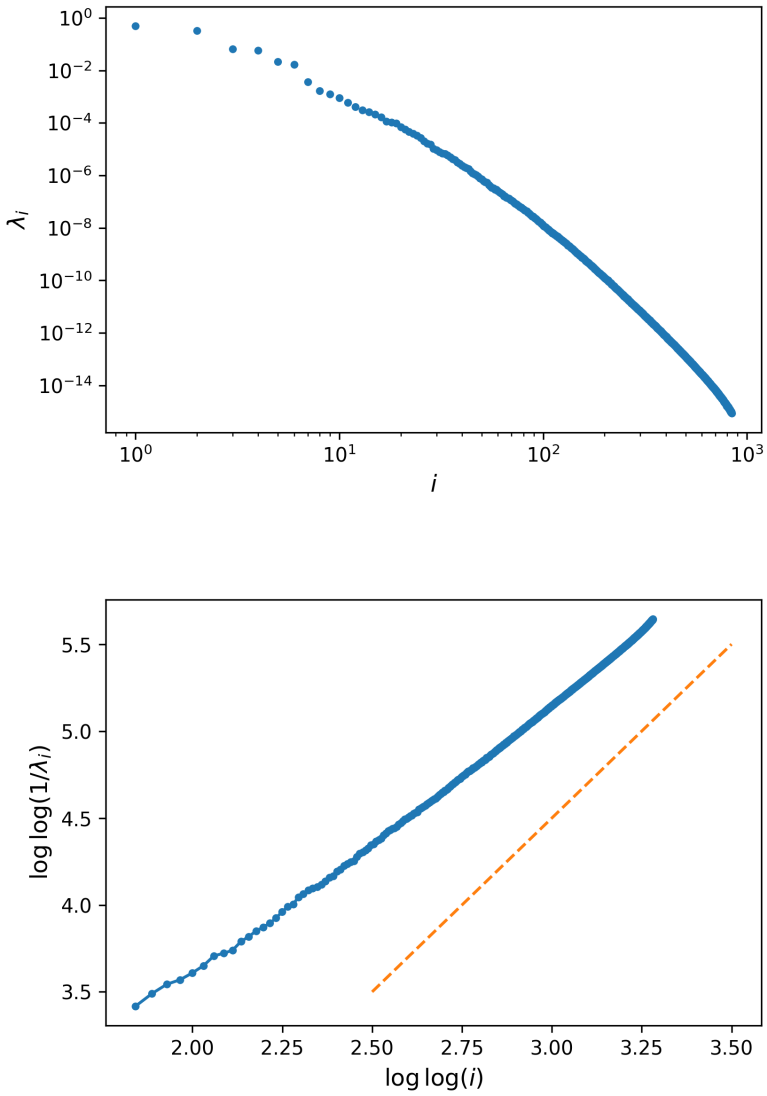


Figure 3.5: Typical half-chain entanglement spectrum $\lambda_1 \geq \lambda_2 \geq \ldots$ observed during the effective 1D dynamics of `CHR`. These plots were generated from an instance with sidelength $L = 44$ after running for 44 iterations, with squared Schmidt values smaller than approximately 10^{-15} truncated. The left figure shows a spectrum of half-chain eigenvalues. The downward curvature in the log-log scale indicates superpolynomial decay. The right figure displays the same data (minus the few largest values) on a loglog-loglog scale. The toy model predicts that the blue curve asymptotes to a straight line with slope two in the right figure, illustrated by the dashed orange line, corresponding to scaling like $\lambda_i \sim 2^{-\Theta(\log^2(i))}$. The plot is qualitatively consistent with this prediction. The spectrum for the brickwork model decays too quickly to obtain as useful statistics without going to much higher numerical precision.

trials for which the bond dimension became large enough for the algorithm to fail, then with 95% confidence, we may conclude that the probability that a random trial fails, p_f, is less than 0.001. Using the bound derived in Section 3.3, we can therefore conclude with 95% confidence that for greater than a 0.9 fraction of 409×409 circuit instances, we can sample from that circuit instance's output distribution with variational distance error less than 0.01. Intuitively, we expect the true simulable fraction to be much larger than this statistical guarantee, as it appears that the entanglement in the effective 1D dynamics only grows extensively for highly structured instances. Note that for both models, the runtime for a fixed truncation error was qualitatively highly concentrated around the mean. We expect that substantially larger instances of both random circuit families could be quickly simulated with more computing power, although 409×409 simulation of the brickwork architecture is already far beyond what could have been achieved by previous simulation methods that we are aware of.

To make this more precise, it is useful to compare our observed runtime with what is possible by previously known methods. The previously best-known method that we are aware of for computing output probabilities for these architectures would be to write the circuit as a tensor network and perform the contraction of the network [115]. The cost of this process scales exponentially in the tree-width of a graph related to the quantum circuit, which for a 2D circuit is thought to scale roughly as the surface area of the minimal cut slicing through the circuit diagram, as in Eq. (3.1). By this reasoning, we estimate that simulating a circuit with brickwork architecture on a 400×400 lattice using tensor network contraction would be roughly equivalent to simulating a depth-40 circuit on a 20×20 lattice with the architecture considered in [115], where the entangling gates are CZ gates. We see that these tasks should be equivalent because the product of the dimensions of the bonds crossing the minimal cut is equal to 2^{200} in both cases: for the brickwork circuit, 100 gates cross the cut if we orient the cut horizontally through the diagram in Figure 3.3 (with $r = 1$) and each gate contributes a factor of 4; meanwhile, for the depth-40 circuit, only one fourth of the unitary layers will contain gates that cross the minimal cut, and each of these layers will have 20 such gates that each contribute a factor of 2 (CZ gates have half the rank of generic gates). The task of simulating a depth-40 circuit on a 7×7 lattice was reported to require more than two hours using tensor network contraction on the 281 petaflop supercomputer Summit [115], and the exponentiality of the runtime suggests scaling this to 20×20 would take many orders of magnitude longer, a task that is decidedly intractable.

The discrepancy between maximal lattice sizes achieved for the two architectures is a result of the fact that the two have very different effective 1D dynamics. In particular, the entanglement of the effective dynamics for the brickwork architecture saturates to a significantly smaller value than that of the cluster state architecture. And even more directly relevant for prospects of

fast simulation, the typical spectrum of Schmidt values across some cut of the effective 1D dynamics for the brickwork architecture decays far more rapidly than that of the 1D dynamics for `CHR`. For this reason, the slower-decaying eigenvalue spectrum of `CHR` was significantly more costly for the runtime of the algorithm. (In fact, the eigenvalue spectrum of the brickwork model decayed sufficiently quickly that we were primarily limited not by the runtime of our algorithm, but by our numerical precision, which could in principle be increased.) But while the slower decay of the spectrum for the `CHR` model necessitated a longer runtime for a given sidelength, it allowed us to study the functional form of the spectrum and in particular compare against the predictions of the toy model of Section 3.3.4 as we discuss below.

While we were computationally limited to probing low-depth and small-size models, our numerical results point toward `SEBD` having an asymptotic running time for both models bounded by $\mathrm{poly}(n, 1/\varepsilon, 1/\delta)$ in order to sample with variational distance ε or compute output probabilities with additive error ε/q^n with probability $1-\delta$, suggesting that Conjecture 3.1 is true. Our numerical evidence for this is as follows:

1. We find that the effective 1D dynamics associated with these random circuit families appear to be in area-law phases, as displayed in Figure 3.4. That is, the entanglement does not grow extensively with the sidelength L, but rather saturates to some constant. We furthermore observe qualitatively identical behavior for some Rényi entropies S_k with $k < 1$. It is known [112] that this latter condition is sufficient to imply that a 1D state may be efficiently described by an MPS, indicating that `SEBD` is efficient for these circuit families and that Conjecture 3.1 is true.

2. For further evidence of efficiency, we study the functional form of the entanglement spectra of the effective 1D dynamics. For the effective 1D dynamics corresponding to `CHR`, we observe superpolynomial decay of eigenvalues (i.e. squared Schmidt values) associated with some cut, displayed in Figure 3.5, indicating that choosing a maximal bond dimension of $D = \mathrm{poly}(1/\epsilon)$ is more than sufficient to incur less than ϵ truncation error per bond. The observed spectrum tends toward a scaling which is qualitatively consistent with the asymptotic scaling of $\lambda_i \sim 2^{-\Theta(\log^2(i))}$ predicted by the toy model of Section 3.3.4 and consistent with our Conjecture 3.1′. Note that this actually suggests that the required bond dimension of `SEBD` may be even smaller than $\mathrm{poly}(1/\epsilon)$, scaling like $D = 2^{\Theta(\sqrt{\log(1/\epsilon)})}$.

While these numerical results may be surprising given the prevalence of average-case hardness conjectures for quantum simulation, they are not surprising from the perspective of the recent works (discussed in previous sections)

that find strong evidence for an entanglement phase transition from an area-law to volume-law phase for 1D unitary-and-measurement processes driven by measurement strengths. Since the effective dynamics of the 2D random shallow circuits we study are exactly such processes, our numerics simply point out that these systems are likely on the area-law side of the transition. (However, no formal universality theorems are known, so the various models of unitary-and-measurement circuits that have been studied are generally not known to be equivalent to each other.) In the case of the brickwork architecture, we are also able to provide independent analytical evidence (Section 3.6.6) that this is the case by showing the "quasi-entropy" $\tilde{S}_2$ for the 1D process is in the area-law phase. We leave the problem of numerically studying the precise relationship between circuit depth, qudit dimension, properties of the associated stat mech models (including "quasi-entropies") as discussed in subsequent sections, and the performance of `SEBD` for future work. In particular, simulations of larger depth and larger qudit local dimension could be used to provide numerical support for Conjecture 3.2, which claims that as these parameters are increased the circuit architectures eventually transition to a regime where our algorithms are no longer efficient.

3.6 Analytical evidence for conjectures from statistical mechanics

3.6.1 Overview

In the previous section, we provided strong numerical evidence that `SEBD` is efficient when acting on certain sufficiently shallow architectures. Here we provide complementary, analytical evidence that bolsters the case for `SEBD`'s (and, in Appendix 3.C, `Patching`'s) efficiency. Our method is based on the technique described in Chapter 2 and developed in Refs. [36, 37, 41, 42, 44, 45], which maps random quantum circuits to classical statistical mechanical systems. We describe how the method can be applied generally to different 2D architectures, but we give special attention to the depth-3 brickwork architecture because it is a worst-case-hard uniform architecture which is simple enough for concrete conclusions to be drawn that act as evidence that the algorithms are efficient. The stat mech method also provides evidence of computational phase transitions as qudit dimension and circuit depth are increased.

The map produces a classical stat mech model for which the entanglement properties of the underlying random circuits are related to thermodynamic properties of the model. In particular, we examine a quantity we call the "quasi-k entanglement entropy," denoted $\tilde{S}_k$, to quantify the entanglement of the 1D state "tracked" by `SEBD` at any given point in time throughout the effective 1D dynamics; the mapping relates $\tilde{S}_k$ to the free energy cost incurred by twisting boundary conditions of the stat mech system. The quasi-k entropy is related but not exactly equal to the Rényi-k entanglement entropy averaged over random circuit instances and measurement outcomes, denoted by $\langle S_k \rangle$. Ideally, we would find rigorous bounds on $\langle S_k \rangle$ (for $0 < k < 1$) for these states throughout the effective 1D dynamics to show that `SEBD` is efficient. We study

the quasi-entropies $\tilde{S}_k$ instead because the stat mech mapping permits for an analytical handle on $\tilde{S}_k$ for integer $k \geq 2$, and the calculations become especially tractable for $k = 2$. Changing the qudit dimension q of the random circuit model corresponds to changing the interaction strengths in the associated stat mech model, which drives a phase transition. This phase transition in the classical stat mech model is accompanied by phase transitions in quasi-entropies. Even though the efficiency of our algorithms is related to different entropic quantities, which are hard to directly analyze, the phase transition in quasi-entropies provides analytical evidence in favor of our conjectures.

In the remaining subsections, we define the quasi-entropy, briefly restate the details of the stat mech map discussed in Chapter 2, apply the map generally to 2D circuits to reason heuristically about order-disorder behavior, and finally conclude by applying it more rigorously to the depth-3 brickwork architecture, where we observe a q-driven order-disorder phase transition in the corresponding stat mech model.

3.6.2 Quasi-entropy

Given an ensemble of pure quantum states, the quasi-entropy is a quantity that is related to the expected amount of entanglement in the state. In our case, the ensemble is generated by a random quantum circuit followed by a projective measurement on some subset of the qudits, and the quasi-entropy is computed as follows.

Suppose we fix a random quantum circuit instance U drawn according to some specified architecture, as well as a known outcome $\mathbf{x}$ for a projective measurement performed on some subset $B \subset [n]$ of the output qudits. Let $\rho = \Pi_{\mathbf{x}} U|0^n\rangle\langle 0^n|U^\dagger \Pi_{\mathbf{x}}$ be the pure output state associated with the instance and measurement outcome, where $\Pi_{\mathbf{x}}$ is the projection of the bits in region B onto the measurement outcome $\mathbf{x}$, and note that the normalization $\mathrm{tr}(\rho)$ is equal to the probability of obtaining the outcome $\mathbf{x}$. Then for any $k \geq 0$ and for some subregion A of the unmeasured qudits $[n] \setminus B$, we define

$$Z_{k,\emptyset} = \mathrm{tr}(\rho)^k \tag{3.17}$$

$$Z_{k,A} = \mathrm{tr}\left(\rho_A^k\right), \tag{3.18}$$

where ρ_A is the reduced density matrix of ρ on A. Letting $\mathbb{E}_U$ denote expectation over choice of instance U and uniformly random measurement outcome $\mathbf{x}$, the quasi-k entropy $\tilde{S}_k(A)$ for the random circuit ensemble is defined as

$$\tilde{S}_k(A) = \frac{1}{1-k} \log \left(\frac{\mathbb{E}_U(\mathrm{tr}(\rho)^k \frac{Z_{k,A}}{Z_{k,\emptyset}})}{\mathbb{E}_U(\mathrm{tr}(\rho)^k)} \right) \tag{3.19}$$

$$= \frac{1}{1-k} \log \left(\frac{\mathbb{E}_U(Z_{k,A})}{\mathbb{E}_U(Z_{k,\emptyset})} \right) \tag{3.20}$$

$$= \frac{F_{k,\emptyset} - F_{k,A}}{1-k}, \tag{3.21}$$

where $F_{k,X} = -\log(\mathbb{E}_U(Z_{k,X}))$ for $X \in \{\emptyset, A\}$ will be associated with the "free energy" of the classical stat mech model that the circuit maps to. Virtually identical quantities were also considered in two other recent works [44, 45].

Note the similarity of the above expression to the average Rényi-k entanglement entropy, given by

$$\langle S_k(A)_\rho \rangle = \frac{\mathbb{E}_U(\mathrm{tr}(\rho) S_k(A)_\rho)}{\mathbb{E}_U(\mathrm{tr}(\rho))} \tag{3.22}$$

$$= \frac{1}{1-k} \frac{\mathbb{E}_U\left(\mathrm{tr}(\rho) \log \frac{Z_{k,A}}{Z_{k,\emptyset}}\right)}{\mathbb{E}_U(\mathrm{tr}(\rho))} . \tag{3.23}$$

Indeed, the two formulas are the same, except that the quasi-entropy weights instances by $\mathrm{tr}(\rho)^k$ instead of $\mathrm{tr}(\rho)$, and takes the logarithm after taking the expectation.

Also note that in the limit $k \to 1$, both $\tilde{S}_k$ and $\langle S_k \rangle$ approach the expected von Neumann entropy, which is defined by the following expression:

$$\langle S(A) \rangle = -\mathop{\mathbb{E}}_U \left(\mathrm{tr}\left(\frac{\rho_A}{\mathrm{tr}(\rho)} \log\left(\frac{\rho_A}{\mathrm{tr}(\rho)} \right) \right) \right) . \tag{3.24}$$

This observation lends some justification to the use of $\tilde{S}_k$ as a proxy for $\langle S_k \rangle$ even when $k \neq 1$. This is further justified by previous work studying random 1D circuits without measurements; in Ref. [42], the growth rate of $\tilde{S}_2$ in random 1D circuits was calculated using the stat mech mapping and no significant difference was found with numerical calculations of $\langle S_2 \rangle$. Moreover, Ref. [36] used the *replica trick* to directly compute $\langle S_2 \rangle$ as a series in powers of $1/q$, where q is the qudit local dimension, and found that the leading term of this expansion agrees with $\tilde{S}_2$, indicating that $\tilde{S}_2$ is a valid substitute for $\langle S_2 \rangle$ in the $q \to \infty$ limit and suggesting it is a good approximation when q is finite.

3.6.3 Stat mech method for the quasi-entropy

For an introduction to the stat mech method for random quantum circuits, we refer the reader back to Chapter 2. There we described how to map kth moment random circuit quantities to partition functions of stat mech systems where particles take on one of $k!$ values, labeled by some element $\nu \in \mathcal{S}_k$ with $\mathcal{S}_k$ denoting the symmetric group. In our case, we would like to compute $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$, which are kth moment quantities since they are linear in $U^{\otimes k} \otimes U^{*\otimes k}$.

Specifically, for $Z_{k,\emptyset}$, we may rewrite Eq. (3.17) as

$$Z_{k,\emptyset} = \sum_{\vec{i} \in [q]^{kn}} \langle \vec{i} | (\Pi_{\mathbf{x}} U |0^n\rangle)^{\otimes k} (\langle 0^n | U^\dagger \Pi_{\mathbf{x}})^{\otimes k} | \vec{i} \rangle \tag{3.25}$$

$$= \sum_{\vec{i} \in [q]^{kn}} \langle \vec{i}, \vec{i} | \Pi_{\mathbf{x}}^{\otimes 2k} U^{\otimes k} \otimes U^{*\otimes k} |0^n\rangle^{\otimes 2k} \tag{3.26}$$

$$= \langle e |^{\otimes n} \Pi_{\mathbf{x}}^{\otimes 2k} U^{\otimes k} \otimes U^{*\otimes k} |0^n\rangle^{\otimes 2k} , \tag{3.27}$$

where e labels the identity permutation and $|e\rangle = \sum_{i_1,\ldots,i_k=0}^{q-1} |i_1,\ldots,i_k\rangle \otimes |i_1,\ldots,i_k\rangle$ follows Eq. (2.14).

For $Z_{k,A}$, we use the identity $\mathrm{tr}\left(\rho_A^k\right) = \mathrm{tr}\left(\rho^{\otimes k} W_{(1\ldots k)}^{(A)}\right)$, where $W_{(1\ldots k)}^{(A)}$ is the operator acting on a k-fold tensor product of the n-qubit system that performs the permutation $(123\ldots k)$ (written in cycle notation) on the k copies of region A, and acts as identity on the copies of region $[n] \setminus A$. We arrive at

$$Z_{k,A} = \sum_{\vec{i}\in[q]^{kn}} \langle\vec{i}|(\Pi_{\mathbf{x}} U|0^n\rangle)^{\otimes k}(\langle 0^n|U^\dagger\Pi_{\mathbf{x}})^{\otimes k} W_{(1\ldots k)}^{(A)}|\vec{i}\rangle \tag{3.28}$$

$$= \sum_{\vec{i}\in[q]^{kn}} \langle\vec{i},\vec{i}|(I^{\otimes nk} \otimes W_{(k\ldots 1)}^{(A)})\Pi_{\mathbf{x}}^{\otimes 2k} U^{\otimes k} \otimes U^{*\otimes k}|0^n\rangle^{\otimes 2k} \tag{3.29}$$

$$= \left(\bigotimes_{a\in A}\langle(1\ldots k)|_a \bigotimes_{c\notin A}\langle e|_c\right) \Pi_{\mathbf{x}}^{\otimes 2k} U^{\otimes k} \otimes U^{*\otimes k}|0^n\rangle^{\otimes 2k}\,, \tag{3.30}$$

where $|(1\ldots k)\rangle$ denotes the state $\sum_{i_1,\ldots,i_k=0}^{q-1} |i_1,i_2,\ldots,i_k\rangle \otimes |i_k,i_1,i_2,\ldots,i_{k-1}\rangle$, following Eq. (2.14).

In Chapter 2, we explained how to map the quantity $\mathbb{E}_U[U^{\otimes k} \otimes U^{*\otimes k}]$ to a partition function of a classical stat mech system, described by Eq. (2.13). Namely, each gate (labeled by integer u) in the circuit diagram gets mapped to two particles, an "incoming" and "outgoing" particle, which can be in one of $k!$ internal states labeled by permutations τ_u and σ_u, respectively. There are interactions between the incoming and outgoing particles originating from the same gate t (denoted by the edge $\langle t\rangle$), and between outgoing particles of one gate u and incoming particles of a subsequent gate v, when v acts immediately after u on one of the same qubits (denoted by the edge $\langle uv\rangle$). These interactions are associated with a weight factor

$$\mathrm{weight}_{\langle t\rangle}(\sigma,\tau) = \mathcal{W}\mathrm{g}(\tau_t\sigma_t^{-1}, q^2)\,, \tag{3.31}$$

where $\mathcal{W}\mathrm{g}$ is the Weingarten function, and

$$\mathrm{weight}_{\langle uv\rangle}(\sigma,\tau) = q^{C(\sigma_u\tau_v^{-1})}\,, \tag{3.32}$$

where $C(\pi)$ returns the number of cycles in the permutation π.

This describes the bulk interactions of the stat mech system. The boundary conditions we impose for the quasi-entropy calculation can be determined from Eqs. (3.27) and (3.30). In both cases, the input state is $|0^n\rangle^{\otimes 2k}$. If gate u is the first gate to act on some qudit a, then the quantities $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$ will pick up a factor of $\langle\tau_u|\left(|0\rangle^{\otimes 2k}\right) = 1$, independent of τ_u. Thus, we have open boundary conditions at the input of the circuit.

Boundary conditions at the end of the circuit are more complex. Suppose gate u is the last gate to act on some qudit a. If $a \in B$, then a is

measured at the end of the circuit, and $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$ pick up factors of $\langle e|\Pi_{\mathbf{x}_a}^{\otimes 2k}|\sigma_u\rangle = 1$; hence there are open boundary conditions for region B. If $a \in [n] \setminus (A \cup B)$, then $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$ both pick up a factor of $\langle e|\sigma_u\rangle = q^{C(\sigma_u)}$. If $a \in A$, then $\mathbb{E}_U(Z_{k,\emptyset})$ picks up the same factor, but $\mathbb{E}_U(Z_{k,A})$ picks up the factor $\langle (1\dots k)|\sigma_u\rangle = q^{C((1\dots k)\circ\sigma_u^{-1})}$ instead. Note that these formulas mirror the right-hand-side of Eq. (3.32). Thus, this boundary condition is equivalent to introducing a new auxiliary particle with internal state $\chi_{a'}$ for each qudit $a' \notin B$ (the prime is used to indicate that it is an auxiliary particle), and adding an interaction $\langle ua'\rangle$ between the new particle and the outgoing node for the final gate u to act on the particle. The weight for this interaction is given by

$$\text{weight}_{\langle ua'\rangle}(\sigma,\tau) = q^{C(\sigma_u\chi_{a'}^{-1})}\,. \tag{3.33}$$

The set of χ values is not an explicit argument of the weight function as these values should be regarded as fixed; the partition function only sums over the possible internal states τ_u and σ_u of incoming and outgoing particles. For the calculation of $\mathbb{E}_U(Z_{k,X})$ for $X \in \{\emptyset, A\}$, the internal state $\chi_{a'}$ of the auxiliary particles is fixed to be the identity e if $a' \notin X$, and fixed to be the k-cycle $(1\dots k)$ if $a' \in X$. The entire map can then be expressed by the equation

$$\mathbb{E}[Z_{k,X}] = \sum_{\sigma,\tau}\prod_u \text{weight}_{\langle u\rangle}(\sigma,\tau)\prod_{\langle uv\rangle}\text{weight}_{\langle uv\rangle}(\sigma,\tau)\prod_{\langle ua'\rangle}\text{weight}_{\langle ua'\rangle}(\sigma,\tau)\,, \tag{3.34}$$

where the right-hand-side depends on X only through the setting of the internal states of the auxiliary particles. This is a partition function—a weighted sum over spin configurations where the weight of each term is given by a product of factors that depend only on the spin values of a pair of particles connected by an edge in the interaction graph. We define the free energy to be the negative logarithm of this partition function (see Eq. (3.21)), mirroring the standard relationship $F = -k_BT\log(Z)$ between the free energy and the partition function from statistical mechanics, with k_BT set to 1. The only difference between $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$ is that in the latter case, the boundary condition at the output of the circuit is "twisted" from e to $(1\dots k)$ only in the region A, and thus $\tilde{S}_k(A)$ measures the free energy cost of this twist.

We provide an example of the interaction graph for the stat mech system that includes the auxiliary nodes in Figure 3.6.

3.6.4 Special case of $k = 2$

When $k = 2$, the symmetric group $\mathcal{S}_k$ has only 2 elements, identity (denoted by $I \equiv e$) and swap (which we denote by by $S \equiv (12)$), so the quantities $\mathbb{E}_U(Z_{2,\emptyset})$ and $\mathbb{E}_U(Z_{2,A})$ map to partition functions of Ising-like classical stat mech models where each node takes on one of two values. Furthermore, as discussed in Chapter 2, in the $k = 2$ case with no measurements, it was shown in Refs. [41, 42] (see also Refs. [36, 37]) that one can get rid of all negative

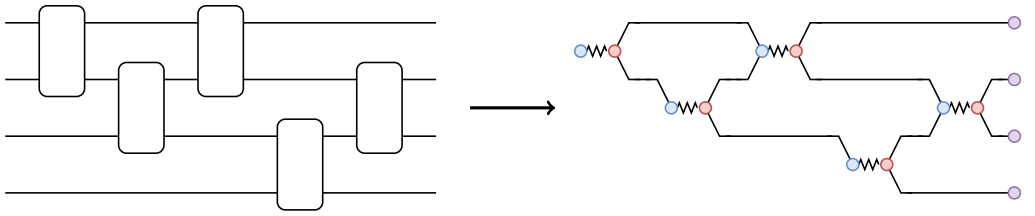


Figure 3.6: Example of stat mech mapping applied to a circuit diagram with 4 qudits and 5 Haar-random gates. Blue nodes are incoming nodes, red nodes are outgoing nodes, and purple nodes are auxiliary nodes. Zigzag edges carry Weingarten weight. Straight carry weight equal to q^C where C is the number of cycles in the product of the two adjacent permutations.

terms in the partition function by decimating half of the nodes, i.e. explicitly performing the sum over the values of the incoming nodes τ in Eq. (3.34). This continues to be true even when there are measurements in between unitaries in the circuit, as discussed in the appendices. However, the decimation causes the two-body interactions to become three-body interactions between any three outgoing particles in internal states σ_{u_1}, σ_{u_2}, σ_{u_3} when unitary u_3 succeeds unitaries u_1 and u_2 and shares a qudit with each. The lack of negative weights for $k = 2$ is convenient because it allows one to view the system as a classical spin model at a real temperature and can therefore be analyzed with well-studied numerical techniques like Monte Carlo sampling.

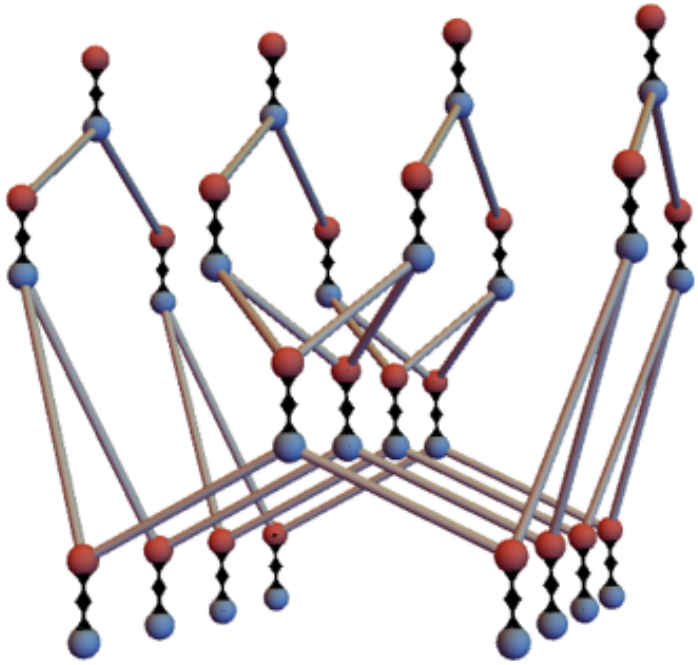
3.6.5 Mapping applied to general 2D circuits

We now apply the mapping directly to a depth-d circuit acting on a $\sqrt{n} \times \sqrt{n}$ lattice of qudits consisting of nearest-neighbor two-qudit Haar-random gates. This is the relevant case for the algorithms presented in this paper. In this section, we will assume for concreteness that the first unitary layer includes gates that act on qudits at gridpoints (i, j) and $(i, j + 1)$ for all odd i and all j, the second layer on (i, j) and $(i, j + 1)$ for all even i and all j, the third layer on (i, j) and $(i + 1, j)$ for all i and all odd j, and the fourth layer on (i, j) and $(i + 1, j)$ for all i and all even j. Subsequent layers then cycle through these four orientations.

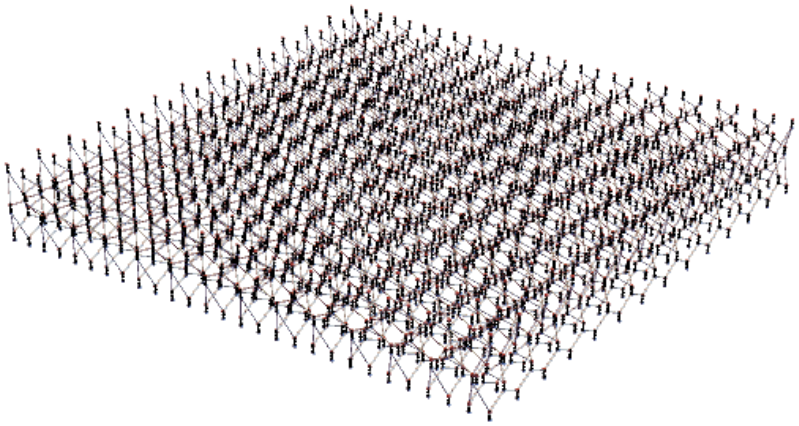
The classical stat mech model

Replacing the unitaries in the circuit diagram with pairs of nodes and connecting them as described previously yields a graph embedded in three dimensions. The nodes in this graph still have degree three, so locally the graph looks similar to the honeycomb lattice (the lattice that arises from a 1+1D circuit as discussed in Refs. [37, 41, 44, 45] and in the appendix), but globally the nodes form a 3D lattice that can be viewed roughly as a $\sqrt{n} \times \sqrt{n} \times d$ slab, although

the details of how these nodes connect is not straightforward to visualize. We have included pictures of the graph in Figure 3.7.



(a) Depth-4 circuit on 4×4 lattice



(b) Depth-5 circuit on 28×28 lattice

Figure 3.7: The graph produced by the stat mech mapping on shallow 2D circuits. (a) A close up view of the graph reveals that the degree of most nodes is three, similar to the honeycomb lattice. (b) A far-away view reveals that globally the graph looks like a two-dimensional slab of thickness roughly d.

Recall that edges between nodes originating from the same unitary are assigned a weight equal to the Weingarten function and edges between suc-

cessive unitaries follow the interaction $\text{weight}_{\langle uv\rangle}(\sigma,\tau) = q^{C(\sigma_u\tau_v^{-1})}$. For $k=2$ this amounts to a ferromagnetic Ising interaction where

$$\text{weight}_{\langle uv\rangle}(\sigma,\tau) = \begin{cases} q^2 & \text{if } \sigma_u\tau_v = I \\ q & \text{if } \sigma_u\tau_v = S\,. \end{cases} \tag{3.35}$$

To analyze the output state, we will divide the n qudits into three groups A, B, and C. We suppose that after the d unitary layers have been performed, a projective measurement is performed on the qudits in region B. Qudits in regions A and C are left unmeasured and we wish to calculate quantities like $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$. The mapping calls for us to introduce an auxiliary node for each unmeasured qudit in the circuit, i.e. an auxiliary node for qudits in regions A and C. For $\mathbb{E}_U(Z_{k,\emptyset})$ all of the auxiliary nodes are set to identity e, while for $\mathbb{E}_U(Z_{k,A})$, the auxiliary nodes for region A are set to the k-cycle $(1\ldots k)$.

Eliminating negative weights via decimation when $k=2$

The quantities $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$ are now given by classical partition functions on this graph with appropriate boundary conditions for the auxiliary nodes in regions A and C. We wish to understand whether this stat mech model is ordered or disordered. We are faced with the issue that the Weingarten function can take negative values and thus some configurations over this graph could have negative weight. For $k=2$, as previously discussed, we can rectify this by decimating all the incoming nodes. The resulting graph has half as many nodes and interactions between groups of three adjacent outgoing particles associated with gates u_1, u_2, and u_3, whenever unitary u_3 acts after u_1 and u_2. There is a simple formula for the weights:

$$\text{weight}_{\langle u_1u_2u_3\rangle}(\sigma) = \begin{cases} 1 & \text{if } \sigma_{u_1} = \sigma_{u_2} = \sigma_{u_3} \\ \frac{1}{q+q^{-1}} & \text{if } \sigma_{u_2} \neq \sigma_{u_3} \\ 0 & \text{if } \sigma_{u_1} \neq \sigma_{u_2} = \sigma_{u_3}. \end{cases} \tag{3.36}$$

Now, all the weights are non-negative. Moreover, the largest weight occurs when all the nodes agree, indicating a generally ferromagnetic interaction between the trio of nodes. If either σ_{u_1} or σ_{u_2} disagrees with the other two values, the weight is reduced by a factor of $q+q^{-1}$. When σ_{u_3} disagrees, the weight is 0; these configurations are forbidden and contribute nothing to the partition function.

Given an assignment of I or S to each node σ_u, we can associate a pattern of domain walls, that is, a set of edges connecting nodes with disagreeing values. These domain walls partition the 2D slab into contiguous domains of adjacent nodes all given the same value.

Allowed domain wall configurations and disorder-order phase transitions

Using this observation, we can understand the kinds of domain wall structures that will appear in configurations that contribute non-zero weight. Recall that the stat mech model occupies a 2D slab of constant thickness in the direction of time, which we orient vertically. In this setting, domain wall structures are membrane-like since the graph is embedded in 3D. Membranes that have upward curvature, shaped like a bowl, are not allowed, because somewhere there would need to be an interaction where the upper node disagrees with the two below it, a situation that leads to 0 weight as in Eq. (3.36). On the other hand, cylindrically shaped domain wall membranes do not have this issue, nor do dome-shaped membranes with downward curvature. These three cases are illustrated in Figure 3.8. The weight of a configuration is reduced

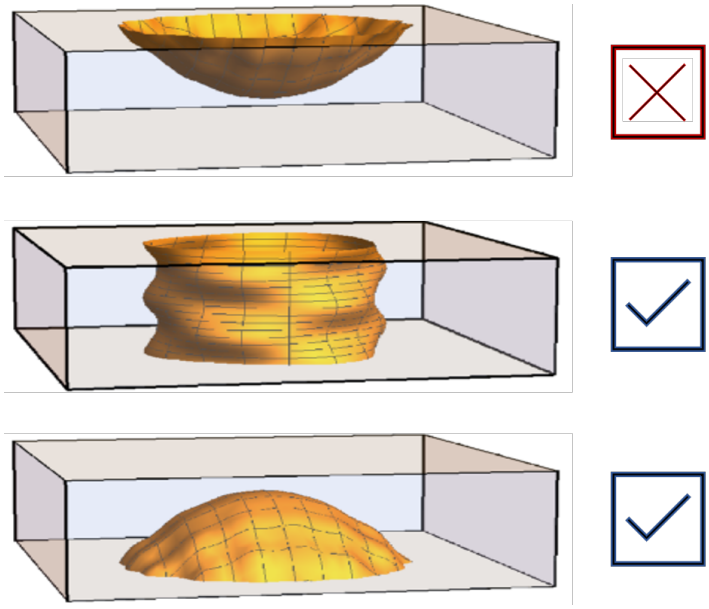


Figure 3.8: Cartoon depiction of forbidden and allowed domain wall structures in the stat mech model for a shallow 2D circuit. Time is oriented vertically. For a particular spin configuration, domain walls mark the boundary between regions assigned e and regions assigned (12), forming a membrane. If this membrane has upward curvature, it is forbidden (contributes 0 weight to the partition function), whereas if it does not have upward curvature, it is allowed. For allowed configurations, the contribution decreases exponentially in the total domain wall area.

by a factor of $q + q^{-1}$ for each unit of domain wall, an effect that acts to minimize the domain wall size when drawing samples from the thermal distribution (energy minimization). On the other hand, larger domain walls have more configurational entropy—there are many ways to cut through the graph with a cylindrically shaped membrane—an effect that acts to bring out more

domain walls in samples from the thermal distribution (entropy maximization). The question is, which of these effects dominates? For a certain setting of the depth d (slab thickness) and local dimension q, is there long-range order, or is there exponential decay of correlations indicating disorder? Generally speaking, increasing depth magnifies both effects: cylindrical domain wall membranes must be longer—meaning larger energy—when the depth is larger; however, longer cylinders also have more ways of propagating through the graph. Meanwhile, increasing q only magnifies the energetic effect since it increases the interaction strength and thus the energy cost of a domain wall unit but leaves the configurational entropy unchanged.

Thus, in the limit of large q we expect the energetic effect to win out and the system to be ordered for any fixed circuit depth d and any circuit architecture. What about small q? Physically speaking, q must be an integer at least 2 since it represents the local Hilbert space dimension of the qudit. However, the statistical mechanical model itself requires no such restriction, and we can allow q to vary continuously in the region $[1, \infty)$. Then for $q \to 1$, the energy cost of one unit of domain wall becomes minimal (but it does not vanish). Depending on the exact circuit architecture and the depth of the circuit, the system may experience a phase transition into the disordered phase once q falls below some critical threshold q_c. The depth-3 circuit with brickwork architecture that we present later in Section 3.6.6 provides an example of such a transition. It is disordered when $q = 2$ and experiences a phase transition as q increases to the ordered phase at a transition point we estimate to be roughly $q_c \approx 6$.

When q is fixed and d is varied, it is less clear what to expect. Suppose for small d, the system is disordered. Then increasing d will amplify both the energetic and entropic effects, but likely not in equal proportions. If the amplification of the energetic effect is stronger with increasing depth, then we expect to transition from the disordered phase to the ordered phase at some critical value of the depth d_c. Without a better handle on the behavior of the stat mech model, we cannot definitively determine if and when this depth-driven phase transition will happen.

However, we have other reasons to believe that there should be a depth-driven phase transition. In particular, we now provide an intuitive argument for why a disorder-order transition in the parameter q should imply a disorder-order transition in the parameter d. Consider fixed d, and another fixed integer $r \geq 1$ such that $d/r \gg 1$. We may group together $r \times r$ patches of qudits to form a "supersite" with local dimension q^{r^2}. Similarly, we may consider a "superlayer" of $O(r)$ consecutive unitary layers. Since $O(r)$ layers is sufficient to implement an approximate unitary k-design on a $r \times r$ patch of qudits (taking $k = O(1)$) [86], we intuitively take each superlayer to implement a Haar-random unitary between pairs of neighboring supersites. Thus, a depth-d circuit acting on qudits of local dimension q is roughly equivalent to a depth-

$O(d/r)$ circuit acting on qudits of local dimension q^{r^2} in the supersite picture. If for a fixed d, we observe a disorder-order phase transition for increasing q, then for fixed q and fixed d/r, we should also observe a disorder-order phase transition with increasing r. Equivalently, we should see a transition for fixed q and increasing d. This logic is not perfect because superlayers do not exactly map to layers of Haar-random two-qudit gates between neighboring supersites, but nonetheless we take it as reason to expect a depth-driven phase transition.

Efficiency of `SEBD` algorithm from stat mech

The efficiency of the `SEBD` algorithm relies on the error incurred during the MPS compression being small. If the inverse error has a polynomial relationship (or better) with the bond dimension of truncation, then the algorithm's time complexity is polynomial (or better) in the inverse error and the number of qudits. This will be the case if the MPS prior to truncation satisfies an area law for the Rényi-k entropy for some $0 < k < 1$. The stat mech mapping is unable to probe these values of k. However, we hypothesize that the behavior of larger values of k is indicative of the behavior for $k < 1$ since the examples where the Rényi-k entropy with $k \geq 1$ satisfies an area law but efficient MPS truncation is not possible require contrived spectrums of Schmidt coefficients. Although some physical processes give rise to situations where the von Neumann and Rényi-k entropies with $k > 1$ exhibit different behavior (see e.g. [116], which showed that for random 1D circuits without measurements but with the unitaries chosen to commute with some conserved quantity, after time t the entropy is $O(t)$ for $k = 1$ but $O(\sqrt{t \log t})$ for $k > 1$), the numerical evidence we gave in Section 3.5, where the scaling of all the Rényi-k entropies appears to be the same, suggests our case is not one of these situations.

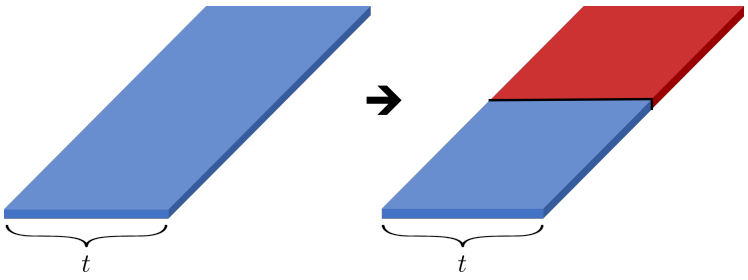
Previously, we discussed how for 1D circuits with alternating unitary and weak measurement dynamics, there has been substantial numerical evidence in prior literature for a phase transition from an area-law phase to a volume-law phase as the parameters of the circuit are changed. There has also been analytical work [44, 45] on this model using the stat mech mapping (and in Appendix 3.A, we use a similar approach to analyze 1D circuits with a different form of weak measurement, inspired by the `CHR` problem discussed earlier, and show there is a q-driven phase transition from a disordered phase to an ordered phase).

The `SEBD` algorithm simulating a 2D circuit of constant depth made from Haar-random gates may be viewed as a system with very similar dynamics—an alternation between entanglement-creating unitary gates and entanglement-destroying weak measurements. However, none of the unitary-and-measurement models that have been previously studied capture the exact dynamics of `SEBD`, one reason being that `SEBD` tracks the evolution of several columns of qudits at once (recall it must include all qudits within the lightcone of the first column). The Haar-random unitaries create entanglement within

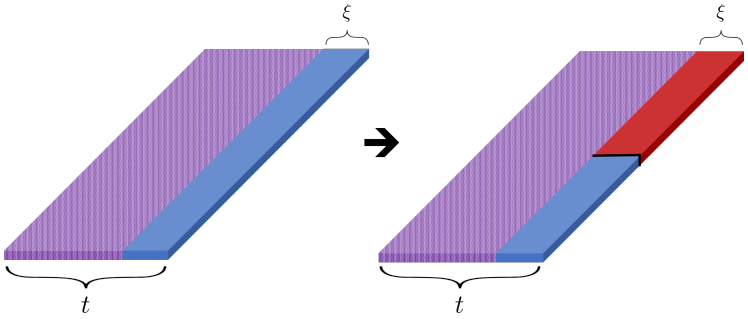
these columns of qudits, but not in the exact way that entanglement is created by Haar-random nearest-neighbor gates acting on a single column. Nonetheless, we expect the story to be the same for the dynamics of SEBD since the main findings of studies of these unitary-and-measurement models have been quite robust to variations in which unitary ensembles and which measurements are being implemented; we expect that varying parameters of the circuit architecture like q and d can lead to entanglement phase transitions, and thus transitions in computational complexity.

Indeed, the discussion from the previous section suggests precisely this fact. When we apply the stat mech mapping directly to 2D circuits instead of to 1D unitary-and-measurement models, we expect disorder-order phase transitions as both q and d are varied. To make the connection to entanglement entropy explicit here, we note that after t steps of the SEBD algorithm, all $\sqrt{n}$ qudits in the first t columns of the $\sqrt{n} \times \sqrt{n}$ lattice have been measured, and we have an MPS representation of the state on columns $t+1$ through $t+r$, where $r = O(d)$ is the radius of the lightcone (which depends on circuit architecture, but cannot be larger than d). To calculate the entropy of the MPS, we take the region A to be the top half of these r columns, and region C to be the bottom half. Region B consists of the first t columns, which experience projective measurements. The prescription for computing $\tilde{S}_2(A)$ calls for determining the free energy cost of twisting the boundary conditions in region A, which creates a domain wall along the $A:C$ border. If the bulk is in the ordered phase, then this domain wall membrane originating at the $A:C$ boundary will penetrate through the graph a distance of t, leading to a domain wall area of $O(td)$. If the bulk is in the disordered phase, it will only penetrate a constant distance, on the order of the correlation length ξ of the disordered stat mech model, before being washed out by the disorder, leading to a domain wall area of only $O(\xi d)$. This is the key observation that connects order-disorder to the quasi-entropy; the observation is inspired by a similar transition for random tensor networks (as opposed to random quantum circuits), studied in [39]. The typical domain wall configurations before and after twisting boundary conditions in the ordered and disordered phases is reflected in the cartoon in Figure 3.9. As elaborated upon in Appendix 3.A, we expect there to be a correspondence between the scaling of the domain wall size and the free energy cost after twisting the boundary conditions of the stat mech model.

This implies that the quasi-entropy $\tilde{S}_2$ is in the area (resp. volume) law phase when the classical stat mech model is in the disordered (resp. ordered) phase. Heuristically we might expect the runtime of the SEBD algorithm to scale like $\mathrm{poly}(n) \exp\left(O(\tilde{S}_2)\right)$, suggesting that the disorder-to-order transition is accompanied by an efficient-to-inefficient transition in the complexity of the SEBD algorithm. Furthermore, near the transition point within the volume-law phase, the quasi-entropy scales linearly with system size but with a small con-



(a) Ordered phase



(b) Disordered phase

Figure 3.9: The stat mech mapping yields nodes arranged within a roughly $\sqrt{n} \times t \times d$ prism. (a) In the ordered phase, twisting the boundary conditions at the right boundary introduces a domain wall between the two phases (indicated by red and blue) that propagates through the bulk for a total area of $O(td)$. (b) In the disordered phase, boundary conditions introduce bias that is noticeable only within a constant $O(\xi)$ distance of the boundary, and the domain wall membrane introduced by twisting the boundary conditions is quickly washed out by the bulk disorder (dotted purple). The total area is $O(\xi d)$.

stant prefactor, suggesting that the SEBD runtime, though exponential, could be considerably better than previously known exponential-time techniques.

3.6.6 Depth-3 2D circuits with brickwork architecture

Now, we turn our attention specifically to the depth-3 brickwork architecture that we also numerically simulated. In this architecture, three layers of two-qudit gates are performed on a 2D lattice of qudits as shown in Figure 3.10(a). Note that this architecture was also introduced in Section 3.4; the architecture we consider here is exactly the "extended brickwork architecture" of that section with the extension parameter r fixed to be one.

As previously discussed in Section 3.4, this structure is known to be universal in the sense that one may simulate any quantum circuit using a brickwork circuit (with polynomial overhead in the number of qudits) by judiciously choosing which two-qudit gates to perform and performing adaptive measure-

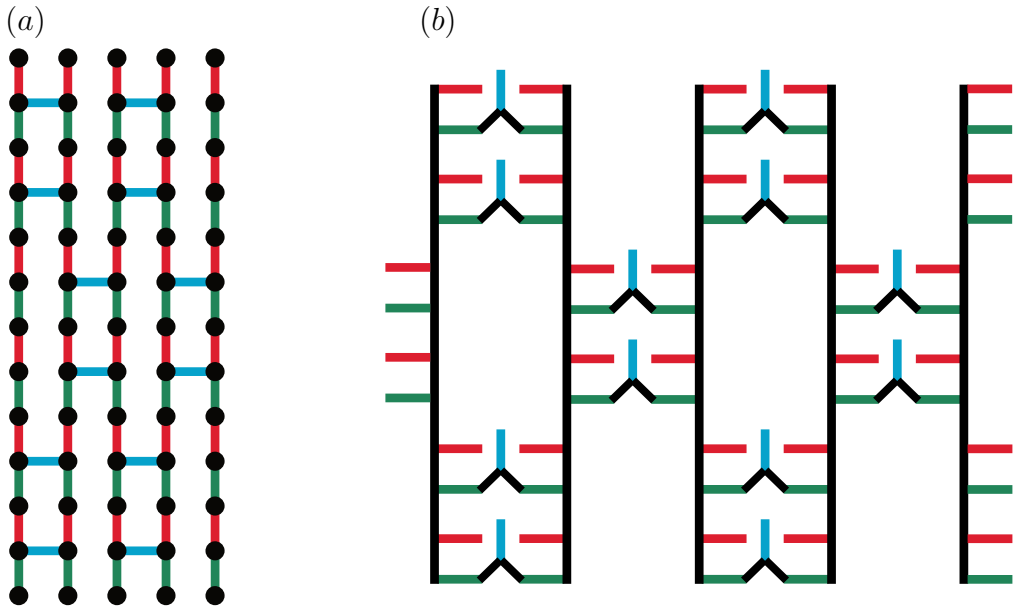


Figure 3.10: Stat mech map for brickwork architecture. (a) The circuit diagram for the brickwork architecture. Qudits lie at location of black dots. Three layers of two-qudit gates act between nearest-neighbor qudits—first qudits linked by a vertical red edge, then vertical green, then horizontal blue. In our `SEBD` simulation of this circuit architecture, we sweep from left to right. (b) Result of stat mech mapping applied to brickwork architecture depicted in (a). Nodes are implied to lie at the endpoints of each edge. Red, green, and blue edges carry Weingarten weight, where the color indicates which gate from (a) it originated from. Black edges carry weight given by $\text{weight}_{\langle uv \rangle}(\sigma, \tau) = q^{C(\sigma_u \tau_v^{-1})}$.

ments [108]. Thus, it is hard to exactly sample or compute the output probabilities of brickwork circuits in the worst case assuming that the polynomial hierarchy does not collapse, and we expect neither the `SEBD` algorithm nor the `Patching` algorithm to be efficient. However, we now give evidence that these algorithms are efficient in the "average-case," where each two-qudit gate is Haar random, by considering the order/disorder properties of the stat mech model that the brickwork architecture maps to.

Stat mech mapping for general k

The stat mech mapping proceeds as previously discussed for 2D circuits, but we will see that the brickwork architecture allows us to make some important simplifications. Each gate in the circuit is replaced by a pair of nodes, which are connected with an edge. Then, the outgoing nodes of the first (red) layer are connected to the incoming nodes of the second (green) layer, and the outgoing nodes of the second (green) layer are connected to the incoming nodes of the

third (blue) layer. The resulting graph is shown in Figure 3.10(b). Edges connecting incoming and outgoing nodes of the same layer are shown in color (red, green, blue) and carry weight equal to the Weingarten function. Edges connecting subsequent layers are black. These edges carry weight given by $\text{weight}_{\langle uv\rangle}(\sigma,\tau) = q^{C(\sigma_u\tau_v^{-1})}$.

To perform the full mapping, we would also add a layer of auxiliary nodes for any unmeasured qudits and connect them to the third layer. However, we are interested primarily in the bulk order-disorder properties of the sytem and suppose that all the qudits, except perhaps those at the boundary of the 2D system, will be measured after the third layer, so we need not consider auxiliary nodes.

Looking at Figure 3.10(b), we see that some of the nodes have degree 1 and connect to the rest of the graph via a (red or blue) Weingarten link. We can immediately decimate these nodes from the graph. For any τ, we have [35]

$$\sum_{\sigma\in\mathcal{S}_k} \mathcal{W}\text{g}(\tau\sigma^{-1},q^2) = \sum_{\sigma\in\mathcal{S}_k} \mathcal{W}\text{g}(\sigma,q^2) = \frac{(q^2-1)!}{(k+q^2-1)!} \tag{3.37}$$

which is independent of τ, so decimating these spins merely contributes the above constant to the total weight. This constant will appear in both the numerator and denominator of quantities like $\mathbb{E}_U(Z_{k,A})/\,\mathbb{E}_U(Z_{k,\emptyset})$, and we ignore them henceforth. The remaining graph can be straightened out, yielding Figure 3.11(a). The fact that Figure 3.11(a) is a graph embedded in a plane that includes only two-body interactions is one upshot of studying the brickwork architecture, as it makes the analysis more straightforward and the stat mech model easier to visualize. This property and the fact that the brickwork architecture is universal for MBQC constitute the primary reasons we studied this architecture in the first place. Architectures with larger depth would lead to stat mech models that cannot be straightforwardly collapsed onto a single plane while maintaining the two-body nature of the interactions.

Simplifications when $k = 2$

As in previous examples, we examine the $k = 2$ case. In this case we might as well decimate all the degree-2 nodes in the graph in Figure 3.11(a). This yields a graph with entirely degree-3 nodes, as shown in Figure 3.11(b). The graph has two kinds of links, both carrying standard Ising interactions. The vertical blue links have weights given by

$$\text{weight}_{\langle uv\rangle}(\sigma) = \begin{cases} q^2(q^2+1) & \text{if } \sigma_u\sigma_v = e \\ q^2(2q) & \text{if } \sigma_u\sigma_v = (12)\,, \end{cases} \tag{3.38}$$

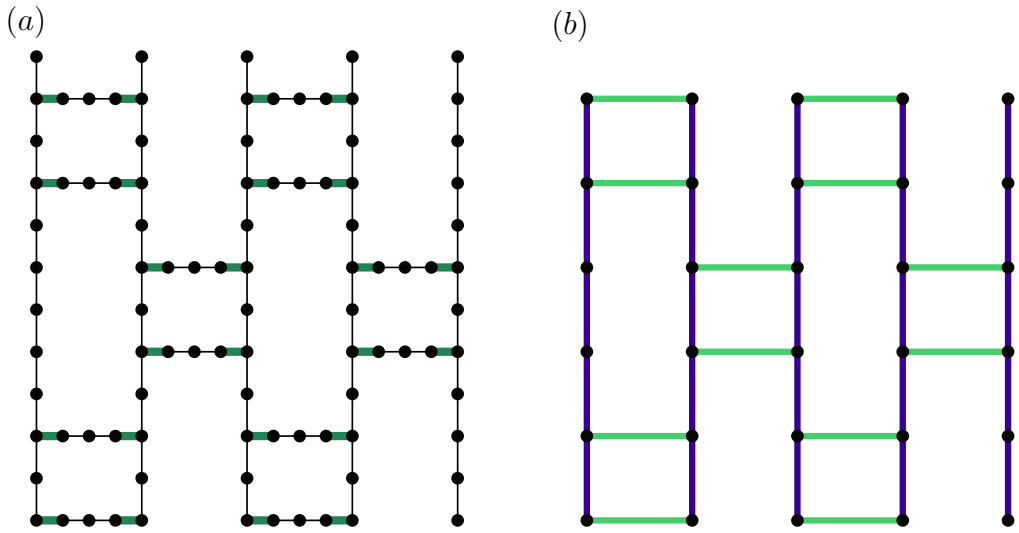


Figure 3.11: Interaction graph for brickwork stat mech system after decimation of some of the particles. (a) The graph that results from decimating degree-1 nodes in Figure 3.10(b). Each thin black link carries weight equal to the function $q^{C(\sigma\tau^{-1})}$ while each thick green link carries weight equal to $\mathcal{W}\mathrm{g}(\sigma\tau^{-1}, q^2)$. (b) The graph that results from decimating nodes of the graph in (a). For $k = 2$, both the horizontal light green and the vertical blue links are ferromagnetic, but have different strengths.

while the horizontal light green links have weights given by

$$\text{weight}_{\langle uv\rangle}(\sigma) = \begin{cases} \frac{1}{q^2(q^4-1)^2}\left(q^6+q^4-4q^3+q^2+1\right) & \text{if } \sigma_u\sigma_v = e \\ \frac{1}{q^2(q^4-1)^2}\left(2q^5-2q^4-2q^2+2q\right) & \text{if } \sigma_u\sigma_v = (12)\,. \end{cases} \tag{3.39}$$

Both of these interactions are ferromagnetic and become stronger as q increases. We may think of the model as the square lattice Ising model for which $1/2$ of the links carry a ferromagnetic interaction of one strength, $1/4$ of the links carry ferromagnetic interactions of another strength, and the final $1/4$ of the links have no interaction at all. The energy functional can be written

$$E/(kT) = -J_{\text{vert}} \sum_{\langle ij\rangle} s_i s_j - J_{\text{horiz}} \sum_{\langle ij\rangle} s_i s_j\,, \tag{3.40}$$

where s_i take on values in $\{+1, -1\}$. For $q = 2$ we have $J_{\text{vert}} = \log(5/4)/2 = 0.112$ and $J_{\text{horiz}} = \log(53/28)/2 = 0.319$. Both of these values are weaker than the critical interaction strength for the square lattice Ising model of $J_{\text{square}} = \log\left(1+\sqrt{2}\right)/2 = 0.441$. This indicates that the graph generated by the stat mech mapping on 2D circuits of depth 3 with brickwork architecture is in the disordered phase when $q = 2$. This remains true for $q = 3$. For $q = 4$, $J_{\text{horiz}} = 0.500 > J_{\text{square}}$, but $J_{\text{vert}} = 0.377 < J_{\text{square}}$. Recall that $1/4$ of the links can be thought to have $J = 0$ since they are missing. Taking this into

account, the value of J averaged over all the links remains below J_{square} for $q = 5$, and slightly exceeds it for $q = 6$.

This indicates that when we run SEBD on these uniform depth-3 circuits with Haar-random gates, the quasi-entropy satisfies $\tilde{S}_2 = O(1)$ (independent of the number of qudits n) when $q = 2$ or $q = 3$ (and probably also for $q = 4$ and $q = 5$). We take this as evidence that SEBD would be efficient for these circuits.

3.7 Future work and open questions

Our work yields several natural follow-up questions and places for potential future work. We list some here.

1. Can ideas from our work also be used to simulate *noisy* 2D quantum circuits? Roughly, we expect that increasing noise in the circuit corresponds to decreasing the interaction strength in the corresponding stat mech model, pushing the model closer toward the disordered phase, which is (heuristically) associated with efficiency of our algorithms. We therefore suspect that if noise is incorporated, there will be a three-dimensional phase diagram depending on circuit depth, qudit dimension, and noise strength. As the noise is increased, our algorithms may therefore be able to simulate larger depths and qudit dimensions than in the noiseless case.

2. Can one approximately simulate random 2D circuits of arbitrary depth? This is the relevant case for Google's quantum computational supremacy experiment [6]. Assuming Conjecture 3.2, our algorithms are not efficient once the depth exceeds some constant, but it is not clear if this difference in apparent complexity for shallow vs. deep circuits is simply an artifact of our simulation method, or if it is inherent to the problem itself.

3. Our algorithms are well defined for all 2D circuits, not only random 2D circuits. Are they also efficient for other kinds of unitary evolution at shallow depths, for example evolution by a fixed local 2D Hamiltonian for a short amount of time?

4. Can we rigorously prove Conjecture 3.1? One way to make progress on this goal would be to find a worst-case-hard uniform circuit family for which it would be possible to perform the analytic continuation of quasi-entropies $\tilde{S}_k$ in the $k \to 1$ limit using the mapping to stat mech models.

5. Can we give numerical evidence for Conjecture 3.2, which claims that our algorithms undergo computational phase transitions? This would require numerically simulating our algorithms for circuit families with increasing local Hilbert space dimension and increasing depth and finding evidence that the algorithms eventually become inefficient.

6. How precisely does the stat mech mapping inform the efficiency of our algorithms? Is the correlation length of the stat mech model associated with the runtime of our simulation algorithms? How well does the phase transition point in the stat mech model (and accompanying phase transition in quasi-entropies) predict the computational phase transition point in the simulation algorithms? If such questions are answered, it may be possible to predict the efficiency and runtime of the simulation algorithms for an arbitrary (and possibly noisy) random circuit distribution via Monte Carlo studies of the associated stat mech model. In this way, the performance of the algorithms could be studied even when direct numerical simulation is not feasible.

7. In the regime where `SEBD` is inefficient, i.e., when the effective 1D dynamics it simulates are on the volume-law side of the entanglement phase transition, is `SEBD` still better than previously known exponential-time methods? Intuitively, we expect this to be the case close to the transition point.

8. Can `SEBD` and/or `Patching` be generalized to simulate shallow circuits in three or higher dimensions? For `SEBD` the natural approach would be to use a PEPS (higher dimensional generalization of MPS) and simulate action of unitary gates and measurements, but PEPS cannot be efficiently contracted or truncated exactly in the same way as MPS.

APPENDIX TO CHAPTER 3

3.A Stat mech mapping for circuits with weak measurements

In Chapter 2, we described the stat mech map in general, and in Section 3.6 we applied it to 2D circuits with projective measurements on some of the qudits *after* the circuit. Here, we generalize the map to allow for weak measurements *during* the circuit. Then, we apply this formalism to a model of 1D random quantum circuits with weak measurements inspired by the CHR problem introduced in the main text. The problem of 1D circuits interspersed with weak measurements was previously studied using this approach in Refs. [44, 45] (however, we analyze a different weak measurement).

3.A.1 Setup and weak measurements

As before, we let our system consist of n qudits of local dimension q. The circuits we consider are specified by a sequence of pairs of qudits (indicating where unitary gates are applied) and single-qudit weak measurements; this sequence can be assembled into a quantum circuit diagram. The single-qudit measurements are each described by a set $\mathcal{M}$ of measurement operators along with a probability distribution μ over the set $\mathcal{M}$. These sets are normalized such that $\operatorname{tr}\big(M^\dagger M\big)$ is constant for all $M \in \mathcal{M}$ and $\mathbb{E}_{M\leftarrow\mu}\, M^\dagger M = \mathbb{I}_q$ where here $\mathbb{I}_q$ denotes the $q \times q$ identity matrix (in other places in this thesis, $\mathbb{I}_q$ has been denoted simply by I). Thus we have $\operatorname{tr}\big(M^\dagger M\big) = q$ for all M. The introduction of a probability measure over $\mathcal{M}$ in our notation, which was also used in [44], is not conventional, but it is equivalent to the standard formulation and will be important for later definitions.

When a measurement is performed, if the state of the system at the time of measurement is σ, the probability of measuring the outcome associated with operator M is $\mu(M)\operatorname{tr}\big(M\sigma M^\dagger\big)$ (Born rule for quantum measurements). For a fixed outcome M, the quantity $\operatorname{tr}\big(M\sigma M^\dagger\big)$ is a function of σ that we refer to as the *relative likelihood* of obtaining the outcome M on the state σ, since it gives the ratio of the probability of obtaining outcome M in the state σ to the probability of obtaining outcome M in the maximally mixed state $\frac{1}{q}\mathbb{I}_q$. After obtaining outcome M, the state is updated by the rule $\sigma \to M\sigma M^\dagger / \operatorname{tr}\big(M\sigma M^\dagger\big)$. Thus a pure initial state remains pure throughout the evolution. For notational convenience and without loss of generality, we will assume that for each u, the uth unitary is immediately followed by single-qudit measurements $(\mathcal{M}_u, \mu_u)$ and $(\mathcal{M}'_u, \mu'_u)$ on the qudits $a_u, a'_u \in [n]$ that are acted upon by the unitary, respectively; in the case no measurement is performed, we may simply take $\mathcal{M}_u$ to consist solely of the identity operator, and in the case that more than one measurement is performed, we may multiply together the sets of measurement operators and their corresponding probability distributions to form a single set describing the overall weak measurement.

Thus, the (non-normalized) output state of the circuit with l unitaries acting on the initial state $|0\dots0\rangle$ can be expressed as

$$\rho = E\,|0\dots0\rangle\langle 0\dots0|\,E^\dagger\,, \tag{3.41}$$

with

$$E = (M_l' M_l U_l)\dots(M_2' M_2 U_2)(M_1' M_1 U_1)\,, \tag{3.42}$$

where each unitary U_u is chosen from the Haar measure over unitaries acting on qudits a_u and a_u', while M_u and M_u' are the measurement operators associated with the measurement outcome obtained upon performing a measurement on qudits a_u and a_u', respectively, following application of unitary U_u.

3.A.2 Generalized interaction weights

As in the main text, we are interested in studying the quasi-entropy $\tilde{S}_k$, where averages are taken over instance U as well as the outcomes of the weak measurements during the circuit. Accordingly, we redefine the expectation value of a quantity Q to be

$$\mathop{\mathbb{E}}_{U}(Q) = \mathop{\mathbb{E}}_{M_1\leftarrow\mu_1}\mathop{\mathbb{E}}_{M_1'\leftarrow\mu_1'}\dots\mathop{\mathbb{E}}_{M_l\leftarrow\mu_1}\mathop{\mathbb{E}}_{M_l'\leftarrow\mu_1'}\int_{U(q^2)} dU_1\dots\int_{U(q^2)} dU_l\; Q\,, \tag{3.43}$$

where $\int_{U(q^2)}$ denotes integration over the Haar measure of the unitary group with dimension q^2. As before, we would like to calculate

$$\tilde{S}_k(A) = \frac{1}{1-k}\log\left(\frac{\mathbb{E}_U(\mathrm{tr}(\rho)^k \frac{Z_{k,A}}{Z_{k,\emptyset}})}{\mathbb{E}_U(\mathrm{tr}(\rho)^k)}\right) \tag{3.44}$$

$$= \frac{1}{1-k}\log\left(\frac{\mathbb{E}_U(Z_{k,A})}{\mathbb{E}_U(Z_{k,\emptyset})}\right) \tag{3.45}$$

$$= \frac{F_{k,\emptyset} - F_{k,A}}{1-k}\,. \tag{3.46}$$

The mapping from the quantities $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$ to partition functions proceeds identically to what was discussed in Chapter 2 and the main text of this chapter—the interaction graph is exactly the same. The only difference is that the weight formula for edges $\langle uv\rangle$ between successive unitaries needs to be updated due to the inclusion of a weak measurement in between unitaries u and v.

Recall that the formula for the partition function comes about by performing Haar integration on each individual two-qudit unitary. Each two-qudit gate in the circuit diagram is then mapped to a weighted sum over terms of the form $|\sigma\rangle^{\otimes 2}\langle\tau|^{\otimes 2}$, which gives rise to the incoming and outgoing particles in the stat mech system. Previously, the $\langle uv\rangle$ edge weights were given by $\langle\tau_v|\sigma_u\rangle$

which evaluates to Eq. (3.32). Now, if weak measurement $(\mathcal{M}, \mu)$ acts between gates u and v, this is modified to

$$\text{weight}_{\langle uv\rangle}(\sigma,\tau) = \mathop{\mathbb{E}}_{M\leftarrow\mu} \langle\tau_v|(M\otimes M^*)^{\otimes k}|\sigma_u\rangle = \mathop{\mathbb{E}}_{M\leftarrow\mu} \text{tr}\big((M^\dagger M)^{\otimes k} W_{\sigma_u\tau_v^{-1}}\big)\,, \tag{3.47}$$

where W_π is the operator that performs the permutation π on the k copies of the system. A similar modification is made to the weights of edges connecting to the auxiliary nodes.

$$\text{weight}_{\langle ua'\rangle}(\sigma,\tau) = \mathop{\mathbb{E}}_{M\leftarrow\mu} \text{tr}\Big((M^\dagger M)^{\otimes k} W_{\sigma_u\chi_{a'}^{-1}}\Big)\,, \tag{3.48}$$

generalizing Eq. (3.33). Later, in Appendix 3.C, we will be interested in expressing entropies of the *classical* output distribution of the circuit in terms of partition functions and to handle this case we will update Eq. (3.48). Note that the quantity $\text{tr}\big(X^{\otimes k}W_\pi\big)$ is equal for all π with the same cycle structure, which corresponds to some partition $\lambda = (\lambda_1, \ldots, \lambda_r)$ of k, where $\sum_i \lambda_i = k$ and $\lambda_1 \geq \ldots \geq \lambda_r > 0$. Then we have

$$\text{tr}\big(X^{\otimes k}W_\pi\big) = \prod_{i=1}^{r} \text{tr}\big(X^{\lambda_i}\big)\,. \tag{3.49}$$

This formula allows us to simplify the weight formulas (3.47) and (3.48) in a few special cases. If no measurement is made, then $\mathcal{M} = \{I\}$ and $\text{weight}_{\langle uv\rangle}(\sigma,\tau) = q^{C(\sigma_u\tau_v^{-1})}$, where $C(\pi)$ is the number of cycles r in the permutation π, recovering the weight equations (3.32) and (3.33) from the main text. On the other hand, if a projective measurement onto one of the q basis states is made, then $\mathcal{M} = \{\sqrt{q}\Pi_m\}_{m=0}^{q-1}$ and μ is the uniform distribution, where $\Pi_m = |m\rangle\langle m|$. Since in this case $\text{tr}\big((M^\dagger M)^w\big) = q^w$ for any power w and any $M \in \mathcal{M}$, we have $\text{weight}_{\langle uv\rangle}(\sigma,\tau) = q^{k-1}$ for any pair σ_u, τ_v.

3.A.3 Mapping applied to 1D circuits with weak measurements

In Section 3.3.3, we discussed the connection between the effective 1D dynamics of our SEBD algorithm and previous work (originating from [57–59] on 1D Haar-random circuits with some form of measurements in between each layer of unitaries.

In this subsection, we apply the stat mech mapping to the 1D with weak measurement model and explain the connection between the area-law-to-volume-law transition that has been observed in numerical simulations and the disorder-to-order thermal transition in the classical stat mech model, which occurs at a non-zero critical temperature T_c. This analysis was first performed in [45] and independently in [44]. The results presented in this section are essentially a reproduction of their analysis but for a different weak measurement, chosen to be relevant for the dynamics of the SEBD algorithm acting on the CHR problem. We include this analysis for two purposes: first, to shed light

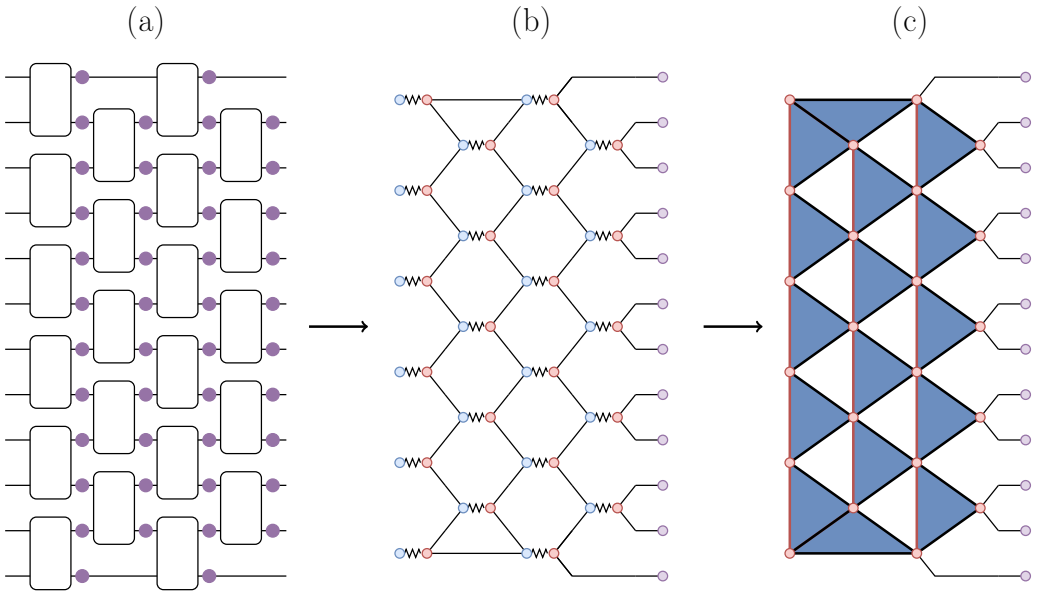


Figure 3.12: Summary of series of maps for Haar-random 1D circuits with weak measurements. (a) The quantum circuit diagram for the unitary plus weak measurement model consists of layers of Haar-random two-qudit gates followed by layers of weak measurements on every qudit, indicated by purple dots. (b) The stat mech mapping results in a model on the honeycomb lattice, where horizontal zigzag links have weight given by the Weingarten function and diagonal straight links have weight that depends on the weak measurement. Blue dots and red dots represent incoming and outgoing nodes, respectively. Light purple dots represent auxiliary nodes. (c) By decimating the incoming (blue) nodes in the honeycomb lattice, we reduce the number of nodes by half and generate a model with three-body interactions living on rightward-pointing triangles, shaded in blue. When $k = 2$ the weights are all positive, and the three-body interaction can be decomposed into an anti-ferromagnetic interaction along vertical (red) links and ferromagnetic interactions along diagonal (black) links.

on the behavior of `SEBD` acting on `CHR`, and second, to serve as a more complete example of the stat mech mapping in action, complementing the more heuristic analysis we give in Section 3.6 of the main text.

Mapping to the honeycomb lattice

Let us assume that our circuit has n qudits of local dimension q arranged on a line with open boundary conditions. A circuit of depth d acts on the qudits where each layer consists of nearest-neighbor two-qudit Haar-random unitaries. In between each layer of unitaries, a weak measurement is performed on every qudit, described by the set $\mathcal{M}$ of measurement operators and a probability

distribution μ over $\mathcal{M}$. The first step of the stat mech mapping is to replace each Haar-random unitary with a pair of nodes and connect these nodes according to the order of the unitaries acting on the qudits. The second step is to introduce a new auxiliary node for each qudit and connect each outgoing node within the final layer of unitaries to the corresponding pair of auxiliary nodes. The resulting graph is the honeycomb lattice, as shown in Figure 3.12(b). We now review what the interactions are on this graph. The horizontal zigzag links in Figure 3.12(b) host interactions that contribute a weight equal to the Weingarten function. When $k = 2$, the interaction depends only on if the pair of nodes agree ($\sigma_u \tau_u^{-1} = e$) or if they disagree ($\sigma_u \tau_u^{-1} = (12)$). In this case the interactions are given explicitly by

$$\text{weight}_{\langle u \rangle}(\sigma) = \mathcal{W}\text{g}(\sigma_u \tau_u^{-1}, q^2) = \begin{cases} \frac{1}{q^4-1} & \text{if } \sigma_u \tau_u^{-1} = e \\ -\frac{1}{q^2(q^4-1)} & \text{if } \sigma_u \tau_u^{-1} = (12). \end{cases} \tag{3.50}$$

Meanwhile, the diagonally oriented links in Figure 3.12(b) host interactions that depend on the details of the weak measurement being applied in between each layer of unitaries, which we now define.

Weak measurement and diagonal weights

The weak measurement we choose is given as follows. First, for a fixed $q \times q$ unitary matrix U, define

$$M_U^{(m)} = \sqrt{q} \cdot \text{diag}(U_{m,\cdot}) \tag{3.51}$$

that is, the $q \times q$ matrix whose diagonal entries are given by the mth row of U, scaled by a factor of $\sqrt{q}$, and whose off-diagonal entries are 0. Define the probability distribution μ_U to be the uniform distribution over the set $\mathcal{M}_U = \{M_U^{(m)}\}_{m=0}^{q-1}$. We can see that $(\mathcal{M}_U, \mu_U)$ forms a valid weak measurement since

$$\sum_{m=0}^{q-1} \mu_U(m) (M_U^{(m)})^\dagger M_U^{(m)} = \sum_{m=0}^{q-1} \text{diag}(|U_{m,\cdot}|^2) = \mathbb{I}_q \tag{3.52}$$

where the last equality follows from the fact that the sum of the squared norms of the entries within a column of a unitary matrix is 1. When $U = \mathbb{I}_q$, the measurement operator $M_U^{(m)}$ is a projector onto the mth basis state (scaled by a factor of $\sqrt{q}$), and the weak measurement is simply a projective measurement onto the computational basis.

The weak measurement that we consider for our analysis will be a mixture of the weak measurement $(\mathcal{M}_U, \mu_U)$ for different U. Formally, we take $\mathcal{M} = \cup_{U \in U(q)} \mathcal{M}_U$. We let the distribution μ over $\mathcal{M}$ be the distribution resulting from drawing U according to the Haar measure, and then drawing M from $\mathcal{M}_U$ uniformly at random.

This weak measurement is seen to exactly reproduce the weak measurement of SEBD acting on CHR in Algorithm 4 when $q = 2$, where the measurement operators were the diagonal matrices

$$M^{(0)} = \begin{pmatrix} \cos(\theta/2) & 0 \\ 0 & e^{-i\phi}\sin(\theta/2) \end{pmatrix} \tag{3.53a}$$

$$M^{(1)} = \begin{pmatrix} \sin(\theta/2) & 0 \\ 0 & e^{i\phi}\cos(\theta/2) \end{pmatrix} \tag{3.53b}$$

with angles (θ, ϕ) drawn according to the Haar measure on the sphere. Indeed, even for $q \neq 2$, this weak measurement arises from a natural generalization of the CHR problem, where one makes Haar-random measurements on a cluster state of higher local dimension, which is created by applying a generalized Hadamard gate to each qudit followed by a generalized CZ gate on each pair of neighboring qudits on the 2D lattice.

To compute the weights on the edges of the stat mech model for $k = 2$, we apply the formula in Eqs. (3.47) and (3.48).

$$\begin{aligned} \text{weight}_{\langle uv\rangle}(\sigma,\tau) &= \int_{U(q)} dU \sum_{m=0}^{q-1} \frac{1}{q} \operatorname{tr}\left(\left((M_U^{(m)})^\dagger M_U^{(m)} \right)^{\otimes 2} W_{\sigma_u \tau_v^{-1}} \right) \\ &= \int_{U(q)} dU q \sum_{m=0}^{q-1} \begin{cases} \operatorname{tr}\left(\operatorname{diag}(|U_{m,\cdot}|^2)\right)^2 & \text{if } \sigma_u\tau_v^{-1} = e \\ \operatorname{tr}(\operatorname{diag}\left(|U_{m,\cdot}|^4\right)) & \text{if } \sigma_u\tau_v^{-1} = (12) \end{cases} \\ &= \begin{cases} q^2 & \text{if } \sigma_u\tau_v^{-1} = e \\ q^2 \cdot w & \text{if } \sigma_u\tau_v^{-1} = (12)\,, \end{cases} \end{aligned} \tag{3.54}$$

with the definition

$$w = \int_{U(q)} dU \sum_m \frac{1}{q} \operatorname{tr}\big(\operatorname{diag}(|U_{m,\cdot}|^4)\big) \tag{3.55}$$

$$= q \int_{U(q)} dU |U_{0,0}|^4 \tag{3.56}$$

$$= q \sum_{\sigma,\tau\in\mathcal{S}_2} \mathcal{W}\mathrm{g}(\sigma\tau^{-1}, q) \tag{3.57}$$

$$= 2q \sum_{\sigma\in\mathcal{S}_2} \mathcal{W}\mathrm{g}(\sigma, q) \tag{3.58}$$

$$= 2q \left(\frac{1}{q^2-1} - \frac{1}{q(q^2-1)} \right) \tag{3.59}$$

$$= \frac{2}{q+1}\,, \tag{3.60}$$

where in the third line we have invoked the Haar integration formula that appears in Eq. (2.10), and then substituted the explicit values for the Weingarten function when $k = 2$. The formula for $\text{weight}_{\langle ua' \rangle}$ is given similarly.

We can see that for all $q > 1$, the weight is larger when the values of the nodes agree than when they disagree, indicating a ferromagnetic Ising interaction. Indeed, the interaction for $k = 2$ will be ferromagnetic regardless of what weak measurement M is made since $\text{tr}\big(M^\dagger M\big)^2 \geq \text{tr}\big((M^\dagger M)^2\big)$ holds for all M. Furthermore, for our choice of weak measurement, the ferromagnetic Ising interaction becomes stronger as q increases.

Eliminating negative weights via decimation when $k = 2$

The possibility of a negative weight on the horizontal edges of the honeycomb lattice in Figure 3.12(b) appears to impede further progress in the analysis since the classical model cannot be viewed as a physical system with real interaction energies at a real temperature. As discussed in the main text, for $k = 2$, this problem may be circumvented by decimating half of the spins; that is, we explicitly perform the sum over $\{\tau_u\}_u$ in the partition function in Eq. (3.34), yielding a new stat mech model involving only the outgoing nodes with assignment σ_u. Since the decimated incoming nodes (except for those in the first layer) each have three neighbors, all three of which are undecimated outgoing nodes, the new model will have a three-body interaction between each such trio of nodes.

We may furthermore observe that, for our choice of weak measurement when $k = 2$, the three-body weight may be re-expressed as the product of three two-body weights acting on the three edges of the triangle. Below we give formulas for the two-body weights; our formulas are a unique decomposition of the three-body interaction up to a shifting of overall constant factors from one link to another. Thus, via decimation we have moved from the honeycomb lattice with two-body interactions to the triangular lattice with two-body interactions, as illustrated in Figure 3.12(c). There are two kinds of two-body interactions on this triangular lattice. Vertically oriented links between nodes u_1 and u_2 host anti-ferromagnetic interactions

$$\text{weight}_{\langle u_1 u_2 \rangle}(\sigma) = \begin{cases} \frac{1}{q^4-1} & \text{if } \sigma_{u_1}\sigma_{u_2} = e \\ \frac{w}{1+q^2}\left((q^2-w^2)(q^2w^2-1)\right)^{-1/2} & \text{if } \sigma_{u_1}\sigma_{u_2} = (12)\,, \end{cases} \tag{3.61}$$

and diagonally oriented links host ferromagnetic interactions, where

$$\text{weight}_{\langle u_1 u_2 \rangle}(\sigma) = \begin{cases} q\sqrt{q^2-w^2} & \text{if } \sigma_{u_1}\sigma_{u_2} = e \\ q\sqrt{w^2q^2-1} & \text{if } \sigma_{u_1}\sigma_{u_2} = (12). \end{cases} \tag{3.62}$$

For all values of the measurement strength p, the ferromagnetic interactions are stronger than the anti-ferromagnetic interaction.

Phase diagram

The model described above for $k = 2$ is exactly the anisotropic Ising model on the triangular lattice. In general this model may be described by its energy functional

$$E/kT = -J_1 \sum_{\langle ij \rangle_1} g_i g_j - J_2 \sum_{\langle ij \rangle_2} g_i g_j - J_3 \sum_{\langle ij \rangle_3} g_i g_j \,, \tag{3.63}$$

where $g_i \in \{+1, -1\}$ are Ising spin variables and the three sums are over links along each of the three triangular axes. This model has been studied and its phase diagram is well understood [117, 118]. In the setting where along two of the axes the interaction strength is equal in magnitude and ferromagnetic, while along the third axis it is weaker in magnitude and antiferromagnetic, the model is known to experience a phase transition as the temperature is varied. At high temperatures, it is in the disordered phase; in other words, samples drawn from the thermal distribution exhibit exponentially decaying correlations between spin values σ_u with a constant correlation length of ξ. At low temperatures, it is in an ordered phase where samples exhibit long-range correlation. At the critical point, the interaction strengths satisfy the equation [117, 118]

$$\sinh(2J_1)\sinh(2J_2) + \sinh(2J_2)\sinh(2J_3) + \sinh(2J_1)\sinh(2J_3) = 1\,. \tag{3.64}$$

For us, the parameter q plays the role of the temperature, and the interaction strengths, derived from Eqs. (3.61) and (3.62), are given by

$$J_1 = J_2 = \frac{1}{4}\log\left(\frac{q^2 - w^2}{w^2 q^2 - 1}\right) \tag{3.65}$$

$$J_3 = -\frac{1}{2}\log\left(\frac{w(q^2-1)}{\sqrt{(q^2-w^2)(q^2w^2-1)}}\right)\,. \tag{3.66}$$

Using these equations, we can solve for the critical point, and we find it to be $q_c = 3.249$. Only integer values of q correspond to valid quantum circuits, so we conclude that the model is disordered when $q = 2$ or $q = 3$ and ordered when $q \geq 4$. We plot this one-dimensional phase diagram in Figure 3.13.

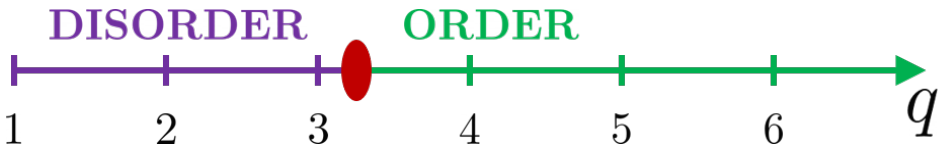


Figure 3.13: Phase diagram showing for which values of q the anisotropic Ising model on the triangular lattice is ordered and disordered. The critical point, indicated by the red dot, occurs at $q_c = 3.249$.

Connection between (dis)order and scaling of entanglement entropy

We expect the scaling of the quantity $\tilde{S}_2 = F_{2,A} - F_{2,\emptyset} = -\log(\mathbb{E}_U(Z_{2,A})/\mathbb{E}_U(Z_{2,\emptyset}))$ to be related to the order or disorder of the model by the following argument. For $\mathbb{E}_U(Z_{2,\emptyset})$, the auxiliary spins are all set to $\chi_a = e$, biasing the bulk spins nearby to prefer e over (12). For $\mathbb{E}_U(Z_{2,A})$, the spins within the region A are twisted so that $\chi_a = (12)$, introducing a domain wall at the boundary. In the ordered phase, the bias introduced at the boundary extends throughout the whole bulk since there is no decay of correlation with distance. The domain wall at the boundary in the calculation of $\mathbb{E}_U(Z_{2,A})$ forces the bulk to separate into two regions with distinct phases separated by a domain wall that cuts through the bulk. The domain wall has length of order $\min(|A|, d)$ where $|A|$ is the number of sites in region A and d is the depth. In the calculation of $\mathbb{E}_U(Z_{2,\emptyset})$, there is no domain wall. The addition of one additional unit of domain wall within a configuration leads the weight of the configuration to decrease by a constant factor, so in the ordered phase we expect $-\log(\mathbb{E}_U(Z_{2,A})/\mathbb{E}_U(Z_{2,\emptyset})) = O(\min(|A|, d))$. Meanwhile, in the disordered phase, there is a natural length scale ξ that boundary effects will penetrate into the bulk. The domain wall at the boundary due to twisted boundary conditions will be washed out by the bulk disorder after a distance on the order of $\xi = O(1)$. Thus we expect $-\log(\mathbb{E}_U(Z_{2,A})/\mathbb{E}_U(Z_{2,\emptyset})) = O(1)$. A cartoon illustrating this logic appears in Figure 3.9 of the main text. For further discussion of the connection between order-disorder properties of the stat mech model and entropic properties of the underlying quantum objects, see Refs. [39, 40, 44, 45].

This logic suggests that, if we take the scaling of $\tilde{S}_2$ to be a good proxy for the scaling of $\langle S_2 \rangle$, the disorder-to-order phase transition in the classical model would be accompanied by an area-law-to-volume-law phase transition in the Rényi-2 entropy of the output of random circuits.

Relationship to numerical simulation of `SEBD` on `CHR`

In Section 3.3.3, with fixed $q = 2$, it was established that the effective dynamics of `SEBD` running on `CHR` are alternating layers of entangling two-qubit CZ gates and weak measurements on every qubit of a 1D line, where the form of the weak measurement is given explicitly. The dynamics we have studied in this section use the same weak measurement, but choose the two-qubit entangling gates to be Haar-random. We have established that the quasi-2-entropy $\tilde{S}_2$ satisfies an area law for this process when $q = 2$, and the statement remains true for $q = 3$ when the weak measurement corresponds to a natural generalization of the `CHR` problem to larger local dimension. For $q = 4$, it is no longer true; the dynamics of $\tilde{S}_2$ satisfy a volume law.

Due to the similarity between the dynamics studied in this section and that of `SEBD` running on `CHR`, our conclusion provides a partial explanation for the numerical observation presented in Section 3.5 that the average entangle-

ment entropy $\langle S_k \rangle$ satisfies an area law when SEBD runs on CHR for $q = 2$ and various values of k.

Additional observations appearing in previous work

The above analysis is essentially a restatement of what appears in recent works by Bao, Choi, and Altman [45] and separately Jian, You, Vasseur, and Ludwig [44], except that here we analyzed a different weak measurement. In particular, Ref. [45] considered the case where a projective measurement occurs with some probability p on each qudit after each layer of unitaries, and otherwise there is no measurement. They made the observation that we describe above that the $k = 2$ mapping can be written as a 2-body anisotropic Ising model on the triangular lattice with an exact solution. Both of these papers went beyond what we have presented here to analyze the $q \to \infty$ limit directly, where they observed that the stat mech model becomes a standard ferromagnetic Potts model on the square lattice for all integers k. For $k = 2$ this is exactly the square lattice Ising model and indeed, we can see from Eq. (3.66) that when $q \to \infty$, $J_3 \to 0$; the anti-ferromagnetic links along one axis vanish leaving a square lattice with exclusively ferromagnetic interactions. The fact that the model becomes tractable for all integers $k \geq 2$ allows these papers to invoke analytic continuation and make sense of the $k \to 1$ limit, where the quasi-entropy $\tilde{S}_k$ exactly becomes the expected von Neumann entropy $\langle S \rangle$.

3.B Patching

We now describe a second algorithm for sampling from the output distributions and computing output probabilities of 2D quantum circuits acting on qudits of local dimension q. While the SEBD algorithm described in the previous section is efficient if the corresponding effective 1D dynamics can be efficiently simulated with TEBD, the algorithm of this section is efficient if the circuit depth d and local dimension q are constant and the conditional mutual information (CMI) of the classical output distribution is exponentially decaying in a sense that we make precise below. In Appendix 3.C we will give evidence that the output distribution of sufficiently shallow random 2D circuits acting on qudits of sufficiently small dimension satisfies such a property with high probability, and the property is not satisfied if the circuit depth or local dimension exceeds some critical constant value.

The algorithm we describe is an adaptation and simplification of the Gibbs state preparation algorithm of Ref. [119]. In that paper, the authors essentially showed that a quantum Gibbs state defined on a lattice can be prepared by a quasipolynomial time quantum algorithm, if the Gibbs state satisfies two properties: (1) exponential decay of correlations and (2) exponentially decaying quantum conditional mutual information for shielded regions. Our situation is simpler than the one considered in that paper, due to the fact that sufficiently separated regions of the lattice are causally disconnected as a re-

sult of the fact that the circuit inducing the distribution is constant-depth and therefore has a constant-radius lightcone. The structure of our algorithm is very similar to theirs, except we can make some simplifications and substantial improvements as a result of the constant-radius lightcone and the fact that we are sampling from a classical distribution rather than a quantum Gibbs state.

Before we describe the algorithm, we set some notation. Let Λ denote the set of all qudits of a $L_1 \times L_2$ rectangular grid (assume that $L_1 \leq L_2 \leq \text{poly}(L_1)$). If A and B are two subsets of qudits of Λ, we define $\text{dist}(A, B) = \min_{i \in A, j \in B} \text{dist}(i, j)$, where $\text{dist}(i, j)$ is the distance between sites i and j as measured by the ∞-norm. There are two primary facts that our algorithm relies on. First, if the circuit has depth d, any two sets of qudits separated by a distance greater than $2d$ have non-overlapping lightcones. Hence, if A and B are two lattice regions separated by distance at least $2d$, and ρ is the quantum state output by the circuit (before measurement), it holds that $\rho_{AB} = \rho_A \otimes \rho_B$ and therefore $\mathcal{D}_{AB} = \mathcal{D}_A \otimes \mathcal{D}_B$ if $\mathcal{D} = \sum_{\mathbf{x}} \mathcal{D}(\mathbf{x}) |\mathbf{x}\rangle\langle\mathbf{x}|$ is the classical output distribution of the circuit and (for example) $\mathcal{D}_A$ denotes the marginal of $\mathcal{D}$ on subregion A. (Note that our notation is slightly different in this section – we now use subscripts on $\mathcal{D}$ to denote marginals, and the dependence of $\mathcal{D}$ on the circuit instance is left implicit.) Second, if the classical CMI $I(X : Z|Y)_p$ of three random variables with joint distribution p_{XYZ} is small, then p_{XYZ} is close to the distribution $p_{X|Y} p_Y p_{Z|Y}$ corresponding to a Markov chain $X - Y - Z$. We state this more formally as the following lemma, which follows from the Pinsker inequality.

Lemma 3.9 (see e.g., [120])**.** *Let X, Y, Z be discrete random variables, and let p_{XYZ} denote their joint distribution. Then*

$$I(X : Z|Y)_p \geq \frac{1}{2 \ln 2} \| p_{XYZ} - p_{X|Y} p_Y p_{Z|Y} \|_1^2.$$

Following [119], we also formally define a notion of CMI decay.

Definition 3.3 (Markov property)**.** *Let p denote a probability distribution supported on Λ. Then p is said to satisfy the* $\delta(l)$-Markov condition *if, for any tripartition of a subregion X of the lattice into subregions $X = A \cup B \cup C$ such that* $\text{dist}(A, C) \geq l$*, we have*

$$I(A : C|B)_p \leq \delta(l). \tag{3.67}$$

Intuitively, our algorithm works by first sampling from the marginal distributions of spatially separated patches on the lattice, and then stitching the patches together to approximately obtain a sample from the global distribution. For a $O(1)$-depth circuit whose output distribution has exponentially decaying CMI, the efficiency of this procedure is guaranteed by the two facts above. We now show this more formally.

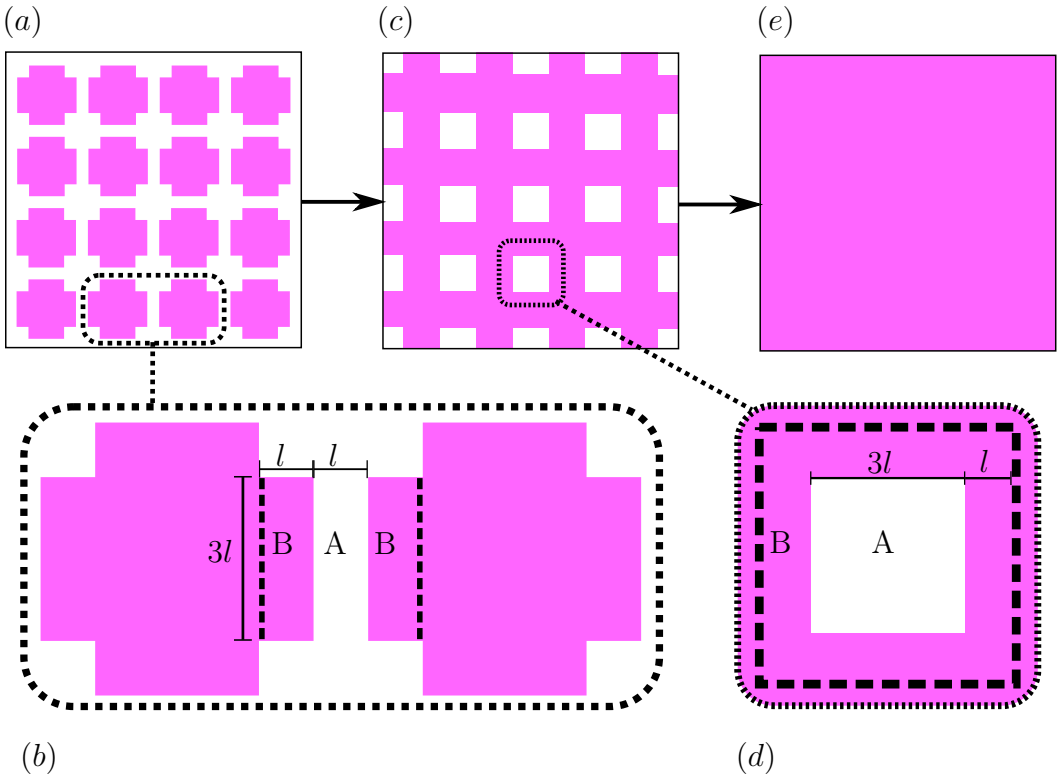


Figure 3.14: `Patching`. Pink represents marginals of the output distribution that have been approximately sampled, while white represents unsampled regions. In (a), the algorithm has sampled from disconnected patches. Figure (b) depicts how the algorithm transitions from configuration (a) to (c). Namely, the algorithm generates a sample from the conditional distribution on A, conditioned on the configuration of region B. Similarly, figure (d) depicts how the "holes" of configuration (c) are filled in. The end result is shown in (e), an approximate sample from the global distribution on the full lattice.

Theorem 3.1. *Suppose C is a 2-local quantum circuit of depth d defined on a 2D rectangular grid Λ of $n = L_1 \times L_2$ qudits, and let $\mathcal{D}(\mathbf{x}) = |\langle \mathbf{x}|C|0\rangle^{\otimes n}|^2$ denote its output distribution. Then if $\mathcal{D}$ satisfies the $\delta(l)$-Markov condition, for any integer $l > 2d$ `Patching` with a length-scale parameter l runs in time $nq^{O(dl)}$ and samples from some distribution $\mathcal{D}'$ that satisfies $\|\mathcal{D}' - \mathcal{D}\|_1 \leq O(1)(n/l^2)\sqrt{\delta(l)}$.*

In particular, if $d = O(1)$, $q = O(1)$, and $\mathcal{D}$ is $\mathrm{poly}(n)e^{-\Omega(l)}$-Markov, then for any polynomial $r(n)$, for some choice of lengthscale parameter `Patching` runs in time $\mathrm{poly}(n)$ and samples from a distribution that is $1/r(n)$-close to $\mathcal{D}$ in total variation distance.

Proof. The algorithm proceeds in three steps, illustrated in Figure 3.14. First, for each square subregion R_i shaded in Figure 3.14(a) with $i \in [O(n/l^2)]$,

sample from $\mathcal{D}_{R_i}$, the marginal distribution of $\mathcal{D}$ on subregion R_i. To do this, first restrict to the qudits and gates in the lightcone of R_i. Sampling from the output distribution on R_i produced by this restricted version of the circuit is equivalent to sampling from the marginal on R_i of the true distribution produced by the full circuit. Since $l > 2d$, this restriction of the circuit is contained in a sublattice of dimensions $O(l) \times O(l)$. Using standard tensor network methods [52], sampling from the output distribution of this restricted circuit on R_i can be performed in time $q^{O(dl)}$. Since there are $O(n/l^2)$ patches, this step can be performed in time $nq^{O(dl)}$. After performing this step, we have prepared the state $\mathcal{D}_{R_1} \otimes \cdots \otimes \mathcal{D}_{R_k} = \mathcal{D}_{R_1,\ldots,R_k}$ where the equality holds because the patches are separated by $l > 2d$ and are therefore mutually independent.

In the second step, we apply "recovery maps" to approximately prepare a sample from the larger, connected lattice subregion S shaded in Figure 3.14(c). The prescription for these recovery maps is given in Figure 3.14(b). Referring to this figure, a recovery map $\mathcal{R}_{B\to AB}$ is applied to generate a sample from subregion A, conditioned on the state of region B. Explicitly, the mapping is given by linearly extending the map $\mathcal{R}_{B\to AB}(|b\rangle\langle b|_B) = \sum_a \mathcal{D}_{A|B}(a|b)\, |a\rangle\langle a|_A \otimes |b\rangle\langle b|_B$. Note that, for a tripartite distribution $\mathcal{D}_{ABC}$, $\mathcal{R}_{B\to AB}(\mathcal{D}_{BC}) = \mathcal{D}_{A|B}\mathcal{D}_B\mathcal{D}_{C|B}$. To implement this recovery map, one can again restrict to gates in the lightcone of region AB and utilize standard tensor network simulation algorithms to generate a sample from the marginal distribution on A, conditioned on the (previously sampled) state of B. The time complexity for this step is again $q^{O(dl)}$. After applying this and $O(n/l^2)$ similar recovery maps, we obtain a sample from a distribution $\mathcal{D}'_S$. By Lemma 3.9, the triangle inequality, and Definition 3.3, the error of this step is bounded as

$$\|\mathcal{D}'_S - \mathcal{D}_S\|_1 \leq O(1)(n/l^2)\sqrt{\delta(l)} = O(1)(n/l^2)\sqrt{\delta(l)}. \tag{3.68}$$

Note that the fact that the errors caused by recovery maps acting on disjoint regions accumulate at most linearly has been referred to previously [119] as the "union property" for recovery maps. The final step is very similar to the previous step. We now apply recovery maps, described by Figure 3.14(d), to fill in the "holes" of the subregion S and approximately obtain a sample from the full distribution $\mathcal{D} = \mathcal{D}_\Lambda$. By a similar analysis, we find that the error incurred in this step is again $O(1)(n/l^2)\sqrt{\delta(l)}$, and therefore the procedure samples from a distribution $\mathcal{D}'_\Lambda$ for which $\|\mathcal{D}'_\Lambda - \mathcal{D}_\Lambda\|_1 \leq O(1)(n/l^2)\sqrt{\delta(l)}$.

The second paragraph of the theorem follows immediately by choosing a suitable $l = \Theta(\log n)$. □

A straightforward application of Markov's inequality implies that a polynomial-time algorithm for sampling with error $1/\operatorname{poly}(n)$ succeeds with high probability over a random circuit instance if the output distribution CMI is exponentially decaying in expectation. We formalize this as the following corollary.

Corollary 3.6. *Let $\mathcal{C}$ be a random circuit distribution. Define $\mathcal{C}$ to be $\delta(l)$-Markov if, for any tripartition of a subregion X of the lattice into subregions $X = A \cup B \cup C$ such that dist$(A, C) \geq l$, we have*

$$\langle I(A : C|B)_{\mathcal{D}} \rangle \leq \delta(l) \tag{3.69}$$

where the angle brackets denote an average over circuit realizations and $\mathcal{D}$ is the associated classical output distribution. Then if $d = O(1), q = O(1)$, and $\mathcal{C}$ is $\text{poly}(n)e^{-\Omega(l)}$-Markov, then for any polynomials $r(n)$ and $s(n)$, `Patching` can run in time $\text{poly}(n)$ and, with probability $1 - 1/s(n)$ over the random circuit realization, sample from a distribution that is $1/r(n)$-close to the true output distribution in variational distance.

Thus, proving that some uniform worst-case-hard circuit family $\mathcal{C}$ is $\text{poly}(n)e^{-\Omega(l)}$-Markov provides another route to proving the part of Conjecture 3.1 about sampling with small total variation distance error. In Section 3.6, we will give analytical evidence that if $\mathcal{C}$ is a random circuit distribution of sufficiently low depth and small qudit dimension, then $\mathcal{C}$ is indeed $\text{poly}(n)e^{-\Omega(l)}$-Markov, and if the depth or qudit dimension becomes sufficiently large, then $\mathcal{C}$ is not $\text{poly}(n)f(l)$-Markov for any $f(l) = o(1)$, supporting Conjecture 3.2, which states that our algorithms exhibit computational phase transitions.

Finally, we note that `Patching` can also be used to estimate specific output probabilities of a random circuit instance C with high probability if C is drawn from a distribution $\mathcal{C}$ that is $\text{poly}(n)e^{-\Omega(l)}$-Markov. This shows that the Markov condition could also be used to prove the second part of Conjecture 3.1 regarding computing output probabilities with small error. This is similar to how `SEBD` can also be used to compute output probabilities, as discussed in Section 3.3.2.

Lemma 3.10. *Let $\mathcal{C}$ be a circuit distribution over constant depth d and constant qudit dimension q 2D circuits on n qudits which is $\text{poly}(n)e^{-\Omega(l)}$-Markov and invariant under application of a final layer of arbitrary single-qudit gates. Then for a circuit instance C drawn from $\mathcal{C}$ and a fixed $\mathbf{x} \in [q]^n$, a variant of `Patching` can be used to output a number $\mathcal{D}'(\mathbf{x})$ in time $\text{poly}(n)$ that satisfies*

$$|\mathcal{D}'(\mathbf{x}) - \mathcal{D}(\mathbf{x})| \leq q^{-n}/r(n) \tag{3.70}$$

with probability $1 - 1/s(n)$ for any polynomials $r(n)$ and $s(n)$, where $\mathcal{D}$ is the output distribution associated with C.

Proof. With probability $1 - 1/\text{poly}(n)$ over the circuit instance C, `Patching` with some choice of lengthscale $l = \Theta(\log n)$ efficiently samples from a distribution $\mathcal{D}'_C$ that is $1/\text{poly}(n)$-close in variational distance to $\mathcal{D}_C$ for any

choice of polynomials. Hence, for an output probability $\mathbf{y}$ chosen uniformly at random and a circuit C drawn from $\mathcal{C}$, it holds that

$$\mathop{\mathbb{E}}_{\mathbf{y}} \mathop{\mathbb{E}}_{C} |\mathcal{D}'(\mathbf{y}) - \mathcal{D}(\mathbf{y})| \leq q^{-n}/\operatorname{poly}(n) \tag{3.71}$$

if $l = c \log n$ and c is a sufficiently large constant. By a nearly identical argument to that used in the proof of Corollary 3.3, due to the invariance of $\mathcal{C}$ under application of a final layer of single qudit gates, for some fixed $\mathbf{x} \in [q]^n$ we also have

$$\mathop{\mathbb{E}}_{C} |\mathcal{D}'(\mathbf{x}) - \mathcal{D}(\mathbf{x})| \leq q^{-n}/\operatorname{poly}(n) \tag{3.72}$$

for any choice of polynomial. Finally, it is straightforward to see that an instance of `Patching` that samples from $\mathcal{D}'$ can also be used to exactly compute $\mathcal{D}'(\mathbf{x})$ for any $\mathbf{x}$. (To do this, the algorithm computes conditional probabilities via tensor network contractions as before, except instead of using these conditional probabilities to sample, it simply multiplies them together similarly to how `SEBD` can be used to compute output probabilities.) Applying Markov's inequality completes the proof. □

3.C Efficiency of `Patching` algorithm from stat mech

We now study the predictions of the stat mech model for the fate of the `Patching` algorithm we introduced in Appendix 3.B. To do so, we in turn study the predictions of the stat mech model for entropic properties of the *classical* output distribution, as `Patching` is efficient if the CMI of the classical output distribution is exponentially decaying with respect to shielded regions.

We have previously applied the stat mech model to study expected entropies of quantum states. However, we now wish to study expected entropies of the classical output distribution. To this end, we now consider the non-unitary quantum circuit consisting of the original, unitary circuit followed by a layer of dephasing channels applied to every qudit. The resulting mixed state is classical (i.e., diagonal in the computational basis) and is exactly equal to the output distribution we want to study. That is, the state after application of the dephasing channels is $\sum_{\mathbf{x}} \mathcal{D}(\mathbf{x}) |\mathbf{x}\rangle\langle\mathbf{x}|$ where $\mathcal{D}$ is the output distribution of the circuit. Note that the application of the dephasing channel is not described in the formalism we have discussed previously, but is easily incorporated. In particular, we need to compute the weights between the auxiliary node with assignment $\chi_{a'}$ and the corresponding outgoing node with assignment σ_u associated with the gate that is the last in the circuit to act on qudit a. We may update Eq. (3.48) and compute the following, recalling the definition of

$|\nu\rangle$ from Eq. (2.14).

$$\begin{aligned}\text{weight}_{\langle ua'\rangle}(\sigma,\tau) &= \langle\sigma_u|\left(\sum_{i=0}^{q-1}|i\rangle\langle i|\otimes|i\rangle\langle i|\right)^{\otimes k}|\chi_{a'}\rangle \\ &= \sum_{i_1,\ldots,i_k}\langle i_1,\ldots,i_k|W_{\sigma_u^{-1}}|i_1,\ldots,i_k\rangle\ \langle i_1,\ldots,i_k|W_{\chi_{a'}}|i_1,\ldots,i_k\rangle\,.\end{aligned} \tag{3.73}$$

We therefore see that $\text{weight}_{\langle ua'\rangle}(\sigma,\tau)$ in this setting is exactly equal to the number of k-tuples of indices $(i_1,\ldots,i_k)$ with $i_j\in[q]$ that are invariant under both permutation operators $\sigma_u,\chi_{a'}\in\mathcal{S}_k$ acting as $\sigma_u\cdot(i_1,\ldots,i_k)=(i_{\sigma(1)},\ldots,i_{\sigma(k)})$. In fact, for our purposes, the auxiliary spin $\chi_{a'}$ will either be set to the identity e or to the k-cycle permutation $(1\ldots k)$. In the former case, the weight reduces to $\text{tr}(W_{\sigma_u})=q^{C(\sigma_u)}$. In the latter case, since the only tuples that are invariant under application of the cycle permutation $(1\ldots k)$ are the q tuples of the form $(x,x,\ldots,x)$ for $x\in[q]$, the weight is simply q for all σ_u. Summarizing,

$$\text{weight}_{\langle ua'\rangle}(\sigma,\tau)=\begin{cases}q^{C(\sigma_u)}, & \chi_{a'}=e\\ q, & \chi_{a'}=(1\ldots k).\end{cases} \tag{3.74}$$

From these expressions, we may immediately note the following facts. First, flipping some auxiliary spin from e to $(1\ldots k)$ cannot increase the weight of a configuration, and hence such a flip corresponds to an increase in free energy. Second, if an auxiliary spin is in the $(1\ldots k)$ configuration, then the auxiliary spin may be effectively removed from the system since in this case the contribution of the auxiliary spin to the weight of a configuration is constant across all configurations.

With these modified weights, we may now compute "quasi-entropies" $\tilde{S}_k(X)$ as before, where now in the $k\to 1$ limit $\tilde{S}_k(X)$ approaches the expected Shannon entropy of the marginal of the output distribution on subregion X, $\langle S(X)_{\mathcal{D}}\rangle$, where the average is over random circuit instances.

Disordered stat mech model suggests `Patching` is successful

We consider the quasi-CMI defined by

$$\tilde{I}_2(A:C|B)=\tilde{S}_2(AB)+\tilde{S}_2(BC)-\tilde{S}_2(B)-\tilde{S}_2(ABC), \tag{3.75}$$

where all quasi-entropies are taken with respect to the collection of classical output distributions that arise from the quantum circuit architecture. This definition is in analogy to the definition of CMI as $I(A:C|B)=S(AB)+S(BC)-S(B)-S(ABC)$ [120]. Note that we may define the quasi-k-CMI $\tilde{I}_k(A:C|B)$ analogously for any nonnegative k, and it holds that $\langle I(A:$

$C|B)_{\mathcal{D}}\rangle = \lim_{k\to 1} \tilde{I}_k(A:C|B)$ where the angle brackets denote an expectation over random circuit instances.

Recalling that $\tilde{S}_2(X) = F_{2,X} - F_{2,\emptyset}$, we may rewrite the quasi-2-CMI as

$$\tilde{I}_2(A:C|B) = (F_{2,AB} - F_{2,B}) - (F_{2,ABC} - F_{2,BC}). \tag{3.76}$$

In stat mech language, the quasi-CMI is essentially the difference in free energy costs of twisting the boundary condition of subregion A in the case where (1) no other spins have boundary conditions, and the case where (2) subregion C also has an imposed boundary condition.

Now, consider some random circuit family $\mathcal{C}$ with associated stat mech model that is in the disordered phase for $k = 2$. For any subregion X of qudits, and partition of X into subregions $X = A \cup B \cup C$, we expect this difference between free energy costs will decay exponentially with the separation between A and C as

$$\tilde{I}_2(A:C|B) \leq \text{poly}(n,q)e^{-\text{dist}(A,C)/\xi} \tag{3.77}$$

where ξ is a correlation length. This is because in the disordered phase of the stat mech model, information about the boundary of region C will be exponentially attenuated as the distance from region C grows. If we take $\tilde{I}_2(A:C|B)$ as a proxy for the average CMI of the output distribution, $\langle I(A:C|B)_{\mathcal{D}}\rangle$, we conclude that the random circuit family $\mathcal{C}$ is $\text{poly}(n,q)e^{-\Theta(l)}$-Markov as defined in Appendix 3.B. The results of that section then show that `Patching` can be used to efficiency sample from the output distribution and estimate output probabilities with high precision and high probability. We take this exponential decay of quasi-2-CMI as evidence that the average CMI also decays exponentially, and therefore that `Patching` is successful. Recall from that main text that the (worst-case-hard) depth-3 brickwork architecture's associated stat mech model is disordered; we therefore expect `Patching` to be capable of efficiently simulating this architecture.

Ordered stat mech model suggests `Patching` is unsuccessful

We first obtain exact, closed form results in the zero-temperature limit of the stat mech model, which corresponds to the $q \to \infty$ limit. However, we expect that qualitatively similar results hold outside of this limit.

As before, consider the stat mech model obtained by applying dephasing channels to all qudits after the application of all gates. Consider some connected, strict subset A of qudits on the original grid. Suppose we are interested in the quasi-entropy $\tilde{S}_k(A) = (F_{k,A} - F_{k,\emptyset})/(k-1)$ of the output distribution on this region. This quantity is given by the free energy cost of twisting the boundary conditions (auxiliary spins) associated with region A from e to $(1 \dots k)$. The auxiliary spins associated with qudits in the complement of A are fixed to be in the identity permutation configuration, e. For both sets of boundary conditions, all non-auxiliary spins will order in the configuration e.

This is because the configuration e maximizes the weights in Eq. (3.74) for spins connected to auxiliary spins in the configuration e, and the weight of a spin connected to an auxiliary spin in the configuration $(1 \dots k)$ is independent of that spin's configuration. Hence, regardless of the configuration of the auxiliary spins, all bulk spins are in the identity permutation configuration in the $q \to \infty$ limit of infinitely strong couplings.

Therefore, twisting a single auxiliary spin from e to $(1 \dots k)$ results in a reduction of the total weight by a factor of $q/q^{C(e)} = q/q^k = q^{1-k}$, corresponding to a free energy increase of $(k-1)\log(q)$. We therefore compute

$$\tilde{S}_k(A) = \frac{F_{k,A} - F_{k,\emptyset}}{k-1} = |A| \log(q). \tag{3.78}$$

Note that this result is exact in the $q \to \infty$ limit. Notably, we find that all integer quasi-entropies are equal in this limit, and so we may trivially perform the analytic continuation to the von Neumann (i.e. Shannon) entropy:

$$\langle S(A) \rangle = \lim_{k \to 1} |A| \log(q) = |A| \log(q). \tag{3.79}$$

Hence, in the $q \to \infty$ limit, the entropy of a strict subregion of the output distribution is maximal.

Now, let X denote the set of *all* qudits. We want to compute $\langle S(X) \rangle$. We again proceed by computing the quasi-entropies:

$$\tilde{S}_k(X) = \frac{F_{k,X} - F_{k,\emptyset}}{k-1}.$$

As before, for each auxiliary spin associated with region X that we "twist," the weight of the configuration is decreased by a factor of q^{1-k} relative to the configuration in which all auxiliary spins are set to e. However, in this case, as opposed to our previous calculation, *all* of the auxiliary spins are twisted. Recall from Eq. (3.74) that the weight between a twisted auxiliary spin and a bulk spin is independent of the value of the bulk spin. Hence, if all auxiliary spins are twisted, the lowest energy state in the bulk is no longer just the configuration in which all spins take the value e—in the absence of a symmetry-breaking boundary condition, there is now a global spin-flip symmetry and the ground space is $k!$-fold degenerate, consisting of all configurations in which all bulk spins are aligned. This symmetry contributes a factor of $k!$ to the partition function and $-\log(k!)$ to the free energy. We hence calculate

$$\tilde{S}_k(X) = |A| \log(q) - \frac{\log(k!)}{k-1}. \tag{3.80}$$

We now perform the analytic continuation to the Shannon entropy:

$$\langle S(X)\rangle = \lim_{k\to 1} \tilde{S}_k(X) \tag{3.81}$$

$$= |A|\log(q) - \lim_{k\to 1} \frac{\log(k!)}{k-1} \tag{3.82}$$

$$= |A|\log(q) - \frac{1-\gamma}{\ln(2)} \tag{3.83}$$

$$\approx |A|\log(q) - 0.61, \tag{3.84}$$

where $\gamma \approx 0.557$ denotes the Euler constant. The expected Shannon entropy of the output distribution is therefore $\frac{1-\gamma}{\ln(2)}$ less bits than maximal in the low-temperature limit, corresponding to $q \to \infty$.

From the above facts, we can immediately compute the expected CMI of the output distribution in this limit. Let (A, B, C) be any partition of the qudits. We have

$$\langle I(A:C|B)_{\mathcal{D}}\rangle \tag{3.85}$$

$$\equiv \langle S(AB)_{\mathcal{D}} + S(BC)_{\mathcal{D}} - S(B)_{\mathcal{D}} - S(ABC)_{\mathcal{D}}\rangle \tag{3.86}$$

$$= [(|A|+|B|)\log(q)] + [(|B|+|C|)\log(q)] \tag{3.87}$$

$$- [(|B|)\log(q)] - [(|A|+|B|+|C|)\log(q) - \frac{1-\gamma}{\ln(2)}]$$

$$= \frac{1-\gamma}{\ln(2)} \approx 0.61. \tag{3.88}$$

We therefore find that in this limit, the expected CMI of the classical output distribution approaches a constant equal to $\frac{1-\gamma}{\ln(2)}$. While this result was derived with respect to the *completely* ordered stat mech model, corresponding to $q \to \infty$, we expect similar behavior for ordered stat mech models in general. In particular, if X denotes the set of all qudits, in the case of an ordered k^{th}-order stat mech model, $\tilde{S}_k(X)$ will similarly receive an extra contribution corresponding to the global spin-flip symmetry, which will also be contributed to the corresponding quasi-CMI $\tilde{I}_k(A:C|B)_{\mathcal{D}}$. Hence, we do not expect the quasi-CMIs to decay when the corresponding stat mech model is in an ordered phase. We take this as evidence that the average CMI does not decay, and therefore that `Patching` is not successful in efficiently sampling from the output distribution with small error.

3.D Relation to worst-to-average-case reductions based on truncated Taylor series

It was shown [25] that for any constant-depth random circuit family with Haar-random gates acting on n qubits for which it is #P-hard to compute output probabilities in the worst case, there does not exist a poly(n)-time algorithm for computing the output probability of some arbitrary output string

$\mathbf{x}$ up to additive error $2^{-\tilde{\Theta}(n^3)}$ with high probability over the circuit realization, unless there exists a poly(n)-time randomized algorithm for computing a #P-hard function. (Note: in even more recent work using the same technique, the error robustness has been improved from $2^{-\tilde{\Theta}(n^3)}$ to $2^{-\Theta(n\log(n))}$ [26, 27].) Essentially, for Haar-random circuits, near-exact average-case computation of output probabilities is as hard as worst-case computation of output probabilities. Our complexity separation in Section 3.4 shows that the error tolerance for this hardness result cannot be improved to $2^{-n}/2^{n^c}$ for any $c < 1$.

This hardness result builds on and improves other prior work [4] on the average-case hardness of random circuit simulation. In particular, the original paper [4] uses a different interpolation scheme than that used in Ref. [25] to perform the worst-to-average-case reduction. Interestingly, as discussed below, we find that the interpolation scheme of Ref. [4] cannot be used to prove hardness results about our algorithms' performance on a shallow random 2D quantum circuit possessing worst-case hardness for computing output probabilities; this essentially is a consequence of how `SEBD` and `Patching` exploit the unitarity of the circuit to be simulated. While this observation may be of technical interest for future work on worst-to-average-case reductions for quantum circuit simulations, the alternative interpolation scheme of Ref. [25] does not suffer from this limitation.

While Refs. [4, 25] prove hardness results for the near-exact computation of output probabilites of random circuits, it is ultimately desirable to prove hardness for the Random Circuit Sampling (RCS) problem of sampling from the output distribution of a random circuit with small error in variational distance, as this is the computational task corresponding to the problem that the quantum computer solves. *A priori*, one might hope that such a result could be proved via such a worst-to-average-case reduction. In particular, it was pointed out in these works that improving the error tolerance of the hardness result to $2^{-n}/\,\mathrm{poly}(n)$ would be sufficient to prove hardness of RCS. Our work rules out such a proof strategy working in general by showing that even improving the error tolerance to $2^{-n}/2^{n^c}$ for any constant $c < 1$ is unachievable. In particular, any proof of the hardness of RCS should be sensitive to the depth and should not be applicable to the worst-case-hard shallow random circuit ensembles that admit approximate average-case classical simulations.

Implications for reductions based on truncated Taylor series

In this section, we discuss the relation between our algorithms (`SEBD` and `Patching`) applied to the computation of output probabilities and the recent result [4] on the hardness of average-case simulation of random circuits based on polynomial interpolation via truncated Taylor series. In particular, we discuss how this polynomial interpolation argument is insufficient to show that the task of even *exactly* computing output probabilities and sampling from the output distribution of a constant-depth Haar-random circuit instance with high probability using our algorithms is classically hard, even

though these circuits possess worst-case hardness. We first briefly review their technique before discussing a limitation in the robustness of the polynomial interpolation scheme. We then discuss how this robustness limitation makes the interpolation scheme inapplicable to our algorithms.

The main point is that our algorithms exploit unitarity (via the fact that gates outside of the lightcone of the qudits currently under consideration are ignored), but the hardness result of Ref. [4] holds with respect to circuit families that are non-unitary, albeit very close to unitary in some sense. Our algorithms are unable to simulate these slightly non-unitary circuits to the precision required for the worst-to-average case reduction, regardless of how well they can simulate Haar-random circuit families. While it is true that in this scheme there is an adjustable parameter K which, when increased, brings the non-unitary circuit family closer to approximating the true Haar-random family, increasing K also increases the degree of the interpolating polynomial. This makes the interpolation more sensitive to errors in such a way that, for any choice of K, the robustness that the interpolation can tolerate is not large enough to overcome the inherent errors that our algorithms make when trying to simulate these non-unitary families. The existence of simulation algorithms like `SEBD` and `Patching`, which exploit the unitarity of the circuit, may present an obstruction to applying worst-to-average-case reduction techniques that obtain a polynomial structure at the expense of unitarity. Note that, as discussed previously, a very recent alternative worst-to-average case reduction [25] based on "Cayley paths" rather than truncated Taylor series does not suffer from this same limitation.

Background: truncated Haar-random circuit ensembles and polynomial interpolation

In this section, we give an overview (omitting some details) of the interpolation technique of Ref. [4] used to show their worst-to-average-case reduction, partially departing from their notation. Suppose U is a unitary operator. Then we define the θ-contracted and K-truncated version of U to be $U'(\theta, K) = U \sum_{k=0}^{K} \frac{(-\theta \ln U)^k}{k!}$. Note that $U'(\theta, \infty) = U e^{-i\theta(-i \ln U)}$ is simply U pulled-back by angle θ towards the identity operator I. Note that $U'(0, \infty) = U$ and $U'(1, \infty) = I$. For $U'(\theta, K)$ for $K < \infty$, the operator that performs this pullback is then approximated by a Taylor series which is truncated at order K. If $K < \infty$, $U'(\theta, K)$ is (slightly) non-unitary.

Suppose C is some circuit family for which computing output probabilities up to error $2^{-\operatorname{poly}(n)}$ is classically hard. Now, for each gate G in C, multiply that gate by $H'(\theta, K)$ with H Haar-distributed and supported on the same qubits as G. This yields some distribution over non-unitary circuits that we call $\mathcal{D}(C, \theta, K)$. Note that if $\theta = 0$, $\mathcal{D}$ exactly becomes the Haar-random circuit distribution with the same architecture as C. When $\theta = 1$, the hard circuit C is recovered up to some small correction due to the truncation. If

K is sufficiently large, we can assume that computing output probabilities for this slightly perturbed version of C is also classically hard.

Fix some circuit A drawn from $\mathcal{H}(C)$, the distribution over circuits with the same architecture as C with gates chosen according to the Haar measure. Let $A(C, \theta, K)$ denote the circuit obtained when the θ-pulled-back and K-truncated gates of A are multiplied with their corresponding gates in C. Note that $A(C, \theta, K)$ is distributed as $\mathcal{D}(C, \theta, K)$. Define the quantity

$$p_0(A, \theta, K) = |\langle 0|A(C, \theta, K)|0\rangle|^2. \tag{3.89}$$

Assuming that the circuit C has m gates, it is easy to verify that $p_0(A, \theta, K)$ may be represented as a polynomial in θ of degree $2mK$. Note also that $p_0(A, 1, \infty) = p_0(C)$, which is assumed to be classically hard to compute.

Now, assume that there exists some classical algorithm $\mathcal{A}$ and some $\epsilon = 1/\operatorname{poly}(n)$ such that, for some fixed $K \leq \operatorname{poly}(n)$ and for all $0 \leq \theta \leq \epsilon$, $\mathcal{A}$ can compute $p_0(A, \theta, K)$ up to additive error $\delta \leq 2^{-n^c}$ for some constant c, with probability $1 - 1/\operatorname{poly}(n)$ over $A(C, \theta, K) \sim \mathcal{D}(C, \theta, K)$. Then, $\mathcal{A}$ may evaluate $p_0(A, \theta, K)$ for $2mK + 1$ evenly spaced values of θ in the range $[0, \epsilon]$ (up to very small error), and construct an interpolating polynomial $q_0(A, \theta, K)$. By a result of Rakhmanov [121], there is some interval $[a, b] \subset [0, \epsilon]$ such that $b - a \geq 1/\operatorname{poly}(n)$ and $|p_0(A, \theta, K) - q_0(A, \theta, K)| \leq 2^{-n^{c'}}$ for $\theta \in [a, b]$ where c' depends on c. One then invokes the following result of Paturi.

Lemma 3.11 ([122]). *Let $p : \mathbb{R} \to \mathbb{R}$ be a real polynomial of degree d, and suppose $|p(x)| \leq \delta$ for all $|x| \leq \epsilon$. Then $|p(1)| \leq \delta e^{2d(1+1/\epsilon)}$.*

Applying this result, we find $|p_0(A, 1, K) - q_0(A, 1, K)| \leq 2^{-n^{c'}} e^{\operatorname{poly}(n,m,K)}$. If c is sufficiently large, then $|p_0(A, 1, K) - q_0(A, 1, K)| \leq 2^{-\operatorname{poly}(n)}$ and the quantity $q_0(A, 1, K)$ is hard to compute classically. But this would be a contradiction, because $q_0(A, 1, K)$ can be efficiency evaluated classically by performing the interpolation.

Hence, this argument shows that for some choice of K and a sufficiently large c depending on K, computing output probabilities of circuits in the truncated families $\mathcal{D}(C, \theta, K)$ with $\theta \leq 1/\operatorname{poly}(n)$ up to error 2^{-n^c} is hard (assuming standard hardness conjectures).

Limitation of the interpolation argument

The above argument shows that the average-case simulation of some family $\mathcal{D}(C, \theta, K)$ of non-unitary circuits which in some sense is close to the corresponding Haar-random circuit family to precision $2^{-\operatorname{poly}(n)}$ is classically hard, if simulating C is classically hard and the polynomial in the exponent is sufficiently large.

We now explain how, based on this argument, we are unable to conclude that exactly computing output probabilities of Haar-random circuits is classically hard.[4] In other words, suppose that with probability $1 - 1/\operatorname{poly}(n)$, some algorithm $\mathcal{A}$ can *exactly* compute output probabilities from the distribution $\mathcal{H}(C)$. We argue that a straightforward application of the above result based on Taylor series truncations and polynomial interpolation is insufficient to compute $p_0(C)$ with small error.

Consider some circuit realization A drawn from $\mathcal{H}(C)$, and assume that we can exactly compute its output probability $p_0(A)$. To use the argument of Ref. [4], we actually need to compute $p_0(A, \theta, K)$ for some fixed value of K and θ in some range $[0, \epsilon]$. We first find an upper bound for ϵ which must be satisfied for the interpolation to be guaranteed to succeed with high probability. To this end, we note that [4] the total variation distance between the distributions $\mathcal{D}(C, \theta, \infty)$ and $\mathcal{D}(C, 0, \infty)$ is bounded by $O(m\theta)$. Hence, if we try to use the algorithm $\mathcal{A}$ to estimate $p_0(A, \theta, \infty)$, the failure probability over random circuit instances could be as high as $O(m\theta)$. Therefore, since the θ values to be evaluated are uniformly spaced on the interval $[0, \epsilon]$, the union bound tells us that the probability that one of the $2mK + 1$ values $p_0(A, \theta, K)$ is erroneously evaluated is bounded by $O(m^2 K\epsilon)$. Hence, in order to ensure that all $2mK+1$ points are correctly evaluated, we should take $\epsilon \leq O(1/m^2 K)$.

Now, assume that we have chosen $\epsilon \leq O(1/m^2 K)$ and all $2mK + 1$ points $p_0(A, \cdot, \infty)$ are correctly evaluated. Let θ be one of the evaluation points. We now must consider the error made by approximating the "probability" associated with the truncated version of the circuit with the probability associated with the untruncated version of the circuit, namely $|p_0(A, \theta, \infty) - p_0(A, \theta, K)|$. This error associated with the truncated Taylor series is upper bounded by $\delta \leq \frac{2^{O(nm)}}{K!}$ [4].

Plugging these values into Lemma 3.11, we find that if we use these values to try to interpolate to the classically hard-to-compute quantity $p_0(C, 1, K)$, the error bound guaranteed by Paturi's lemma is no better than $\frac{2^{O(nm)}}{K!} \exp\left(O(2mK(1 + m^2 K))\right)$, which diverges in the limit $n \to \infty$ for any scaling of m and K. Hence, the technique of Ref. [4] is insufficient to show that exactly computing output probabilities of circuits drawn from the Haar-random circuit distribution $\mathcal{H}_C$ with high probability is hard.

Intuitively, the limitation arises because there is a tradeoff between the amount of truncation error incurred and the degree of the interpolating polynomial. As the parameter K is increased, the truncation error is suppressed, but the degree of the interpolating polynomial is increased, making the interpolation more sensitive to errors.

[4] A simplified and slightly weaker version of our argument was also reported in [25].

Inapplicability to `SEBD` and `Patching`

To summarize the findings above, the argument of Ref. [4] for the hardness of computing output probabilities of random circuits applies not directly to Haar-random circuit distributions, but rather to distributions over slightly non-unitary circuits that are exponentially close to the corresponding Haar distributions in some sense. We argued that the interpolation scheme cannot be straightforwardly applied to circuits that are truly Haar-random, and therefore it cannot be used to conclude that simulating truly Haar-random circuits, even exactly, is classically hard.

A priori, it is not obvious whether this limitation is a technical artifact or a more fundamental limitation of the interpolation scheme. In particular, one might imagine that if some algorithm $\mathcal{A}$ is capable of exactly simulating Haar-random circuit families, some modified version of the algorithm $\mathcal{A}'$ might be capable of simulating the associated truncated Haar-random circuit families, at least up to the precision needed for the interpolation argument to work. If this were the case, then the hardness argument *would* be applicable.

However, `SEBD` and `Patching` appear to be algorithms that *cannot* be straightforwardly used to efficiently simulate truncated Haar-random circuit families to the precision needed for the interpolation to work, even under the assumption that they can efficiently, exactly simulate Haar-random circuit families. This is because the efficiency of these algorithms crucially relies on the existence of a constant-radius lightcone for constant-depth circuits. The algorithm is able to ignore all qubits and gates outside of the lightcone of the sites currently being processed. However, the lightcone argument breaks down for non-unitary circuits. If the gates are non-unitary and we want to perform an exact simulation, we are left with using Markov-Shi or some other general-purpose tensor network contraction algorithm, with a running time of $2^{O(d\sqrt{n})}$ for a depth-d circuit on a square grid of n qubits.

Consider what happens if one tries to use one of these algorithms to compute output "probabilities" for a slightly non-unitary circuit coming from a truncated Haar-random distribution $\mathcal{D}(C, \theta, K)$, and then use these computed values to interpolate to the hard-to-compute value $p_0(C, 1, K)$ via the interpolating polynomial of degree $2mK$ proposed in Ref. [4]. Even without any other sources of error, when one of these algorithms ignores gates outside of the current lightcone, it is essentially approximating each gate outside the lightcone as unitary. This causes an incurred error bounded by $2^{O(nm)}/K!$ for the computed output probability. Then, by an argument essentially identical to the one appearing in the previous section, one finds that this error incurred just from neglecting gates outside the lightcone is already large enough to exceed the error permitted for the polynomial interpolation to be valid. We conclude that this worst-to-average-case reduction based on truncated Taylor series expansions cannot be used to conclude that it is hard for `SEBD` or `Patching`

to exactly simulate worst-case-hard shallow Haar-random circuits with high probability.

3.E Deferred proofs

Lemma 3.1 (restated)**.** *Let ϵ_i denote the sum of the squares of all singular values discarded in the compression during iteration i of the simulation of a circuit C with output distribution $\mathcal{D}_C$ by* `SEBD` *with no bond dimension cutoff, and let Λ denote the sum of all singular values discarded over the course of the algorithm. Then the distribution $\mathcal{D}'_C$ sampled from by* `SEBD` *satisfies*

$$\frac{1}{2}\|\mathcal{D}'_C - \mathcal{D}_C\|_1 \leq \mathbb{E}\sum_{i=1}^{L_2}\sqrt{2\epsilon_i} \leq \sqrt{2}\,\mathbb{E}\,\Lambda, \tag{3.90}$$

where the expectations are over the random measurement outcomes.

Proof. We rely upon a well-known fact about the error caused by truncating the bond dimension of a MPS, which we state in Lemma 3.12.

Lemma 3.12 (follows from Ref. [123])**.** *Suppose the MPS $|\psi\rangle$ is compressed via truncation of small singular values, and ϵ is the sum of the squares of the discarded singular values. Then if $|\psi^{(t)}\rangle$ is the truncated version of the MPS after normalization,*

$$\||\psi\rangle\langle\psi| - |\psi^{(t)}\rangle\langle\psi^{(t)}|\|_1 \leq \sqrt{8\epsilon}. \tag{3.91}$$

The second inequality follows from the fact that $\sqrt{\sum_i x_i^2} \leq \sum_i x_i$ for $x_i \geq 0$. To prove the first inequality, we start by considering the version of the algorithm with no truncation, which we have argued samples exactly from $\mathcal{D}$. Let $\mathcal{N}_t$ denote the TPCP map corresponding to the application of gates that have come into the lightcone of `column` t and the measurement of `column` t. That is,

$$\mathcal{N}_t(\rho) = \sum_{\mathbf{x}_t} \Pi_t^{\mathbf{x}_t} V_t \rho V_t^\dagger \Pi_t^{\mathbf{x}_t}, \tag{3.92}$$

where $\mathbf{x}_t$ indexes (classical) outcome strings of `column` t. Note that $\mathcal{N}_t(\rho)$ is a classical-quantum state for which the sites corresponding to the first t columns are classical, and the quantum register consists of sites which are in the lightcone of `column` t but not in the first t columns. Define $\rho_t = \mathcal{N}_{t-1}(\rho_{t-1})$ and $\rho_1 = |0\rangle\langle 0|^{\otimes L_1}_{\texttt{column 1}}$, so that ρ_{L_2+1} is a classical state exactly corresponding to output strings on the $L_1 \times L_2$ grid distributed according to $\mathcal{D}$.

Now consider the "truncated" version of the algorithm, which is defined similarly except we use σ_t to denote the state of the algorithm immediately after the truncation at the beginning of iteration t. That is, we define

$$\sigma_t = (T_t \circ \mathcal{N}_{t-1})(\sigma_{t-1}), \tag{3.93}$$

where T_t denotes the mapping corresponding to the MPS truncation and subsequent renormalization at the beginning of iteration t, and we define $\sigma_1 = T_1(\rho_1) = \rho_1$ (there is no truncation at the beginning of the first iteration since the initial state is a product state).

We now have

$$\begin{aligned} \|\mathcal{D}_C - \mathcal{D}'_C\|_1 &= \|\rho_{L_2+1} - \sigma_{L_2+1}\|_1 && (3.94)\\ &\leq \|\rho_{L_2+1} - \mathcal{N}_{L_2}(\sigma_{L_2})\|_1 + \|\mathcal{N}_{L_2}(\sigma_{L_2}) - \sigma_{L_1+1}\|_1 && (3.95)\\ &\leq \|\rho_{L_2} - \sigma_{L_2}\|_1 + \|\mathcal{N}_{L_2}(\sigma_{L_2}) - \sigma_{L_2+1}\|_1, && (3.96) \end{aligned}$$

where the first inequality follows from the triangle inequality, and the second from contractivity of TPCP maps. Applying this inequality recursively yields

$$\begin{aligned} \|\mathcal{D}_C - \mathcal{D}'_C\|_1 &\leq \sum_{i=1}^{L_2} \|\mathcal{N}_i(\sigma_i) - \sigma_{i+1}\|_1 && (3.97)\\ &= \sum_{i=1}^{L_2-1} \|\mathcal{N}_i(\sigma_i) - (T_{i+1} \circ \mathcal{N}_i)(\sigma_i)\|_1 && (3.98) \end{aligned}$$

where we also used the fact that no truncation occurs after $\mathcal{N}_{L_2}$ is applied (i.e. T_{L_2+1} acts as the identity). Now, note that $\|\mathcal{N}_i(\sigma_i) - (T_{i+1} \circ \mathcal{N}_i)(\sigma_i)\|_1$ is exactly the expected error in 1-norm caused by the truncation in iteration $i+1$. (This is true because of the following fact about classical-quantum states: $\left\|\mathbb{E}_i |i\rangle\langle i|_C \otimes (|\psi_i\rangle\langle\psi_i|_Q - |\phi_i\rangle\langle\phi_i|_Q)\right\|_1 = \mathbb{E}_i \||\psi_i\rangle\langle\psi_i| - |\phi_i\rangle\langle\phi_i|\|_1$ where $\{|i\rangle_C\}_i$ is an orthonormal basis for the Hilbert space associated with register C.) By Lemma 3.12, this quantity is bounded by $\mathbb{E}\sqrt{8\epsilon_{i+1}}$. Substituting this bound into the summation yields the desired inequality. □

Lemma 3.5 (restated)**.** *Let $\lambda_1 \geq \lambda_2 \geq \cdots$ denote the half-chain Schmidt values after at least $n/2$ iterations of the toy model process. Then with probability at least $1-\delta$ the half-chain Schmidt values indexed by $i \geq i^* = \exp\Big(\Theta(\sqrt{\log(n/\delta)})\Big)$ obey the asymptotic scaling*

$$\lambda_i \propto \exp\big(-\Theta(\log^2(i))\big). \tag{3.99}$$

Furthermore, upon truncating the smallest Schmidt coefficients up to a truncation error of ϵ, with probability at least $1-\delta$, the half-chain Schmidt rank r of the post-truncation state obeys the scaling

$$r \leq \exp\Big(\Theta\Big(\sqrt{\log(n/\epsilon\delta)}\Big)\Big). \tag{3.100}$$

Proof. Suppose that an EPR pair is measured $2t$ times, corresponding to each of the two qubits being measured t times. A calculation shows that the probability of obtaining s M_1 outcomes is given by a mixture of two binomial

distributions. Letting S be the random variable denoting the number of M_1 outcomes, we find that $\Pr[S = s]$ is given by

$$\frac{1}{2}\Pr\left[B_{2t,\sin^2(\theta/2)} = s\right] + \frac{1}{2}\Pr\left[B_{2t,\cos^2(\theta/2)} = s\right], \tag{3.101}$$

where $B_{n,p}$ denotes a binomial random variable associated with n trials and success probability p. If after the $2t$ measurements we obtain outcome M_1 s times, the post-measurement state is given by (up to normalization)

$$|00\rangle + \tan^{2(t-s)}(\theta/2)|11\rangle. \tag{3.102}$$

Note that s can be assumed to be generated by sampling from either $B_{2t,\sin^2(\theta/2)}$ or $B_{2t,\cos^2(\theta/2)}$ with probability $1/2$ each. In the former case, the post-measurement state may be written as

$$|00\rangle + \tan^{2\left(t-B_{2t,\sin^2(\theta/2)}\right)}(\theta/2)|11\rangle = |00\rangle + \tan^{2t\cos(\theta)-2X_{2t,\sin^2(\theta/2)}}(\theta/2)|11\rangle \tag{3.103}$$

where we have defined the random variable $X_{2t,\sin^2(\theta/2)}$ via $B_{n,p} = np + X_{n,p}$. That is, the random variable $X_{n,p}$ is distributed as a binomial distribution shifted by its mean. Now, defining $\gamma = (\tan(\theta/2))^{2\cos(\theta)}$ and $X'_{n,p} = X_{n,p}/\cos(\theta)$, we may write the post-measurement state as

$$|00\rangle + \gamma^{t-X'_{2t,\sin^2(\theta/2)}}|11\rangle. \tag{3.104}$$

We assume WLOG that $0 < \theta < \pi/2$, so that $0 < \gamma < 1$. Similarly, if s is drawn from $B_{2t,\cos^2(\theta/2)}$, then the post-measurement state is given by

$$|00\rangle + \gamma^{-t-X'_{2t,\cos^2(\theta/2)}}|11\rangle. \tag{3.105}$$

Note that, under a relabeling of basis states $0 \leftrightarrow 1$, the post-measurement state in this case is

$$|00\rangle + \gamma^{t-X'_{2t,\sin^2(\theta/2)}}|11\rangle, \tag{3.106}$$

where we used the fact that $-X'_{2t,\cos^2(\theta/2)}$ is distributed identically to $X'_{2t,\sin^2(\theta/2)}$. Since we will be interested in studying the entanglement spectrum of this process, which is invariant under such local basis changes, we may assume WLOG that the random post-measurement state after $2t$ measurements is given by $|00\rangle + \gamma^{t-X'_{2t,\sin^2(\theta/2)}}|11\rangle$.

We can then model the final state as

$$\bigotimes_t |00\rangle + \gamma^{t-X'_{2t,\sin^2(\theta/2)}}|11\rangle \tag{3.107}$$

up to normalization. This allows an estimate of the tradeoff between rank, truncation error, and associated probability of success.

Let $Q(\ell)$ denote the number of "strict partitions" of ℓ, i.e. the number of ways of writing $\ell = t_1 + t_2 + \dots$ for positive integers $t_1 < t_2 < \dots$. Precise asymptotics are known for $Q(\ell)$ (see https://oeis.org/A000009 and [124]):

$$Q(\ell) = \exp\Big(\Theta(\sqrt{\ell})\Big). \tag{3.108}$$

By expanding Eq. (3.107) as a superposition over computational basis states, we obtain the unnormalized Schmidt coefficients $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots$; each coefficient in the expansion gives an unnormalized Schmidt coefficient. There are $Q(\ell)$ unnormalized Schmidt coefficients that are distributed as $\gamma^{\ell - X'_{2\ell,\sin^2(\theta/2)}}$, where we used the fact that $X'_{t_1,\sin^2(\theta/2)} + X'_{t_2,\sin^2(\theta/2)}$ is distributed as $X'_{t_1+t_2,\sin^2(\theta/2)}$. We say that these $Q(\ell)$ coefficients live in sector ℓ. For a fixed probability p, let $K_{\ell,p}$ denote the smallest positive integer for which, with probability at least $1-p$, all sector-ℓ coefficients lie in the range $[\gamma^{\ell+K_{\ell,p}}, \gamma^{\ell-K_{\ell,p}}]$. By the union bound, to upper bound $K_{\ell,p}$ it suffices to find an integer a for which

$$\Pr\Big[\Big|X'_{2\ell,\sin^2(\theta/2)}\Big| \geq a\Big] \leq \frac{p}{Q(\ell)} = p\exp\Big(-\Theta(\sqrt{\ell})\Big). \tag{3.109}$$

By Hoeffding's inequality, we have $\Pr\Big[\Big|X'_{2\ell,\sin^2(\theta/2)}\Big| \geq a\Big] \leq \exp(-\Theta(a^2/\ell))$; this yields the bound

$$K_{\ell,p} \leq \Theta\left(\sqrt{\ell\log(1/p) + \ell\sqrt{\ell}}\right). \tag{3.110}$$

Furthermore, note that since there are $\Theta(n^2)$ sectors, by the union bound, with probability at least $1-\delta$, for each sector j, all coefficients lie in the range $[\gamma^{j+K_{j,p}}, \gamma^{j-K_{j,p}}]$ if we take p to be $p = \delta/\Theta(n^2)$. We make this choice of p and assume for the remainder of the argument that all coefficients of sector j lie in the given range, which is true with probability at least $1-\delta$. We also note the following fact which will be used below: if ℓ and p are related as $\ell \geq \Theta(\log(1/p))$, then $K_{\ell,p} = O(\ell)$.

Still working with the unnormalized state of Eq. (3.107), we now study the scaling between the Schmidt index i and corresponding coefficient $\tilde{\lambda}_i$ for i in the regime $i \geq \exp\Big(\Theta(\sqrt{\log(1/p)})\Big)$. Note that $\tilde{\lambda}_i = \gamma^\ell$ for some integer ℓ. We first lower bound ℓ. Note that the lower bound is achieved if, for each sector j, all coefficients in that sector are equal to $\gamma^{j-K_{j,p}}$. In this case, the exponent ℓ is equal to $\ell' - K_{\ell',p}$, where ℓ' is the smallest integer such that

$$i \leq \sum_{j=1}^{\ell'} Q(\ell') = \exp\Big(\Theta(\sqrt{\ell'})\Big). \tag{3.111}$$

Rearranging, we see that $\ell' = \Theta(\log^2(i)) \geq \Theta(\log(1/p))$, and hence $\ell = \Theta(\log^2(i))$ since $\ell' - K_{\ell',p} = \Theta(\ell')$. Similarly , an upper bound on ℓ is achieved if, for each sector j, all coefficients in that sector are equal to $\gamma^{j+K_{j,p}}$. In this case, ℓ is equal to $\ell' + K_{\ell',p}$, where ℓ' is defined as above. This yields a matching upper bound for ℓ of $\Theta(\log^2(i))$. We therefore have the scaling $\ell = \Theta(\log^2(i))$, which, using the fact that $\tilde{\lambda}_i = \gamma^\ell$ yields

$$\tilde{\lambda}_i = \exp\big(\Theta(-\log^2(i))\big),\ i \geq \exp\Big(\Theta(\sqrt{\log(1/p)})\Big). \tag{3.112}$$

Noting that λ_i is proportional to $\tilde{\lambda}_i$ via $\lambda_i = \frac{1}{N}\tilde{\lambda}_i$ with $N = \sqrt{\sum_i \tilde{\lambda}_i^2}$, this shows the first statement of the lemma.

Now, suppose that for some $i \geq i^* = \exp\Big(\Theta(\sqrt{\log(1/p)})\Big)$, we truncate all Schmidt coefficients with index $\geq i$. The incurred truncation error is

$$\epsilon = \sum_{j \geq i} \lambda_j^2 < \sum_{j \geq i} \tilde{\lambda}_j^2 = \exp\big(-\Theta(\log^2(i))\big) \tag{3.113}$$

where the inequality holds because the unnormalized state has norm strictly greater than one (i.e. $N > 1$). Rearranging, this becomes

$$i \leq \exp\Big(\Theta(\sqrt{\log(1/\epsilon)})\Big). \tag{3.114}$$

Hence, if we truncate the state at the end of the process up to a truncation error of ϵ, the rank r of the post-truncation state is bounded by

$$r \leq \max\left(\exp\Big(\Theta(\sqrt{\log(1/\epsilon)})\Big), \exp\Big(\Theta(\sqrt{\log(1/p)})\Big)\right) \tag{3.115}$$

$$= \exp\left(\Theta\left(\sqrt{\log\left(\frac{n}{\epsilon \cdot \delta}\right)}\right)\right) \tag{3.116}$$

as desired, where we used the relation $p = \delta/\Theta(n^2)$. □

Lemma 3.7 (restated)**.** *Suppose a 1D random circuit C is applied to qubits $\{1, \ldots, n\}$ consisting of a layer of 2-qubit Haar-random gates acting on qubits $(k, k+1)$ for odd $k \in \{1, \ldots, n-1\}$, followed by a layer of 2-qubit Haar-random gates acting on qubits $(k, k+1)$ for even $k \in \{1, \ldots, n-1\}$. Suppose the qubits of region $B = \{i, i+1, \ldots, j\}$ for $j \geq i$ are measured in the computational basis, and the outcome b is obtained. Then, letting $|\psi_b\rangle$ denote the post-measurement pure state on the unmeasured qubits, and letting $A = \{1, 2, \ldots, i-1\}$ denote the qubits to the left of B,*

$$\mathbb{E}\, S(A)_{\psi_b} \leq c^{|B|} \tag{3.117}$$

for some universal constant $c < 1$, where the expectation is over measurement outcomes and choice of random circuit C.

Proof. We will use a smaller technical lemma, which we state and prove below.

Lemma 3.13. *Let $|\psi\rangle_{AB}$ be some state on subsystems A and B with subsystem B a qubit, and let $|H\rangle_{CD}$ be some two-qubit Haar-random state on subsystems C and D. Suppose a Haar-random two-qubit gate U is applied to subsystems B and C. If subsystem B is measured in the computational basis and outcome b is obtained, then the von Neumann entropy of the post-measurement state $|\psi_b\rangle_{ABCD}$ in subsystem A satisfies*

$$\mathop{\mathbb{E}}_{b,H,U} S(A)_{\psi_b} \leq c \cdot S(A)_\psi \tag{3.118}$$

for some constant $c < 1$, where the expectation is over the random measurement outcome, the random state $|H\rangle_{CD}$, and the Haar-random unitary U.

Proof. Consider the Schmidt decomposition $|\psi\rangle_{AB} = \sqrt{p}|e_1\rangle_A|f_1\rangle_B + \sqrt{1-p}|e_2\rangle_A|f_2\rangle_B$ where we assume WLOG that $p \geq 1/2$. We also assume that $p < 1$, because the statement is trivially true for any value of c when $p = 1$. Note that the entanglement entropy of this state is simply $S(A)_\psi = H_2(p)$ where $H_2(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function. Let $M_0 = (\Pi_0 \otimes I)U$ and $M_1 = (\Pi_1 \otimes I)U$ denote the measurement operators acting on subsystems B and C, where Π_i denotes the projector onto the computational basis state $|i\rangle$ and U is the Haar-random unitary applied to subsystems B and C. Let X denote a random variable equal to 1 with probability p and equal to 2 with probability $1-p$. Let Y denote the measurement outcome of $\{M_0, M_1\}$ when applied to the state $|e_X\rangle_A|f_X\rangle_B|H\rangle_{C,D}$. The probability of obtaining measurement outcome b on the original state is simply $\Pr(Y = b)$, and the post-measurement state after obtaining outcome b is

$$\begin{aligned}
&\frac{1}{\sqrt{\Pr(Y=b)}}\Big(\sqrt{p \cdot \Pr(Y=b|X=1)}|e_1\rangle_A|b\rangle_B|\phi_{b,1}\rangle_{C,D} \\
&\quad + \sqrt{(1-p)\cdot \Pr(Y=b|X=2)}|e_2\rangle_A|b\rangle_B|\phi_{b,2}\rangle_{C,D}\Big) \\
&= \sqrt{\Pr(X=1|Y=b)}|e_1\rangle_A|b\rangle_B|\phi_{b,1}\rangle_{C,D} \\
&\quad + \sqrt{\Pr(X=2|Y=b)}|e_2\rangle_A|b\rangle_B|\phi_{b,2}\rangle_{C,D}
\end{aligned} \tag{3.119}$$

where $|\phi_{b,j}\rangle_{C,D}$ are normalized states on subsystems C and D. Define

$$\epsilon = \min_b |\langle\phi_{b,1}|\phi_{b,2}\rangle|^2. \tag{3.120}$$

Letting $\rho_{A,b}$ denote the reduced density matrix on subsystem A of the post-measurement state after obtaining measurement outcome b, the maximal eigenvalue of this matrix is lower bounded as $\lambda_{\max}(\rho_{A,b}) \geq \Pr(X=1|Y=b) + \epsilon \Pr(X=2|Y=b)$. (To see this, observe that the reduced density matrix on CD is $\sigma = \Pr(X=1|Y=b)\,|\phi_{b,1}\rangle\langle\phi_{b,1}| + \Pr(X=2|Y=b)\,|\phi_{b,2}\rangle\langle\phi_{b,2}|$, and the

maximal eigenvalue is lower bounded as $\lambda_{\max}(\rho_{A,b}) = \lambda_{\max}(\sigma) \geq \langle \phi_{b,1}|\sigma|\phi_{b,1}\rangle \geq \Pr(X = 1|Y = b) + \epsilon \Pr(X = 2|Y = b))$. Furthermore, note that

$$\mathop{\mathbb{E}}_{Y} \lambda_{\max}(\rho_{A,Y}) \geq \mathop{\mathbb{E}}_{Y} [\Pr(X = 1|Y) + \epsilon \Pr(X = 2|Y)] \tag{3.121}$$

$$= p + \epsilon(1 - p). \tag{3.122}$$

Now, using concavity of the binary entropy function, we have

$$\mathop{\mathbb{E}}_{Y} S(A)_{\psi_Y} = \mathop{\mathbb{E}}_{Y} H_2(\lambda_{\max}(\rho_{A,Y})) \tag{3.123}$$

$$\leq H_2(\mathop{\mathbb{E}}_{Y} \lambda_{\max}(\rho_{A,Y})) \tag{3.124}$$

$$\leq H_2(p + \epsilon(1 - p)). \tag{3.125}$$

Consider the ratio $r(p, \epsilon) = \frac{H_2(p+\epsilon(1-p))}{H_2(p)}$. We want to argue that for any $\epsilon > 0$, $r(p, \epsilon)$ is bounded away from one on the interval $p \in [1/2, 1)$. This statement is clearly true for any p bounded away from one since H_2 is monotonically decreasing on the interval $[1/2, 1)$. Furthermore, it is straightforward to show $\lim_{p\to 1} r(p, \epsilon) = 1 - \epsilon$. Hence, we have

$$\frac{\mathbb{E}_Y S(A)_{\psi_Y}}{S(A)_{\psi}} \leq r(p, \epsilon) \leq c(\epsilon) \tag{3.126}$$

where $c(\epsilon) < 1$ unless $\epsilon = 0$. We now average both sides over the choice of Haar-random state on CD as well as the Haar-random unitary U acting on BC. Since the event $\epsilon > 0$ occurs with nonzero probability (in fact, with probability one), we have the strict inequality $\mathbb{E}_{H,U}[c(\epsilon)] = c < 1$, from which the desired inequality follows. □

We may assume that $i \neq 0$ and $j \neq n$ since in these cases we trivially have $S(\rho_A(b)) = 0$. The post-measurement state may be constructed as follows. Apply all gates in the lightcone of qubit i, then measure qubit i. Now apply all gates in the lightcone of qubit $i + 1$ not previously applied, then measure qubit $i + 1$. Assume that qubits are introduced only when they come into the lightcone under consideration. Iterate until all qubits in region B have been measured. Finally, apply any gates that have not yet been applied. It is straightforward to verify that this is equivalent to applying all gates of the circuit before performing the measurement of region B, in the sense that the measurement statistics are the same, and the post-measurement state given some outcome b is the same.

By Lemma 3.13, after the first iteration we are left with the state $|\psi\rangle_{LR}|b_{i_1}\rangle_{i_1}$, such that $\mathbb{E}S(L)_\psi \leq c$ for some constant $c < 1$. In all iterations, we let L denote the current subsystem to the left of the measured qubits, and R denote the subsystem to the right of the measured qubits. Now consider the second iteration. Depending on whether i was even or odd, R may consist of one or two qubits immediately after the measurement of i. In the former

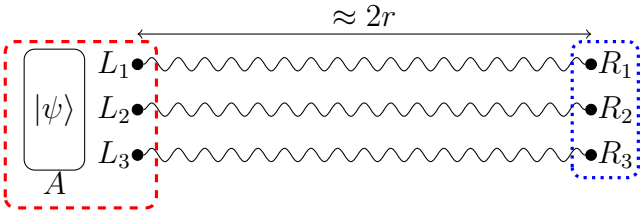


Figure 3.15: Illustration of the state after the qubits of columns $i, i+1, \ldots, j$ have been measured, but before gates in the lightcone of registers A and L have been performed. In each row i, we are left with a post-measurement bipartite state $|\phi_i\rangle_{L_i R_i}$ depicted by a wavy line. The expected entanglement entropy $S(L_i)_{\phi_i}$ decays exponentially in r. The final state of interest $|\psi'\rangle$ is obtained by applying local unitaries to the qubits in the dashed red box before measuring all of these qubits in the computational basis, inducing the final state $|\psi'\rangle$ on $R = R_1 \cup \cdots \cup R_L$. By concavity of the von Neumann entropy, the expected entanglement entropy of $|\psi'\rangle$ across the cut defined by the dotted blue box is upper bounded by the entanglement entropy across this cut before the unitaries and measurements in the dashed red box are performed.

case, we may apply Lemma 3.13 again, obtaining $\mathbb{E}S(L)_\psi \leq c^2$ after the measurement of qubit $i+1$, and obtaining a two-qubit subsystem to the right of the measured qubits. In the latter case, as a consequence of concavity of von Neumann entropy, we have $\mathbb{E}S(L)_\psi \leq c$ after measurement, and are left with a one-qubit subsystem to the right of the measured qubits. Iterating this process, after all qubits of subregion B have been measured, we are left with some state $|\psi\rangle_{LR}$ such that $\mathbb{E}\, S(L)_\psi \leq c^{|B|/2} \leq c'^{|B|}$ where $c' = \sqrt{c} < 1$. Finally, local unitary gates are applied to $|\psi\rangle_{LR}$ to obtain the final post-measurement state on the entire chain. Since each unitary is applied to only the left of region B or only the right of region B, the entanglement entropy across the (A, A^c) cut is unaffected by these gates, and remains bounded by $c^{|B|}$ in expectation. □

Lemma 3.8 (restated)**.** *Let C be an instance of* `Brickwork`(L, r, v)*. Then, with probability at least $1 - 2^{-\Theta(r)}$ over the circuit instance,* `SEBD` *running with maximal bond dimension cutoff $D = \Theta(1)$ and truncation error parameter $\epsilon = 2^{-\Theta(r)}$ can be used to (1) sample from the output distribution of C up to error $n2^{-\Theta(r)}$ in variational distance and (2) compute the output probability of an arbitrary output string up to additive error $n2^{-\Theta(r)}/2^n$ in runtime $\Theta(n)$.*

Proof. Suppose the state stored by `SEBD` immediately before entering into a 1-local region is $|\psi\rangle_A$, defined on register A. After another $O(r)$ iterations of `SEBD`, just before the end of the 1-local region, denote the new one-dimensional state stored by `SEBD` as $|\psi'\rangle$. Note that $|\psi'\rangle$ is a random state, depending on both the random choices of gates in the 1-local region and the random measurement outcomes. We now bound the expected entanglement entropy of $|\psi'\rangle$ across an arbitrary cut.

To this end, we observe that the random final state $|\psi'\rangle$ may be equivalently generated as follows. Instead of iterating SEBD as usual for $O(r)$ iterations, we first introduce a contiguous block of qubits that lie in the 1-local region. In particular, for all rows, we introduce all qubits that lie in columns $\{i, i+1, \ldots, j\}$. Here, i is chosen to be the leftmost column such that the lightcone of column i does not contain qubits in register A. Similarly, j is chosen to be the rightmost column such that the lightcone of qubits in column j does not contain vertical gates. Note that $|i - j| = \Theta(r)$.

We next apply all gates in the lightcone of the qubits of columns $\{i, i+1, \ldots, j\}$, before measuring these qubits in the computational basis. Note that in this step, we are effectively performing a set of L one-dimensional depth-2 Haar-random circuits, and then measuring $\Theta(r)$ intermediate qubits for each of the L instances. For each instance, we are left with a (generically entangled) pure state between a "left" and "right" subsystem, as illustrated in Figure 3.15. Let L_i (R_i) denote the left (right) subsystem associated with row i, and let $|\phi_i\rangle_{L_i R_i}$ denote the associated post-measurement pure state on these subsystems. By Lemma 3.7, it follows that the expected entanglement entropy for any 1D instance obeys $\mathbb{E}\, S(L_i)_{\phi_i} \leq 2^{-\Theta(r)}$ where the expectation is over random circuit instance and measurement outcomes.

The next step is to apply all gates in the lightcone of the qubits of registers A and $L = \cup_i L_i$ before measuring these registers, inducing a (random) 1D post-measurement pure state on subsystem $R = \cup_i R_i$. It is straightforward to verify that the distribution of the random 1D pure state $|\psi'\rangle_R$ obtained via this procedure is identical to that obtained from repeatedly iterating SEBD through column j[5]. Indeed, the procedures are identical up to performing commuting gates and commuting measurements in different orders, which does not affect the measurement statistics or post-measurement states.

Our goal is now to bound the entanglement entropy $S(R_1 R_2 \ldots R_k)_{\psi'}$ in expectation across an arbitrary cut of the post-measurement 1D state. Such a bound follows from the concavity of the von Neumann entropy. Let $\rho_{R_1,\ldots,R_k}$ denote the reduced density matrix on these subsystems before the measurements on A and L are performed. Let $\rho^x_{R_1,\ldots,R_k}$ denote the reduced density matrix on these subsystems after the measurements on A and L are performed and the outcome x is obtained; note that the final state ψ' implicitly depends on x. Now, letting $\Pr[x]$ denote the probability of obtaining outcome x, we have the relation $\sum_x \Pr[x] \rho^x_{R_1 \cdots R_k} = \rho_{R_1 \cdots R_k}$. To see this, observe that for any set of measurement operators $\{M^x\}_x$ satisfying $\sum_x M^{x\dagger} M^x = I$, we have $\rho_{R_1 \cdots R_k} = \mathrm{tr}_{\setminus R_1 \cdots R_k}(|\psi'\rangle\langle\psi'|) = \sum_x \mathrm{tr}_{\setminus R_1 \cdots R_k}\left(M^x |\psi'\rangle\langle\psi'| M^{x\dagger}\right) =$

[5]Strictly speaking, we are actually studying a version of SEBD that only performs the MPS compression step at the end of a 1-local region. Since 1-local operations cannot increase the bond dimension of the associated MPS, the algorithm can forego the compression steps during the 1-local regions without incurring a bond dimension increase.

$\sum_x \Pr[x] \frac{\mathrm{tr}_{\setminus R_1 \cdots R_k}\left(M^x |\psi'\rangle\langle\psi'| M^{x\dagger}\right)}{\mathrm{tr}\left(M^x |\psi'\rangle\langle\psi'| M^{x\dagger}\right)} = \sum_x \Pr[x] \rho^x_{R_1 \cdots R_k}$. Now,

$$\begin{aligned}
&\sum_x \Pr[x] S(R_1 \ldots R_k)_{\psi'} \\
&= \sum_x \Pr[x] S(\rho^x_{R_1,\ldots,R_k}) && (3.127)\\
&\leq S\left(\sum_x \Pr[x] \rho^x_{R_1,\ldots,R_k}\right) && (3.128)\\
&= S(\rho_{R_1,\ldots,R_k}) && (3.129)\\
&= \sum_{i=1}^k S(R_i)_{\phi_i}\,, && (3.130)
\end{aligned}$$

where the first line follows by definition, the second line follows from concavity of the von Neumann entropy, the third line uses the relation we discussed previously, and in the final line we used the fact that $\rho_{R_1,\ldots,R_k}$ is a product state. Hence, we see that for any fixed set of gates and for any outcomes of the measurements of qubits in columns $i, i+1, \ldots, j$, the expected entanglement entropy of the final 1D state ψ' on R across any cut is bounded by the entropy across that cut before the measurements on subregions A and L. Taking the expectations of both sides of this result with respect to the random gates and measurement outcomes of the qubits in columns $i, i+1, \ldots, j$, we finally obtain

$$\mathbb{E}\, S(R_1 \ldots R_k)_{\psi'} \leq L 2^{-\Theta(r)}\,, \tag{3.131}$$

where we used the fact that $k < L$ and $\mathbb{E}\, S(R_i)_{\phi_i} \leq 2^{-\Theta(r)}$. We now use the fact that the largest eigenvalue $\lambda_{\max}(R_1 \cdots R_k)$ of the reduced density matrix is lower bounded as $\lambda_{\max}(R_1 \cdots R_k)_{\psi'} \geq 2^{-S(R_1 \cdots R_k)_{\psi'}}$; this follows from the fact that Shannon entropy upper bounds min-entropy. Using this inequality as well as Jensen's inequality, we have the bound

$$\begin{aligned}
\mathbb{E}\, \lambda_{\max}(R_1 \cdots R_k) &\geq \mathbb{E}\, 2^{-S(R_1 \cdots R_k)_{\psi'}} && (3.132)\\
&\geq 2^{-\mathbb{E}\, S(R_1 \cdots R_k)_{\psi'}} && (3.133)\\
&\geq 2^{-L 2^{-\Theta(r)}} && (3.134)\\
&\geq 1 - L 2^{-\Theta(r)}. && (3.135)
\end{aligned}$$

Therefore, if we truncate all but the largest Schmidt coefficient across the $R_k : R_{k+1}$ cut, we incur an expected truncation error upper bounded by $L 2^{-\Theta(r)}$. Hence, by Markov's inequality, we incur a truncation error upper bounded by $L 2^{-\Theta(r)}$ with probability at least $1 - 2^{-\Theta(r)}$.

Therefore, if we run `SEBD` using a *per bond* truncation error of $\epsilon = L 2^{-\Theta(r)}$ and a maximum bond dimension cutoff of $D = O(1)$, the failure probability will be upper bounded by $Lv 2^{-\Theta(r)}$; here we used the union bound to upper

bound the probability that any of the $O(Lv)$ bonds over the course of the algorithm becomes larger than the cutoff D. Hence, by Corollary 3.1, for at least $1 - 2^{-\Theta(r)}$ fraction of random circuit instances, `SEBD` can sample from the output distribution with variational distance error $Lv2^{-\Theta(r)} < n2^{-\Theta(r)}$. Similarly, by Corollary 3.3, for at least $1 - 2^{-\Theta(r)}$ fraction of circuit instances, `SEBD` can compute the probability of the all-zeros output string up to additive error $n2^{-\Theta(r)}/2^n$.

Since the runtime of `SEBD` is $O(nD^3)$ when acting on qubits as discussed previously, and D is chosen to be constant for the version of the algorithm used here, the runtime is $O(n)$. □

Chapter 4

ANTI-CONCENTRATION OF RANDOM QUANTUM CIRCUITS IN LOGARITHMIC DEPTH

This chapter has been adapted from joint work with Nicholas Hunter-Jones and Fernando G. S. L. Brandão in Ref. [125].

4.1 Motivation

In the previous chapter, we presented a classical algorithm for approximately simulating shallow 2D random quantum circuits (RQCs). The main motivation there was to scrutinize claims that RQCs should be hard to simulate, which have been the basis for recent experiments aiming to demonstrate exponential quantum advantage [6, 7]. Beyond this application, RQCs have notably been used to study the onset of quantum chaos and dynamical spread of entanglement in strongly interacting quantum systems [41, 42, 87], including information processing in black holes [126].

The utility of RQCs in these situations derives from a myriad of quantitative properties they have been shown to possess. For example, RQCs quickly generate entanglement [85, 87, 127], lead to fast scrambling and decoupling of quantum information [128, 129], and act as efficient encoding circuits for good quantum error-correcting codes [130]. When the circuits are geometrically local, they lead to ballistic spreading of local operators [41, 42]. Furthermore, they form approximate unitary designs, that is, despite being composed of local gates, they efficiently approximate a global random unitary transformation up to any polynomial number of moments [37, 86, 131–133]. Meanwhile, as discussed extensively in Chapter 3, computing transition amplitudes of RQCs has been shown to be just as difficult as for arbitrary quantum circuits [4, 25–27], a fact that has been used to suggest that classical simulation of RQCs should require exponential time.

In this chapter, we focus on another property of random quantum circuits called *anti-concentration*. Roughly speaking, when we measure the circuit's output state in the computational basis, anti-concentration is the property that the distribution over measurement outcomes is fairly well spread across all possible outcomes, and not too concentrated onto just one or only a small portion of those outcomes. Quantitatively, our definition of anti-concentration depends on the *collision probability*, the probability that measurement outcomes from two independent copies of the circuit agree. An RQC architecture is said to be anti-concentrated if the collision probability is at most a constant factor larger than its minimal value. Understanding when this is the case is particularly important for knowing when RQCs are hard to classically

simulate. On the one hand, anti-concentration is a necessary ingredient in most formal hardness arguments for RQC simulation [4, 19, 22, 134–138]. On the other hand, certain classical algorithms for simulating RQCs require anti-concentration in order to be efficient, for example, the algorithms discussed in Refs. [30, 31] for noisy circuit simulation and the algorithm in Ref. [29] that spoofs the linear cross-entropy benchmarking metric introduced in Ref. [6].

In most previous work where RQC anti-concentration is needed, it has been asserted as an implication of the 2-design property (see, e.g., Refs. [136, 139]). However, the 2-design property is much stronger than what is required for anti-concentration. It was shown that n-qubit RQCs on a fully connected architecture form approximate 2-designs after roughly $O(n)$ depth [131], and this was later shown to also apply to geometrically local RQCs in 1D and improved to $O(n^{1/D})$ in D spatial dimensions [86]. However, recent work by Barak, Chou, and Gao [29]—using a similar method to the one presented here—showed that for 1D RQCs the collision probability converges in depth $O(\log(n))$, much faster than the 2-design depth of $O(n)$. They also conjectured that 2D RQCs anti-concentrate in depth $O(\sqrt{\log(n)})$.

In this chapter, we prove sharp bounds on the number of gates needed for anti-concentration in two RQC architectures. For 1D RQCs, we confirm the $O(\log(n))$ upper bound on the anti-concentration depth in Ref. [29], and add a lower bound that matches the upper bound even up to the constant prefactor of the $\log(n)$. We also show that an $\Omega(\log(n))$ lower bound on the depth needed for anti-concentration holds regardless of which RQC architecture we use, which refutes the conjecture from Ref. [29] that 2D RQCs anti-concentrate in $O(\sqrt{\log(n)})$ depth. We then consider a fully connected (i.e. not geometrically local) RQC architecture, where each gate acts on a pair of qudits chosen randomly among all $n(n-1)/2$ possible such pairs. We show that, for qubits (local dimension $q = 2$), $5n\log(n)/6$ gates are necessary and sufficient (up to subleading corrections) for anti-concentration to be achieved, which settles a conjecture in Ref. [86].

Our method employs a form of the stat mech method for RQCs, which we discussed generally in Chapter 2, and then applied to shallow 2D RQCs in Chapter 3, although its use in this chapter is completely self-contained. We show how the technique converts the collision probability into a weighted sum over bit assignments to each location in the circuit diagram, which can be viewed as a partition function for an Ising-like statistical mechanical model. The bit assignments can also be interpreted as a Markov chain, and the number of gates needed for anti-concentration ultimately translates into the time needed for certain expectation values to converge under the dynamics of the Markov chain. This method not only yields sharp quantitative bounds, it also produces an appealing qualitative explanation on how and why the collision probability reaches its limiting value, which allows for effective heuristic reasoning even in architectures that we have not explicitly considered here.

The main takeaways from our work are twofold. First, we have shown that anti-concentration is generally achieved much faster than the 2-design property. The fact that anti-concentration occurs in $\Theta(n \log(n))$ circuit size both in 1D and for the fully connected architecture—these being two opposite extremes of geometric locality—suggests that anti-concentration may require only $\Theta(n \log(n))$ size for *any* reasonably well-connected architecture. This comes in sharp contrast to the situation for unitary designs, where the scaling of the size needed with n is highly dependent on the architecture. Second, the fact that we can prove tight upper and lower bounds attests to the power of the stat mech method and suggests it might be similarly useful in other situations.

The structure of this chapter is as follows: in Section 4.2, we briefly define the collision probability and anti-concentration, so that we can precisely state our technical results in Section 4.3; in Section 4.4, we comment on related work and implications of our technical results; in Section 4.5, we overview the correspondence between collision probability and stat mech system, and discuss the intuition behind log-depth anti-concentration that we glean from it; finally, in Section 4.6, we recap the takeaways of the chapter and open problems for the future. Following the outlook section, several appendices appear with more formal definitions and proofs of all the technical claims in the chapter.

4.2 Setup and definition of anti-concentration

In the previous chapter, we discussed various constant-depth 2D RQC architectures, which were families of quantum circuit diagrams indexed by the circuit size n, where the actual gates in the diagram were always chosen randomly from the Haar measure. In this chapter, we extend this concept to deeper circuits and define an RQC architecture as simply an instruction set on how to draw a circuit diagram given both the number of qudits n (each with local Hilbert space dimension q) and the size s of the circuit. The two architectures we consider specifically in this chapter are the *1D architecture* (with periodic boundary conditions), where qudits are arranged in a ring and alternating layers of nearest-neighbor two-qudit gates are performed, and the *complete-graph architecture*, where each two-qudit gate is chosen uniformly at random among all $n(n-1)/2$ possible qudit pairs. Formal definitions of these architectures appear in Appendix 4.A. Note that in the formal analysis, we also include a layer of n single-qudit gates at the beginning of the circuit, which are not counted toward the circuit size s. These gates might be regarded as fixing the local basis for the initial input state.

The associated RQC ensemble for an RQC architecture is formed by following this instruction set to obtain a circuit diagram, and then choosing the value of each gate in the diagram independently from the Haar measure. Choices for each gate determines an overall $q^n \times q^n$ unitary U that the circuit implements. Each instance U is associated with an output probability

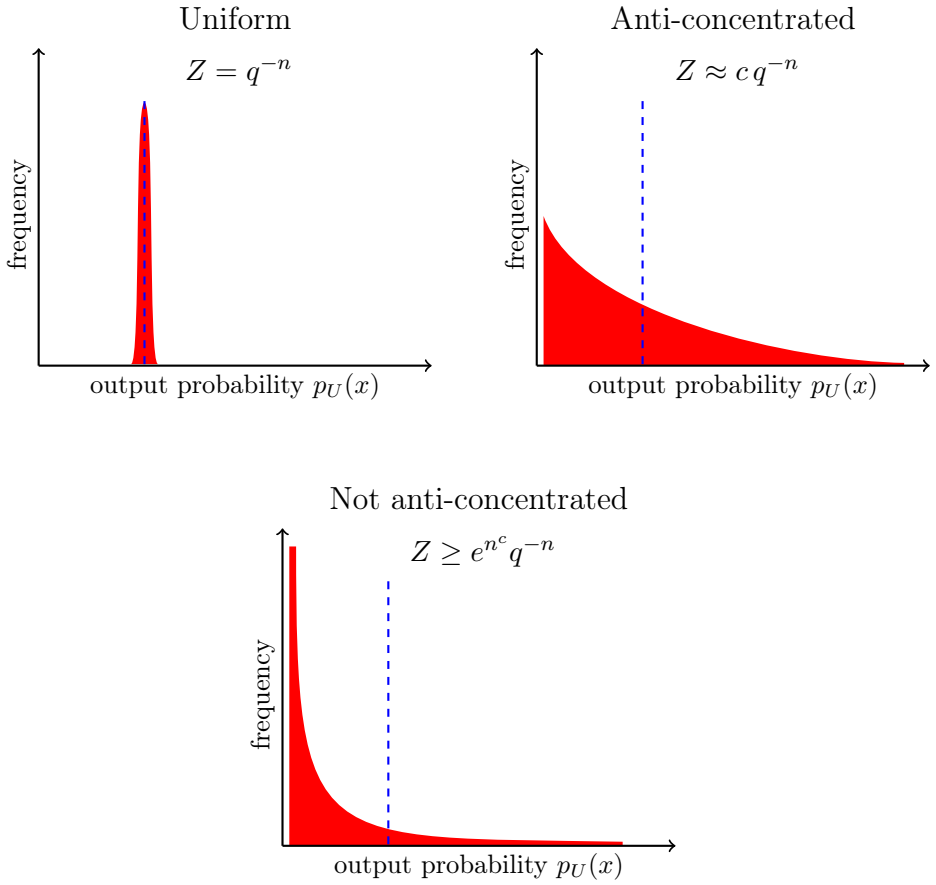


Figure 4.1: A caricature of anti-concentration. For the uniform distribution, which is completely anti-concentrated, all q^n outcomes are allocated probability mass q^{-n} (dashed blue line) and the collision probability is $Z = q^{-n}$. For globally Haar-random unitaries, the output probabilities are on average q^{-n} but have some non-zero variance, and the collision probability is $Z \approx 2q^{-n}$. Whenever $Z \approx cq^{-n}$ for some c independent of n, we call the distribution anti-concentrated. For low-depth RQCs, the mean output probability is q^{-n}, but the variance is much larger, and the collision probability is much larger than q^{-n}. Most of the probability mass is concentrated onto a few measurement outcomes, with the remainder of the outcomes being assigned a very small amount of mass, leading to the frequency of circuit instances for which $p_U(0^n)$ is close to 0 to be large.

distribution p_U over q^n possible computational basis measurement outcomes $x \in [q]^n$, (where $[q] = \{0, 1, \ldots, q-1\}$): $p_U(x) = |\langle x|U|0^n\rangle|^2$. Note that this distribution is referred to by the notation p_{ideal} in Chapter 1 and Chapter 5.

Anti-concentration tries to capture the notion that the probability mass in p_U is well spread out over all the outcomes. The uniform distribution, where each output is allocated q^{-n} fraction of the total probability mass, is the ultimate anti-concentrated distribution because the mass is exactly equally spread, but we say a distribution is still anti-concentrated as long as the average fluctuations from uniform are no larger than $O(q^{-n})$. This definition is captured precisely by the *collision probability*, which is $\sum_x p_U(x)^2$. The collision probability gives the probability that measurement outcomes from two independent copies of the circuit are identical. It is also proportional to the second moment (and thus is related to the variance) of the output probability of a randomly chosen bit string. If p_U is the uniform distribution, then the collision probability is q^{-n}, its minimal possible value. For an RQC architecture at a specified qubit number n and circuit size s, we consider the collision probability averaged over the randomly chosen circuit instances U, defined by the expression

$$Z = \mathop{\mathbb{E}}_{U} \left[\sum_{x \in [q]^n} p_U(x)^2 \right] = q^n \mathop{\mathbb{E}}_{U} \left[p_U(0^n)^2 \right] , \tag{4.1}$$

where the second equality holds because by symmetry, each of the q^n terms in the sum yields the same number under expectation as long as at least one Haar-random gate acts on each qudit.

We say an RQC architecture with n qudits and s gates is anti-concentrated if there is a constant α (independent of n) with $0 < \alpha \leq 1$ for which $Z \leq \alpha^{-1} q^{-n}$, i.e. that the collision probability is only a constant factor larger than its minimal value. In particular, our theorem statements roughly correspond to the choice $\alpha = 1/4$, but other choices of α would yield the same results up to leading order. If desired, Markov's inequality can then be used to bound the fraction of the randomly chosen U whose collision probability is larger than some constant multiple of Z. Moving forward, for convenience, when we say collision probability we will mean the average collision probability Z.

Very shallow circuit architectures are not anti-concentrated: there are expected to be some output probabilities x for which $p_U(x)$ is exponentially larger than the mean of q^{-n}. As the circuit gets deeper, we expect the probability distribution to become closer to uniform, but even at infinite depth, when the circuit unitary U becomes a globally Haar-random $q^n \times q^n$ unitary, the output distribution still does not become completely uniform. In this case, the output distribution will typically follow a Porter-Thomas distribution[1], and Z can be

[1] In the Porter-Thomas distribution, the frequency at which $p_U(x) = p$ is proportional to $\exp(-p/q^{-n})$, illustrated roughly in the middle diagram of Figure 4.1.

exactly computed as

$$\lim_{s\to\infty} Z = Z_H = \frac{2}{q^n + 1}, \tag{4.2}$$

roughly twice as large as the minimal value of q^{-n} associated with the uniform distribution. This statement is proved using the techniques described later. Refer to Figure 4.1 for a graphical illustration of these cases.

While one could capture the notion of anti-concentration with a different definition, the definition we choose is useful and relevant because it has concrete ramifications in all of the previously mentioned applications of anti-concentration. For example, one implication of our definition (by application of the Paley-Zygmund inequality) is that if $Z \leq \alpha^{-1} q^{-n}$, then for any $0 \leq \beta \leq 1$,

$$\Pr_U[p_U(x) \geq \beta q^{-n}] \geq \alpha(1-\beta)^2\,, \tag{4.3}$$

meaning that for at least a constant fraction of the circuit instances the probability of a given measurement outcome x is at least a constant fraction β of the mean measurement probability q^{-n}. This sort of inequality is the relevant one for turning good additive approximations into good multiplicative approximations (with reasonable probability), employed in, for example, Refs. [4, 19, 22, 134–138] in order to argue that it is hard to classically sample output distributions for a large fraction of instances up to small total variation distance error (for more details, see Section 4.4.2). In fact, equations like Eq. (4.3) are sometimes taken to be the definition of anti-concentration [136], which is a weaker definition than ours since, in principle, it can hold even in cases where our definition does not.

4.3 Overview of contributions

Architecture	s_{AC} **upper bound**	s_{AC} **lower bound**
general	$O(n^2)$	$\Omega(n\log(n))$
1D	$c_{\text{1D}}\, n\log(n) + O(n)$	$c_{\text{1D}}\, n\log(n) - O(n)$
complete-graph	$c_{\text{cg}}\, n\log(n) + O(n)$	$c_{\text{cg}}\, n\log(n) - O(n)$

Table 4.1: Summary of results: upper and lower bounds on the circuit size s_{AC} at which anti-concentration is achieved for different random circuit architectures. The constants are given by $c_{\text{1D}} = (2\log((q^2+1)(2q)))^{-1}$ and $c_{\text{cg}} = (q^2+1)/(2(q^2-1))$, where q is the qudit local dimension ($q = 2$ for qubits).

We show that the collision probability is given by a discrete sum, which we interpret as the expectation value of a certain stochastic process or as the partition function of an Ising-like statistical mechanical system. The corre-

spondence between the collision probability and the discrete sum is described in Section 4.5, and a complete derivation is provided in Appendix 4.B.

Analyzing our expression, we derive rigorous upper and lower bounds on the collision probability generally and for two specific architectures. These bounds are stated here and the proofs are provided in the Appendices. These bounds are then used to form upper and lower bounds quoted in Table 4.1 on the anti-concentration size s_{AC}, defined as the minimum circuit size required such that $Z \leq 2Z_H$. The constant 2 in that definition is arbitrary but a different constant would only lead to linear-in-n changes to s_{AC}, which would be subleading and would not affect any of the statements in Table 4.1. All logarithms in this thesis are natural logarithms.

Collision probability upper bounds

Our upper bounds take the following form:

$$Z \leq Z_H \left(1 + e^{-\frac{2a}{n}(s-s^*)}\right) , \tag{4.4}$$

where the constant a is independent of n and depends on the circuit architecture, and s^* is a function of n that also depends on the architecture. Thus, if the anti-concentration size s_{AC} is defined to be the minimum size s such that $Z \leq 2Z_H$, then we have $s_{AC} \leq s^*$. Specifically, we have the following results, which are restated here as theorems, and proved rigorously in the Appendices.

First, we consider the *1D architecture* with periodic boundary conditions, where the qudits are arranged on a ring and alternating layers of $n/2$ nearest-neighbor Haar-random gates are applied.

Theorem 4.1. *For the 1D architecture, Eq. (4.4) holds with*

$$a = \log\left(\frac{q^2+1}{2q}\right) \tag{4.5}$$

$$s^* = \frac{1}{2a} n \log(n) + n\left(\frac{1}{2a}\log(e-1) + \frac{1}{2}\right) \tag{4.6}$$

whenever $s \geq s^*$.

Since this depth of the 1D architecture is given by $d = 2s/n$, we can define $d^* = 2s^*/n = a^{-1}\log(n) + O(1)$ for 1D and conclude that the "anti-concentration depth" d_{AC} satisfies $d_{AC} \leq d^* = O(\log(n))$.

Similarly, we show an upper bound for the *complete-graph architecture*, where each gate acts on a random pair of qudits without regard for their spatial proximity.

Theorem 4.2. *For the complete-graph architecture, Eq. (4.4) holds with*

$$a = \frac{(q-1)^2}{2(q^2+1)} \tag{4.7}$$

$$s^* = \frac{q^2+1}{2(q^2-1)} n \log(n) + cn \tag{4.8}$$

whenever $s \geq s^*$, for a constant c that is independent of n.

A size-s circuit diagram chosen randomly from the complete-graph architecture will have depth at most $O(s \log(n)/n)$ with high probability [129], meaning that $O(\log(n)^2)$ depth is typically sufficient for anti-concentration in the complete-graph architecture.

We also consider general architectures. We define a property called *regularly connected*, which applies to an RQC architecture when for any partition of qubits into two sets, there will be a gate in the circuit that couples the two sets at least once every $O(n)$ gates. The precise definition is given later in Definition 4.5. Nearly all natural architectures have this property, including standard architectures in D spatial dimensions for any D.

Theorem 4.3. *If an architecture is regularly connected, then Eq. (4.4) holds with $a = \Theta(1)$ and $s^* = \Theta(n^2)$.*

This corresponds to $\Theta(n)$ gates per qudit. This result is weaker than our specific results for the 1D and complete-graph architectures, and we conjecture that it can be strengthened to $\Theta(\log(n))$ gates per qudit.

Conjecture 4.1. *Theorem 4.3 can be improved to $s^* = \Theta(n \log(n))$.*

Collision probability lower bounds

Our lower bounds on the collision probability take the form

$$Z \geq \frac{Z_H}{2} \exp\left(A e^{\log(n) - Bs/n}\right), \tag{4.9}$$

for constants A and B that are independent of n. (The lower bound for the complete-graph architecture takes a different but very similar form.) This form implies that if s grows with n like $s \approx fn \log(n)/B$ for some $f < 1$, then we have $Z/Z_H \geq \frac{1}{2} e^{An^{1-f}}$, which becomes arbitrarily large as $n \to \infty$, meaning the architecture is not anti-concentrated. This puts a lower bound on the anti-concentration size s_{AC} of $s_{AC} \geq n \log(n)/B - O(n)$.

Specifically, we show a general lower bound, as well as specific lower bounds for the 1D and complete-graph architectures.

Theorem 4.4. *For any RQC architecture with s two-qudit gates, the following holds:*

$$Z \geq \frac{Z_H}{2} \exp\left(\frac{\log(q)}{q+1} e^{\log(n) - 2\log\left(q^2+1\right)s/n}\right). \tag{4.10}$$

This has the consequence that if s_{AC} and d_{AC} are defined as the minimum size and minimum depth for which $Z \leq 2Z_H$, then

$$s_{AC} \geq \left(2\log\left(q^2+1\right)\right)^{-1} n\log(n) - O(n) \tag{4.11}$$

$$d_{AC} \geq \left(\log\left(q^2+1\right)\right)^{-1} \log(n) - O(1)\,. \tag{4.12}$$

We improve on the general lower bound for the two specific architectures we consider.

Theorem 4.5. *For the 1D architecture, there exists a constant A such that*

$$Z \geq \frac{Z_H}{2} \exp\left(A e^{\log(n) - 2as/n}\right), \tag{4.13}$$

where $a = \log((q^2+1)/(2q))$ is the same as for the upper bound in Eq. (4.5).

This implies that in 1D

$$s_{AC} \geq (2a)^{-1} n\log(n) - O(n) \tag{4.14}$$

$$d_{AC} \geq a^{-1}\log(n) - O(1)\,, \tag{4.15}$$

which is tight with the upper bound up to subleading corrections.

Theorem 4.6. *For the complete-graph architecture,*

$$Z \geq \frac{Z_H}{2} \exp\left(\frac{\log(q)}{q+1} e^{\log(n) + \log\left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right)s}\right). \tag{4.16}$$

Although a slightly different form than the other lower bounds, this still yields the conclusion

$$s_{AC} \geq \frac{q^2+1}{2(q^2-1)} n\log(n) - O(n)\,, \tag{4.17}$$

which is tight with the upper bound up to subleading corrections. When $q = 2$ (qubits), the prefactor of the $n\log(n)$ is $5/6$, settling a conjecture proposed in Ref. [86].

The upper and lower bounds together allow us to conclude that $s_{AC} = \Theta(n\log(n))$ for both the 1D architecture and the complete-graph architecture, and in fact we have matching upper and lower bounds on the constant prefactor of the $n\log(n)$.

We note that, for $q \geq 5$, our results have the counter-intuitive implication that the 1D architecture anti-concentrates faster than the complete-graph architecture, even though it is geometrically local. We argue that this is an artifact of the definition of the models, and can be explained by the fact that the qudit pairs acted upon by the gates in the complete-graph architecture are chosen randomly, while the qudit pairs in the 1D architecture are not random; in fact, in the latter case they are optimally packed into layers of $n/2$ non-overlapping gates. As q increases, anti-concentration becomes arbitrarily fast for the 1D architecture (the coefficient of the $n \log(n)$ decreases like $1/\log(q)$). Meanwhile, for the complete-graph architecture, no matter how large q is, there will always be some minimum number of gates—roughly $n \log(n)/2$—needed simply to guarantee that all the qudits have been involved in the circuit with high probability. We suspect that a parallelized version of the complete-graph architecture would anti-concentrate with a slightly better constant than the 1D architecture.

4.4 Related work and implications

Here we highlight a few relevant previous works and emphasize how our results fit in.

- Harrow and Mehraban [86] studied how quickly RQCs form approximate unitary t-designs and anti-concentrate for various architectures. For geometrically local circuits, they showed that the approximate t-design property is achieved after only $O(n^{1/D})$ depth in D spatial dimensions, the first work to break the $O(n)$ barrier for designs. Since anti-concentration follows from the approximate 2-design property, their work implies an $O(n^{1/D})$ upper bound on the anti-concentration depth. We show that for $D = 1$, the anti-concentration depth is actually $\Theta(\log(n))$ and we conjecture that this is also the case for $D \geq 2$, but we do not prove this, so the $O(n^{1/D})$ bound remains the best known for $D \geq 2$.

  They also considered the question of anti-concentration in the complete-graph architecture and showed an upper bound on the anti-concentration size of $O(n \log(n)^2)$ and a lower bound of $\Omega(n \log(n))$. They used heuristic reasoning to conjecture that (for $q = 2$) the anti-concentration size should be $5n \log(n)/3$, up to leading order. We are able to show that to leading order the anti-concentration size for the complete-graph architecture is $5n \log(n)/6$. This is off by a factor of 2 from the conjecture stated in their paper, which we suspect is due to a minor error in their heuristic reasoning.

- Barak, Chou, and Gao [29] developed a classical algorithm for shallow RQCs that achieves a non-negligible score on the Linear Cross-Entropy Benchmarking (XEB) metric despite not performing a full simulation of the RQCs. The Linear XEB metric was used by Google to verify its 2019 quantum computational supremacy experiment [6]. Barak, Chou, and Gao show that if a depth-d RQC architecture in D spatial dimensions has collision probability

Z, their algorithm achieves a score of ϵ with high probability after a total runtime $(2^n Z) \cdot \exp\left(\epsilon\, 15^{-d}\right) \cdot \mathrm{poly}(n, 2^{d^D})$ (here $q = 2$). They prove that $Z = O(2^{-n})$ after $d = \Omega(\log(n))$ for 1D RQCs, which is equivalent to our Theorem 4.1. This shows that their algorithm achieves a $\epsilon \geq 1/\mathrm{poly}(n)$ score in polynomial time for logarithmic depth 1D RQCs. For 2D RQCs, they conjecture that $Z = O(2^{-n})$ after depth $d = O(\sqrt{\log(n)})$, which would imply their algorithm achieves $\epsilon \geq 1/\mathrm{poly}(n)$ score in polynomial time at that depth. Our Theorem 4.4 contradicts their conjecture by showing generally that $2^n Z \geq \exp\left(n^{1-o(1)}\right)$ when d is sublogarithmic.

- Our method performs expectations over individual gates in the RQC using formulas for Haar integration, a strategy that has also been used on similar problems in the past. Many works have used this strategy to form a random walk over Pauli strings with wide-ranging applications [85, 86, 128–131, 140–142]. Our analysis applies this strategy in a distinct way that more closely resembles a series of works that interpret the resulting expression as the partition function of classical statistical mechanical models [36–42, 44, 45, 47], which are discussed more extensively in Chapter 2. Here, we analyze those partition functions using a Markov chain analysis, but our Markov chain has different transition rules compared to the Pauli string Markov chain.

4.4.1 Connection to 2-design

Anti-concentration for random quantum circuits (as well as some Hamiltonian models) is often established as a consequence of the convergence to approximate unitary 2-designs, where approximately reproducing the first two moments of the Haar measure allows one to bound the RQC collision probability. For both 1D and complete-graph RQCs, size $O(n^2)$ circuits (of linear depth) form approximate 2-designs and therefore anti-concentrate. There are a number of definitions of approximate unitary designs utilizing different norms; we briefly comment on the definitions and requirements for anti-concentration in this architecture.

As we review in Appendix 4.F, defined in terms of the diamond norm, ε-approximate 2-designs have a collision probability upper bounded by Z_H up to additive error. In order to achieve anti-concentration, ε must be taken to be exponentially small (i.e. we require $\varepsilon = 1/q^{2n}$). Ref. [133] introduced a stronger notion of approximate design in terms of the complete positivity of the difference in channels. Under this strong definition, 2-designs bound the collision probability up to relative error with respect to the Haar value, and thus anti-concentrate. A much weaker definition of approximate design is the operator norm of the moment operators, often called the tensor product expander (TPE) condition. Interestingly, TPEs also bound the collision probability up to additive error, but again the error needs to be exponentially small to achieve anti-concentration.

Random quantum circuits on the 1D architecture form ε-approximate 2-designs, in both diamond norm and the stronger definition, when the circuit size is $O(n(n + \log(1/\varepsilon)))$. Moreover, 1D random circuits actually form ε-approximate TPEs in constant depth, when the circuit size is $O(n \log(1/\varepsilon))$. But again, anti-concentration requires that ε be taken to be $\varepsilon = 1/q^{2n}$, thus mandating linear depth. So to establish that the collision probability is bounded up to a relative error, as in the definition of anti-concentration, using unitary 2-designs or a general bound on the moments necessitates linear depth. For non-local RQCs defined on a complete-graph, the best known upper bounds on the approximate 2-design depth are the same as for the 1D architecture. But it has been conjectured that this may be improved for non-local RQCs, which would close the gap between the 2-design time and the depth required for anti-concentration.

To further emphasize the distinction between anti-concentration and unitary 2-designs, we note that anti-concentration can be achieved for specific short-depth circuits without generating entanglement across the system (indeed, a circuit consisting of a single layer of single-qubit Hadamards suffices). Moreover, for an ensemble of random quantum circuits, anti-concentration can be equivalently phrased as the statement that certain matrix elements of second moment operator $\mathbb{E}_U[U^{\otimes 2} \otimes U^{*\otimes 2}]$ reach the Haar value of $2/q^{2n}$ after some depth. Whereas the approximate 2-design condition gives that $\mathbb{E}_U[\langle\psi|U^{\otimes 2} \otimes U^{*\otimes 2}|\psi\rangle]$ is small for all states $|\psi\rangle$, even those that are entangled across the tensor copies. As we show in Appendix 4.F, there are necessarily some states which require linear depth to equilibrate to the minimal Haar value, at least for RQCs on the 1D architecture.

4.4.2 Implications for arguments on hardness of simulation

Anti-concentration is a key ingredient in hardness-of-simulation arguments [4, 19, 22, 25–27, 134–138, 143] that underlie quantum computational supremacy proposals. In this section we review its role in those arguments (briefly covered previously in Chapter 1) and the implications our results have in this context.

The starting point for these hardness arguments is the long-known observation that the answer to a hard classical problem can be encoded into the output probability $p_U(x)$ of a quantum circuit U.[2] Thus, exactly computing $p_U(x)$ for arbitrary U and x should not be possible in classical polynomial time. This remains true even if one only needs to compute $p_U(x)$ up to some constant relative error. The ultimate goal in the context of quantum computational supremacy is to show that there is no polynomial-time classical algorithm that approximately simulates random circuits (or at least to give extremely convincing evidence in favor of this conclusion). More precisely, the approximate simulation task is to produce samples from a distribution p'_U for

[2] See Chapter 1 for discussion on how this is done.

which

$$\frac{1}{2}\|p_U - p'_U\|_1 = \frac{1}{2}\sum_x |p_U(x) - p'_U(x)| = \varepsilon \tag{4.18}$$

for some small $\varepsilon = O(1)$, and to do this for a large fraction of U drawn randomly from some random ensemble. Turning the starting point into the ultimate goal requires a few steps (some of which rely on conjecture). Anti-concentration is one of these steps.

The primary role anti-concentration plays is to turn a small *additive* difference $|p_U(x) - p'_U(x)|$ for most x into a small *relative* difference $r(x)$ for most x, where

$$r(x) = \frac{|p_U(x) - p'_U(x)|}{p_U(x)}\,. \tag{4.19}$$

If Eq. (4.18) is obeyed, then the value of $|p_U(x) - p'_U(x)|$ is on the order of ε/q^n for most x. Meanwhile, the mean value of $p_U(x)$ for random x is exactly $1/q^n$. If $p_U(x)$ is anti-concentrated, then for most x, $p_U(x)$ will be within a constant factor of the mean, as shown in the middle diagram of Figure 4.1, and $r(x) = O(\epsilon)$ will hold for most x. However, if $p_U(x)$ is not anti-concentrated, then $p_U(x)$ will be much smaller than the mean for most x, as depicted in the right diagram of Figure 4.1. This means that without anti-concentration, $r(x) \gg \epsilon$ for most x, which is problematic because the hard classical problems encoded into $p_U(x)$ are no longer hard when the relative error is extremely large, so anti-concentration appears to be necessary if there is any hope of completing the hardness argument using existing techniques.

Even if anti-concentration holds, more is needed to show hardness of approximately simulating RQCs. One must turn hardness of computing $p_U(x)$ into hardness of *sampling* from p_U and also turn hardness for arbitrary U into hardness for a *random* U. There are techniques that work for each of these steps individually, but currently they do not work together simultaneously, and thus an additional conjecture must be made.

Our work puts sharp bounds on the number of gates needed for anti-concentration to hold in multiple RQC architectures, which constrains when these hardness arguments have the potential to work. Our finding that the number of gates per qudit needed for anti-concentration grows only logarithmically in n in the 1D and complete-graph architectures implies that RQC-based quantum computational supremacy might be achievable at a shallower circuit depth than previously believed. For example, Google's 2019 quantum computational supremacy experiment was based on 2D RQC's of depth exceeding the $\sqrt{n}$ diameter of the qubit array [5, 6]. The fact that 1D circuits anti-concentrate in $\Theta(\log(n))$ depth is evidence that 2D circuits should have the same scaling (if anything, anti-concentration should happen faster in 2D). Thus a similar quantum computational supremacy experiment might be equally defensible at $\Theta(\log(n))$ depth instead of $\Omega(\sqrt{n})$ depth. We note, however, that

there are other reasons to want to go to larger depth (e.g. classical simulation via tensor network methods becomes harder at larger depths).

Without anti-concentration, the hardness-of-simulation arguments appear to break down, but this does not generally imply that simulation is easy. On this topic, we refer to the results of Chapter 3, where we proposed a classical algorithm for solving the approximate simulation problem for 2D RQCs. We proved that the algorithm is efficient for a certain constant-depth (and thus not anti-concentrated) 2D RQC architecture, but it is conjectured to become inefficient once the depth exceeds a larger constant threshold. Thus, the complexity of the algorithm transitions to inefficient before the circuits become anti-concentrated, suggesting that in 2D there could be a regime where the RQCs are too shallow to be anti-concentrated, but classical simulation is still hard.

4.5 Summary of method and intuition for logarithmic convergence

The main technical contribution of our work is to derive a correspondence between the collision probability and a discrete sum (which can be interpreted as the partition function of a classical statistical mechanical model or as the expectation of a Markov chain), and then to derive rigorous upper and lower bounds on the sum. Here we describe the correspondence along with a brief example for a simple random quantum circuit in Figure 4.2. We also explain why this correspondence leads us to expect anti-concentration to be achieved after $\Theta(n \log(n))$ gates in most architectures. In Appendix 4.B, we give a more careful derivation of this correspondence, and then in Appendix 4.C, Appendix 4.D, and Appendix 4.E, we use it to rigorously prove the upper and lower bounds quoted in Table 4.1.

Recall that we wish to compute the collision probability, defined as

$$Z = q^n \mathop{\mathbb{E}}_U \left[p_U(0^n)^2 \right] \tag{4.20}$$

$$= q^n \mathop{\mathbb{E}}_U \left[\langle 0^n|^{\otimes 2} U^{\otimes 2} |0^n\rangle\langle 0^n|^{\otimes 2} U^{\dagger \otimes 2} |0^n\rangle^{\otimes 2} \right] , \tag{4.21}$$

where U is the unitary enacted by the random quantum circuit. The Haar measure uniformly covers the unitary group so, intuitively speaking, taking the expectation over application of a Haar-random gate removes much of the bias in the quantum state; we use a technique that allows us to effectively keep track of only n bits of information about the n-qudit state after the application of (two copies of) each Haar-random gate. Instead of 0 or 1, our bits take values I or S, because they are associated with the identity and swap operations on two qudit copies.

In particular, if V is a $q \times q$ Haar-random matrix and σ is an operator on two copies of a q-dimensional Hilbert space, then the quantity $\mathbb{E}_V[V^{\otimes 2} \sigma V^{\dagger \otimes 2}]$ is equal to a linear combination of the identity operation I on the two copies of the Hilbert space, and the swap operation S on the two copies of the Hilbert

space. Specifically, it is given by

$$\left(\frac{\operatorname{tr}(\sigma) - q^{-1}\operatorname{tr}(\sigma S)}{q^2-1}\right) I + \left(\frac{\operatorname{tr}(\sigma S) - q^{-1}\operatorname{tr}(\sigma)}{q^2-1}\right) S\,. \tag{4.22}$$

This well-known formula is derived in Appendix 4.B; it is also a direct consequence of Eq. (2.10) from Chapter 2.

By applying the formula to each of the two-qudit Haar-random gates sequentially, the state (which begins in $|0^n\rangle\langle 0^n|^{\otimes 2}$) evolves as a sum over n-fold tensor products of identity and swap operations. Each of these n-fold tensor products is labeled by an n-bit vector that we call a *configuration* $\vec{\nu} \in \{I, S\}^n$. For a circuit with s two-qudit gates, each term in the resulting sum is then associated with a length-$(s+1)$ sequence of configurations $\gamma = (\vec{\gamma}^{(0)}, \ldots, \vec{\gamma}^{(s)})$, which we call a *trajectory*. Each trajectory γ has a certain non-negative coefficient in the sum, allowing us to write

$$Z = \frac{1}{(q+1)^n} \sum_{\gamma} \operatorname{weight}(\gamma) \tag{4.23}$$

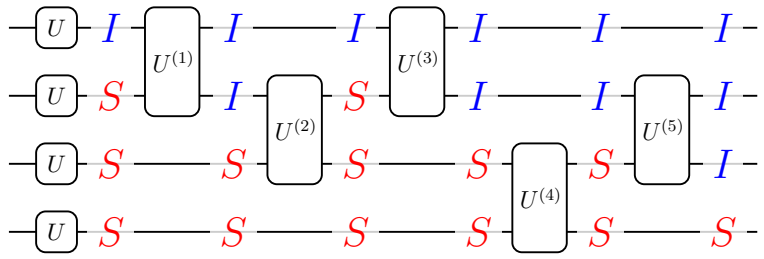
for a fairly simple weighting function, described as follows and derived more carefully in Appendix 4.B.

First of all, the weight for most trajectories is simply 0. In order for a trajectory to have non-zero weight, it must obey the following rules. If the gate at time step t acts on qudits a and b, then the configuration values $\gamma_a^{(t)}, \gamma_b^{(t)} \in \{I, S\}$ at positions a and b must be equal, either both I or both S. Thus if the values disagreed at the previous time step, i.e. $\gamma_a^{(t-1)} \neq \gamma_b^{(t-1)}$, one of the bits must be flipped during the transition from $\vec{\gamma}^{(t-1)}$ to $\vec{\gamma}^{(t)}$. If the values at positions a and b already agreed at time step $t-1$, they must remain unchanged from time step $t-1$ to time step t. Moreover, the bit values at the other $n-2$ positions must also remain unchanged from time step $t-1$ to time step t.

For trajectories that obey these rules, the weight begins at 1, but each time a bit flip occurs, the weight is reduced by a constant factor of $2/5$ (for qubits; generally $q/(q^2+1)$). Thus, the most significant terms in the weighted sum are the terms with the fewest bit flips along the trajectory.

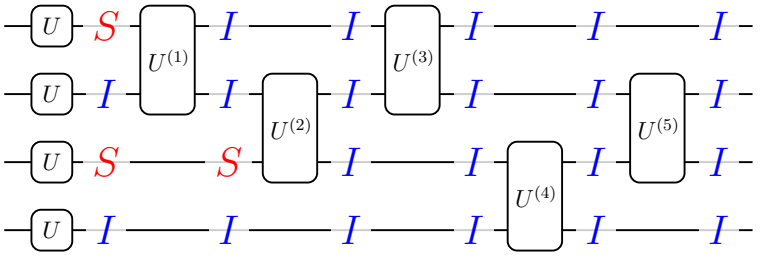
The expression for Z as a weighted sum can alternatively be interpreted as a partition function for an Ising-like classical statistical mechanical model since it is a weighted sum over "spin" configurations for spins with two possible values, or it can be interpreted as the expectation of a certain quantity over a simple Markov chain that generates the sequence $(\vec{\gamma}^{(0)}, \ldots, \vec{\gamma}^{(s)})$. We take the latter approach in our application of the method to prove our upper and lower bounds. See Figure 4.2 for an example of two trajectories for a simple RQC, along with a calculation of their weight.

$$\text{collision probability} = Z = \frac{1}{(q+1)^n} \sum_{\gamma} \text{weight}(\gamma)$$



$$\gamma = \left(\vec{\gamma}^{(0)}, \quad \vec{\gamma}^{(1)}, \quad \vec{\gamma}^{(2)}, \quad \vec{\gamma}^{(3)}, \quad \vec{\gamma}^{(4)}, \quad \vec{\gamma}^{(5)} \right)$$

$$\text{weight}(\gamma) = \frac{2}{5} \times \frac{2}{5} \times \frac{2}{5} \times 1 \times \frac{2}{5} = \frac{16}{625}$$



$$\gamma = \left(\vec{\gamma}^{(0)}, \quad \vec{\gamma}^{(1)}, \quad \vec{\gamma}^{(2)}, \quad \vec{\gamma}^{(3)}, \quad \vec{\gamma}^{(4)}, \quad \vec{\gamma}^{(5)} \right)$$

$$\text{weight}(\gamma) = \frac{2}{5} \times \frac{2}{5} \times 1 \times 1 \times 1 = \frac{4}{25}$$

Figure 4.2: Two example trajectories for a quantum circuit diagram with $n = 4$ qubits and $s = 5$ gates. Each gate displayed is chosen randomly from the Haar measure over single or two-qubit unitaries. The collision probability Z is expressed as a weighted sum over trajectories $\gamma = (\vec{\gamma}^{(0)}, \ldots, \vec{\gamma}^{(s)})$, which are length-$(s+1)$ sequences of assignments ("configurations") of I or S to each of the n qudits. When the input bits to a gate are assigned opposite values, one must be switched at the next configuration in the sequence. These bit flips happen at gates 1, 2, 3, and 5 in the first example, and at gates 1 and 2 in the second example. Each bit flip results in a reduction of the weight by a factor $2/5$ (when $q = 2$). In the second example, the trajectory reaches one of the fixed point configurations where all n values agree; this is not the case in the first example. Trajectories that quickly reach a fixed point generally have larger weights and make up most of the contribution to the collision probability.

The correspondence given in Eq. (4.23) is powerful because we have a good sense of what to expect from the weighted sum over trajectories, and we can draw conclusions that were not obvious from the definition of the collision probability itself. For example, we can straightforwardly analyze the infinite circuit-size limit. In this limit, each positive-weight trajectory γ will be forced to keep flipping bits (each time a two-qudit gate acts on a disagreeing pair of bits) until it reaches a fixed point, either I^n or S^n, in which case bits can no longer be flipped since all the bits agree. Let $Q(x)$ be the total weight of all trajectories that begin at a configuration with x S assignments and $n-x$ I assignments. At some point in the circuit, a disagreeing pair of bits will be acted upon by a gate, and one of the bits must flip, sending the number of S assignments either to $x-1$ or $x+1$ and reducing the weight by $q/(q^2+1)$. Since there are an infinite number of gates, the following recursion relation must be obeyed

$$Q(x) = \frac{q}{q^2+1}\left(Q(x-1) + Q(x+1)\right) , \tag{4.24}$$

which, by imposing the boundary conditions $Q(0) = Q(n) = 1$, has the unique solution

$$Q(x) = \frac{q^x + q^{n-x}}{q^n+1} . \tag{4.25}$$

Moreover, for each x, there are $\binom{n}{x}$ configurations, each contributing weight $Q(x)$, so

$$\lim_{s\to\infty} Z = \frac{\sum_{x=0}^{n}\binom{n}{x}(q^x+q^{n-x})}{(q+1)^n(q^n+1)} = \frac{2}{q^n+1} = Z_H \tag{4.26}$$

reproducing the value Z_H that would be obtained if the random quantum circuit were one large $q^n \times q^n$ Haar-random transformation instead of a series of $q^2 \times q^2$ two-qudit gates. (The fact that a $q^n \times q^n$ Haar-random transformation yields Z_H is a direct consequence of Eq. (4.22) with the substitution $q \to q^n$.) This conclusion makes sense since a random circuit with an infinite number of 2-local Haar-random gates should enact a global Haar-random transformation.

When the circuit size is a finite number s, we have $Z > Z_H$, corresponding to the fact that many trajectories have not yet reached a fixed point and are overweighted compared to their contribution to Z_H. As the circuit size increases, more of the trajectories get closer to the fixed point and Z approaches Z_H. The point at which anti-concentration is achieved is intimately connected with the point at which most of the weight can be accounted for by trajectories that have reached a fixed point. A depiction of this process at $n = 60$ is given in Figure 4.3.

Our quantitative challenge is to understand, for a certain RQC architecture, how quickly these trajectories approach the fixed points, and consequently how quickly Z approaches Z_H, as the circuit size increases. Recall that we define the *anti-concentration size* s_{AC} to be the circuit size (as a function of the number of qudits n) needed for Z to be only a constant factor

larger than Z_H. Perhaps surprisingly, we find in multiple architectures that $s_{AC} = \Theta(n \log(n))$, corresponding to only $\Theta(\log(n))$ gates per qudit. We can explain this observation heuristically by generating trajectories γ at random with probability proportional to weight(γ) (in the statistical mechanical interpretation, this corresponds to drawing samples from the thermal distribution). For typical trajectories generated in this fashion, each additional layer of $\Theta(n)$ gates will cause the trajectory to move a constant fraction of the way closer to terminating at a fixed point. Since trajectories typically begin on the order of n bit flips away from the fixed point (i.e. the initial configuration typically has $\Theta(n)$ I assignments and $\Theta(n)$ S assignments), $\Omega(\log(n))$ layers are necessary and sufficient for typical trajectories to get within a constant distance from the fixed point.

This heuristic statement is perhaps confirmed most clearly in the complete-graph architecture, where qudit pairs are chosen uniformly at random. Here let $x \ll n$, and suppose that the current configuration at time step t has value S at x of the n positions and value I at the other $n - x$ positions. If we perform gates on $n/2$ random pairs of qudits, we will expect roughly x of those pairs to couple an I value with an S value. Each time this happens, a bit must be flipped and there is an opportunity for the trajectory to move closer to the fixed point I^n. Thus, we expect the number of S values in the configuration at time step $t+n/2$ to have decreased by an amount proportional to x. After $\Theta(n \log(n))$ gates, we expect the trajectory to be at (or very close to) the fixed point I^n with high probability. Fewer gates would leave most trajectories too far from the fixed point for anti-concentration to have been reached. In Figure 4.3, we illustrate the convergence of typical trajectories and the correspondent convergence of Z for the complete-graph architecture at $n = 60$.

We prove that a similar situation occurs even if the gates are arranged in a 1D fashion, and we fully expect that this situation applies for nearly all natural[3] architectures, including circuits on D-dimensional lattices for $D > 1$. We formalize this in Conjecture 4.1. We believe Conjecture 4.1 firstly because anti-concentration should intuitively only be faster when the circuit becomes more connected, and the 1D architecture is perhaps the least connected a natural architecture can be, as it takes $\Omega(n^2)$ gates for information to travel across the diameter of the qudit array. Secondly, the above intuitive argument about the convergence of typical trajectories to a fixed point in $O(n \log(n))$ gates should apply to any natural architecture. Specifically, if you choose a configuration with x S assignments at random, and you apply a layer of $\Theta(n)$ two-qudit gates, with high probability you will have formed $\Theta(x)$ disagreeing

[3]One can construct contrived architectures that do not quickly anti-concentrate by partitioning the qudits into many subsets and only rarely choosing a gate that couples qudits from different subsets. We define a property we call *regularly connected* to rule out this kind of situation. We prove that it implies anti-concentration in $O(n^2)$ gates and conjecture this can be improved to $O(n \log(n))$.

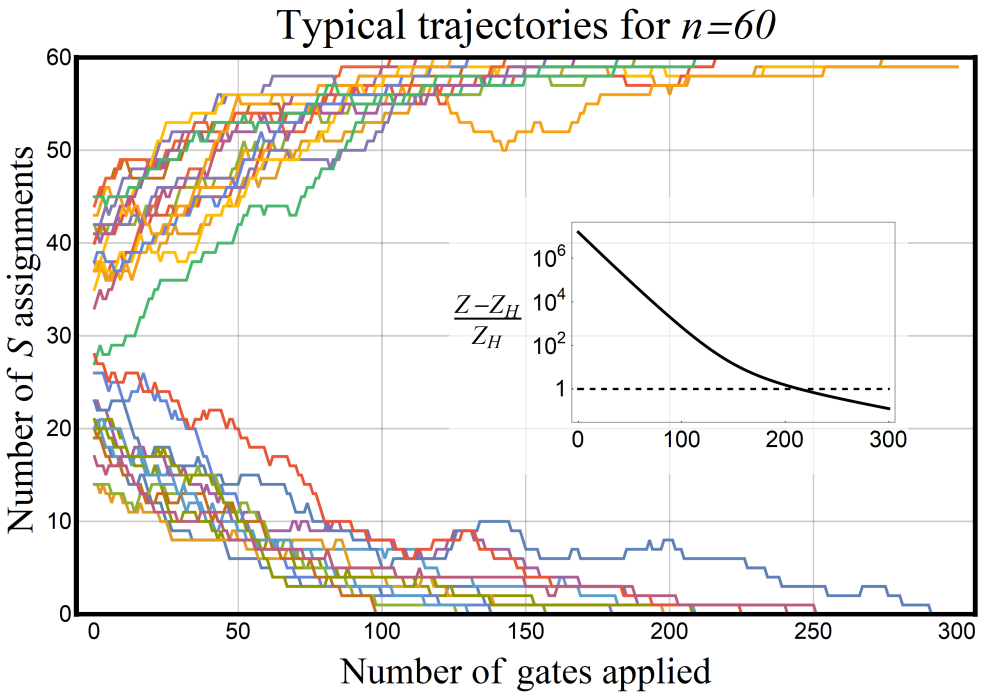


Figure 4.3: Thirty trajectories generated randomly for the complete-graph architecture at $n = 60$. A trajectory γ is chosen with probability proportional to weight(γ) in the $s \to \infty$ limit, and then the number of S assignments (out of 60) are plotted for the first 300 time steps. The trajectories rapidly approach either the fixed point I^n with 0 S assignments, or the fixed point S^n with 60 S assignments, but not all have reached the fixed point within 300 time steps. The distance of a typical trajectory from the nearest fixed point decays exponentially with time, with characteristic time scale $\Theta(n)$. Thus, it takes $\Theta(n \log(n))$ gates for most typical trajectories to have reached the fixed point. Inset: As trajectories approach the fixed points, the collision probability Z (which can be efficiently numerically calculated for the complete-graph architecture) approaches Z_H. Anti-concentration is defined as the point where it falls beneath $2Z_H$ (dashed line), which occurs at $s = 214$ for $n = 60$.

pairs and moved the trajectory a constant fraction of the way to the nearest fixed point. The difficulty in proving Conjecture 4.1 lies in characterizing what happens in the low-probability event that this is not the case.

Indeed, our rigorous proofs for the 1D and complete-graph architectures have to deal with the fact that it is not sufficient to examine only typical trajectory behavior. In particular, the collective contribution of trajectories at the tails of what is allowed are tricky to bound. These rigorous bounds are provided in the Appendices. Nonetheless, heuristic reasoning about typical trajectory behavior ultimately gives accurate predictions about the collision probability in these cases.

4.6 Outlook

In a quantum computer, quantum information is ultimately accessed by making measurements of the output state and obtaining samples from the associated output distribution over measurement outcomes. In many applications, it is desirable to choose our quantum computation completely at random, the only constraint being the arrangement of the different gates, and thus it is important to characterize the output distribution over measurement outcomes in random quantum circuits, and how it depends on the underlying circuit architecture.

One feature of the output distribution is that, for very shallow circuits, there are a relatively small number of very "heavy" measurement outcomes that are exponentially more likely than average to be obtained, a fact that inhibits the design of certain classical simulation algorithms, but also in other cases prevents potential proofs that no good simulation algorithms exist. As the circuit gets deeper, the probability mass gradually anti-concentrates and eventually becomes fairly well spread out over all possible measurement outcomes. We have developed a framework to quantitatively understand this situation; we map the anti-concentration process to the equilibration of a simple stochastic process (an alternative interpretation of the stochastic process is the partition function of a statistical mechanical model). The stochastic process allows for effective qualitative reasoning, but also produces sharp quantitative anti-concentration upper and lower bounds.

Both sides of our bounds have meaningful and surprising takeaways. On the one hand, the fact that only $O(n \log(n))$ gates are needed to achieve anti-concentration in geometrically local and non-local architectures contradicts the intuition that anti-concentration should not occur until information has had time to spread across the entire system. In fact, up to a constant factor, the anti-concentration time does not appear to be sensitive to exact connectivity structure of the circuit. While we only rigorously consider two architectures, our work gives strong evidence that any natural architecture anti-concentrates in $O(n \log(n))$ gates (which typically corresponds to $O(\log(n))$ depth). In cases where anti-concentration is a desirable property, our work gives explicit bounds on how many gates are needed, and the fact that this number is relatively small will come as welcome news in practical situations where the gates are noisy or otherwise costly to implement.

On the other hand, by showing that $\Omega(n \log(n))$ gates are necessary for anti-concentration (and computing the optimal constant prefactor in our two specific scenarios), we have cleared up some confusion about very shallow circuits. Increasing the depth causes the anti-concentration process to begin, but our lower bound implies that the phenomenon of very heavy measurement outcomes will remain for any architecture of constant depth. Even the 2D circuits of depth $O(\sqrt{\log(n)})$ (for which the lightcone volume is $O(\log(n))$)

considered in Ref. [29] cannot be anti-concentrated, as had been speculated in that work.

We conclude with some other specific open problems inspired by our work.

- We have proved that the anti-concentration size is $\Theta(n\log(n))$ for the 1D and complete-graph architectures. We believe this is true for most other natural architectures and formally conjecture in Conjecture 4.1 that this follows from our "regularly connected" definition.

- A sharp anti-concentration analysis for 2D and higher dimensional geometrically local architectures would be particularly valuable since, unlike in 1D, $\Theta(\log(n))$-depth 2D circuits can perform universal quantum computation (indeed, $\Omega(1)$-depth is sufficient [53]), and 2D circuits form the basis for Google's 2019 quantum computational supremacy experiment [6].

- We suspect the constant prefactor of $(2\log(q^2+1))^{-1}$ in the general lower bound in Theorem 4.4 could be improved. What is its optimal value? That is, can we show an improved general lower bound and then find an RQC architecture $\mathcal{A}_{\text{fast}}$ that has a matching upper bound. This would show that $\mathcal{A}_{\text{fast}}$ is the fastest anti-concentrator. A candidate for $\mathcal{A}_{\text{fast}}$ is the architecture where each layer of $n/2$ gates is formed by choosing a random partition of the n qudits into $n/2$ pairs.

- Are there other problems involving second moment calculations over RQCs where our techniques would produce sharp upper and lower bounds? One such problem could be the 2-design time for RQCs in various architectures.

APPENDIX TO CHAPTER 4

4.A Formal definitions

4.A.1 Random quantum circuits (RQCs)

Here we establish some precise definitions for the terms in this paper. Throughout, we consider systems of n qudits of local Hilbert space dimension q, with basis states $\{|0\rangle, |1\rangle, \ldots, |q-1\rangle\}$. Loosely speaking, a quantum circuit is a sequence of unitary transformations called gates, which each typically involve only a few of the n qudits, acting on the initial state $|0\rangle^{\otimes n} \equiv |0^n\rangle$. Formally, we let a *quantum circuit diagram* of circuit size s be specified by a length-s sequence $(A^{(1)}, \ldots, A^{(s)})$ of non-empty subsets of $[n] = \{0, 1, 2, \ldots, n-1\}$, indicating for each gate which qudits participate in that gate. Since we consider circuits consisting only of two-qudit gates, we require $|A^{(t)}| = 2$ for all t. We also make the assumption that the circuit begins with a single-qudit gate on each of the n qudits at the beginning of the circuit, without counting these n gates toward the circuit size. This sequence can be turned into a diagram as in Figure 4.2 (ignoring the overlaid I and S), where the gate sequence is $(\{0,1\}, \{1,2\}, \{0,1\}, \{2,3\}, \{1,2\})$. Note that the single-qudit unitaries are each displayed with the symbol U but will not necessarily be the same unitary. The circuit *depth* d of a circuit diagram is the minimum number of layers of non-overlapping gates needed to implement all s gates in the circuit, or formally, the smallest integer such that there exists a sequence $0 = s_0 < s_1 < s_2 < \ldots < s_d = s$ where $A^{(t)} \cap A^{(t')} = \emptyset$ whenever $s_j < t < t' \leq s_{j+1}$.

Once a circuit diagram has been chosen, a *quantum circuit instance* is generated by additionally specifying a length-$s+n$ sequence of unitary matrices $(U^{(-n+1)}, \ldots, U^{(-1)}, U^{(0)}, U^{(1)}, \ldots, U^{(s)})$ where $U^{(-j)}$ is a $q \times q$ (single-qudit) matrix for each $j = 0, \ldots, n-1$ and $U^{(t)}$ is a $q^2 \times q^2$ (two-qudit) matrix for each $t = 1, \ldots, s$. We denote the global $q^n \times q^n$ unitary operator implemented by the circuit by U, where

$$U = U^{(s)}_{A^{(s)}} U^{(s-1)}_{A^{(s-1)}} \ldots U^{(2)}_{A^{(2)}} U^{(1)}_{A^{(1)}} U^{(0)}_{\{0\}} \ldots U^{(-n+1)}_{\{n-1\}}, \tag{4.27}$$

with V_X indicating the action of the $q^{|X|} \times q^{|X|}$ unitary V on the qudits in subregion $X \subset [n]$ tensored with the identity operation on the qudits in the complement of X.

In this work, we will always assume that projective computational basis measurements are performed on all n qudits at the end of the circuit. Thus, a quantum circuit instance U has a corresponding classical probability distribution p_U over possible measurement outcomes $x \in [q]^n$, as follows:

$$p_U(x) = |\langle x|U|0^n\rangle|^2 . \tag{4.28}$$

Random quantum circuits will refer to situations when, once a circuit diagram has been fixed, the actual unitary gates $U^{(t)}$ that determine the circuit

instance are each randomly chosen independently from some distribution over the unitary group. In this paper, we always take this distribution to be the Haar measure, but since our techniques rely on calculating expectations over quantities with only two copies of each $U^{(t)}$, our results also apply when the gates are drawn from any 2-design, such as the Clifford group. Note that Google's quantum computational supremacy experiment [6] drew gates from another distribution that is not a 2-design. Heuristically speaking, as long as the distribution lacks any bias or symmetries, we expect properties like anti-concentration to be the same as in the Haar-random case.

4.A.2 Random quantum circuit architectures

An *architecture* for random quantum circuits is simply a procedure for choosing a circuit diagram. Formally, we define it to be a (possibly randomized) classical algorithm that, given parameters n and s, computes a circuit diagram of size s on n qudits. Given an architecture and parameters n and s, we let the expectation of some quantity Q, denoted $\mathbb{E}_U[Q]$, refer to the expectation over the process of first choosing a circuit diagram according to the architecture, and then choosing a circuit instance by randomly generating each gate in the circuit diagram independently from the Haar measure. Next, we define the two architectures we consider.

Definition 4.1 (Complete-graph architecture)**.** *Circuit diagrams of size s on n qudits are generated by choosing s gates each uniformly at random from the set of all two-qudit gates, i.e. $A^{(t)}$ is chosen uniformly from $\{\{a, b\} : a, b \in [n], a \neq b\}$.*

Note that if it could be guaranteed that every qudit would eventually participate in at least one gate, the distribution over circuit instances would be equivalent if we omitted the first layer of n single-qudit gates (defined to be part of every architecture), a fact that follows from the invariance of the Haar measure; the single-qudit gates could be absorbed into the two-qudit Haar-random gates that act directly before or after without changing the distribution over the two-qudit gates. However, in the complete-graph architecture there is a chance that a qudit does not participate in any two-qudit gates, although for sufficiently large circuit size the probability of this vanishes.

Definition 4.2 (1D architecture)**.** *Assume that n is even and $d = 2s/n$ is an integer. The circuit diagram of size s on n qudits is generated by alternating between the two types of layers of $n/2$ non-overlapping nearest-neighbor two-qudit gates on a ring. That is, for each $t = 1, \ldots, n/2$, if j is even, then $A^{(t+jn/2)} = \{2t-2, 2t-1\}$, and if j is odd, then $A^{(t+jn/2)} = \{2t-1, 2t\}$, where index n is identified with index 0 to enforce periodic boundary conditions.*

4.A.3 Collision probability and anti-concentration

Anti-concentration is a concept that describes a classical probability distribution for which the probability mass is not too concentrated onto a small

number of outcomes of the random variable. The uniform distribution is the ultimate anti-concentrated distribution, as the probability mass is allocated evenly over every possible outcome, but we would still like the term anti-concentrated to apply to some non-uniform distributions if the probability mass is fairly well spread over many of the outcomes. There are multiple ways to make this quantitative. For the purposes of this paper, we choose one way—the collision probability—that mirrors previous work on anti-concentration of quantum circuit outputs and suffices for the applications we discuss in the introduction.

Let X be a discrete random variable and M be the set of outcomes of X. We can form another random variable p, where p is equal to $\Pr[X = x]$ for an x chosen uniformly at random from M. Since $\sum_x \Pr[X = x] = 1$, we have $\mathbb{E}[p] = 1/|M|$ no matter how X is distributed. We define the collision probability for X to be

$$Z = \sum_{x \in M} \Pr[X = x]^2 \tag{4.29}$$

$$= \mathbb{E}[p^2] \cdot |M| \tag{4.30}$$

$$= \delta(p)\,|M| + |M|^{-1}\,, \tag{4.31}$$

which is the probability that two identical independent copies of X will be equal to each other—hence *collision* probability. If the distribution over X is the uniform distribution, then the distribution over p is the point distribution on the value $|M|^{-1}$, the collision probability takes its minimal value $Z = |M|^{-1}$, and $\delta(p) = 0$. If X is non-uniform but still somewhat anti-concentrated, then p will not always be $|M|^{-1}$ but it will usually be close, and this will be reflected by a collision probability that is greater, but not too much greater than $|M|^{-1}$. Formally, we make the following definition.

Definition 4.3 (Anti-concentrated)**.** *We say that a random variable X over a set M of outcomes is α-anti-concentrated for $0 < \alpha \leq 1$ if*

$$Z = \sum_{x \in M} \Pr[X = x]^2 \leq \frac{1}{|M|\alpha}\,. \tag{4.32}$$

Thus a distribution is 1-anti-concentrated if and only if it is the uniform distribution.

In our setting, the random variable X is the measurement outcome of a random quantum circuit instance, which is distributed according to the distribution p_U over the outcome set $[q]^n$. Example distributions of p_U for RQC outputs in the uniform, the non-uniform but still anti-concentrated, and the not anti-concentrated case are shown in the caricature in Figure 4.1. A random quantum circuit architecture for specified n and s is understood as an ensemble over many different U, only some of which will have output distributions p_U that are α-anti-concentrated for a certain choice of α. We would like

to say that the architecture as a whole is anti-concentrated if typical circuit instances drawn from the architecture are anti-concentrated, acknowledging that not every instance will be. We also require this to hold for the same constant α as n increases, with s increasing like some function $s(n)$. Formally, we accomplish this by averaging the collision probability over the random circuit instance, as follows.

Definition 4.4 (Anti-concentrated RQC architecture)**.** *We say that a random quantum circuit architecture is α-anti-concentrated for $0 < \alpha \leq 1$ at circuit size $s(n)$ if there exists n_0 such that whenever $n \geq n_0$*

$$Z = \mathop{\mathbb{E}}_{U} \left[\sum_{x \in [q]^n} p_U(x)^2 \right] \leq (\alpha q^n)^{-1} , \tag{4.33}$$

where $\mathbb{E}_U$ denotes drawing circuit instances according to the architecture over n qudits with circuit size $s(n)$. Generally, we say that the architecture is anti-concentrated at size $s(n)$ if there exists a constant $\alpha > 0$ independent of n for which it is α-anti-concentrated at that size.

RQC architectures for which every qudit experiences at least one gate, which includes all the architectures introduced above, will have a symmetry over the q^n measurement outcomes in the sense that the quantity $p_U(x)$ is distributed identically (over circuit instances) for every x. In this case each term in the sum in Eq. (4.33) will have the same contribution and we can write simply

$$Z = q^n \mathop{\mathbb{E}}_{U} \left[p_U(0^n)^2 \right] . \tag{4.34}$$

The anti-concentration of an architecture implies that most of the instances drawn from that architecture have good anti-concentration properties: Given an architecture at a certain size and a bound on its collision probability $Z \leq \alpha^{-1} q^{-n}$, we can use Markov's inequality to assert that at least a $1 - \beta$ fraction of instances have collision probability at most $q^{-n}\left(1 + (\alpha^{-1} - 1)\beta^{-1}\right)$. In practice, we expect the collision probability of individual instances to be even more clustered near the mean collision probability than this analysis indicates, but proving that this is the case would seem to require computing higher moments like $\mathbb{E}_U[p_U(0^n)^k]$ for $k > 2$.

As discussed in the main text, an important implication of an α-anti-concentrated architecture is that for any β with $0 \leq \beta \leq 1$ and sufficiently large n

$$\Pr_{U}[p_U(x) \geq \beta q^{-n}] \geq (1 - \beta)^2 \alpha , \tag{4.35}$$

which follows directly from the Paley-Zygmund inequality. This inequality indicates that whenever an architecture is anti-concentrated, at least a constant fraction of the outcomes will be allocated an amount of mass that is within

a constant factor β of the mean mass; it cannot be the case that all but a vanishing fraction of the outcomes are allocated a vanishing fraction of the mean mass.

4.B Framework for analysis: Random quantum circuits as a stochastic process

This section gives more details on the correspondence discussed in Section 4.5 from the main text. The key idea in our analysis of the collision probability of RQCs is to perform the Haar expectation over each local unitary individually. This is possible due to explicit formulas for expectations under action by a Haar-random unitary. We use these formulas to re-express the collision probability, originally an integral over many continuously varying unitary matrices drawn from the Haar measure, as a weighted discrete sum, which is then analyzed using combinatorial and stochastic methods. This weighted sum can also be interpreted as the partition function of a classical statistical mechanical Ising-like model, as described in Chapter 2, or as the expectation value of a simple stochastic process. Figure 4.4 depicts these equivalent representations of the problem. In this section, we explain this method and derive the important formulas that will apply generally for any RQC architecture, which are then used in later sections to prove our main results.

4.B.1 Averaging individual unitaries over the Haar measure

The quantity of interest for anti-concentration is the expected collision probability, which is proportional to a second moment over choice of unitary operator U, as illustrated in the following equation, where we recall that $|0^n\rangle\langle 0^n|^{\otimes 2}$ is two copies of the circuit input state

$$Z = q^n \mathop{\mathbb{E}}_U \left[\left(\langle 0^n | U | 0^n \rangle \langle 0^n | U^\dagger | 0^n \rangle \right)^2 \right] \tag{4.36}$$

$$= q^n \operatorname{tr} \left[\mathop{\mathbb{E}}_U \left[U^{\otimes 2} |0^n\rangle\langle 0^n|^{\otimes 2} U^{\dagger \otimes 2} \right] |0^n\rangle\langle 0^n|^{\otimes 2} \right] . \tag{4.37}$$

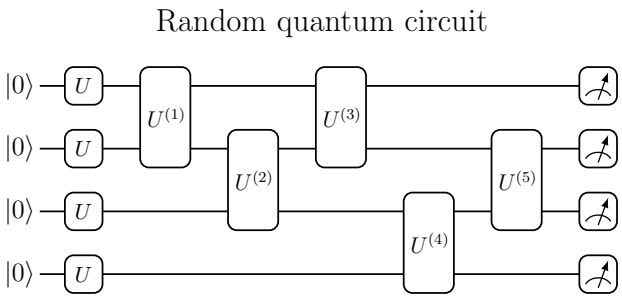
Moreover, for a fixed quantum circuit diagram, the unitary U is given by Eq. (4.27) as a product of single-qudit unitaries $U^{(-j)}$ acting on qudit j for $j = 0, \ldots, n-1$ and two-qudit unitaries $U^{(t)}$ acting on some pair of qudits $A^{(t)} \subset [n]$ for $t = 1, \ldots, s$. Each unitary is independently chosen according to the Haar measure, and its expectation can be evaluated separately. Let

$$M^{(t)}[\rho] = \mathop{\mathbb{E}}_{U^{(t)}} \left[U^{(t)}_{A^{(t)}}{}^{\otimes 2} \rho \, U^{(t)}_{A^{(t)}}{}^{\dagger \otimes 2} \right] . \tag{4.38}$$

Then we can write

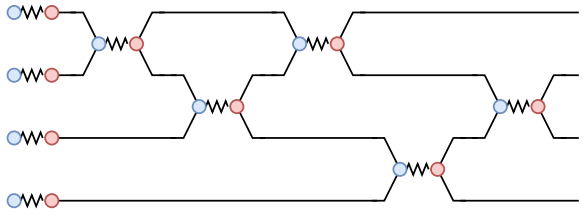
$$\mathop{\mathbb{E}}_U \left[U^{\otimes 2} |0^n\rangle\langle 0^n|^{\otimes 2} U^{\dagger \otimes 2} \right] \tag{4.39}$$

$$= M^{(s)} \circ M^{(s-1)} \circ \cdots \circ M^{(1)} \circ M^{(0)} \circ \ldots \circ M^{(-n+1)} \left[|0^n\rangle\langle 0^n|^{\otimes 2} \right] . \tag{4.40}$$



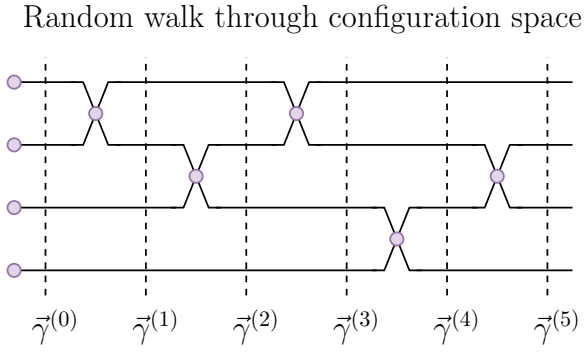
Collision probability

$$Z = q^n \mathop{\mathbb{E}}_{U} \left[p_U(0^n)^2 \right]$$



Partition function

$$Z = q^n \sum_{\sigma,\tau} \prod_{\langle tt' \rangle} \text{weight}_{\langle tt' \rangle}(\sigma, \tau)$$



Sum over trajectories

$$Z = \frac{1}{(q+1)^n} \sum_{\gamma} \prod_{t=1}^{s} M_{\vec{\gamma}^{(t)}\vec{\gamma}^{(t-1)}}$$

Figure 4.4: Diagram depicting the equivalent ways to interpret the expected value of the collision probability for random quantum circuits. Top: a random quantum circuit of size five. Middle: the reinterpretation as the partition function of a classical statistical mechanics model with local Ising-like particles as in Chapter 2. Bottom: another interpretation as a stochastic process of evolving configurations.

When an architecture is itself a mixture over randomly chosen circuit diagrams, such as the complete-graph architecture, the overall quantity in Eq. (4.39) is a mixture over terms of the form in Eq. (4.40).

The remainder of this subsection illustrates how the action of $M^{(t)}$ can be evaluated, ultimately allowing us to arrive at the expression for Z given in Eq. (4.59). In the other subsections of this section, we explain how that equation can be interpreted as a partition function of a classical statistical mechanical model or as the expectation over simple stochastic process.

When the local unitaries are drawn from the Haar measure (or any exact 2-design), the expression $M^{(t)}[\rho]$ can be evaluated in a simple way. Generally, for σ a $q^2 \times q^2$ Hermitian operator, and with V chosen from the Haar measure over the set of $q \times q$ unitaries, we define

$$M[\sigma] = \mathop{\mathbb{E}}_{V} \left[V^{\otimes 2} \sigma V^{\dagger \otimes 2} \right] \tag{4.41}$$

and observe that, for any unitary W and any σ

$$M[\sigma] W^{\otimes 2} = \mathop{\mathbb{E}}_{V} \left[V^{\otimes 2} \sigma (W^\dagger V)^{\dagger \otimes 2} \right] = \mathop{\mathbb{E}}_{V} \left[(WV)^{\otimes 2} \sigma V^{\dagger \otimes 2} \right] = W^{\otimes 2} M[\sigma] \,, \tag{4.42}$$

where the second equality follows from the invariance of the Haar measure under the substitution $V \to WV$. A mathematical fact from Schur-Weyl duality (see [144, 145]) is that any operator on k copies of a system that commutes with $W^{\otimes k}$ for any unitary W must be a linear combination of permutation operators over the k systems. Here we have $k = 2$ and thus the only permutation operators are the identity operation I and the swap operation S, which can be defined as the operator satisfying $S|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle$ for any $|\psi\rangle, |\phi\rangle$. Letting $M[\sigma] = \alpha I + \beta S$, we make the following calculations

$$\operatorname{tr}[M[\sigma]] = \operatorname{tr}[\sigma] = \alpha q^2 + \beta q \tag{4.43}$$

$$\operatorname{tr}[M[\sigma] S] = \operatorname{tr}[\sigma S] = \alpha q + \beta q^2 \,, \tag{4.44}$$

which determine α and β and allow us to write

$$M[\sigma] = \frac{\operatorname{tr}(\sigma) - q^{-1} \operatorname{tr}(\sigma S)}{q^2 - 1} I + \frac{\operatorname{tr}(\sigma S) - q^{-1} \operatorname{tr}(\sigma)}{q^2 - 1} S \,. \tag{4.45}$$

The unitaries $U^{(-j)}$ are $q \times q$ (single-qudit) that act on qudit j. Two copies of the input state on qudit j is $|0\rangle\langle 0|^{\otimes 2}_{\{j\}}$. Denote two copies of the input state on the other $n-1$ qudits by $\rho_{[n]\setminus\{j\}}$. Using Eq. (4.45), we then find

$$M^{(-j)} \left[\rho_{[n]\setminus j} \otimes |0\rangle\langle 0|^{\otimes 2}_{\{j\}} \right] = \rho_{[n]\setminus\{j\}} \otimes \frac{1}{q(q+1)} \left(I + S \right)_{\{j\}} \,, \tag{4.46}$$

meaning that $M^{(-j)}$ simply replaces the state on qudit j as a uniform sum over operators I and S. Hence

$$M^{(0)} \circ M^{(-1)} \circ \cdots \circ M^{(-n+1)} \left[|0^n\rangle\langle 0^n|^{\otimes 2} \right] = \bigotimes_{j=0}^{n-1} \left(\frac{1}{q(q+1)} (I+S)_{\{j\}} \right) \tag{4.47}$$

$$= \frac{1}{q^n(q+1)^n} \sum_{\vec{\gamma} \in \{I,S\}^n} \bigotimes_{j=0}^{n-1} \gamma_j \,. \tag{4.48}$$

We call each $\vec{\gamma} \in \{I, S\}^n$ a *configuration*. The above equation states that the expected value of two copies of the state after application of all the single-qudit unitaries is precisely a uniform sum over all identity/swap configurations of the n sites.

Now, we need to examine the action of $M^{(t)}$ for $t > 0$. In this case, the unitaries are $q^2 \times q^2$ and act on the qudit pair $A^{(t)}$. We can use Eq. (4.45) by replacing $q \to q^2$ and sending $I \to I \otimes I$, the identity operation on two copies of two qudits, and $S \to S \otimes S$, the swap operation on two copies of two qudits. We assume that the input state is a product state $\rho_{[n]\setminus A^{(t)}} \otimes \rho_{A^{(t)}}$ and see that

$$M^{(t)}[\rho_{[n]\setminus A^{(t)}} \otimes \rho_{A^{(t)}}] = \rho_{[n]\setminus A^{(t)}} \otimes (g_I (I \otimes I)_{A^{(t)}} + g_S (S \otimes S)_{A^{(t)}}) \,, \tag{4.49}$$

where

$$g_I = \frac{\mathrm{tr}(\rho_{A^{(t)}}) - q^{-2} \,\mathrm{tr}\left(\rho_{A^{(t)}} (S \otimes S)\right)}{q^4 - 1} \tag{4.50}$$

$$g_S = \frac{\mathrm{tr}\left(\rho_{A^{(t)}} (S \otimes S)\right) - q^{-2} \,\mathrm{tr}(\rho_{A^{(t)}})}{q^4 - 1} \,. \tag{4.51}$$

Since the two qudit gates act after the single-qudit gates, the input state to $M^{(t)}$ will always be a sum of tensor products of I and S, so we only need to evaluate the above expression when $\rho_{A^{(t)}}$ is either $I \otimes I$, $I \otimes S$, $S \otimes I$, or $S \otimes S$. Doing so, we arrive at

$$M^{(t)} \left[\rho_{[n]\setminus A^{(t)}} \otimes (I \otimes I)_{A^{(t)}} \right] = \rho_{[n]\setminus A^{(t)}} \otimes (I \otimes I)_{A^{(t)}} \tag{4.52}$$

$$M^{(t)} \left[\rho_{[n]\setminus A^{(t)}} \otimes (S \otimes S)_{A^{(t)}} \right] = \rho_{[n]\setminus A^{(t)}} \otimes (S \otimes S)_{A^{(t)}} \tag{4.53}$$

$$M^{(t)} \left[\rho_{[n]\setminus A^{(t)}} \otimes (I \otimes S)_{A^{(t)}} \right] = M^{(t)} \left[\rho_{[n]\setminus A^{(t)}} \otimes (S \otimes I)_{A^{(t)}} \right] \tag{4.54}$$

$$= \rho_{[n]\setminus A^{(t)}} \otimes \left(\frac{q}{q^2+1} (I \otimes I)_{A^{(t)}} + \frac{q}{q^2+1} (S \otimes S)_{A^{(t)}} \right) . \tag{4.55}$$

Thus, if ρ is a linear combination of configurations in $\{I, S\}^n$, $M^{(t)}[\rho]$ will also be a linear combination of configurations, with coefficients that transform

linearly under application of $M^{(t)}$. For configurations $\vec{\gamma}, \vec{\nu} \in \{I, S\}^n$, we let $M^{(t)}_{\vec{\nu}\vec{\gamma}}$ be the matrix element of this linear transformation defined such that

$$M^{(t)} \left[\bigotimes_{j=0}^{n-1} \gamma_j \right] = \sum_{\vec{\nu} \in \{I,S\}^n} M^{(t)}_{\vec{\nu}\vec{\gamma}} \bigotimes_{j=0}^{n-1} \nu_j \,. \tag{4.56}$$

Suppose $U^{(t)}$ acts on qudits $A^{(t)} = \{a, b\} \subset [n]$. Then from Eqs. (4.52), (4.53), (4.55), we have

$$M^{(t)}_{\vec{\nu}\vec{\gamma}} = \begin{cases} 1 & \text{if } \gamma_a = \gamma_b \text{ and } \vec{\gamma} = \vec{\nu} \\ \frac{q}{q^2+1} & \text{if } \gamma_a \neq \gamma_b \text{ and } \nu_a = \nu_b \text{ and } \gamma_c = \nu_c \; \forall c \in [n] \setminus \{a, b\} \\ 0 & \text{otherwise} \end{cases} \tag{4.57}$$

Importantly, $M^{(t)}_{\vec{\nu}\vec{\gamma}}$ is always non-negative. The way to think about the above equation is to notice three things. First, the input configuration $\vec{\gamma}$ and the output configuration $\vec{\nu}$ must agree on all indices that are not involved in the gate, i.e. for all indices $c \notin \{a, b\}$; otherwise the matrix element is 0. Second, if the two input values involved in the gate agree, i.e. if $\gamma_a = \gamma_b$ then $\nu_a = \nu_b = \gamma_a = \gamma_b$ must hold (in which case the matrix element is 1); otherwise it is 0. Third, if the two input values disagree, then one of them must be flipped so that the two output values agree (in which case the matrix element is reduced to $q/(q^2 + 1)$); otherwise it is 0.

Note also that

$$\operatorname{tr}\left[\left(\bigotimes_{j=0}^{n-1} \gamma_j \right) |0^n\rangle\langle 0^n|^{\otimes 2} \right] = 1 \tag{4.58}$$

for all $\vec{\gamma} \in \{I, S\}^n$. Thus from Eq. (4.37), we find

$$Z = \frac{1}{(q+1)^n} \sum_{\gamma \in \{I,S\}^{n \times (s+1)}} \prod_{t=1}^{s} M^{(t)}_{\vec{\gamma}^{(t)}\vec{\gamma}^{(t-1)}} \tag{4.59}$$

$$=: \frac{1}{(q+1)^n} \sum_{\gamma} \operatorname{weight}(\gamma) \,. \tag{4.60}$$

which was the expression quoted in Eq. (4.23) from the main text. In the above equation, the sum is over length-$(s + 1)$ sequences of configurations, which we call a *trajectory* $\gamma = (\vec{\gamma}^{(0)}, \ldots, \vec{\gamma}^{(s)})$ and the weight of each term is given by the product of the matrix elements for each step in the trajectory. This final equation is depicted graphically in the right part of Figure 4.4.

4.B.2 Collision probability as statistical mechanical partition function

The expression for the collision probability in Eq. (4.59) can be interpreted as a partition function for a classical statistical mechanical model by thinking

of each $\gamma_j^{(t)}$ as an Ising spin variable with the association $\{I, S\} \leftrightarrow \{+1, -1\}$. A trajectory γ is then a configuration of the Ising spins, and Z is a weighted sum over all the spin configurations. Moreover, the weight is always non-negative and is given by a product of factors $M^{(t)}_{\vec{\gamma}^{(t)}\vec{\gamma}^{(t-1)}}$ that can be determined by examining a small number of the spin values. This means that the energy functional over spin configurations of the classical Ising-like model is always real and can be broken up into local terms that depend on the local dimension q and which qudits are acted upon at each step in the circuit.

The statistical mechanics interpretation has been a useful one for similar problems in the past, where certain RQC moment quantities can be exactly rewritten as the partition sum over spin configurations of a lattice model, as depicted in the central diagram in Figure 4.4. We can arrive at the formulation as in Eq. (4.59) from the lattice model by summing over a subset of the spins and reinterpreting the resulting nodes as 4-body interaction vertices.

This exact rewriting of RQC moment quantities has been used to compute, for instance, correlation functions [41], Rényi entropies [36], and the distance to forming an approximate design [37]. Moreover, thermal phase transitions in the classical model can be related to phase transitions of entanglement-entropy-like quantities for the output state of the RQC [44, 45, 47]. The interpretation is particularly intriguing when considering analogous quantities to Z for higher moments. The collision probability is a second moment quantity, and the resulting stat mech model has Ising-like variables with two possible values. Quantities related to the kth moment will map to classical stat mech models that have $k!$ possible values, one for each element of the symmetric group $\mathcal{S}_k$. However, one challenge of computing higher-moment quantities is that the weights in the partition function can be negative (corresponding to non-real values of the energy for certain spin configurations), complicating many strategies for bounding its behavior, including the strategies employed in the rest of this paper.

4.B.3 Unbiased random walk

We can build from the formula for Z in Eq. (4.59) and re-express it in terms of a length-s unbiased random walk through configuration space $\{I, S\}^n$, which we denote P_u. At time step 0, a configuration $\vec{\gamma}^{(0)}$ is chosen uniformly at random, i.e. the initial distribution is the uniform distribution in configuration space, denoted Λ_u. Then configuration $\vec{\gamma}^{(t+1)}$ at time step $t+1$ is generated from the configuration $\vec{\gamma}^{(t)}$ at time step t as follows: letting $A^{(t)} = \{a, b\}$, if the ath and bth bits of $\vec{\gamma}^{(t)}$ agree, then the configuration is left unchanged at time step $t+1$; if they disagree, either the value at a or the value at b is flipped each with probability 1/2 to form $\vec{\gamma}^{(t+1)}$. The weight is reduced each time a bit is flipped. Thus we can write

$$Z = \frac{2^n}{(q+1)^n} \mathop{\mathbb{E}}_{P_u, \Lambda_u} \left[\left(\frac{2q}{q^2+1} \right)^{(\#\text{ of bit flips during walk})} \right], \tag{4.61}$$

where $\mathbb{E}_{P_u,\Lambda_u}$ indicates the expectation over the choosing a length-s walk as described above, where the initial distribution is Λ_u. This is seen to be equivalent to Eq. (4.59) since the probability of a certain trajectory occurring is given by $q^{-n}(1/2)^{\#\text{ of bit flips}}$ and thus each trajectory contributes exactly the same amount toward Z, once the probability of observing the trajectory is accounted for.

4.B.4 Biased random walk

A potential problem with the unbiased random walk picture is that the weight of a particular walk is related to the number of bit flips that occur during that walk; it depends not only where the walk begins and ends but also on how it got there. To fix this issue, we can form an equivalent *biased* random walk denoted P_b. In this case, the initial distribution Λ_b is not uniform over $\{I,S\}^n$, rather, the probability of choosing $\vec{\gamma}^{(0)} = \vec{\nu}$ is proportional to $q^{-|\vec{\nu}|}$, where $|\vec{\nu}|$ is the Hamming weight of $\vec{\nu}$ (number of S entries). Specifically, we have

$$\Lambda_b(\vec{\nu}) = \frac{q^n}{(q+1)^n} q^{-|\vec{\nu}|}\,. \tag{4.62}$$

The dynamics of P_b are the same as P_u except that when the two bits involved in a gate disagree, it chooses to flip the S to I with probability $q^2/(q^2+1)$ and I to S with probability $1/(q^2+1)$. Thus, it is biased in the I direction. Then we can express

$$Z = \frac{1}{q^n} \mathop{\mathbb{E}}_{P_b,\Lambda_b} \left[q^{|\vec{\gamma}^{(s)}|} \right]\,. \tag{4.63}$$

Note that the quantity being averaged is exponentially large in the Hamming weight of its final ending point, making the quantity sensitive to the probability that the biased walk stays far from the all I configuration. The biased walk is observed to be equivalent to the unbiased walk simply by noting that, once the probability of observing a certain trajectory is included, every trajectory contributes the same amount to Z for both walks. The exponential weighting underneath the expectation in the biased walk exactly cancels the bias in the probability of observing a certain walk.

4.B.5 Computing sums over trajectories

Throughout our analysis, we will need to compute weighted sums over various trajectories, or, relatedly, compute probabilities that the biased and unbiased walks end in a certain place. We use the following lemma. The key takeaway is that (perhaps surprisingly), in the limit of infinite size, the contribution of all trajectories originating from a certain initial configuration *depends only on the Hamming weight* of that initial configuration, and not the configuration itself. Moreover, this contribution can be calculated. This lemma is a more precise and generalized version of the recursive calculation of $Q(x)$ in Section 4.5 in the main text.

Lemma 4.1. *Fix an infinite-size circuit diagram, that is, an infinite sequence of qudit pairs $A = (A^{(1)}, A^{(2)}, \ldots)$. Also fix integers $0 \leq x, y, m \leq n$ such that $y \leq x < y + m$, as well as an initial configuration $\vec{\gamma}^{(0)}$ such that $|\vec{\gamma}^{(0)}| = x$. For each $s \geq 0$, let $\mathcal{T}_s$ be the set of length-s trajectories that*

(1) begin at configuration $\vec{\gamma}^{(0)}$

(2) have a non-zero contribution to Z for the circuit diagram $(A^{(1)}, \ldots, A^{(s)})$ formed by truncating A to length s

(3) end at any configuration $\vec{\gamma}^{(s)}$ for which $|\vec{\gamma}^{(s)}| = y$, and

(4) satisfy $y < |\vec{\gamma}^{(t)}| < y + m$ for all $t = 0, 1, 2, \ldots, s-1$.

Let $\mathcal{T} = \bigcup_{s=0}^{\infty} \mathcal{T}_s$. Then

$$\sum_{\gamma \in \mathcal{T}} \left(\frac{q}{q^2+1} \right)^{(\#\text{ of bit flips during } \gamma)} = \frac{1}{1-q^{-2m}} \left(q^{-(x-y)} - q^{-2m+x-y} \right) . \quad (4.64)$$

Proof. First, we claim that the sum should depend only on x, y, and m, and not on $\vec{\gamma}^{(0)}$ (other than through its dependence on x). To see this, note that there is a one-to-one correspondence between trajectories in $\mathcal{T}$ and sequences of Hamming weights $(x, x_1, \ldots, x_{s'-1}, y)$ with the property that either $x_t = x_{t+1} + 1$ or $x_t = x_{t+1} - 1$ for every t (no consecutive duplicates). This is seen by (1) the fact that given a trajectory in $\mathcal{T}$, one can generate such a sequence by taking the Hamming weight of each configuration in the sequence and removing consecutive duplicates and (2) the fact that given such a Hamming weight sequence one can generate a unique trajectory by starting with $\vec{\gamma}^{(0)}$, evolving the trajectory according to the circuit diagram A, and always choosing whether to flip I to S or S to I so that the order of Hamming weights prescribed by the sequence is followed. Thus, the sum over trajectories in $\mathcal{T}$ may be replaced by a sum over Hamming weight sequences, which does not depend on $\vec{\gamma}^{(0)}$, except through its Hamming weight x.

For each x in the interval $[y, y+m]$, let the expression on the left-hand-side of the lemma be given by $Q(x)$. Then for each x in $[y+1, y+m-1]$, we have the recursion relation

$$Q(x) = \frac{q}{q^2+1} (Q(x-1) + Q(x+1)) , \quad (4.65)$$

since the first bit flip will either send x to $x-1$ or to $x+1$ and in either case a factor of $q/(q^2+1)$ is incurred. The recursion relation gives rise to a general solution of the form

$$Q(x) = Fq^x + Gq^{-x} \quad (4.66)$$

for some constants F and G. This is a unique solution since all values can be generated once two consecutive values are specified, and the specification

of two consecutive values also uniquely specifies F and G. To find F and G in this case, we must also impose the boundary conditions $Q(y) = 1$ and $Q(y+m) = 0$, since if $x = y$ the only trajectory in $\mathcal{T}$ is the length-0 trajectory $(\vec{\gamma}^{(0)})$, and if $x = y + m$, $\mathcal{T}$ is the empty set. By specifying these boundary conditions we can solve for F and G and verify the statement of the lemma. □

Corollary 4.1. *Fix non-negative integers x, y, m such that $y \leq x < y + m$. For the biased walk, if the starting configuration has Hamming weight x, the probability that the walk reaches a configuration with Hamming weight y before it reaches a configuration with Hamming weight $y + m$ is given by*

$$\frac{q^{x-y}}{1 - q^{-2m}} \left(q^{-(x-y)} - q^{-2m+x-y} \right) . \tag{4.67}$$

Proof. The transition rules of the biased walk prescribe that transitions upward in Hamming weight occur with probability $1/(q^2 + 1)$, and transitions downward in Hamming weight occur with probability $q^2/(q^2 + 1)$. Thus the probability of a series of transitions in which the initial Hamming weight is x, the final Hamming weight is y, and the number of times a bit flip occurs is b is precisely $q^{x-y}(q/(q^2 + 1))^b$. The sum over all paths weighted by their probability is then precisely the sum in the left-hand-side of Lemma 4.1 scaled by q^{x-y}, yielding the corollary. □

Corollary 4.2. *If we begin at a trajectory $\vec{\gamma}^{(0)}$ with $|\vec{\gamma}^{(0)}| = x$ and allow the biased walk to evolve until it ends at one of the fixed points I^n or S^n, then the probability the trajectory ends at I^n is given by*

$$P_I(x) = \frac{1}{1 - q^{-2n}} \left(1 - q^{-2n+2x} \right) \tag{4.68}$$

and the probability it ends at S^n is given by

$$P_S(x) = \frac{q^{-2n+2x}}{1 - q^{-2n}} \left(1 - q^{-2x} \right) . \tag{4.69}$$

Proof. Termination at I^n corresponds to the cases where Hamming weight 0 is hit before Hamming weight n. Thus the equation for $P_I(x)$ follows from Corollary 4.1 with $y = 0$ and $m = n$. We have $P_S(x) = 1 - P_I(x)$ since the trajectory must terminate at one fixed point or the other. □

4.B.6 Sanity check: Infinite circuit size convergence to Haar value

The Markov chain has two stationary distributions, at configurations I^n and S^n. In the infinite circuit size limit, the biased walk will converge to a mixture of these two fixed-point configurations, where the amount of mass at each fixed-point depends only on the Hamming weight of the initial configuration, as described in Corollary 4.2. Using the expressions for P_I and P_S, we

find that, in the infinite circuit size limit,

$$Z = \frac{1}{q^n} \sum_{\vec{\gamma}^{(0)}} \Lambda_b(\vec{\gamma}^{(0)}) \mathop{\mathbb{E}}_{P_b, \vec{\gamma}^{(0)}} \left[q^{|\vec{\gamma}^{(s)}|} \right] \tag{4.70}$$

$$= \frac{1}{(q+1)^n} \sum_{\vec{\gamma}^{(0)}} q^{-|\vec{\gamma}^{(0)}|} (P_I(|\vec{\gamma}^{(0)}|) + q^n P_S(|\vec{\gamma}^{(0)}|)) \tag{4.71}$$

$$= \frac{1}{(q+1)^n (1-q^{-2n})} \sum_{x=0}^{n} \binom{n}{x} q^{-x} (1 - q^{-2n+2x} + q^{-n+2x} - q^{-n}) \tag{4.72}$$

$$= \frac{1}{(q+1)^n (1-q^{-2n})} \left(\frac{(q+1)^n}{q^n} - \frac{(q+1)^n}{q^{2n}} + \frac{(q+1)^n}{q^n} - \frac{(q+1)^n}{q^{2n}} \right) \tag{4.73}$$

$$= \frac{(2q^{-n} - 2q^{-2n})(q+1)^n}{(q+1)^n (1-q^{-2n})} \tag{4.74}$$

$$= \frac{2}{q^n + 1} = Z_H \,, \tag{4.75}$$

where Z_H is the Haar value. This outcome is expected since in the infinite circuit size limit the distribution over random unitaries formed from Haar-random local components will approach the distribution over n-qudit Haar random unitaries.

4.C Bounds for general architectures

4.C.1 Upper bound on collision probability

In order to have a meaningful upper bound, we need the architecture to satisfy basic connectivity requirements; for example, if the architecture performs gates on the same pair of qudits over and over again, Z will never decrease and the output distribution never become anti-concentrated. We need to rule out this sort of architecture.

Recall that an RQC architecture is a (possibly randomized) procedure for choosing a length-s sequence $(A^{(1)}, \ldots, A^{(s)})$ of pairs of qudit indices on which to perform a Haar-random gate.

Definition 4.5 (Regularly connected)**.** *We say an RQC architecture is h-regularly connected if for any n, any t, any subsequence $A = (A^{(1)}, \ldots, A^{(t)})$ and any proper subset $R \subset [n]$ of qudit indices, there is at least a $1/2$ probability that, conditioned on the first t gates in the gate sequence being A, there exists some index t' for which $t < t' \leq t + hn$, $A^{(t')} \cap R \neq \emptyset$, and $A^{(t')} \not\subset R$.*

The above definition requires that given any partition of the qudits into two sets, we should expect at least one gate to couple a qudit from one set with a qudit from the other set after only a linear number of gates. Note that both the 1D and the complete-graph architecture have this property. In 1D, it only takes two layers, or n gates, to guarantee having performed a gate

that crosses any partition one might choose. Similarly, in the complete-graph architecture, the probability that a randomly chosen gate crosses a partition is at least $1/n$ (which happens if the partition splits the indices into a set with one index and a set with the other $n-1$ indices), and the probability of having crossed the partition becomes large after $\Theta(n)$ gates. Most natural architectures we might consider have this property. One architecture that is not regularly connected is the hypercube architecture, where $n = 2^D$ qudits lie at the vertices of a D-dimensional hypercube, and D layers of gates are performed by cycling through each set of parallel edges. In this architecture, it would take $nD/2 = \Theta(n \log(n))$ gates to guarantee that any partition has been crossed.

Assuming the regularly connected property, we can show a weak upper bound on the collision probability.

Theorem 4.3 (restated)**.** *If an RQC architecture is h-regularly connected, then the collision probability satisfies*

$$Z \leq Z_H \left(1 + e^{-\frac{2a}{n}(s-s^*)}\right) , \tag{4.76}$$

where

$$a = (2h)^{-1} \log\left(\frac{2(q^2+1)}{(q+1)^2}\right) \tag{4.77}$$

$$s^* = (2a)^{-1} \log\left(\frac{2q}{q+1}\right) n^2 + O(n) . \tag{4.78}$$

Proof. We use the expression given to us by the unbiased walk in Eq. (4.61)

$$Z = \frac{2^n}{(q+1)^n} \mathop{\mathbb{E}}_{P_u, \Lambda_u} \left[\left(\frac{2q}{q^2+1}\right)^{(\#\text{ of bit flips during walk})}\right] . \tag{4.79}$$

Define $Z^{(t)}$ to be the value of the collision probability, given above via the biased walk, after t time steps, so $Z = Z^{(s)}$ and $Z^{(0)} = 2^n/(q+1)^n$.

Consider a given trajectory produced by the unbiased walk up to time step t, $\gamma = (\vec{\gamma}^{(0)}, \ldots, \vec{\gamma}^{(t)})$. If $\vec{\gamma}^{(t)} = I^n$ or $\vec{\gamma}^{(t)} = S^n$ then the walk has reached a fixed point and will never again change. From the calculation in Appendix 4.B.6, we know that the sum of all the weights of all walks *of any length* that reach a fixed point is precisely Z_H. Since the weights are non-negative, this implies that the sum over walks that have reached it before time step t is less than Z_H, and hence the combined weight of trajectories that have *not* reached a fixed point by time step t is at least $Z^{(t)} - Z_H$. Meanwhile, if $\vec{\gamma}^{(t)}$ is not at a fixed point, then we can consider the proper subset $R \subset [n]$ of sites with value S. By the h-regularly connected property, there is at least a $1/2$ chance that one of the gates between time step $t+1$ and $t+hn$ matches an index in R

with one in the complement of R. When this happens, a bit must be flipped and the weight of that trajectory is reduced by factor $2q/(q^2+1)$. Thus, the following must hold

$$Z^{(t+hn)} - Z_H \leq \left(\frac{1}{2} + \frac{1}{2}\frac{2q}{q^2+1}\right)\left(Z^{(t)} - Z_H\right) . \tag{4.80}$$

Moreover, we know that $Z^{(0)} = 2^n/(q+1)^n$, so

$$\begin{aligned} Z^{(s)} &\leq Z_H + \left(\frac{(q+1)^2}{2(q^2+1)}\right)^{s/(hn)}\left(\frac{2^n}{(q+1)^n} - Z_H\right) && (4.81)\\ &\leq Z_H + \left(\frac{(q+1)^2}{2(q^2+1)}\right)^{s/(hn)}\left(\frac{2^n}{(q+1)^n}\right) && (4.82)\\ &= Z_H\left(1 + \frac{2^n(q^n+1)}{2(q+1)^n}\left(\frac{(q+1)^2}{2(q^2+1)}\right)^{s/(hn)}\right) && (4.83)\\ &\leq Z_H\left(1 + \frac{2^n q^n}{(q+1)^n}\left(\frac{(q+1)^2}{2(q^2+1)}\right)^{s/(hn)}\right) && (4.84)\\ &\leq Z_H(1 + e^{-\frac{2a}{n}(s-s^*)}) , && (4.85) \end{aligned}$$

where

$$\begin{aligned} a &= (2h)^{-1}\log\left(\frac{2(q^2+1)}{(q+1)^2}\right) = \Theta(1) && (4.86)\\ s^* &= (2a)^{-1}\log\left(\frac{2q}{q+1}\right)n^2 = \Theta(n^2) . && (4.87) \end{aligned}$$

□

Note that we have made no attempt to optimize the constant prefactor of the $\Theta(n^2)$ or the value of a. Indeed, we conjecture that Theorem 4.3 could be improved so that $s^* = \Theta(n\log(n))$, which would be a dramatic improvement that implies the fundamental scaling of the anti-concentration size is independent of the architecture's connectivity, so long as it satisfies the regularly connected property.

4.C.2 Lower bound on collision probability

In this section, we prove an $\Omega(n\log(n))$ lower bound on the circuit size needed for anti-concentration in general circuit architectures. This also implies an $\Omega(\log(n))$ lower bound on the anti-concentration depth.

Theorem 4.4 (restated)**.** *For any RQC architecture of size s on n qudits with local dimension q, the collision probability satisfies*

$$Z \geq \frac{Z_H}{2}\exp\left(\frac{\log(q)}{q+1}\exp\left(\log(n) - \frac{2s}{n}\log\left(q^2+1\right)\right)\right) . \tag{4.88}$$

Corollary 4.3. *For a given RQC architecture, let s_{AC} be the minimum circuit size, as a function of n, such that $Z \leq 2Z_H$. Then it must hold that*

$$s_{AC} \geq \left(2\log\left(q^2+1\right)\right)^{-1} n\log(n) - O(n)\,. \tag{4.89}$$

Proof. This statement follows directly from the bound in Eq. (4.88). □

Corollary 4.4. *For a given RQC architecture, let d_{AC} be the minimum circuit depth, as a function of n, such that $Z \leq 2Z_H$. Then it must hold that*

$$d_{AC} \geq \left(\log\left(q^2+1\right)\right)^{-1} \log(n) - O(1)\,. \tag{4.90}$$

Proof. Each layer can have at most $n/2$ gates so it must hold that $d_{AC} \geq 2s_{AC}/n$. □

Proof of Theorem 4.4. We use the framework of the biased random walk, given by the expression for Z in Eq. (4.63). For each of the n sites, there is some initial probability that it starts with value S, and then each gate involving that site has some chance of flipping it to value I. However, there will always be some minimum probability that even after many gates, the value has not yet been flipped to I. This constitutes the idea behind our lower bound.

Given an index $j \in [n]$, we compute a lower bound on the probability that $\gamma_j^{(t)} = S$ for all $t = 0, 1, \ldots, s$, (i.e. the jth bit begins with value S and is never flipped to I), as a function of the number of gates s_j that act on qudit j

$$\Pr_{P_b,\Lambda_b}\left[\gamma_j^{(t)} = S \;\; \forall t \in \{0,\ldots,s\}\right] \geq \frac{1}{q+1}\left(\frac{1}{q^2+1}\right)^{s_j}\,, \tag{4.91}$$

since there is a $1/(q+1)$ chance that $\gamma_j^{(0)} = S$ when we draw $\vec{\gamma}^{(0)}$ from Λ_b, and the probability it does not flip after each gate is at least $1/(q^2+1)$. This holds for each j, and thus we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{P_b,\Lambda_b}\left[|\vec{\gamma}^{(s)}|\right] &= \sum_{j=1}^{n} \Pr_{P_b,\Lambda_b}\left[\gamma_j^{(s)} = S\right] && (4.92)\\
&\geq \sum_{j=1}^{n} \Pr_{P_b,\Lambda_b}\left[\gamma_j^{(t)} = S \;\; \forall t \in \{0,\ldots,s\}\right] && (4.93)\\
&\geq \frac{1}{q+1}\sum_{j=1}^{n}\left(\frac{1}{q^2+1}\right)^{s_j}\,. && (4.94)
\end{aligned}$$

Since each of the s gates in the circuit diagram acts on two indices, it must hold that $\sum_j s_j = 2s$, and given this constraint, the minimum of the final expression above occurs when all the s_j are equal, and thus

$$\mathop{\mathbb{E}}_{P_b,\Lambda_b}\left[|\vec{\gamma}^{(s)}|\right] \geq \frac{1}{q+1} n \left(\frac{1}{q^2+1}\right)^{2s/n}\,. \tag{4.95}$$

By convexity of the exponential function, we have $\mathbb{E}[q^x] \geq q^{\mathbb{E}[x]}$, and hence

$$Z = \frac{1}{q^n} \mathop{\mathbb{E}}_{P_b, \Lambda_b} [q^{|\vec{\gamma}^{(s)}|}] \tag{4.96}$$

$$\geq \frac{1}{q^n} \exp\left(\log(q) \frac{n}{q+1} \left(\frac{1}{q^2+1} \right)^{2s/n} \right) \tag{4.97}$$

$$\geq \frac{Z_H}{2} \exp\left(\frac{\log(q)}{q+1} \exp\left(\log(n) - \frac{2s}{n} \log\left(q^2+1\right) \right) \right) . \tag{4.98}$$

□

4.D Bounds for the 1D architecture

We now focus specifically on the 1D architecture defined formally in Definition 4.2. We assume periodic boundary conditions, although it would be possible to consider open boundary conditions as well. In 1D, the qudits are arranged in a geometrically local fashion and it is fruitful to think of a configuration $\vec{\gamma} \in \{I, S\}^n$ as being composed of contiguous *domains*, consecutive sites where all the values are I or all the values are S. We then identify *domain walls* as locations where one domain ends and another begins. Gates that couple qudits in different domains cause one of the values to flip, which moves the domain wall separating those domains one unit to the left or one unit to the right. The notation for talking formally about this is discussed in the next subsection, and then the upper and lower bounds on Z are proved.

4.D.1 Domain walls and notation

In 1D, configurations $\vec{\gamma} \in \{I, S\}^n$ are associated with a set of domain wall locations. We let

$$DW(\vec{\gamma}) = \{e \in \{0, 1, 2, \ldots, n-1\} : \gamma_e \neq \gamma_{e+1}\} \tag{4.99}$$

be the set of domain wall positions for a configuration $\vec{\gamma}$, where γ_0 is identified with γ_n when there are periodic boundary conditions. For each set of domain wall locations there are exactly two configurations that map to it, since choosing $\gamma_0 = I$ or $\gamma_0 = S$ determines the value of all other sites.

A configuration trajectory $\gamma = (\vec{\gamma}^{(0)}, \ldots, \vec{\gamma}^{(s)})$ is then associated with a sequence of sets of domain wall locations $G = (g^{(0)}, \ldots, g^{(s)})$ where $g^{(t)} = DW(\vec{\gamma}^{(t)})$. We call G a *domain wall trajectory*. Domain wall trajectories with non-zero contribution to the collision probability Z obey the following rules: when there is a domain wall at position e and a gate acts on qudits $\{e, e+1\}$, the domain wall must move to position $e-1$ or $e+1$ (at the cost of a reduction in the weight) and may annihilate with another domain wall if there is already a domain wall at the new position. However, pairs of domain walls cannot be created; the number of domain walls that exist throughout the domain wall trajectory is non-increasing, and a particular domain wall can be uniquely tracked throughout each step of the trajectory (either until the final step or

until its annihilation). Let $\mathcal{G}$ be the set of all domain wall trajectories that obey these rules. Any domain wall trajectory $G \in \mathcal{G}$ will have the property that when t is odd, e is even for all $e \in g^{(t)}$, and when t is even (but non-zero), e is odd for all $e \in g^{(t)}$. This is because odd (even) numbered layers couple qubits $\{2j-1, 2j\}$ ($\{2j, 2j+1\}$) meaning domain walls must lie between qudit positions $2j$ and $2j+1$ (between qudit positions $2j-1$ and $2j$) for some j.

By converting the sum over trajectories in Eq. (4.59) to a sum over domain wall trajectories, we can express Z by the equation

$$Z = \frac{2}{(q+1)^n} \sum_{G \in \mathcal{G}} \text{weight}(G)\,, \tag{4.100}$$

where the weight is given as follows, recalling that $A^{(t)}$ is the pair of qudit indices involved in the tth gate, which in 1D is always $A^{(t)} = \{j, j+1\}$ for some j.

$$\text{weight}(G) = \prod_{t=1}^{s} M^{(t)}_{g^{(t-1)} g^{(t)}} \tag{4.101}$$

$$M^{(t)}_{g^{(t-1)} g^{(t)}} \begin{cases} \frac{q}{q^2+1} & \text{if } \min(A^{(t)}) \in g^{(t-1)} \\ 1 & \text{otherwise}\,. \end{cases} \tag{4.102}$$

In other words, if the gate on qudits $\{j, j+1\}$ and there is a domain wall at position j, then the weight is reduced by a factor $q/(q^2+1)$ (and the domain wall must move to position $j-1$ or position $j+1$, possibly annihilating if a domain wall already exists at that position).

Given two domain wall trajectories G and G', we will consider the combined domain wall trajectory

$$G \sqcup G' = (g^{(0)} \sqcup g'^{(0)}, \ldots, g^{(s)} \sqcup g'^{(s)})\,, \tag{4.103}$$

where $\sqcup$ is the disjoint union and is defined only under the assumption $g^{(t)} \cap g'^{(t)} = \emptyset$ for all t.

The upshot of thinking about trajectories this way is that if $H = G \sqcup G'$, then

$$\text{weight}(H) = \text{weight}(G) \cdot \text{weight}(G')\,. \tag{4.104}$$

In particular, we will find it useful to decompose a domain wall trajectory G into $G = G_U \sqcup G_0$ where G_U is a domain wall trajectory with a conserved number of domain walls throughout the trajectory, and G_0 is a trajectory for which $|G_0^{(s)}| = 0$, i.e. all the domain walls have annihilated by the end of the trajectory. This decomposition is unique, and an example is shown in Figure 4.5. Let $\mathcal{G}_U$ and $\mathcal{G}_0$ be the subsets of $\mathcal{G}$ that have no annihilations and that have no surviving domain walls at the end of the circuit, respectively. Let $\mathcal{G}_{U,k}$ be the subset of $\mathcal{G}_U$ with k domain walls. When the boundary conditions are periodic, k must be even for $G_{U,k}$ to be non-empty.

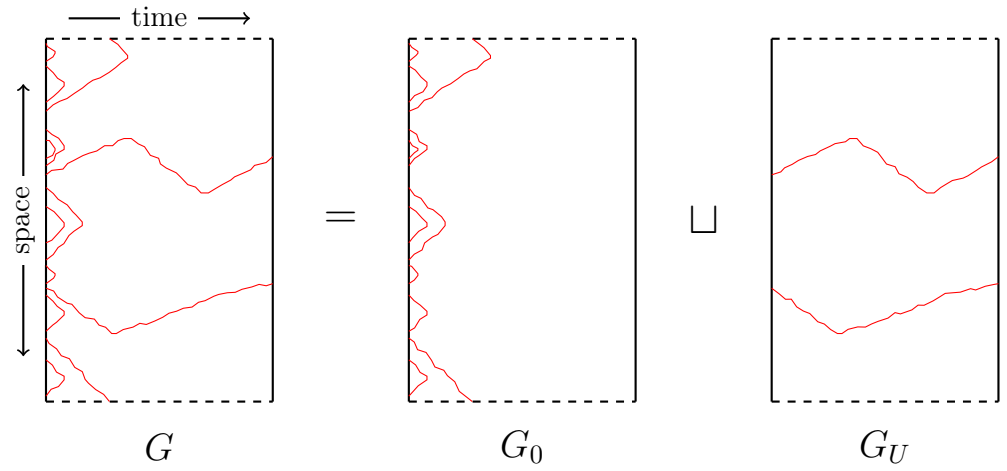


Figure 4.5: Cartoon illustrating unique decomposition of a domain wall trajectory G into a disjoint union of one part, G_0, where all domain walls annihilate prior to the end of the circuit, and another part, G_U, where no domain walls annihilate.

4.D.2 Collision probability upper bound

Theorem 4.1 (restated)**.** *For the 1D architecture, let*

$$a = \log\left(\frac{q^2+1}{2q}\right) \tag{4.105}$$

$$s^* = \frac{1}{2a} n \log(n) + n\left(\frac{1}{2a}\log(e-1) + \frac{1}{2}\right) = (2a)^{-1} n \log(n) + O(n)\,. \tag{4.106}$$

Then,

$$Z \leq Z_H(1 + e^{-\frac{2a}{n}(s-s^*)}) \tag{4.107}$$

whenever $s \geq s^*$. The circuit depth d is $d = 2s/n$, so we may define $d^* = 2s^*/n$ and equivalently conclude

$$Z \leq Z_H(1 + e^{-a(d-d^*)})\,. \tag{4.108}$$

Note that when $s < s^*$, an upper bound on Z can still be inferred from this method. The essence of the proof of Theorem 4.1 is the same as the proof of the statement proved in [29], although we have expressed it here within our notation and framework.

Proof. We use the formula in Eq. (4.100), which expresses Z as a weighted sum over domain wall trajectories. Each domain wall trajectory $G = (g^{(0)}, \ldots, g^{(s)})$ can be associated with an integer $k = |g^{(s)}|$, the number of domain walls that remain unannihilated at the end of the trajectory. Due to periodic boundary conditions, k must be even, and let $k_0 = k/2$. Let $\mathcal{G}_k \subset \mathcal{G}$ be the associated set of length-s domain wall trajectories, and let $\mathcal{G}_{U,k} \subset \mathcal{G}_k$ be the subset

containing domain wall trajectories that have a conserved number of domain walls throughout. As discussed in the previous subsection, it is possible to uniquely decompose $H \in \mathcal{G}_k$ into $H = G \sqcup G'$ where $G \in \mathcal{G}_{U,k}$ and $G' \in \mathcal{G}_0$.

Suppose we fix a domain wall configuration $g^{(0)}$ for the initial time step at the beginning of the circuit with k domain walls. There are $\binom{n}{k}$ such configurations. The total weight of all the trajectories in $\mathcal{G}_{U,k}$ that begin at this configuration is at most $(2q/(q^2+1))^{k(d-1)}$ since each domain wall must move either left or right (introducing a factor of 2) during each of the d layers of gates, except for possibly the first layer (if the domain wall begins at an even position it does not move during the first layer), and each time one moves it incurs a weight reduction of $q/(q^2+1)$. This does not account for the rule that the k domain walls cannot intersect, but it still yields an upper bound on the total weight.

Meanwhile, the sum of the weights of all domain wall trajectories in $\mathcal{G}_0$ approaches $Z_H(q+1)^n/2$ from below as depth increases. This follows from the analysis in Appendix 4.B.6 where it was shown that the sum over all trajectories that eventually reach a fixed point is exactly $Z_H(q+1)^n$, but at a finite depth, not every trajectory will have reached a fixed point, so only a subset of the terms are included in the sum. Due to the fact that each domain wall configuration corresponds to 2 equal-weight trajectories through $\{I, S\}^n$ the sum of the weights of all the domain wall trajectories in $\mathcal{G}_0$ can be at most $Z_H(q+1)^n/2$.

Collecting these observations and recalling that $k = 2k_0$, we have

$$Z = \frac{2}{(q+1)^n} \sum_{k_0=0}^{n/4} \sum_{G \in \mathcal{G}_{2k_0}} \text{weight}(G) \tag{4.109}$$

$$= \frac{2}{(q+1)^n} \sum_{k_0=0}^{n/4} \sum_{G \in \mathcal{G}_{U,2k_0}} \sum_{\substack{G' \in \mathcal{G}_0 \\ G \cap G' = \emptyset}} \text{weight}(G) \cdot \text{weight}(G') \tag{4.110}$$

$$\leq \left(\sum_{k_0=0}^{n/4} \sum_{G \in \mathcal{G}_{U,2k_0}} \text{weight}(G) \right) \left(\frac{2}{(q+1)^n} \sum_{G' \in \mathcal{G}_0} \text{weight}(G') \right) \tag{4.111}$$

$$\leq \left(\sum_{k_0=0}^{n/4} \binom{n}{2k_0} \left(\frac{2q}{q^2+1} \right)^{2k_0(d-1)} \right) (Z_H) \tag{4.112}$$

$$= Z_H \sum_{k_0=0}^{n/4} \binom{n}{2k_0} (e^{-a})^{2k_0(d-1)} \leq Z_H (1 + e^{-a(d-1)})^n \tag{4.113}$$

$$\leq Z_H (1 + (e-1) n e^{-a(d-1)}) \tag{4.114}$$

$$= Z_H \left(1 + \exp\left(\log(n) - da + \log(e-1) + a\right)\right) \tag{4.115}$$

$$\leq Z_H \left(1 + \exp\left(-a(d - d^*)\right)\right) , \tag{4.116}$$

where Eq. (4.114) holds so long as $d \geq d^*$, based on the following small lemma.

Lemma 4.2. *If $b, c > 0$ and $cb \leq 1$, then*

$$(1+c)^b \leq 1 + cb(e-1)\,. \tag{4.117}$$

Proof.

$$(1+c)^b = \sum_{k=0}^{b} \binom{b}{k} c^k = 1 + cb \sum_{k=1}^{b} \binom{b}{k} \frac{c^{k-1}}{b} \tag{4.118}$$

$$\leq 1 + cb \sum_{k=1}^{b} \binom{b}{k} b^{-k} \leq 1 + cb\left((1+b^{-1})^b - 1\right) \tag{4.119}$$

$$\leq 1 + cb(e-1)\,. \tag{4.120}$$

□

□

4.D.3 Collision probability lower bound

Theorem 4.5 (restated)**.** *Consider the 1D architecture. There are constants A and A' such that as long as $s^* - s \geq A'n$, the collision probability satisfies*

$$Z \geq \frac{Z_H}{2} \exp\left(A e^{\log(n) - \frac{2a}{n}s}\right)\,, \tag{4.121}$$

where a and s^* are the same as in Theorem 4.1.

In our proof, the constant A is explicit but very small, on the order of e^{-10}, and $A' \approx -\log(A)$. The value of A could certainly be improved with some attempt at optimization.

Corollary 4.5. *For the 1D architecture, if we define s_{AC} and d_{AC} to be the smallest circuit size and circuit depth for which $Z \leq 2Z_H$, then*

$$\left| s_{AC} - \left(2\log\left(\frac{q^2+1}{2q}\right)\right)^{-1} n\log(n) \right| \leq O(n) \tag{4.122}$$

$$\left| d_{AC} - \left(\log\left(\frac{q^2+1}{2q}\right)\right)^{-1} \log(n) \right| \leq O(1)\,. \tag{4.123}$$

Proof. Theorem 4.1 implies that

$$s_{AC} \leq s^* = (2a)^{-1} n\log(n) + O(n)\,. \tag{4.124}$$

Meanwhile, Theorem 4.5 implies that if

$$\begin{aligned} s &\leq (2a)^{-1} n \log(n) - \max\left((2a)^{-1} \log\bigl(\log(4) A^{-1}\bigr), A' \right) n \\ &= (2a)^{-1} n \log(n) - O(n)\,, \end{aligned} \tag{4.125}$$

then $Z \geq 2Z_H$. Hence $s_{AC} \geq (2a)^{-1} n \log(n) - O(n)$. Together these imply that $|s_{AC} - (2a)^{-1} n \log(n)| = O(n)$. □

Proof of Theorem 4.5. Eq. (4.100) expresses Z as a weighted sum over domain wall trajectories. Heuristically, when $d < d^*$ we expect that the output distribution will *not* be anti-concentrated and that domain wall trajectories drawn at random with probability proportional to its weight will usually have many domain walls that never annihilate. To lower bound Z, we will sum over the set of configurations with k unannihilated domain walls for a particularly chosen value of k.

For a fixed value of the depth d, define

$$n_H = \frac{e^{(d-1)a}}{2(e-1)}\,. \tag{4.126}$$

We chose n_H to be exactly half the value of n for which a depth-d circuit would be anti-concentrated. Heuristically, we expect on the order of $n/2n_H$ unannihilated domain walls in typical configurations.

Let k be an even integer to be specified later. Let $\mathcal{H}_k \subset \mathcal{G}_{U,k}$ be the set that contains any domain wall trajectory $H = (h^{(0)}, \ldots, h^{(s)})$ for which

(1) H has k domain walls at each time step (none annihilate)

(2) For each of the k domain walls in the initial configuration $h^{(0)}$, the nearest domain wall in both directions is at most n_H positions away.

Now, temporarily fix some $H \in \mathcal{H}_k$. It has k domain walls which move around throughout the trajectory. We let $e_{H,j,t}$ be the location of the jth domain wall at time step t in the trajectory H. We then define the set $\mathcal{J}_{H,j} \subset \mathcal{G}_0$, for $j = 1, \ldots, k$ to be the set of domain wall trajectories for which (1) all of the domain walls annihilate before time step s, and (2) the position e_t of any domain wall at time step t satisfies

$$e_{H,j,t} < e_t < e_{H,j+1,t}\,. \tag{4.127}$$

In other words, all of the domain walls fall between the jth and $(j+1)$th domain walls of H. This ensures that H is disjoint from any $J_j \in \mathcal{J}_{H,j}$.

Specifying a trajectory $H \in \mathcal{H}_k$ as well as $J_j \in \mathcal{J}_{H,j}$ for each $j = 1, \ldots, k$, determines a unique trajectory $H' = H \sqcup J_{H,1} \sqcup \ldots \sqcup J_{H,k}$. This decomposition

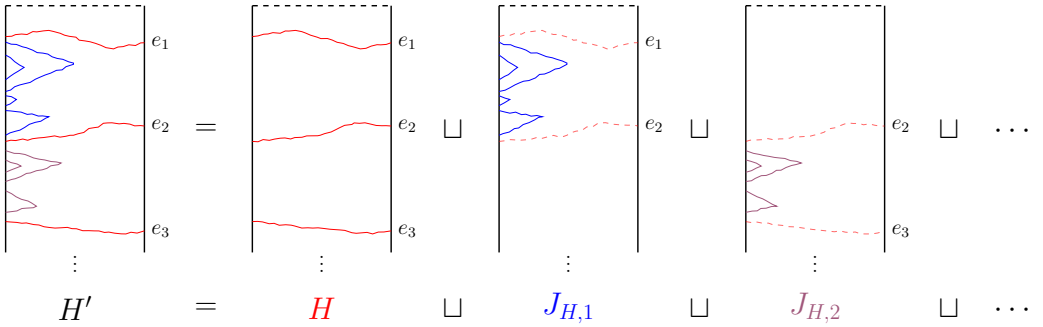


Figure 4.6: Outline of the main idea of the proof of Theorem 4.5. We choose a domain wall trajectory H which has k domain walls that never annihilate and such that the distance between consecutive domain walls is always at most n_H. We then choose domain wall trajectories $J_{H,1}, \ldots, J_{H,k}$ such that the domain walls of $J_{H,j}$ lie between the jth and $(j+1)$th domain walls of H and all annihilate before the end of the circuit. The domain wall configuration H' is the disjoint union of H and $J_{H,j}$ for $j = 1, \ldots, k$. We can lower bound the collision probability by lower bounding the weighted sum over the contribution from all H' formed this way.

is illustrated in Figure 4.6. Thus, if we perform the weighted sum only over the set of H' formed this way, we will arrive at a lower bound to Z, as follows:

$$Z = \frac{2}{(q+1)^n} \sum_{H \in \mathcal{G}} \text{weight}(H) \tag{4.128}$$

$$\geq \frac{2}{(q+1)^n} \left(\sum_{H \in \mathcal{H}_k} \text{weight}(H) \right) \prod_{j=1}^{k} \left(\sum_{J_j \in \mathcal{J}_{H,j}} \text{weight}(J_j) \right) . \tag{4.129}$$

The quantities in parentheses can be bounded with the following two lemmas, whose proofs are delayed until after the proof of the Theorem.

Lemma 4.3. *If $4d \leq \lfloor n/k \rfloor$ and $n_H/2 \geq \lceil n/k \rceil$ hold, then the set $\mathcal{H}_k$ satisfies*

$$\sum_{H \in \mathcal{H}_k} \textit{weight}(H) \geq \left(\frac{1}{2} \left\lfloor \frac{n}{k} \right\rfloor \right)^k \left(\frac{2q}{q^2+1} \right)^{dk} . \tag{4.130}$$

Lemma 4.4. *Fix a value of H and j. Suppose that the jth and $(j+1)$th domain walls of the initial configuration of H lie at positions e and $e + X - 1$ (mod n), respectively, for some positive integer $X < n$. Then*

$$\left(\sum_{J \in \mathcal{J}_{H,j}} \textit{weight}(J) \right) \geq \frac{1}{c} \left(\frac{q+1}{q} \right)^X , \tag{4.131}$$

where $c = 3e^{10}$.

The sum of the domain length X for each of the domains is simply n. Thus the $((q+1)/q)^X$ factors cancel the $1/(q+1)^n$ prefactor for Z, and we have

$$Z \geq q^{-n} \left(\frac{1}{2} \left\lfloor \frac{n}{k} \right\rfloor \right)^k c^{-k} \left(\frac{2q}{q^2+1} \right)^{dk} = q^{-n} \left(\frac{1}{2} \left\lfloor \frac{n}{k} \right\rfloor \right)^k c^{-k} e^{-adk} \tag{4.132}$$

for any k that satisfies $4d \leq \lfloor n/k \rfloor$ and $n_H/2 \leq \lceil n/k \rceil$.

Now we choose a value of k to maximize the right-hand-side of the above equation. In the limit of large n, the requirement that k is an even integer will have negligible effect. In our analysis, we handle this requirement by defining k' to be a real number and k to be the smallest even integer larger than k', and then we make a few rather crude bounds on the floor and ceiling of quantities like n/k, which are not asymptotically tight, but good enough for our purposes. We choose

$$k' = \frac{n \left(\frac{2q}{q^2+1} \right)^d}{8ce} = \frac{ne^{-da}}{8ce} = \frac{n}{n_H} \frac{e^{-a}}{16e(e-1)c} \tag{4.133}$$

$$k = \text{smallest even integer greater than } k' . \tag{4.134}$$

Note that n/k' is at least $8ce$, which is very large, meaning $\lceil n/2k' \rceil /2 \leq n/2k' \leq 2\lfloor n/2k' \rfloor$ certainly holds. For finite n, we can say that as long as $k' \geq 1$, then $k' \leq k \leq 2k'$ will hold. The requirement $k' \geq 1$ translates into

$$d \leq a^{-1}(\log(n) - \log(8ce)) , \tag{4.135}$$

which, by recalling $s = nd/2$ and that $s^* \geq (2a)^{-1} n \log(n)$, can be re-expressed as

$$s^* - s \geq A'n \tag{4.136}$$

with $A' = (2a)^{-1} \log(8ce(e-1)) + 2^{-1}$, which is assumed to hold in the theorem statement. This implies that

$$\left\lfloor \frac{n}{k} \right\rfloor \geq \left\lfloor \frac{n}{2k'} \right\rfloor \geq \frac{n}{4k'} . \tag{4.137}$$

Inspection of the formula for k' reveals that the relation $4d \leq \lfloor n/k \rfloor$ holds for any d and n. Moreover, we have $n_H/2 = (ne^{-a})/(k' 32e(e-1)c) \leq \lceil n/k \rceil$, so the second relation holds as well.

Recall that $Z_H = 2/(q^n+1) \leq 2q^{-n}$. Plugging in the above bound on $\lfloor n/k \rfloor$ into Eq. (4.132), we find

$$Z \geq \frac{Z_H}{2}\exp(k) \geq \frac{Z_H}{2}\exp(k') \tag{4.138}$$

$$\geq \frac{Z_H}{2}\exp\left(\frac{ne^{-da}}{8ce}\right) \tag{4.139}$$

$$= \frac{Z_H}{2}\exp\left(\frac{1}{8ce}e^{\log(n)-da}\right) \tag{4.140}$$

$$= \frac{Z_H}{2}\exp\left(\frac{1}{8ce}e^{\log(n)-\frac{2as}{n}}\right) \tag{4.141}$$

$$= \frac{Z_H}{2}\exp\left(Ae^{\log(n)-\frac{2as}{n}}\right) \tag{4.142}$$

for A defined to equal $1/8ce$. Note that this value of A is quite small (on the order of e^{-10}), but with some optimization could likely be made much larger.

□

Now we provide the delayed proofs of the two lemmas.

Proof of Lemma 4.3. Each term in the sum on the left-hand-side is non-negative, so we make a lower bound by summing over a subset of the terms. To do so, we can split the n indices up into k nearly equal-size segments of length at most $\lceil n/k \rceil$, which is less than $n_H/2$ by assumption. Then for each of these segments, we choose the location of a single domain wall that is at least distance d from each edge of the segment. This will generate a unique initial domain wall configuration that satisfies criteria (2) of $\mathcal{H}_k$, since any pair of consecutive domain walls is closer than n_H apart. The total number of choices is at least

$$\left(\left\lfloor \frac{n}{k} \right\rfloor - 2d\right)^k \tag{4.143}$$

which, by the assumption $4d \leq \lfloor n/k \rfloor$, is at least $(\lfloor n/k \rfloor/2)^k$.

Once the initial k domain wall locations have been chosen, we examine how they can propagate through the circuit. Each layer of gates will force each of the k domain walls to move in one of two directions, and the weight is reduced by a factor $(q/(q^2+1))^k$, except for the first layer, where some of the domain walls may not move if they begin at an even index. Since, by construction, there are no instances where domain walls start within a distance of $2d$ of any other domain wall, there is no chance of domain walls crossing. Thus, we find that for each initial set of k locations chosen in the manner outlined above, the combined weight of all possible trajectories is at least $(2q/(q^2+1))^{kd}$. This proves the lemma. □

Proof of Lemma 4.4. Consider an alternative 1D qudit system with periodic boundary conditions consisting of X sites by identifying site $e+X$ with site e and ignoring all other sites. Because $H \in \mathcal{H}_k$, we can be assured that $X \leq n_H$. Let $\mathcal{J}'_{H,j}$ be the set of all domain wall trajectories on the size-X system. Let $\mathcal{J}'_{H,j,l}$ be the subset that have $l = 2l_0$ domain walls on the last time step. Because the collision probability, denoted Z_X, for this X-qudit system must satisfy $Z_X \geq Z_{H,X}$, and here $Z_{H,X} = 2/(q^X+1)$, it must be the case that

$$Z_X = \frac{2}{(q+1)^X}\left(\sum_{J'\in\mathcal{J}'_{H,j}} \text{weight}(J)\right) = \frac{2}{(q+1)^X}\sum_{l_0=0}^{X/4}\left(\sum_{J'\in\mathcal{J}'_{H,j,2l_0}} \text{weight}(J)\right) \tag{4.144}$$

$$\geq \left(\frac{2}{q^X+1}\right) =: Z_{H,X}\,. \tag{4.145}$$

We can upper bound the contribution of all the terms with $l_0 > 0$ in the above expression by the method that yielded the upper bound in Theorem 4.1. The sum of those terms is upper bounded by the second term in Eq. (4.114), that is

$$\frac{2}{(q+1)^X}\sum_{l_0=1}^{X/4}\left(\sum_{J'\in\mathcal{J}'_{H,j,2l_0}} \text{weight}(J)\right) \leq Z_{H,X}(e-1)Xe^{-a(d-1)} \tag{4.146}$$

$$= \left(\frac{2}{q^X+1}\right)\frac{X}{2n_H} \leq \frac{1}{2}\left(\frac{2}{q^X+1}\right), \tag{4.147}$$

since $X \leq n_H$. Combining Eqs. (4.145) and (4.147), we find a lower bound on the $l_0 = 0$ term

$$\left(\sum_{J'\in\mathcal{J}'_{H,j,0}} \text{weight}(J)\right) \geq \left(\frac{(q+1)^X}{2}\right)\left(\frac{2}{q^X+1}\right)\left(1-\frac{1}{2}\right) \tag{4.148}$$

$$= \left(\frac{q+1}{q}\right)^X\left(\frac{1}{2}\frac{q^X}{q^X+1}\right) \tag{4.149}$$

$$\geq \left(\frac{q+1}{q}\right)^X\left(\frac{1}{3}\right), \tag{4.150}$$

where the last inequality follows since $q \geq 2$ and $X \geq 1$ must be true. Now, every domain wall trajectory in $\mathcal{J}_{H,j}$ will also be in $\mathcal{J}'_{H,j,0}$, but the converse will not be true. Some trajectories in the latter set will have one or more domain walls that intersect with either the jth or the $(j+1)$th domain wall of H at some time step, which is not allowed within the former set. Thus, the sum over the domain wall trajectories in $\mathcal{J}_{H,j}$ will be smaller than the sum over those in $\mathcal{J}'_{H,j,0}$. However, we argue that the difference is at most some constant

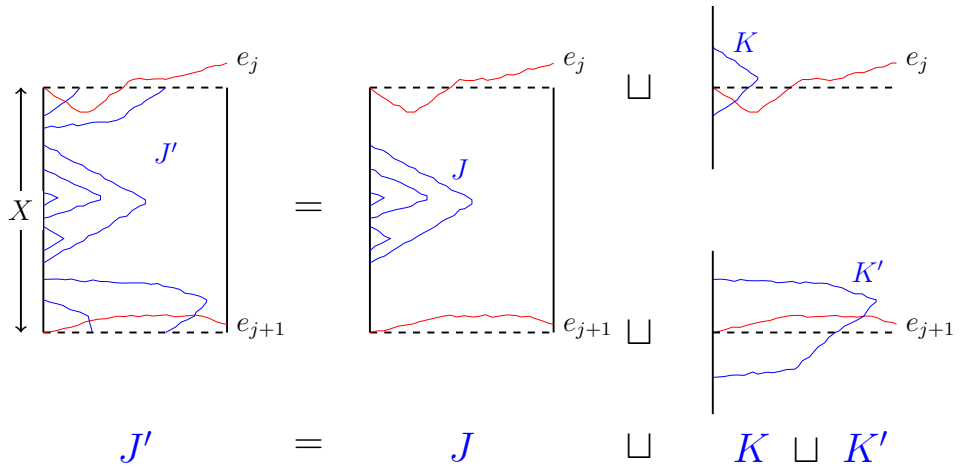


Figure 4.7: Outline of the argument in the proof of Lemma 4.4 that the sum over domain wall trajectories in $\mathcal{J}_{H,j}$ is at least the sum in $\mathcal{J}'_{H,j,0}$ divided by some constant factor, expressed in Eq. (4.159). Every trajectory in $J' \in \mathcal{J}'_{H,j,0}$ can be decomposed into a trajectory $J \in \mathcal{J}_{H,j}$ a trajectory $K \in \mathcal{B}_{H,j}$, and a trajectory $K' \in \mathcal{B}_{H,j+1}$, where each domain wall in K intersects the jth domain wall of H, and each domain wall of K' intersects the $(j+1)$th domain wall of H. Because the combined weight of all possible K and K' is only a constant factor, independent of n, the combined weight of all possible J cannot be more than a constant factor smaller than the combined weight of all possible J'. Note that the system with X sites has periodic boundary conditions in this figure.

factor by the following argument, which is also described in Figure 4.7. Let $\mathcal{B}_{H,j}$ be the set of all trajectories in which every domain wall either intersects the jth domain wall of H at some time step t, or it annihilates with a domain wall that previously intersected with the jth domain wall of H. Then any trajectory in $\mathcal{J}'_{H,j,0}$ can be formed as the disjoint union of a trajectory in $J \in \mathcal{J}_{H,j}$, a trajectory in $K \in \mathcal{B}_{H,j}$ and a trajectory in $K' \in \mathcal{B}_{H,j+1}$, to account for the parts that intersect the jth and $(j+1)$th domain walls. Given J', the choice of J for this decomposition is unique, but there may be multiple choices of (K, K') for which it holds. Note also that a trajectory in $\mathcal{B}_{H,j}$ can be decomposed into individual domain wall pairs that coincide with the jth domain wall of H at some time step t and annihilate at some time step t'. The combined weight of all such pairs, given fixed coincidence point at $e_{H,j,t}$ is at most $(2q/(q^2+1))^{2t'}$. Summing over $t' \geq t$, we find that the combined weight for all possible domain wall pairs coinciding at time step t is at most

$$\frac{\left(\frac{2q}{q^2+1}\right)^{2t}}{1-\left(\frac{2q}{q^2+1}\right)^2} = \frac{(q^2+1)^2}{(q^2-1)^2}\left(\frac{2q}{q^2+1}\right)^{2t}. \tag{4.151}$$

There can be many domain wall pairs that intersect the jth domain wall of H, but for each value of t there will either be no intersection (in which case the factor is 1) or one intersection (in which case the factor is at most the above quantity). Thus we can take the product over including or not including a domain wall at each value of t and find:

$$\sum_{K\in\mathcal{B}_{H,j}} \text{weight}(K) \leq \prod_{t=1}^{s} \left(1 + \frac{(q^2+1)^2}{(q^2-1)^2}\left(\frac{2q}{q^2+1}\right)^{2t}\right) . \tag{4.152}$$

This implies

$$\left(\sum_{J'\in\mathcal{J}'_{H,j,0}} \text{weight}(J)\right) \tag{4.153}$$

$$\leq \left(\sum_{J\in\mathcal{J}_{H,j}} \text{weight}(J)\right)\left(\sum_{K\in\mathcal{B}_{H,j}} \text{weight}(K)\right)\left(\sum_{K'\in\mathcal{B}_{H,j+1}} \text{weight}(K')\right) \tag{4.154}$$

$$\leq \left(\sum_{J\in\mathcal{J}_{H,j}} \text{weight}(J)\right)\cdot \prod_{t=1}^{s} \left(1 + \frac{(q^2+1)^2}{(q^2-1)^2}\left(\frac{2q}{q^2+1}\right)^{2t}\right)^2 \tag{4.155}$$

$$\leq \left(\sum_{J\in\mathcal{J}_{H,j}} \text{weight}(J)\right)\cdot \exp\left(2\sum_{t=1}^{s} \frac{(q^2+1)^2}{(q^2-1)^2}\left(\frac{2q}{q^2+1}\right)^{2t}\right) \tag{4.156}$$

$$\leq \left(\sum_{J\in\mathcal{J}_{H,j}} \text{weight}(J)\right)\cdot \exp\left(\frac{8q^2}{(q^2-1)^2}\frac{1}{1-\left(\frac{2q}{q^2+1}\right)^2}\right) \tag{4.157}$$

$$= \left(\sum_{J\in\mathcal{J}_{H,j}} \text{weight}(J)\right)\cdot \exp\left(\frac{8q^2(q^2+1)^2}{(q^2-1)^4}\right) \tag{4.158}$$

$$\leq \left(\sum_{J\in\mathcal{J}_{H,j}} \text{weight}(J)\right)\cdot e^{10} , \tag{4.159}$$

where the last inequality follows since $q \geq 2$ and the function of q inside the exp is monotonically decreasing. Combining the above with Eq. (4.150), we arrive at

$$\left(\sum_{J\in\mathcal{J}_{H,j}} \text{weight}(J)\right) \geq \left(\frac{q+1}{q}\right)^X \left(\frac{1}{3e^{10}}\right) = \left(\frac{q+1}{q}\right)^X c^{-1} . \tag{4.160}$$

□

4.E Bounds for the complete-graph architecture

4.E.1 Proof intuition and guide

In the following sections, we complete the proofs for upper and lower bounds of the complete-graph architecture, defined formally in Definition 4.1. The first insight about the complete-graph architecture is that all configurations with the same Hamming weight are equivalent, as there is a symmetry upon permutation of the qudits. Thus, trajectories through configuration space $\{I, S\}^n$ are reduced to trajectories through Hamming weight space $\{0, 1, \ldots, n\}$.

Our upper bound will use the framework of the unbiased walk, and the lower bound will use the biased walk. Recall we can use the unbiased walk to express the collision probability Z as a sum over all possible paths that the trajectory might take, working from Eq. (4.61)

$$Z = \frac{1}{(q+1)^n} \sum_{\vec{\gamma}^{(0)}} \mathop{\mathbb{E}}_{P_u, \vec{\gamma}^{(0)}} \left[\left(\frac{2q}{q^2+1} \right)^{(\#\text{ of bit flips during walk})} \right] \tag{4.161}$$

$$= \frac{1}{(q+1)^n} \sum_{x=0}^{n} \binom{n}{x} \mathop{\mathbb{E}}_{P_u, x} \left[\left(\frac{2q}{q^2+1} \right)^{(\#\text{ of bit flips during walk})} \right], \tag{4.162}$$

where the $\binom{n}{x}$ comes from the fact that this is the number of initial configurations with Hamming weight x. For the complete-graph case, P_u takes on a simple form: if the current configuration is x, the chances that the configuration changes on the next time step is precisely the chances of finding mismatching values upon drawing a random pair of indices in $[n]$, which is given by $\frac{2x(n-x)}{n(n-1)}$, and if it does change, it is equally likely to become $x-1$ or to become $x+1$. For the biased walk P_b, everything is the same except that when the configuration changes, it is biased to travel to $x-1$ with probability $q^2/(q^2+1)$. Also, in the biased case, the initial configuration is not a uniform choice over all configurations but instead distributed according to Λ_b, and the expectation in the above equation is replaced with $\mathbb{E}[q^{|\vec{\gamma}^{(s)}|}]$. Thus larger Hamming weight configurations are exponentially more significant in their contribution to Z.

To get an intuition for what we expect, we first think about the biased walk, which is what we use for the lower bound. Here, the peak of the probability mass in the initial configuration Λ_b starts around Hamming weight $x = n/(q+1)$. On average, the walk lingers for $n(n-1)/2x(n-x)$ time steps before moving, which is approximately equal to $n/2x$ when x is close to 0. Due to the bias, and due to the time required to wait, the effective speed of the biased walk is

$$v_{\text{eff}}(x) = \frac{2x(n-x)}{n(n-1)} \left(\frac{q^2}{q^2+1} - \frac{1}{q^2+1} \right) \approx \frac{2x}{n} \frac{q^2-1}{q^2+1} \tag{4.163}$$

in the direction of 0, since each time it moves, it has a $q^2/(q^2+1)$ chance of moving one unit closer to 0, but a $1/(q^2+1)$ chance of moving one unit

farther away from 0. Thus, in expectation, the time it takes for the peak of the probability mass to reach value 0 is

$$\sum_{x=1}^{n/(q+1)} \frac{1}{v_{\text{eff}}(x)} \approx \frac{q^2+1}{q^2-1}\frac{1}{2} n \log(n) \approx: s^* \,, \tag{4.164}$$

noting that $\sum_x \frac{1}{x} \approx \log(n)$.

This strongly suggests that s^* time steps are *necessary* for anti-concentration, as any less time would mean the peak of the distribution over Hamming weights at the end of the circuit will be located at some Hamming weight $y > 0$ and as a result, it will receive a significant amount of weight q^y in its contribution to Z. This is the intuition for our lower bound.

The biased walk also gives intuition for why there is a matching upper bound. If the circuit size is a little bigger than the lower bound, we expect the peak of the distribution to have terminated at the fixed point at 0. It is still possible that the tail of the distribution, which will not yet have reached the fixed point, is too fat to for anti-concentration to have been achieved; each unit farther away from 0 results in a factor of q larger contribution to Z, so we need the tail to be exponentially decaying if we want to be able to ignore it. This is essentially what we are able to show, albeit in a way where it might not be completely clear that this is what we have done. Intuitively, one reason we expect this exponentially decaying tail is because the effective speed slows down as you get closer to zero. This gives the tail of the distribution, which is sitting further away from 0, time to "catch up," as its effective speed is faster.

To actually perform the upper bound, we turn back to the unbiased walk. To be clear and to match the progression in the full proof, we introduce the concept of a *reduced* path (equivalently, "reduced walk") as a walk that never stands still at a certain configuration. If its Hamming weight at time step t is y, then its Hamming weight at time step $t+1$ will move to $y-1$ or $y+1$. For each walk, we can form a corresponding reduced walk simply by removing consecutive duplicates from the sequence of configurations. Another way to look at it is that given a fixed reduced walk, the actual walk will linger at each location for a certain number of time steps before continuing. In the limit of large circuit size s, there is enough time for the actual walk to linger as long as it would like at each step, and any reduced walk will successfully be "completed" by the actual walk. In this limit, $Z = Z_H$ where $Z_H = 2/(q^n+1)$ is the Haar value. Away from this limit, there is some probability that some reduced paths will not be completed.

We are able to express the difference between Z and Z_H as a sum over all reduced paths, including reduced paths that do not terminate at Hamming weight 0 or Hamming weight n, where the summand is proportional to the

probability that the reduced path is not completed within s time steps:

$$(q+1)^n(Z - Z_H) \tag{4.165}$$

$$= \sum_{\text{red path } \phi} \left(\frac{q}{q^2+1}\right)^{\text{length of } \phi} \frac{(q-1)^2}{2q} \Pr[\phi \text{ not completed in } s \text{ time steps}]\,. \tag{4.166}$$

The next key insight is to use a Chernoff bound to bound the probability of a reduced walk not being completed. If L is the length of the walk, the Chernoff bound states (for any constant $a > 0$)

$$\Pr[L > s] \leq \frac{\mathbb{E}[e^{aL}]}{e^{as}}\,, \tag{4.167}$$

but this is particularly useful because, for fixed ϕ, L is itself a sum of independent random variables $L_{\phi^{(i)}}$, the number of time steps the walk waits on step i. Thus

$$\mathbb{E}[e^{aL}] = \prod_i \mathbb{E}[e^{aL_{\phi^{(i)}}}] \tag{4.168}$$

and because each random variable $L_{\phi^{(i)}}$ is exponentially distributed, we can calculate $\mathbb{E}[e^{aL_{\phi^{(i)}}}]$ exactly. For the purposes of the proof sketch, denote

$$T_x = \mathbb{E}[e^{aL_x}]\,, \tag{4.169}$$

which will depend only on the Hamming weight x of the configuration the walk is at. (The walk will wait longer when it is near 0 or n than when it is near $n/2$.) This dependence appears to be a problem as it is unclear how to actually perform the following sum over all possible ϕ. (The constant a for each reduced walk ϕ, denoted a_ϕ, will be specified later.)

$$(q+1)^n(Z - Z_H) = \frac{(q-1)^2}{2q} \sum_{\text{red path } \phi} e^{-a_\phi s} \prod_{i=1}^{\text{length of } \phi} \left(\frac{q}{q^2+1} T_{\phi^{(i)}}\right)\,. \tag{4.170}$$

To proceed, we break ϕ up into subpaths that inch closer and closer to 0 and n. We can write ϕ as the concatenation of ϕ_x, ϕ_{x-1}, …, ϕ_w, where ϕ_v begins at either v or $n-v$ and only reaches $v-1$ or $n-v+1$ for the first time on the very last step. Then, w is the minimum Hamming weight distance from one of the fixed points (0 or n) the reduced walk ϕ ever reaches. Because the walk ϕ_v spends all its time between v and $n-v$, the expectation T_y for all of the y within one of these ϕ_v walks will be less than or equal to T_v (the walk moves slower when its closer to 0 or n), and we can write

$$\begin{aligned} &\frac{2q}{(q-1)^2}(q+1)^n(Z - Z_H) \\ &\leq \sum_{x=0}^{n} \binom{n}{x} \sum_{w=0}^{\min(x,n-x)} e^{-a_w s} \prod_{v=w}^{x} \left[\sum_{\phi_v} \left(\frac{q}{q^2+1} T_v\right)^{\text{length of } \phi_v}\right]\,. \end{aligned} \tag{4.171}$$

Here the $\binom{n}{x}$ comes in as the number of configurations with Hamming weight x, and a_ϕ has changed to a_w because we will choose it so that it only depends on the end point w of ϕ.

The above equation is huge progress because we already know how to perform the sums in brackets. Essentially, the factor of T_x simply changes the *effective* value of q; we may define $\bar{q}$ to satisfy

$$\bar{q}/(\bar{q}^2+1) = qT_x/(q^2+1)\,. \tag{4.172}$$

Then we can use the formulas for sums over paths that we have already developed in Lemma 4.1 to perform the sums. What we find is that, for the values of a_w that we can choose, we must allocate roughly $(q^2+1)n/2(q^2-1)x$ time steps for s such that $e^{-a_w s}$ can cancel out the value of the sum in brackets for ϕ_x. Note that this is precisely the inverse of the effective speed we defined before. Then, for all the sums to be canceled from $v=1$ to $v=n/2$, we must allocate

$$\sum_{v=1}^{n/2} \frac{q^2+1}{2(q^2-1)}\frac{n}{x} \approx \frac{q^2+1}{2(q^2-1)} n\log(n) \tag{4.173}$$

time steps. Fundamentally, the $\log(n)$ factor becomes necessary because the walk waits longer and longer as it gets closer and closer to the fixed points. In the full analysis, we find a term linear in n is also necessary to fully anti-concentrate, but our analysis of the constant prefactor for the linear term is not tight.

4.E.2 Preliminaries

Trajectories

For the complete-graph architecture, we may keep track of only the Hamming weight of a certain configuration. Thus, our random walks are over the set $\{0,1,\ldots,n\}$. A trajectory γ is now a sequence of integers $(\gamma^{(0)},\ldots,\gamma^{(s)})$. Generally speaking, if $t > s$ for a sequence of length s, let $\gamma^{(t)}$ return $\gamma^{(s)}$. A sequence is valid if for every t, $|\gamma^{(t)} - \gamma^{(t-1)}| \leq 1$ and such that if 0 or n appears, it appears only once at the very end of the sequence. Let Γ be the set of all valid trajectories.

For any valid trajectory γ, the unbiased random walk associates a non-zero probability:

$$\Pr_{P_u}[\gamma] = \prod_{t=1}^{s} P_u[\gamma^{(t)}|\gamma^{(t-1)}]\,, \tag{4.174}$$

where

$$P_u[y|x] = \begin{cases} \frac{x(n-x)}{n(n-1)} & \text{if } |y-x|=1 \\ 1-\frac{2x(n-x)}{n(n-1)} & \text{if } y=x \\ 0 & \text{otherwise} \end{cases}\,. \tag{4.175}$$

We can make the same definition for the biased random walk by replacing P_u with P_b where

$$P_b[y|x] = \begin{cases} \frac{2q^{1+x-y}}{q^2+1}\frac{x(n-x)}{n(n-1)} & \text{if } |y-x|=1 \\ 1-\frac{2x(n-x)}{n(n-1)} & \text{if } y=x \\ 0 & \text{otherwise} \end{cases} . \tag{4.176}$$

For $P \in \{P_u, P_b\}$ and any subset $\Upsilon \subseteq \Gamma$, we let $\Pr_P[\Upsilon] = \sum_{\gamma\in\Upsilon} \Pr_P[\gamma]$ be the total probability assigned to paths in Υ.

Conditional probabilities and expectations

For any $\gamma \in \Upsilon$, we may also define the conditional probability

$$\Pr_P[\gamma|\Upsilon] = \frac{\Pr_P[\gamma]}{\Pr_P[\Upsilon]} , \tag{4.177}$$

which indicates drawing from the subset Υ with probability proportional to that assigned by the (unbiased or biased) random walk. This also allows us to naturally define conditional expectation values for some quantity Q computed from γ

$$\mathop{\mathbb{E}}_P[Q[\gamma]|\gamma\in\Upsilon] = \sum_{\gamma\in\Upsilon} \frac{\Pr_P[\gamma]}{\Pr_P[\Upsilon]} Q[\gamma] . \tag{4.178}$$

Trajectory concatenation and other operations

For any trajectory $\gamma = (\gamma^{(0)}, \ldots, \gamma^{(s)}) \in \Gamma$, let $L[\gamma] = s$ be the length of the trajectory, and let $\gamma^{(L)}$ be shorthand for $\gamma^{(L[\gamma])}$. The statement $w \in \gamma$ returns true if there exists some t for which $\gamma^{(t)} = w$. Then, we let

$$S_w[\gamma] = \begin{cases} \min\left(\{t : \gamma^{(t)} = w\}\right) & \text{if } w \in \gamma \\ -1 & \text{if } w \notin \gamma \end{cases} \tag{4.179}$$

be the first time step along γ for which the trajectory reaches w. We also let

$$M[\gamma] = \max_{0\le t\le L[\gamma]} \gamma_t \tag{4.180}$$

$$m[\gamma] = \min_{0\le t\le L[\gamma]} \gamma_t \tag{4.181}$$

be the maximum and minimum Hamming weight the trajectory passes through.

We can naturally concatenate two trajectories γ_1 and γ_2 if $\gamma_1^{(L)} = \gamma_2^{(0)}$ to form a trajectory $\gamma = \gamma_1 \cdot \gamma_2$ of length $L[\gamma_1] + L[\gamma_2]$. We will say that $\gamma_A \subset \gamma_C$ if there exists some γ_B for which $\gamma_A \cdot \gamma_B = \gamma_C$.

For any trajectory γ, we let $\tilde{\gamma}$ be the flipped trajectory

$$\tilde{\gamma} = (n - \gamma^{(0)}, n - \gamma^{(1)}, \ldots, n - \gamma^{(s)}) . \tag{4.182}$$

In general, if v is an integer with $0 \leq v \leq n$, then let $\tilde{v} = \min(v, n - v)$.

Similarly, let $\bar{\gamma}$ return the reversed trajectory.

$$\bar{\gamma} = (\gamma^{(s)}, \ldots, \gamma^{(0)}) \,. \tag{4.183}$$

Moreover, let $\gamma^{[t]}$ return the trajectory γ truncated to length t, or simply return γ if $t \geq L[\gamma]$, i.e.

$$\gamma^{[t]} = \begin{cases} \gamma & \text{if } t \geq L[\gamma] \\ (\gamma^{(0)}, \ldots, \gamma^{(t)}) & \text{if } t < L[\gamma] \end{cases} . \tag{4.184}$$

Let $\gamma^{[L-t]}$ be shorthand for $\gamma^{[L[\gamma]-t]}$. More generally, let

$$\gamma^{[a,b]} = (\gamma^{(a)}, \gamma^{(a+1)}, \ldots, \gamma^{(b)}) \,. \tag{4.185}$$

Important subsets of Γ

We now define various subsets of Γ. Let

$$\Gamma_x = \{\gamma \in \Gamma : \gamma^{(0)} = x\} \tag{4.186}$$

be the subset of trajectories that begin at x, and let

$$\Gamma^w = \{\gamma \in \Gamma : S_w[\gamma] = L[\gamma]\} \tag{4.187}$$

be the set of trajectories that reach w for the first time and immediately terminate. We make the natural combination of these

$$\Gamma_x^w = \{\gamma \in \Gamma : \gamma_0 = x, S_w[\gamma] = L[\gamma]\} \,. \tag{4.188}$$

Of particular importance are sets where $w = 0$ or $w = n$, which include valid trajectories that terminate at one of the fixed points of the random walk. Define

$$\Gamma_x^* = \Gamma_x^0 \cup \Gamma_x^n \tag{4.189}$$

and note that $\Pr_{P_u}[\Gamma_x^*] = \Pr_{P_b}[\Gamma_x^*] = 1$, a statement that intuitively makes sense since walks will eventually reach either 0 or n with probability 1. Adding the superscript w to any set Υ restricts to walks for which $L[\gamma] = S_w[\gamma]$.

When any walk in Υ can be concatenated with any walk in Υ' we let

$$\Upsilon \cdot \Upsilon' = \{\gamma \cdot \gamma' : \gamma \in \Upsilon, \gamma' \in \Upsilon'\} \,. \tag{4.190}$$

Additionally, we let

$$\tilde{\Upsilon} = \{\tilde{\gamma} : \gamma \in \Upsilon\} \tag{4.191}$$

$$\bar{\Upsilon} = \{\bar{\gamma} : \gamma \in \Upsilon\} \,. \tag{4.192}$$

Reduced trajectories

We also introduce the concept of a *reduced* trajectory, which we sometimes refer to synonymously as a reduced walk, which is a valid trajectory for which $|\gamma^{(t)} - \gamma^{(t-1)}| = 1$ for all t; that is, the reduced walk never stands still. We let the set of reduced walks be Ψ, and let all sub and superscripts restrict Ψ in the same way they restricted Γ. For any $\gamma \in \Gamma$ we can associate a reduced walk $\psi \in \Psi$ by removing consecutive duplicates from γ. Under this definition, we let $R[\gamma] = \psi$. For any $\psi \in \Psi$, we let

$$\Gamma_\psi = \{\gamma \in \Gamma : R[\gamma] = \psi, R[\psi^{[L-1]}] \neq \psi\}, \tag{4.193}$$

where the second condition acts to include only trajectories γ whose final configuration appears only once (i.e. when the final configuration is removed, the reduced sequence changes).

Under dynamics by either the unbiased or biased walk, it is easy to calculate the probability associated with Γ_ψ:

$$\Pr_{P_u}[\Gamma_\psi] = \left(\frac{1}{2}\right)^{L[\psi]} \tag{4.194}$$

$$\Pr_{P_b}[\Gamma_\psi] = q^{\psi^{(0)} - \psi^{(L)}} \left(\frac{q}{q^2+1}\right)^{L[\psi]} . \tag{4.195}$$

Finally, define the following subsets of Ψ:

$$\Lambda_x = \Psi^{x-1}_{x|n-x+1} \cup \Psi^{n-x+1}_{x|x-1} \tag{4.196}$$

$$\Xi_w = \{\psi \in \Psi_w : m(\psi) \geq w, M(\psi) \leq n - w\}, \tag{4.197}$$

where the subset $\Psi^c_{a|b}$ is defined as follows:

$$\Psi^w_{x|z} = \begin{cases} \{\psi \in \Psi^w_x : M[\psi] < z\} & \text{if } w < x < z \\ \{\psi \in \Psi^w_x : m[\psi] > z\} & \text{if } z < x < w \\ \emptyset & \text{otherwise} \end{cases} . \tag{4.198}$$

In words, the set $\Psi^w_{x|z}$ includes reduced walks that begin at x and end at w without ever reaching z. Thus Λ_x is the set of reduced walks that start at x and end at $x-1$ without ever reaching $n-x+1$ or end at $n-x+1$ without ever reaching $x-1$. The set Ξ_w is the set of reduced walks of any finite length that start at w, but never reach either $w-1$ or $n-w+1$.

4.E.3 Upper bound proof

Theorem 4.2 (restated). *For the complete-graph architecture with circuit size s on n qudits with local Hilbert space dimension q*

$$Z \leq Z_H \left(1 + e^{-\frac{2a}{n}(s-s^*)}\right) , \tag{4.199}$$

as long as $s \geq s^*$*, where*

$$s^* = \frac{q^2+1}{2(q^2-1)} n \log(n) + O(n) \tag{4.200}$$

$$a = \frac{(q-1)^2}{2(q^2+1)} \,. \tag{4.201}$$

Proof. In this proof, we will be working with expressions for the collision probability Z. It will take several steps to manipulate the original expression into the form we need, so we will move back and forth between updating the expression and developing the tools needed to justify these updates.

We start by expressing

$$Z = \frac{1}{(q+1)^n} \sum_{x=0}^{n} \binom{n}{x} \mathop{\mathbb{E}}_{P_u} \left[\left(\frac{2q}{q^2+1} \right)^{L\left[R\left[\gamma^{[s]}\right]\right]} \,\middle|\, \gamma \in \Gamma_x^* \right] . \tag{4.202}$$

This is seen to be equivalent to Eq. (4.61) as follows. There are $\binom{n}{x}$ initial configurations with Hamming weight x, and generating a length-s trajectory beginning at x with the unbiased Markov chain is equivalent to randomly choosing a trajectory γ from Γ_x^*, which begins at x and ends at a fixed point (0 or n), with probability proportional to that assigned by the unbiased walk, and then truncating the walk to length s, denoted by $\gamma^{[s]}$. Then $R[\gamma^{[s]}]$ is the reduced trajectory, where consecutive duplicates are removed, and $L[R[\gamma^{[s]}]]$ is the length of that reduced trajectory, or in other words, the total number of bit flips that have occurred within the first s time steps.

Moving ahead, we observe that drawing γ from Γ_x^* is equivalent to first drawing a reduced trajectory ψ from Ψ_x^* and then drawing γ from Γ_ψ, so we can rewrite

$$Z = \frac{1}{(q+1)^n} \sum_{x=0}^{n} \binom{n}{x} \sum_{\psi \in \Psi_x^*} \Pr_{P_u}[\Gamma_\psi] \mathop{\mathbb{E}}_{P_u} \left[\left(\frac{2q}{q^2+1} \right)^{L\left[R\left[\gamma^{[s]}\right]\right]} \,\middle|\, \gamma \in \Gamma_\psi \right] . \tag{4.203}$$

Now, note the following general statement about any integer-valued random variable X such that $0 \leq X \leq M$. For any function f, we have

$$\mathbb{E}[f(X)] = \sum_{m=0}^{M} \Pr[X = m] f(m) = \sum_{m=0}^{M} \left(\Pr[X < m+1] - \Pr[X < m] \right) f(m) \tag{4.204}$$

$$= f(M) + \sum_{m=1}^{M} \Pr[X < m] \left(f(m-1) - f(m) \right) . \tag{4.205}$$

Taking $X = L[R[\gamma^{[s]}]]$ for γ drawn at random from Γ_ψ and $f(X) = (2q/(q^2+1))^X$, we find

$$\mathop{\mathbb{E}}_{P_u}\left[\left(\frac{2q}{q^2+1}\right)^{L[R[\gamma^{[s]}]]} \;\middle|\; \gamma \in \Gamma_\psi\right] \tag{4.206}$$

$$= \left(\frac{2q}{q^2+1}\right)^{L[\psi]} + \sum_{m=1}^{L[\psi]} \Pr_{P_u}\left[L[R[\gamma^{[s]}]] < m \;\middle|\; \gamma \in \Gamma_\psi\right] \left(\frac{2q}{q^2+1}\right)^{m-1} \frac{(q-1)^2}{q^2+1} \tag{4.207}$$

$$= \left(\frac{2q}{q^2+1}\right)^{L[\psi]} + \sum_{m=1}^{L[\psi]} \Pr_{P_u}\left[L[\gamma] > s \;\middle|\; \gamma \in \Gamma_{\psi^{[m]}}\right] \left(\frac{2q}{q^2+1}\right)^{m-1} \frac{(q-1)^2}{q^2+1}, \tag{4.208}$$

where the last line follows since the conditions $L[R[\gamma^{[s]}]] < m$ with $\gamma \in \Gamma_\psi$ and $L[\gamma] > s$ with $\gamma \in \Gamma_{\psi^{[m]}}$ both correspond to deciding if the configuration has changed at least m times within the first s steps.

The quantity

$$\frac{1}{(q+1)^n} \sum_{x=0}^{n} \binom{n}{x} \sum_{\psi \in \Psi_x^*} \Pr_{P_u}[\Gamma_\psi] \left(\frac{2q}{q^2+1}\right)^{L[\psi]} \tag{4.209}$$

is precisely equal to Z_H, as this represents the limit of infinite size where all trajectories terminate at one of the fixed points (see Appendix 4.B.6). Thus, also noting that $\Pr_{P_u}[\Gamma_\psi] = 2^{-L[\psi]}$, we have

$$\frac{(q+1)^n(2q)}{(q-1)^2}(Z - Z_H) \tag{4.210}$$

$$= \frac{2q}{q^2+1} \sum_{x=0}^{n} \binom{n}{x} \sum_{\psi \in \Psi_x^*} \Pr_{P_u}[\Gamma_\psi] \sum_{m=1}^{L[\psi]} \Pr_{P_u}\left[L[\gamma] > s \;\middle|\; \gamma \in \Gamma_{\psi^{[m]}}\right] \left(\frac{2q}{q^2+1}\right)^{m-1} \tag{4.211}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \sum_{\psi \in \Psi_x^*} \Pr_{P_u}[\Gamma_\psi] \sum_{m=1}^{L[\psi]} \Pr_{P_u}\left[L[\gamma] > s \;\middle|\; \gamma \in \Gamma_{\psi^{[m]}}\right] \left(\frac{2q}{q^2+1}\right)^{m} \tag{4.212}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \sum_{\psi \in \Psi_x^*} 2^{-L[\psi]} \sum_{m=1}^{L[\psi]} \Pr_{P_u}\left[L[\gamma] > s \;\middle|\; \gamma \in \Gamma_{\psi^{[m]}}\right] \left(\frac{2q}{q^2+1}\right)^{m}. \tag{4.213}$$

Now, in the first line below, by associating $\phi = \psi^{[m]}$, we reorder and regroup the sums: instead of summing over paths ψ that end at a fixed point and then all intermediate points $m = 1, \ldots, L[\psi]$ along the path, we first sum over all m, all ϕ (not necessarily ending at a fixed point) of length m, and then all ψ for which $\phi \subset \psi$ (recall this means that the first $L[\phi]$ entries in the trajectory

ψ are equal to ϕ). In the second line, we note that the sums over m and ϕ of length m is just a sum over all ϕ (of any length). In the third line, we note that the total probability of all the walks ψ for which $\phi \subset \psi$ is just $2^{-L[\phi]}$.

$$\frac{2q}{(q-1)^2}(q+1)^n(Z - Z_H) \tag{4.214}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \sum_{m=1}^{\infty} \sum_{\substack{\phi \in \Psi_x \\ L[\phi]=m}} \left(\sum_{\substack{\psi \in \Psi_x^* \\ \phi \subset \psi}} 2^{-L[\psi]} \right) \Pr_{P_u}\left[L[\gamma] > s \;\middle|\; \gamma \in \Gamma_\phi \right] \left(\frac{2q}{q^2+1} \right)^{L[\phi]} \tag{4.215}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \sum_{\phi \in \Psi_x} \left(\sum_{\substack{\psi \in \Psi_x^* \\ \phi \subset \psi}} 2^{-L[\psi]} \right) \Pr_{P_u}\left[L[\gamma] > s \;\middle|\; \gamma \in \Gamma_\phi \right] \left(\frac{2q}{q^2+1} \right)^{L[\phi]} \tag{4.216}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \sum_{\phi \in \Psi_x} \left(2^{-L[\phi]} \right) \Pr_{P_u}\left[L[\gamma] > s \;\middle|\; \gamma \in \Gamma_\phi \right] \left(\frac{2q}{q^2+1} \right)^{L[\phi]} \tag{4.217}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \sum_{\phi \in \Psi_x} \Pr_{P_u}\left[L[\gamma] > s \;\middle|\; \gamma \in \Gamma_\phi \right] \left(\frac{q}{q^2+1} \right)^{L[\phi]} . \tag{4.218}$$

Now we examine the final expression. The difference between Z and Z_H is a sum over Ψ_x, which includes all reduced paths ϕ that start at x and may or may not terminate at 0 or n. The statement $L[\gamma] > s$ is true if the number of time steps it takes to complete this reduced path is at least s, i.e. the probability that *the path does not finish within s time steps.* As a sanity check, when s becomes infinite, we expect this probability to become zero for any path as there would be enough time for any path to finish, and in this case $Z = Z_H$ as expected. This expression represents progress because we will be able to bound the probability of a certain path being completed using a Chernoff bound.

For any random variable X and for any constant $a > 0$,

$$\Pr[X > k] \leq \frac{\mathbb{E}[e^{aX}]}{e^{ak}} . \tag{4.219}$$

We use this bound with $X = L[\gamma]$, $k = s$, and yet-to-be-specified constants $a_\phi > 0$

$$\frac{2q(q+1)^n}{(q-1)^2}(Z - Z_H) \leq \sum_{x=0}^{n} \binom{n}{x} \sum_{\phi \in \Psi_x} e^{-a_\phi s} \mathop{\mathbb{E}}_{P_u}\left[e^{a_\phi L[\gamma]} \;\middle|\; \gamma \in \Gamma_\phi \right] \left(\frac{q}{q^2+1} \right)^{L[\phi]} . \tag{4.220}$$

The Chernoff bound has the additional benefit that $\mathbb{E}[e^{aX}]$ separates when X is the sum of independent random variables. In particular, once ϕ is fixed, $L[\gamma]$ is the sum of exponentially distributed random variables corresponding

to how many time steps the path γ waits at each position along the reduced path ϕ.

This is seen formally by noting that

$$\phi = (\phi^{(0)}, \phi^{(1)}) \cdot (\phi^{(1)}, \phi^{(2)}) \cdot \ldots \cdot (\phi^{(L-1)}, \phi^{(L)}) \tag{4.221}$$

$$\Gamma_\phi = \Gamma_{(\phi^{(0)},\phi^{(1)})} \cdot \Gamma_{(\phi^{(1)},\phi^{(2)})} \cdot \ldots \cdot \Gamma_{(\phi^{(L-1)},\phi^{(L)})}\,, \tag{4.222}$$

and meanwhile, for any collection of subsets Υ_m,

$$\mathop{\mathbb{E}}_{P_u}\left[e^{aL[\gamma]} \mid \gamma \in \Upsilon_1 \cdot \ldots \cdot \Upsilon_M\right] = \prod_{m=1}^{M} \mathop{\mathbb{E}}_{P_u}\left[e^{aL[\gamma]} \mid \gamma \in \Upsilon_m\right]. \tag{4.223}$$

For any $r = 0, \ldots, L[\phi] - 1$, we can evaluate

$$\mathop{\mathbb{E}}_{P_u}\left[e^{aL[\gamma]} \mid \gamma \in \Gamma_{(\phi^{(r)},\phi^{(r+1)})}\right] = \sum_{t=1}^{\infty} \left(1 - \lambda_{\phi^{(r)}}^{-1}\right)^{t-1} \lambda_{\phi^{(r)}}^{-1} e^{at} \tag{4.224}$$

$$= \frac{1}{1 - \lambda_{\phi^{(r)}}(1 - e^{-a})}\,, \tag{4.225}$$

where

$$\lambda_v = \frac{n(n-1)}{2v(n-v)} \tag{4.226}$$

is the expected amount of time the walk will wait at Hamming weight v before moving to $v+1$ or $v-1$, and hence

$$\mathop{\mathbb{E}}_{P_u}\left[e^{a_\phi L[\gamma]} \,\middle|\, \gamma \in \Gamma_\phi\right] = \prod_{r=0}^{L[\phi]-1} \frac{1}{1 - \lambda_{\phi^{(r)}}(1 - e^{-a_\phi})}. \tag{4.227}$$

We have made some progress at evaluating the bound on Z, but at this point it remains unclear how to perform the sum over paths $\phi \in \Psi_x$. To do so, first we will decompose paths ϕ into a series of subpaths that inch closer and closer to the fixed points at 0 and n. In particular, we will decompose a path ϕ as a concatenation of subpaths drawn from Λ_v for various v and one final subpath drawn from Ξ_w, as described in the following lemma. Recall from Eqs. (4.196) and (4.197) that these subsets of Ψ are defined by where they start, where they end, and/or some maximum or minimum point at which they ever reach.

Lemma 4.5. *Suppose $\phi \in \Psi_x$. Let $\tilde{x} = \min(x, n-x)$ and let $w = \min(m(\phi), n - M(\phi))$. Then there is a unique sequence of trajectories $(\phi_v)_{v=w}^{\tilde{x}}$ with $\phi_v \in \Lambda_v$ for $v = w+1, \ldots, \tilde{x}$ and $\phi_w \in \Xi_w$ and such that*

$$\phi = \alpha_{\tilde{x}} \cdot \alpha_{\tilde{x}-1} \cdot \ldots \cdot \alpha_w\,, \tag{4.228}$$

where for each v either $\alpha_v = \phi_v$ or $\alpha_v = \tilde{\phi}_v$, depending on whether α_{v+1} terminates at v or at $n - v$.

Proof. Let r_v be the minimum r such that $\phi^{(r)} = v$ or $\phi^{(r)} = n - v$. Then for each $v = w + 1, \ldots, \tilde{x}$, we can define

$$\alpha_v = \phi^{[r_v, r_{v-1}]} \tag{4.229}$$

and

$$\alpha_w = \phi^{[r_w, L[\phi]]} \,. \tag{4.230}$$

Then, each α_v begins at either v or $n - v$ and terminates upon reaching either $v - 1$ or $n - v + 1$ for the first time. Hence it is a member of Λ_v or $\tilde{\Lambda}_v$, but not both. Finally, α_w is a member of Ξ_w because it begins at w and never reaches either $w - 1$ or $n - w + 1$, since this would contradict the definition of w. □

We will use the notation $\tilde{v} = \min(v, n - v)$ for any integer v throughout the remainder of the proof. The above lemma allows us to replace the sum over $\phi \in \Psi_x$ with sums over w from 0 to $\tilde{x}$ and sums over $\phi_v \in \Lambda_v$, $\phi_w \in \Xi_w$. The summand is a product of factors $(1 - \lambda_{\phi^{(r)}}(1 - e^{-a_\phi}))^{-1}$, each of which can be collected within just one of the sums. Moreover, the fact that these products are invariant under reversing the path, i.e.

$$\prod_{r=0}^{L[\psi]-1} f(\psi^{(r)}) = \prod_{r=0}^{L[\psi]-1} f(\tilde{\psi}^{(r)}) \tag{4.231}$$

for any function f, means that it is unimportant that α_v can equal ϕ_v or $\tilde{\phi}_v$ as both yield the same result.

We will choose a_ϕ so that it only depends on $w = \min(m(\phi), n - M(\phi))$, denoted henceforth by a_w. Collecting these observations, and noting that the $L[\phi]$ factors of $q/(q^2 + 1)$ can each be allocated to one of the steps taken in ϕ, we find that

$$\frac{(q+1)^n (2q)}{(q-1)^2} \, (Z - Z_H) \tag{4.232}$$

$$\leq \sum_{x=0}^{n} \binom{n}{x} \sum_{w=0}^{\tilde{x}} e^{-a_w s} \prod_{v=w}^{\tilde{x}} \left(\sum_{\substack{\phi_v \in \Lambda_v \\ \text{or} \\ \phi_w \in \Xi_w}} \prod_{r=0}^{L[\phi_v]-1} \frac{q}{q^2+1} \frac{1}{1 - \lambda_{\phi_v^{(r)}}(1 - e^{-a_w})} \right) \tag{4.233}$$

$$\leq \sum_{x=0}^{n} \binom{n}{x} \sum_{w=0}^{\tilde{x}} e^{-a_w s} \prod_{v=w}^{\tilde{x}} \left(\sum_{\substack{\phi_v \in \Lambda_v \\ \text{or} \\ \phi_w \in \Xi_w}} \left(\frac{q}{q^2+1} \frac{1}{1 - \lambda_v(1 - e^{-a_w})} \right)^{L[\phi_v]} \right) , \tag{4.234}$$

where in the final line, we used the fact that by definition of $\phi_v \in \Lambda_v$ or $\phi_v \in \Xi_v$, $v \leq \phi_v^{(r)} \leq n - v$ for all $r < L[\phi]$ and also $\lambda_v \geq \lambda_{v'}$ whenever $\tilde{v} < \tilde{v}'$.

This form is very useful because we know how to perform sums in parentheses, using the strategy we first saw in Lemma 4.1. The values of these sums are given by the following lemma, whose proof is delayed until after the main proof.

Lemma 4.6. *Given v and a parameter a that satisfies $1 \le e^a \le (1 - \frac{(q-1)^2}{q^2+1}\frac{1}{\lambda_v})^{-1}$, the following identities hold:*

$$\sum_{\alpha\in\Lambda_v}\left(\frac{q}{q^2+1}\frac{1}{1-\lambda_v(1-e^{-a})}\right)^{L[\alpha]} = \bar{q}_{v,a}^{-1}\frac{1+\bar{q}_{v,a}^{-n+2v}}{1+\bar{q}_{v,a}^{-n+2v-2}} \tag{4.235}$$

$$\sum_{\alpha\in\Xi_v}\left(\frac{q}{q^2+1}\frac{1}{1-\lambda_v(1-e^{-a})}\right)^{L[\alpha]} = \frac{\bar{q}_{v,a}^2+1}{(\bar{q}_{v,a}-1)^2}\left(1-\bar{q}_{v,a}^{-1}\frac{1+\bar{q}_{v,a}^{-n+2v}}{1+\bar{q}_{v,a}^{-n+2v-2}}\right), \tag{4.236}$$

where $\bar{q}_{v,a}$ is defined in Definition 4.6.

Definition 4.6. *Given x and parameter a satisfying $1 \le e^a \le (1-\frac{(q-1)^2}{q^2+1}\frac{1}{\lambda_x})^{-1}$, let*

$$\bar{q}_{x,a} = \left(\frac{q^2+1}{2q}\right)\left(1-\lambda_x(1-e^{-a})\right)\left(1+\sqrt{1-\frac{4q^2}{(q^2+1)^2(1-\lambda_x(1-e^{-a}))^2}}\right). \tag{4.237}$$

Note that $\bar{q}_{x,a}$ satisfies the equation

$$\frac{\bar{q}_{x,a}}{\bar{q}_{x,a}^2+1} = \frac{q}{q^2+1}\frac{1}{1-\lambda_x(1-e^{-a})} \tag{4.238}$$

and that

$$\bar{q}_{x,a}^{-1} = \left(\frac{q^2+1}{2q}\right)\left(1-\lambda_x(1-e^{-a})\right)\left(1-\sqrt{1-\frac{4q^2}{(q^2+1)^2(1-\lambda_x(1-e^{-a}))^2}}\right). \tag{4.239}$$

This lemma allows us to define

$$V_{v,a} = \frac{\bar{q}_{v,a}^2+1}{(\bar{q}_{v,a}-1)^2}\left(1-\bar{q}_{v,a}^{-1}\frac{1+\bar{q}_{v,a}^{-n+2v}}{1+\bar{q}_{v,a}^{-n+2v-2}}\right), \tag{4.240}$$

and state

$$\frac{(q+1)^n(2q)}{(q-1)^2}(Z-Z_H) \tag{4.241}$$

$$\le \sum_{x=0}^{n}\binom{n}{x}\sum_{w=0}^{\tilde{x}}e^{-sa_w}\left(\prod_{v=w+1}^{\tilde{x}}\bar{q}_{v,a_w}^{-1}\frac{1+\bar{q}_{v,a_w}^{-n+2v}}{1+\bar{q}_{v,a_w}^{-n+2v-2}}\right)V_{w,a_w} \tag{4.242}$$

$$\le \sum_{x=0}^{n}\binom{n}{x}\sum_{w=0}^{\tilde{x}}e^{-sa_w}\left(\prod_{v=w+1}^{\tilde{x}}\tilde{q}_{v,a_w}^{-1}\right)e^{\frac{q^2}{q^2-1}}V_{w,a_w}, \tag{4.243}$$

where the second line follows from the observation that (also noting $\bar{q}_{v,a} < q$ for all v, a)

$$\prod_{v=w+1}^{\tilde{x}} \frac{1+\bar{q}_{v,a_w}^{-n+2v}}{1+\bar{q}_{v,a_w}^{-n+2v-2}} \leq \prod_{m=0}^{\infty} \frac{1+q^{-2m}}{1+q^{-2m-2}} = \prod_{m=0}^{\infty} \left(1+q^{-2m}\frac{1-q^{-2}}{1+q^{-2m-2}}\right) \tag{4.244}$$

$$\leq \prod_{m=0}^{\infty} \left(1+q^{-2m}\right) \leq \prod_{m=0}^{\infty} \exp\left(q^{-2m}\right) = \exp\left(\sum_{m=0}^{\infty} q^{-2m}\right) = \exp\left(\frac{q^2}{q^2-1}\right). \tag{4.245}$$

To continue, we will make choices for a_w and show upper bounds for the various factors in the above expression. For $w > 0$, we make the specification for a_w that

$$\eta_w = 1 - e^{-a_w} = \frac{(q-1)^2}{q^2+1}\frac{1}{2\lambda_w} = \frac{(q-1)^2}{q^2+1}\frac{w(n-w)}{n(n-1)} \tag{4.246}$$

and that $\eta_0 = \eta_1$. This choice implies

$$a_w \geq \frac{(q-1)^2}{q^2+1}\frac{1}{2\lambda_w} = \frac{(q-1)^2}{q^2+1}\frac{w(n-w)}{n(n-1)}. \tag{4.247}$$

Moreover, it implies that, so long as $w \leq x \leq n-w$,

$$1-\lambda_x(1-e^{-a_w}) \geq 1-\lambda_w(1-e^{-a_w}) = 1-\frac{(q-1)^2}{2(q^2+1)} = \frac{(q+1)^2}{2(q^2+1)}, \tag{4.248}$$

which by Definition 4.6 implies that

$$q \geq \bar{q}_{x,a_w} \geq \frac{(q+1)^2}{4q}. \tag{4.249}$$

When this is the case, we have

$$V_{w,a_w} \leq \frac{\bar{q}_{w,a_w}^2+1}{(\bar{q}_{w,a_w}-1)^2}\left(1-\bar{q}_{w,a_w}^{-1}\right) \tag{4.250}$$

$$= \frac{\bar{q}_{w,a_w}^2+1}{(\bar{q}_{w,a_w}-1)\bar{q}_{w,a_w}} \tag{4.251}$$

$$\leq \frac{q^2+1}{\left(\frac{(q+1)^2}{4q}-1\right)\frac{(q+1)^2}{4q}} \tag{4.252}$$

$$= \frac{(q^2+1)(4q)^2}{(q-1)^2(q+1)^2} \tag{4.253}$$

$$\leq \frac{320}{9}, \tag{4.254}$$

where the last line follows for all $q \geq 2$, which will be true for any physically realizable circuit. This takes care of the final factor in Eq. (4.243). What remains are the factors of $\bar{q}_{v,a_w}^{-1}$. To handle these we will use the following bound, whose proof is delayed to the next section.

Lemma 4.7. *With $\bar{q}_x$ defined as in Definition 4.6 and as long as $1 \leq e^a \leq (1 - \frac{(q-1)^2}{q^2+1}\frac{1}{\lambda_x})^{-1}$,*

$$\bar{q}_{x,a}^{-1} \leq q^{-1} \exp\left(a\left(\lambda_x \frac{q^2+1}{q^2-1} + \lambda_x^2(1-e^{-a})\frac{(q^2+1)^4}{(q^2-1)^3}\right)\right) . \tag{4.255}$$

We also need the following observation, which holds under the assumption that $1 \leq j < k \leq \frac{n}{2}$

$$\sum_{r=j+1}^{k} \lambda_r \leq \frac{n}{2} \sum_{r=j+1}^{k} \left(\frac{1}{r} + \frac{1}{n-r}\right) \leq \frac{n}{2}\left(\int_j^k d\rho \frac{1}{\rho} + \int_{n-k-1}^{n-j-1} d\rho \frac{1}{\rho}\right) \tag{4.256}$$

$$= \frac{n}{2}\left(\log(k/j) + \log\big((n-j-1)/(n-k-1)\big)\right) \tag{4.257}$$

$$\leq \frac{n}{2}\left(\log(k/j) + \log(2)\right) < \frac{n}{2}\left(\log(2k/j) + 1\right) . \tag{4.258}$$

$$\sum_{r=j+1}^{k} \lambda_r^2 \leq \frac{n^2}{4} \sum_{r=j+1}^{k} \frac{n^2}{r^2(n-r)^2} \leq n^2 \sum_{r=j+1}^{k} \frac{1}{r^2} \leq \frac{n^2}{j} < \frac{\pi^2 n^2}{6j} . \tag{4.259}$$

Similarly, for the case where $j = 0$, we find

$$\sum_{r=1}^{k} \lambda_r \leq \frac{n}{2}\big(\log(2k) + 1\big) , \qquad \sum_{r=1}^{k} \lambda_r^2 \leq \frac{\pi^2 n^2}{6} . \tag{4.260}$$

Now we can write the following, where $\bar{w} = \max(w, 1)$

$$\exp\left(-\frac{q^2}{q^2-1}\right) \frac{9(q+1)^n (2q)}{320(q-1)^2} (Z - Z_H) \tag{4.261}$$

$$\leq \sum_{x=0}^{n} \binom{n}{x} \sum_{w=0}^{\tilde{x}} e^{-s a_w} \left(\prod_{v=w+1}^{\tilde{x}} \bar{q}_{v,a_w}^{-1}\right) \tag{4.262}$$

$$\leq \sum_{x=0}^{n} \binom{n}{x} \sum_{w=0}^{\tilde{x}} e^{-s a_w} \left(\prod_{v=w+1}^{\tilde{x}} q^{-1} \exp\left(\frac{q^2+1}{q^2-1} a_w \lambda_v + \frac{(q^2+1)^4}{(q^2-1)^3} a_w \eta_w \lambda_v^2\right)\right) \tag{4.263}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \sum_{w=0}^{\tilde{x}} e^{-s a_w} q^{-\tilde{x}+w} \exp(Q) , \tag{4.264}$$

where

$$Q = \frac{q^2+1}{q^2-1} a_w \sum_{v=w+1}^{\tilde{x}} \lambda_v + \frac{(q^2+1)^4}{(q^2-1)^3} a_w \eta_w \sum_{v=w+1}^{\tilde{x}} \lambda_v^2 \tag{4.265}$$

$$\leq \frac{q^2+1}{q^2-1} a_w \frac{n}{2} \left(\log \frac{2\tilde{x}}{\bar{w}} + 1 \right) + \frac{(q^2+1)^4}{(q^2-1)^3} a_w \eta_w \frac{n^2\pi^2}{6\bar{w}} \tag{4.266}$$

$$\leq a_w \left(\frac{q^2+1}{q^2-1} \frac{n}{2} \left(\log \frac{2\tilde{x}}{\bar{w}} + 1 \right) + \frac{(q^2+1)^3(q-1)^2}{(q^2-1)^3} \frac{n\pi^2(n-w)}{6(n-1)} \right) . \tag{4.267}$$

Now we are in a position to nearly complete the proof. We choose

$$s^* = \frac{1}{2} \frac{q^2+1}{q^2-1} n \log(n) + cn \,, \tag{4.268}$$

where

$$c = Q_1 + Q_2 + Q_3 \,, \tag{4.269}$$

with

$$Q_1 = \frac{q^2+1}{2(q^2-1)} + \frac{(q^2+1)^3(q-1)^2}{(q^2-1)^3} \frac{\pi^2}{6} \tag{4.270}$$

$$Q_2 = \frac{q^2+1}{(q-1)^2} \left(\log \left(\frac{320(q-1)(q^n+1)}{9q^n} \right) + \frac{q^2}{q^2-1} \right) \tag{4.271}$$

$$Q_3 = \frac{4(q^2+1)\log(q)}{(q-1)^2} \,. \tag{4.272}$$

The $O(n\log(n))$ term in s^* will be necessary to cancel the $O(n\log(2\tilde{x}/w))$ term in Eq. (4.264). Meanwhile, Q_1 is necessary to cancel the remaining terms on the right-hand-side of Eq. (4.264), Q_2 is used to cancel factors on the left-hand-side of Eq. (4.264), and finally Q_3 is vital for canceling the q^w factor, as follows:

$$\frac{q^n+1}{q^n}(q+1)^n(Z - Z_H) \tag{4.273}$$

$$\leq \frac{q-1}{q} \sum_{x=0}^{n} \binom{n}{x} \sum_{w=0}^{\tilde{x}} e^{-(s-s^*)a_w} q^{-\tilde{x}+w} \exp\left(-\frac{4w(n-w)}{n-1} \log(q) \right) \tag{4.274}$$

$$\leq \frac{q-1}{q} \sum_{x=0}^{n} \binom{n}{x} \sum_{w=0}^{\tilde{x}} e^{-(s-s^*)a_w} q^{-\tilde{x}-w} \tag{4.275}$$

$$\leq e^{-(s-s^*)a_1} \frac{q-1}{q} \sum_{x=0}^{n} \binom{n}{x} q^{-\tilde{x}} \sum_{w=0}^{\tilde{x}} q^{-w} \tag{4.276}$$

$$\leq e^{-(s-s^*)a_1} \sum_{x=0}^{n} \binom{n}{x} q^{-\tilde{x}} \tag{4.277}$$

$$\leq e^{-(s-s^*)a_1} \sum_{x=0}^{n} \binom{n}{x} \left(q^{-x} + q^{-n+x}\right) \tag{4.278}$$

$$= 2e^{-(s-s^*)a_1} \left(\frac{q+1}{q}\right)^n . \tag{4.279}$$

It was in Eq. (4.276), where we used $a_1 \leq a_w$ to pull the exponential out from the sum, that the assumption $s \geq s^*$ was necessary. This is the only place it has been needed. As $Z_H = 2/(q^n + 1)$, we then have

$$Z \leq Z_H(1 + e^{-(s-s^*)a_1}) , \tag{4.280}$$

where $a_1 = (q-1)^2/(n(q^2+1))$. Defining $a = na_1/2$, this completes the proof of the upper bound. □

4.E.4 Lower bound proof

Theorem 4.6 (restated)**.** *For the complete-graph architecture of size s on n qudits with local dimension q, the collision probability satisfies*

$$Z \geq \frac{Z_H}{2} \exp\left(\frac{\log(q)}{q+1} \exp\left(\log(n) + s \log\left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right)\right)\right) . \tag{4.281}$$

Corollary 4.6. *For the complete-graph architecture, let s_{AC} be the minimum circuit size, as a function of n, such that*

$$Z \leq 2Z_H . \tag{4.282}$$

Then it must hold that

$$\left| s_{AC} - \frac{q^2+1}{2(q^2-1)} n \log(n) \right| = O(n) . \tag{4.283}$$

Proof. The upper bound on Z in Theorem 4.2 implies

$$s_{AC} \leq \frac{q^2+1}{2(q^2-1)} n \log(n) + O(n) . \tag{4.284}$$

Meanwhile, since $s = s_{AC}$ implies $Z \leq 2Z_H$, the bound in Theorem 4.6 implies that

$$\exp\left(\frac{\log(q)}{q+1} \exp\left(\log(n) + s_{AC} \log\left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right)\right)\right) \leq 4 \tag{4.285}$$

and thus

$$s_{AC} \geq \frac{\log(n) - \log\left(\frac{(q+1)\log(4)}{\log(q)}\right)}{-\log\left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right)} \tag{4.286}$$

$$\geq \left(\log(n) - \log\left(\frac{(q+1)\log(4)}{\log(q)}\right)\right)\left(\frac{n(q^2+1)}{2(q^2-1)} - 1\right) \tag{4.287}$$

$$= \frac{n(q^2+1)}{2(q^2-1)} \log(n) - O(n) , \tag{4.288}$$

where we have used the general inequality $-1/\log(1-u) \geq 1/u - 1$. □

Proof of Theorem 4.6. The structure of the proof is very similar to Theorem 4.4 for general architectures. We use the framework of the biased random walk.

Let $x\gamma^{(t)}$. The transition rule is such that

$$\gamma^{(t+1)} = \begin{cases} x & \text{with probability } 1 - \frac{2x(n-x)}{n(n-1)} \\ x-1 & \text{with probability } \frac{2x(n-x)}{n(n-1)} \frac{q^2}{q^2+1} \\ x+1 & \text{with probability } \frac{2x(n-x)}{n(n-1)} \frac{1}{q^2+1} \end{cases} \tag{4.289}$$

and so

$$\mathop{\mathbb{E}}_{P_b}[\gamma^{(t+1)}|\gamma^{(t)} = x] = x - \frac{2x(n-x)}{n(n-1)} \frac{q^2-1}{q^2+1} \tag{4.290}$$

$$\geq x\left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right). \tag{4.291}$$

As this is true for all x, when we have some probability distribution Λ over values of x, it still holds that

$$\mathop{\mathbb{E}}_{P_b}[\gamma^{(t+1)}] = \sum_{x=0}^{n} \Pr_{\Lambda}[\gamma^{(t)} = x] \mathop{\mathbb{E}}_{P_b}[\gamma^{(t+1)}|\gamma^{(t)} = x] \tag{4.292}$$

$$\geq \sum_{x=0}^{n} \Pr_{\Lambda}[\gamma^{(t)} = x] x \left(1 - \frac{2(q^2+1)}{n(q^2-1)}\right) \tag{4.293}$$

$$= \left(1 - \frac{2(q^2+1)}{n(q^2-1)}\right) \sum_{x=0}^{n} \Pr_{\Lambda}[\gamma^{(t)} = x] x \tag{4.294}$$

$$= \left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right) \mathop{\mathbb{E}}_{\Lambda}[\gamma^{(t)}], \tag{4.295}$$

and by applying this equation recursively from the starting distribution Λ_b, we find

$$\mathop{\mathbb{E}}_{P_b,\Lambda_b}[\gamma^{(s)}] \geq \left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right)^s \mathop{\mathbb{E}}_{\Lambda_b}[\gamma^{(0)}] \tag{4.296}$$

$$= \left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right)^s \frac{n}{q+1}. \tag{4.297}$$

By convexity, we have $\mathbb{E}[q^x] \geq q^{\mathbb{E}[x]}$, and hence

$$Z = \frac{1}{q^n} \mathop{\mathbb{E}}_{P_b,\Lambda_b}[q^{|\vec{\gamma}^{(s)}|}] \tag{4.298}$$

$$\geq \frac{1}{q^n} \exp\left(\log(q) \frac{n}{q+1} \left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right)^s\right) \tag{4.299}$$

$$\geq \frac{Z_H}{2} \exp\left(\frac{\log(q)}{q+1} \exp\left(\log(n) + s\log\left(1 - \frac{2(q^2-1)}{n(q^2+1)}\right)\right)\right). \tag{4.300}$$

□

4.E.5 Delayed proofs of lemmas

Proof of Lemma 4.6. The first equation will follow fairly straightforwardly from Lemma 4.1. Note that the factor in parentheses on the left-hand-side is $\bar{q}_{v,a}/(\bar{q}_{v,a}^2+1)$ as defined in Definition 4.6, as well as the fact that Λ_v contains walks that start at v and end at $v-1$ or $n-v+1$. The walks that start at v and end at $v-1$ are covered by Lemma 4.1 with $x \to v$, $y \to v-1$, and $m \to n-2v+2$. The walks that start at v and end at $n-v+1$ are equivalent to walks starting at $n-v$ and ending at $v-1$, and are thus covered by Lemma 4.1 with $x \to n-v$, $y \to v-1$, and $m \to n-2v+2$. Summing the results from these two substitutions yields the quantity

$$\begin{aligned}&\frac{1}{1-\bar{q}_{v,a}^{-2(n-2v+2)}}\left(\bar{q}_{v,a}^{-1}-\bar{q}_{v,a}^{-2n+4v-3}\right)+\frac{1}{1-\bar{q}_{v,a}^{-2(n-2v+2)}}\left(\bar{q}_{v,a}^{-n+2v-1}-\bar{q}_{v,a}^{-n+v-3}\right)\\&=\bar{q}_{v,a}^{-1}\frac{1+\bar{q}_{v,a}^{-n+2v}}{1+\bar{q}_{v,a}^{-n+2v-2}}\,,\end{aligned}\tag{4.301}$$

which proves that the first equation in the Lemma is correct.

The second equation is not a direct application of Lemma 4.1, but it can be shown by a similar method. Define

$$\Xi_{x,v}=\{\psi\in\Psi_x : v\le m(\psi), M(\psi)\le n-v\}\tag{4.302}$$

so $\Xi_v=\Xi_{v,v}$. Moreover, for fixed v, let

$$I(x)\sum_{\alpha\in\Xi_{x,v}}\left(\frac{q}{q^2+1}\frac{1}{1-\lambda_v(1-e^{-a})}\right)^{L[\alpha]}=\sum_{\alpha\in\Xi_{x,v}}\left(\frac{\bar{q}_{v,a}}{\bar{q}_{v,a}^2+1}\right)^{L[\alpha]}.\tag{4.303}$$

The function $I(x)$ obeys the recursion relation

$$I(x)=1+\frac{\bar{q}_{v,a}}{\bar{q}_{v,a}^2+1}\left(I(x-1)+I(x+1)\right)\,,\tag{4.304}$$

since there is one term in the sum corresponding to the length-0 trajectory, which contributes 1, but all other terms appear either in $I(x-1)$ or $I(x+1)$ reduced by factor $\bar{q}_{v,a}/(\bar{q}_{v,a}^2+1)$. The general solution to this recursion relation is

$$I(x)=\frac{\bar{q}_{v,a}^2+1}{(\bar{q}_{v,a}-1)^2}+A\bar{q}_{v,a}^{x}+B\bar{q}_{v,a}^{-x}\tag{4.305}$$

for some constants A and B. Here is where we rely on boundary conditions. We must have $I(v-1)=I(n-v+1)=0$ since these sums do not include

any terms. This allows us to solve for A and B and find

$$A = -\frac{\bar{q}_{v,a}^2 + 1}{(\bar{q}_{v,a} - 1)^2} \frac{\bar{q}_{v,a}^{-n+v-1}}{1 + \bar{q}_{v,a}^{-n+2v-2}} \tag{4.306}$$

$$B = -\frac{\bar{q}_{v,a}^2 + 1}{(\bar{q}_{v,a} - 1)^2} \frac{\bar{q}_{v,a}^{v-1}}{1 + \bar{q}_{v,a}^{-n+2v-2}} \tag{4.307}$$

$$I(v) = \frac{\bar{q}_{v,a}^2 + 1}{(\bar{q}_{v,a} - 1)^2} \left(1 - \bar{q}_{v,a}^{-1} \frac{1 + \bar{q}_{v,a}^{-n+2v}}{1 + \bar{q}_{v,a}^{-n+2v-2}}\right) . \tag{4.308}$$

□

Proof of Lemma 4.7. Define $\eta = \lambda_x(1 - e^{-a})$ and let $\zeta = 1 - (1 - \eta)^2$. Thus, $1 - \eta = \sqrt{1 - \zeta}$.

We can write

$$\bar{q}_{x,a}^{-1} = q^{-1} \left(\frac{q^2 + 1}{2}\right) \left(\sqrt{1 - \zeta}\right) \left(1 - \sqrt{1 - \frac{4q^2}{(q^2 + 1)^2(1 - \zeta)}}\right) \tag{4.309}$$

$$= q^{-1} \left(\frac{q^2 + 1}{2}\right) \left(\sqrt{1 - \zeta} - \frac{q^2 - 1}{q^2 + 1} \sqrt{1 - \frac{(q^2 + 1)^2}{(q^2 - 1)^2} \zeta}\right) \tag{4.310}$$

$$\leq q^{-1} \left(\frac{q^2 + 1}{2}\right) \left(1 - \frac{1}{2}\zeta - \frac{q^2 - 1}{q^2 + 1} \left(1 - \frac{(q^2 + 1)^2}{2(q^2 - 1)^2} \zeta - \frac{(q^2 + 1)^4}{2(q^2 - 1)^4} \zeta^2\right)\right) \tag{4.311}$$

$$= q^{-1} \left(1 + \frac{1}{2} \frac{q^2 + 1}{q^2 - 1} \zeta + \frac{(q^2 + 1)^4}{4(q^2 - 1)^3} \zeta^2\right) \tag{4.312}$$

$$\leq q^{-1} \exp\left(\frac{1}{2} \frac{q^2 + 1}{q^2 - 1} \zeta + \frac{(q^2 + 1)^4}{4(q^2 - 1)^3} \zeta^2\right) \tag{4.313}$$

$$\leq q^{-1} \exp\left(\frac{q^2 + 1}{q^2 - 1} a\lambda_x + \frac{(q^2 + 1)^4}{(q^2 - 1)^3} a\eta\lambda_x\right) , \tag{4.314}$$

which is equal to the lemma statement, where in the first inequality we utilized $1 - \frac{u}{2} - \frac{u^2}{2} \leq \sqrt{1 - u} \leq 1 - \frac{u}{2}$, in the second inequality we used $1 + u \leq \exp(u)$, and in the third inequality we used $\zeta \leq 2\eta \leq 2a\lambda_x$. The condition on a is necessary to ensure that $\bar{q}_x$ is real. □

4.F Approximate 2-designs and anti-concentration

In this appendix, we clarify the relation between approximate unitary 2-designs and anti-concentration. As we discussed in the text, forming a unitary 2-design is a sufficient condition for anti-concentration.

First, we recall some definitions. The k-fold channel of an operator $\mathcal{O}$ with respect to a probability distribution μ on the unitary group $\mathcal{U}(q^n)$ is defined as

$$\Phi_\mu^{(k)}(\mathcal{O}) \int d\mu(U)\, U^{\otimes k}(\mathcal{O}) U^{\dagger \otimes k} . \tag{4.315}$$

We denote the channel with respect to the Haar measure on the unitary group as $\Phi_H^{(k)}$. The diamond norm of a quantum channel Φ is defined as $\|\Phi\|_\diamond \sup_{\psi,D} \|\Phi\otimes\mathcal{I}_D(\psi)\|_1$, where $\mathcal{I}_D$ is the identity channel on a D-dimensional ancilla and ψ is a state on the entire system.

Definition 4.7 (Approximate designs)**.** *A probability distribution μ on $\mathcal{U}(q^n)$ is an ε-approximate unitary k-design if the k-fold channels obey*

$$\left\|\Phi_\mu^{(k)} - \Phi_H^{(k)}\right\|_\diamond \leq \varepsilon\,. \tag{4.316}$$

For a given k, if $\varepsilon = 0$, we say that the distribution forms an exact k-design.

A weaker notion of approximate design involves the operator norm of the moment operators, sometimes referred to as the tensor product expander (TPE) condition. The vectorization isomorphism uniquely maps channels to operators, with which we can define the kth moment operator from the k-fold channel for a probability distribution μ on the unitary group $\mathcal{U}(q^n)$ as

$$\widehat{\Phi}_\mu^{(k)}\mathrm{vec}\big(\Phi_\mu^{(k)}\big) = \int d\mu(U)\, U^{\otimes k}\otimes U^{*\otimes k}\,. \tag{4.317}$$

For convenience we denote $U^{\otimes k}\otimes U^{*\otimes k}$ by $U^{\otimes k,k}$.

Definition 4.8 (Weak approximate designs)**.** *A probability distribution μ on $\mathcal{U}(q^n)$ is a weak ε-approximate unitary k-design if the kth moment operators obey*

$$\left\|\widehat{\Phi}_\mu^{(k)} - \widehat{\Phi}_H^{(k)}\right\|_\infty \leq \varepsilon\,. \tag{4.318}$$

The expectation of the collision probability for completely Haar-random unitaries is $Z_H = \mathbb{E}_H[Z] = 2/(q^n+1) \leq 2/q^n$, and thus anti-concentrates with $\alpha = 1/2$ as defined in Definition 4.4. But as the collision probability is a second moment quantity, where $p_U(x)^2 = |\langle x|U|0^n\rangle|^4$, for an exact unitary 2-design μ, we find that

$$Z = \mathop{\mathbb{E}}_{\mu}\left[\sum_x p_U(x)^2\right] = \mathop{\mathbb{E}}_{H}\left[\sum_x p_U(x)^2\right] = \frac{2}{q^n+1} \tag{4.319}$$

and thus also 1/2-anti-concentrates, where $\mathbb{E}_H[\cdot]$ denotes the expectation with respect to the Haar measure on the unitary group.

Proposition 4.1. *An ε-approximate 2-design μ with $\varepsilon = 1/q^{2n}$ has a collision probability of $Z = \mathbb{E}_\mu[\sum_x p_U(x)^2] \leq 3/q^n$ and is thus a 1/3-anti-concentrator. Moreover, the same holds for a weak ε-approximate 2-design (TPE) μ with $\varepsilon = 1/q^{2n}$.*

Proof. For an ε-approximate 2-design in diamond norm, we find

$$\mathop{\mathbb{E}}_{\mu}\left[\sum_x p_U(x)^2\right] \tag{4.320}$$

$$= q^n \mathop{\mathbb{E}}_{\mu}\left[|\langle x|U|0^n\rangle|^4\right] - \mathop{\mathbb{E}}_{H}\left[|\langle x|U|0^n\rangle|^4\right] + \mathop{\mathbb{E}}_{H}\left[|\langle x|U|0^n\rangle|^4\right] \tag{4.321}$$

$$= q^n \operatorname{tr}\left(|x\rangle\langle x|^{\otimes 2}\Big(\mathop{\mathbb{E}}_{\mu}\left[U^{\otimes 2}(|0^n\rangle\langle 0^n|)U^{\dagger\otimes 2}\right] - \mathop{\mathbb{E}}_{H}\left[U^{\otimes 2}(|0^n\rangle\langle 0^n|)U^{\dagger\otimes 2}\right]\Big)\right) + \frac{2}{q^n+1} \tag{4.322}$$

$$\leq q^n \Big\| |x\rangle\langle x|^{\otimes 2}\big(\Phi_{\mu}^{(2)}(|0^n\rangle\langle 0^n|) - \Phi_{H}^{(2)}(|0^n\rangle\langle 0^n|)\big)\Big\|_1 + \frac{2}{q^n+1} \tag{4.323}$$

$$\leq q^n \big\| |x\rangle\langle x|^{\otimes 2}\big\|_{\infty} \big\|(\Phi_{\mu}^{(2)} - \Phi_{H}^{(2)})(|0^n\rangle\langle 0^n|)\big\|_1 + \frac{2}{q^n(q^n+1)} \tag{4.324}$$

$$\leq \frac{2}{(q^n+1)} + q^n\varepsilon \leq \frac{3}{q^n}\,, \tag{4.325}$$

where we wrote the difference in terms of the 2-fold channels, in the second to last line used Hölder's inequality, and in the last line used the definition of the diamond norm and the definition of an ε-approximate 2-design.

Given the definition of an approximate design in terms of the diamond norm, we must take the error to be exponentially small. Thus, for an approximate 2-design μ with $\varepsilon = 1/q^{2n}$, the collision probability is $Z \leq 3/q^n$ and thus $1/q^{2n}$-approximate unitary 2-designs in diamond norm anti-concentrate with $\alpha = 1/3$.

For a weak ε-approximate 2-design in operator norm (TPE), we proceed similarly,

$$\mathop{\mathbb{E}}_{\mu}\left[\sum_x p_U(x)^2\right] = q^n\left(\mathop{\mathbb{E}}_{\mu}\left[|\langle x|U|0^n\rangle|^4\right] - \mathop{\mathbb{E}}_{H}\left[|\langle x|U|0^n\rangle|^4\right] + \mathop{\mathbb{E}}_{H}\left[|\langle x|U|0^n\rangle|^4\right]\right) \tag{4.326}$$

$$= q^n \operatorname{tr}\left(|0^n x\rangle\langle 0^n x|^{\otimes 2}\Big(\mathop{\mathbb{E}}_{\mu}\left[U^{\otimes 2,2}\right] - \mathop{\mathbb{E}}_{H}\left[U^{\otimes 2,2}\right]\Big)\right) + \frac{2}{q^n+1} \tag{4.327}$$

$$\leq q^n \Big\|\widehat{\Phi}_{\mu}^{(2)} - \widehat{\Phi}_{H}^{(2)}\Big\|_{\infty} + \frac{2}{q^n+1} \tag{4.328}$$

$$\leq \frac{2}{q^n+1} + q^n\varepsilon\,, \tag{4.329}$$

where we wrote the difference in terms of the 2-fold moment operators, in the second to last line used Hölder's inequality, and in the last line used the definition of a weak ε-approximate 2-design. Again, we must take the error to be exponentially small. For $\varepsilon = 1/q^{2n}$, the collision probability is $Z \leq 3/q^n$ and thus $1/q^{2n}$-approximate unitary 2-designs in operator norm anti-concentrate with $\alpha = 1/3$. □

As n-qudit RQCs on the 1D architecture are known to form ε-approximate unitary 2-designs in $O(n + \log(1/\varepsilon))$ depth [133, 146], anti-concentration for 1D random circuits in linear depth is an immediate corollary. Moreover, an n-independent upper bound on the spectral gap for the 1D architecture [133], implies that they form weak approximate 2-designs in $O(\log(1/\varepsilon))$ depth. By Proposition 4.1, where we must take $\varepsilon = 1/q^{2n}$, this again requires linear depth for 1D RQCs.

For non-local RQCs on the complete-graph architecture, the best known upper bounds on the 2-design circuit size are $O(n^2)$ [131]. However, it has been conjectured that this can be improved to $O(n \log(n))$, in which case anti-concentration and 2-designs could occur at the same depth for the complete-graph circuit architecture.

To argue that anti-concentration must be distinct from the 2-design property, we consider lower bounds on the 2-design depth for RQCs on the 1D architecture. The spectral gap of the second moment of a probability distribution ν on the unitary group is defined as $g(\nu)\|\widehat{\Phi}_{\nu}^{(2)} - \widehat{\Phi}_{H}^{(2)}\|_{\infty}$. Ref. [133] proved an n-independent bound on the spectral gap for 1D RQCs. This implies that the behavior of the spectral gap for 1D RQCs of depth d must be $g(\nu_{\,1\mathrm{D\,RQC}}) = (1 - 1/c)^d$, for some constant $c > 1$. Further recalling that the operator norm can be written as $\|M\|_{\infty} = \max_y \langle y|M|y\rangle$, this implies that some states requires linear depth in order to become small. Specifically, there is some state $|y\rangle$ on the 4-fold space which requires the 1D circuit depth to be at least $d = \Omega(n)$ in order for the second moment operator for 1D RQCs $\mathbb{E}_U \left[\langle y|U^{\otimes 2,2}|y\rangle\right]$ to approach the minimal Haar value.

Chapter 5

APPROXIMATION OF NOISY RANDOM QUANTUM CIRCUITS AS IDEAL CIRCUITS WITH WHITE NOISE

This chapter is based on joint work with Nicholas Hunter-Jones and Fernando Brandão. It was adapted into a standalone article [147] that appeared publicly shortly after the defense of this thesis.

5.1 Motivation

There is a fundamental tradeoff in quantum computation between computation size and error rate. Naturally, the longer the computation, the lower the physical error rate must be to maintain a high probability of an errorless computation. Once the error rate is beneath a constant threshold, the theory of fault tolerance and quantum error correction [24, 96] may be employed to push the probability of a *logical* error arbitrarily close to zero, despite the prevalence of many physical errors during the computation; however, error correction comes at the cost of additional qubits and gates. These overheads, while acceptable in an asymptotic sense, are likely to be overwhelming in the near and intermediate term. This inspires the idea of an upcoming Noisy Intermediate-Scale Quantum (NISQ) era [3], where hardware capabilities are good enough to perform non-trivial quantum tasks on dozens or hundreds of qubits, but quantum error correction, which might require thousands or millions of qubits, remains beyond reach.

In this chapter, we study a model of NISQ devices performing random computations and prove a precise sense in which, for typical circuit instances, local errors are quickly scrambled and can be treated as white noise. For some applications, this phenomenon makes it possible for the signal of the noiseless computation to be extracted by repetition despite a large overall chance that at least one error occurs.

Our local error model assumes that each two-qubit gate in the quantum circuit is followed by a pair of gate-independent single-qubit unital noise channels acting on the two qubits involved in the gate. For simplicity and ease of analysis, we assume that each of these noise channels is identical, but we fully expect the takeaways from our work to apply when the noise strength is allowed to vary from location to location. For concreteness, we can consider the depolarizing channel with error probability ϵ in this introduction. In this case, the fidelity of the noisy computation with respect to the ideal computation is expected to be roughly equal to the probability that no errors occur. We see that, for a circuit with s two-qubit gates, this quantity, denoted here

by $F = (1-\epsilon)^{2s}$, is close to 1 only if the quantity $2\epsilon s$—the average number of errors—satisfies $2\epsilon s \ll 1$.

However, this high-fidelity requirement is quite restrictive in practice. Already for circuits with 50 qubits at depth 20, the error rate ϵ must be on the order of 10^{-4} for the whole computation to run without error at least 90% of the time. This error rate is more than an order of magnitude smaller than what has been achievable in recent experiments on superconducting qubit systems of that size [6, 7, 148]. Indeed, in their landmark 2019 quantum computational supremacy experiment [6], a group at Google performed random circuits on 53 qubits of depth 20, but the fidelity of the computation was $F \approx 0.002$, meaning at least one error occurs in all but a tiny fraction of the trials. Similar experiments at the University of Science and Technology of China on 56 [7] and 60 [148] qubits reported even smaller fidelities of 0.0007 [7] and 0.0004 [148]. This would not be an issue if one could determine when a trial is errorless. (In this case, one could just repeat the experiment $1/F$ times.) However, error-detection requires overheads similar to error-correction.

Rather, low-fidelity random circuit sampling experiments and their claim of quantum computational supremacy benefit from a key assumption [5, 6]: when at least one error does occur, the output of the experiment is well approximated by *white noise*, that is, the output is random and uncorrelated with the ideal (noiseless) output. When this is the case, the signal of diminished size F can, at least for some applications, be extracted from the white noise using $O(1/F^2)$ trials, as we explain later. Specifically, for quantum computational supremacy, the *white-noise assumption* is that the distribution p_{noisy} over measurement outcomes of their noisy device is close to what we call the "white-noise distribution"

$$p_{\text{wn}} = F p_{\text{ideal}} + (1-F) p_{\text{unif}}\,, \tag{5.1}$$

with p_{ideal} the ideal distribution and p_{unif} the uniform[1] distribution. In particular, for the approximation to be non-trivial, we demand that the total variation distance between p_{noisy} and p_{wn} be a small fraction of F, that is

$$\frac{1}{2}\|p_{\text{wn}} - p_{\text{noisy}}\|_1 \ll F\,. \qquad \text{(white-noise assumption)} \tag{5.2}$$

This demand is necessary because we expect that p_{noisy} also decays toward p_{unif} such that $\frac{1}{2}\|p_{\text{noisy}} - p_{\text{unif}}\|_1 = \Theta(F)$, and thus p_{unif} is a trivial approximation for p_{noisy} with error $\Theta(F)$.

[1] In Google's experiment, there was biased noise during readout (they measure $|0\rangle$ more often than $|1\rangle$) that would lead the appropriate definition of white noise to be slightly non-uniform (see Supplementary Material of [6]). We believe most of our analysis could be straightforwardly generalized to account for this kind of end-of-circuit non-unital error (although mid-circuit non-unital errors would likely complicate our method). However, the goal of our work is to study the complexity and behavior of low-fidelity random circuit experiments in an idealized sense, rather than the actual implementation of such ideas in recent superconducting experiments specifically.

Prior to their experiment, the Google group provided numerical evidence [5] in favor of the white-noise assumption[2] for randomly chosen circuits by showing that the output distribution of random circuits of depth 40 on 20 qubits (arranged in a 2D lattice) subject to a local Pauli error model approaches the uniform distribution, and that the fidelity of p_{noisy} with respect to p_{ideal} appears to decay exponentially, consistent with $p_{\text{noisy}} \approx p_{\text{wn}}$. However, their analysis did not specifically estimate the distance[3] between p_{noisy} and p_{wn}. The white-noise condition in Eq. (5.2) requires that the distance between p_{noisy} and p_{wn} decrease as the expected number of errors increases and F decays, so quantifying the differences between the distributions is vital for determining how well the white-noise approximation is obeyed.

In this chapter, we prove rigorous bounds on the error in the white-noise approximation. Our work applies for the complete-graph architecture, where gates act between random pairs of qudits, as well as architectures made of complete layers of $n/2$ parallel gates. We show that, for Pauli noise channels, the approximation remains good as long as (1) $\epsilon^2 s \ll 1$, (2) the ideal output distribution has the anti-concentration property, and (3) $\epsilon \ll 1/(n \log(n))$. We believe that condition (3) could be relaxed to read $\epsilon < c/n$ for some universal constant $c = O(1)$. Condition (1) is a quadratic improvement over the condition $\epsilon s \ll 1$ needed for high fidelity. For circuits with $\epsilon < 0.01$, as is the case with Google's hardware, thousands of gates could potentially be implemented before condition (1) fails. Note that our technical statements hold for non-Pauli error channels as well, but the results mentioned above only follow for incoherent noise channels. Our method can be applied to analyze coherent noise but, consistent with expectations, it suggests that the white-noise approximation does not hold in that case.

By putting the white-noise approximation for random quantum circuits on stronger theoretical footing, our work has several applications. First, the white-noise assumption is an ingredient in formal complexity-theoretic arguments that the task accomplished on noisy devices running random quantum circuits is hard for classical computers (allowing the declaration of quantum computational supremacy) [6]. We complement our main result by showing in Section 5.C that classically sampling from the white-noise distribution within total variation distance ηF is, in a certain complexity-theoretic sense, equivalent up to a factor of F (which is optimal) to sampling from the ideal output distribution within total variation distance $O(\eta)$. This makes low-fidelity ex-

[2]Note that Ref. [5] proposed the stronger ansatz that the output quantum state is a combination of the ideal output state and the maximally mixed state, which implies (but is not necessary for) the statement $p_{\text{noisy}} \approx p_{\text{wn}}$ about classical probability distributions over measurement outcomes.

[3]Ref. [5] did not specifically formulate the assumption as in Eq. (5.2), where we demand that the allowed approximation error decrease with the fidelity, but we argue that the approximation is only meaningful when this is true. For example, in Appendix 5.C we argue that such precision is necessary to make a stronger complexity-theoretic argument for quantum computational supremacy.

periments where errors are common nearly as defensible for quantum computational supremacy as high-fidelity experiments where errors are rare, at least in principle. Second, our result lends theoretical justification to the usage [6, 7, 148] of the linear cross-entropy metric proposed in Ref. [6] to benchmark noise in random circuit experiments and verify that hardware has correctly performed the quantum computational supremacy task. Indeed, as a side result, we show that, for both incoherent *and* coherent noise, the metric decays precisely as $e^{-2s\epsilon \pm O(s\epsilon^2)}$ when ϵ is sufficiently small; this also suggests that the linear cross entropy benchmark can be used to accurately estimate the underlying noise rate ϵ [50].

Beyond Google's random circuit experiment, our work suggests that other scenarios where the white-noise assumption holds may be advantageous in the NISQ era, as one can eschew error-correction and nonetheless perform a fairly long quantum computation, as long as one is willing to repeat the experiment $1/F^2$ times. One example of a scenario where the assumption may hold is quantum simulation of fixed chaotic Hamiltonians, since they are also believed to be efficient at scrambling errors.

The structure of this chapter is as follows: in Section 5.2, we describe our setup and in particular our model for local noise within a random quantum circuit; in Section 5.3, we precisely state our results; in Section 5.4, we discuss further implications and how our results fit in with prior work; in Section 5.5, we give an overview of the intuition behind our result and the method we use in our proofs, which is based on a map from random quantum circuits to certain stochastic processes, which can also be interpreted as partition functions of statistical mechanical systems. We conclude the main text with an outlook in Section 5.6. The rigorous proofs and details behind the map to stochastic processes then appear in the appendix sections.

5.2 Noise model and random quantum circuits

Here we describe our model of noisy random quantum circuits. Let the circuit act on n qudits, each with local Hilbert space dimension q. As in previous chapters, we define a random quantum circuit architecture as an instruction set for how to draw a quantum circuit diagram given the number of qudits n and the circuit size s. In Chapter 4, we explicitly considered the 1D architecture, where the qudits are arranged in a ring and $2s/n$ alternating layers of $n/2$ nearest-neighbor gates are applied. We also considered the complete-graph architecture, where the pair of qudits acted upon by each gate is chosen independently and uniformly at random among all possible pairs (thus, the circuit diagram itself is random). In this chapter, our results work both for the complete-graph architecture and for any architecture that can be decomposed into layers of $n/2$ parallel gates that also have the *regularly connected* property that we defined in Definition 4.5 of Chapter 4. Briefly, an architecture is h-regularly connected if, whenever we partition the qudits into two sets A and the complement of A, the circuit diagram will typically include

a gate coupling a qudit from A and a qudit from the complement of A at least once every hn gates. This is a natural condition that allows us to categorically exclude architectures that are designed to prevent scrambling.

Given an architecture and parameters n and s, we can generate a circuit instance by choosing the circuit diagram according to the instruction set and then choosing each of the gates in the circuit at random according to the Haar measure. Each instance is associated with an output probability distribution p_{ideal} over q^n possible computational basis measurement outcomes $x \in [q]^n$, (where $[q] = \{0, 1, \ldots, q-1\}$) that would be sampled if the circuit were implemented noiselessly. Note that in the formal analysis, we include a layer of n (also Haar-random) single-qudit gates at the beginning and end of the circuit without counting these $2n$ gates toward the circuit size; these might be regarded as fixing the local basis for the input product state and the measurement of the output.

5.2.1 Including local noise

In this chapter, we augment this setup by inserting single-qudit noise channels into the circuit diagram, which act on qudits involved in a multi-qudit gate immediately following the gate, as shown in the example in Figure 5.1. Thus, the core assumption is that the noise is local, i.e. independent from qudit

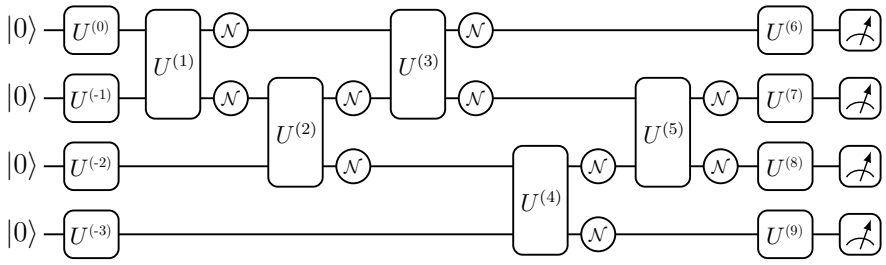


Figure 5.1: Example of a noisy quantum circuit diagram on $n = 4$ qudits with $s = 5$ two-qudit gates. A pair of single-qudit noise channels $\mathcal{N}$ follow each two-qudit gate. The circuit begins and ends with a layer of noiseless single-qudit gates.

to qudit. We assume that each noise channel $\mathcal{N}$ is a unital and completely positive trace-preserving map.

For a given noise channel, there are only two parameters that matter for our analysis, the *average infidelity* and the *unitarity* of the channel. The average infidelity for a channel $\mathcal{N}$ is defined as

$$r = 1 - \int dV \, \mathrm{tr}\left[V|\psi\rangle\langle\psi|V^\dagger \mathcal{N}(V|\psi\rangle\langle\psi|V^\dagger)\right] , \tag{5.3}$$

where the integral is over the Haar-measure on $q \times q$ unitary matrices V and $|\psi\rangle\langle\psi|$ is any pure state. The average infidelity is one measure of the overall

noise strength of the channel $\mathcal{N}$. Following Refs. [149, 150], the *unitarity* is defined for unital channels as

$$u = \frac{q}{q-1}\left(\int dV \,\mathrm{tr}\left[\mathcal{N}\left(V|\psi\rangle\langle\psi|V^\dagger\right)^2\right] - \frac{1}{q}\right). \tag{5.4}$$

The unitarity is the expected purity of the output state under random choice of input state, scaled to have minimum value of 0 and maximum value of 1.

Examples: depolarizing, dephasing, and rotation channels

It is instructive to consider explicitly the following three channels. First, the depolarizing channel

$$\mathcal{N}_{depo}(\rho) = (1-\gamma)\rho + \gamma\frac{I}{q} = (1-\epsilon)\rho + \frac{\epsilon}{q^2-1}\sum_{i=1}^{q^2-1} P_i\rho P_i^\dagger\,, \tag{5.5}$$

where $\gamma = \epsilon q^2/(q^2-1)$, $\{P_i\}_{i=1}^{q^2-1}$ is the set of single-qudit Pauli matrices (appropriately generalized to higher q), and I is the $q \times q$ identity matrix. There are two ways to think of the channel: first, with probability $1-\gamma$ doing nothing and with probability γ resetting the state to the maximally mixed state on that qudit; second, with probability $1-\epsilon$ doing nothing and with probability ϵ choosing a Pauli operator at random to apply to the qudit.

We can also consider the dephasing channel

$$\mathcal{N}_{deph}(\rho) = (1 - \frac{q}{q-1}\epsilon)\rho + \frac{q}{q-1}\epsilon\sum_{i=0}^{q-1}|i\rangle\langle i|\rho|i\rangle\langle i|\,, \tag{5.6}$$

which represents doing nothing with probability $1-q\epsilon/(q-1)$ and performing a measurement in the computational basis with probability $q\epsilon/(q-1)$.

Finally, we can consider a coherent noise channel, for example the rotation channel

$$\mathcal{N}_{rot}(\rho) = e^{-i\theta|0\rangle\langle 0|}\rho e^{i\theta|0\rangle\langle 0|}\,. \tag{5.7}$$

The average infidelity and unitary of these channels are given in Table 5.1.

5.2.2 Output distributions of the quantum circuit

Suppose the locations of the s two-qudit gates have been fixed, with gate t acting on qudits $\{i_t, j_t\}$. Then a circuit instance is specified by a sequence $(U^{(-n+1)}, \ldots, U^{(s+n)})$, where $U^{(t)}$ is a $q^2 \times q^2$ (two-qudit) unitary matrix if $1 \le t \le s$ and a $q \times q$ (single-qudit) unitary matrix otherwise. Accordingly, for each t, let

$$\mathcal{U}^{(t)}(\sigma) = \left(I_{[n]\setminus\{i_t,j_t\}} \otimes U^{(t)}_{\{i_t,j_t\}}\right)\sigma\left(I_{[n]\setminus\{i_t,j_t\}} \otimes U^{\dagger(t)}_{\{i_t,j_t\}}\right) \tag{5.8}$$

channel	**ave. infidelity** r	**unitarity** u
depolarizing, Eq. (5.5)	$\frac{q}{q+1}\epsilon$	$(1-\frac{q^2}{q^2-1}\epsilon)^2$
dephasing, Eq. (5.6)	$\frac{q}{q+1}\epsilon$	$1-\frac{q^2}{q^2-1}(2\epsilon-\frac{q}{q-1}\epsilon^2)$
rotation, Eq. (5.7)	$\frac{2(q-1)}{q(q+1)}(1-\cos(\theta))$	1

Table 5.1: Average infidelity and unitarity for three different single-qudit noise channels.

denote the unitary channel that acts with $U^{(t)}$ on qudits i_t and j_t and as identity on the other qudits. To account for noise, let

$$\widetilde{\mathcal{U}}^{(t)} = \begin{cases} \left(I_{[n]\setminus\{i_t,j_t\}} \otimes \mathcal{N}_{\{i_t\}} \otimes \mathcal{N}_{\{j_t\}}\right) \circ \mathcal{U}^{(t)} & \text{if } 1 \le t \le s \text{ (two-qudit)} \\ \mathcal{U}^{(t)} & \text{otherwise (single-qudit)} \end{cases} \tag{5.9}$$

be the channel that applies noise channels after applying the unitary gate. Now we can define the ideal and noisy output distributions by

$$p_{\text{ideal}}(x) = \operatorname{tr}\left[|x\rangle\langle x|\, \mathcal{U}^{(s+n)} \circ \cdots \circ \mathcal{U}^{(-n+1)}\left(|0^n\rangle\langle 0^n|\right)\right] \tag{5.10}$$

$$p_{\text{noisy}}(x) = \operatorname{tr}\left[|x\rangle\langle x|\, \widetilde{\mathcal{U}}^{(s+n)} \circ \cdots \circ \widetilde{\mathcal{U}}^{(-n+1)}\left(|0^n\rangle\langle 0^n|\right)\right]. \tag{5.11}$$

Our work compares the distribution p_{noisy} to the white-noise distribution p_{wn}, defined by

$$p_{\text{wn}}(x) = F p_{\text{ideal}}(x) + (1-F)q^{-n} \tag{5.12}$$

for some choice of F. The white-noise distribution is a mixture of the ideal distribution and the uniform distribution. In the analysis we treat F as a free parameter, and we choose it such that our bound on the distance between p_{noisy} and p_{wn} is minimized. The total variation distance between two distributions p_1 and p_2 is defined as

$$\text{TVD}(p_1, p_2) = \frac{1}{2}\|p_1 - p_2\|_1 = \frac{1}{2}\sum_x |p_1(x) - p_2(x)|. \tag{5.13}$$

A comment on the various kinds of randomness in our setup

There are multiple types of randomness in our analysis, and in understanding our result it is important to keep track of how they interplay. First of all, the noiseless circuit instance U is generated randomly by choosing each gate to be Haar random. The choice of U determines an ideal *pure* output state. Second of all, for each fixed choice of U, the noise channels may introduce randomness that makes the noisy output state *mixed*. When the noise is depolarizing noise, this might be regarded as the insertion of a randomly chosen pattern of Pauli errors. Lastly, the measurement of the state in the computational basis

gives rise to a random measurement outcome drawn from a certain classical probability distribution, p_{ideal} if we are considering the noiseless circuit, and p_{noisy} if we are considering the noisy circuit. The important thing to remember is that we are primarily concerned with thinking about *fixed* instances U and the interplay between the resulting probability distributions p_{ideal}, p_{noisy} and p_{wn} for that instance. Then, we make a statement about these distributions that holds in expectation over random choice of U. If desired, one could then use Markov's inequality to form bounds on the fraction of instances U for which the white-noise approximation must be good.

5.3 Overview of contributions

The main result of this chapter is a proof that the output distribution p_{noisy} of the quantum circuit with local noise is very close to the white noise distribution p_{wn} if the noise is sufficiently weak. In proving that result, we also prove a statement about the expected fidelity in noisy random quantum circuits, and another statement about the speed at which p_{noisy} approaches the uniform distribution.

For all statements, the notation $\mathbb{E}_U$ denotes expectation over choice of Haar-random single-qudit and two-qudit gates. Additionally, our technical results assume that the random quantum circuit architecture is either the complete-graph architecture, or that it is composed of parallel layers (Definition 5.1) and has the h-regularly connected property (Definition 4.5) for some $h = O(1)$, which means at least one out of every hn gates is expected to act across any division of the qudits into two groups. Standard lattice architectures with periodic boundary conditions in any spatial dimension have these properties. We only utilize the assumption of layers in one specific place of our analysis, and we believe that a version of our results should hold for any natural random quantum circuit architecture.

In the rest of this section, we state our results for general noise channels, deferring the proofs to Appendix 5.B, but first we summarize the contributions specifically applied to the depolarizing channel.

5.3.1 Fidelity decay

Define the quantity

$$\bar{F} = \frac{\mathbb{E}_U \left[\sum_x p_{\text{noisy}}(x)(q^n p_{\text{ideal}}(x) - 1) \right]}{\mathbb{E}_U \left[\sum_x p_{\text{ideal}}(x)(q^n p_{\text{ideal}}(x) - 1) \right]} . \tag{5.14}$$

The quantity $\bar{F}$ is designed to quantify the fidelity of the noisy quantum device with respect to the ideal computation; when $p_{\text{noisy}}(x)$ and $p_{\text{ideal}}(x)$ are viewed as random variables in the instance U, $\bar{F}$ is equal to their covariance, normalized by the variance of p_{ideal}. Note also that the numerator of $\bar{F}$ is the expected score on the linear cross-entropy benchmark, as proposed in Ref. [6], using

Fidelity decay	$\bar{F} = e^{-2s\epsilon \pm O(s\epsilon^2)}$
Approach to uniform	$\mathbb{E}_U \left[\frac{1}{2}\|p_{\text{noisy}} - p_{\text{unif}}\|_1\right] \leq e^{-2s\epsilon + O(s\epsilon^2)}$
Distance from p_{wn} for $F = \bar{F}$	$\mathbb{E}_U \left[\frac{1}{2}\|p_{\text{noisy}} - p_{\text{wn}}\|_1\right] \leq O(F\epsilon\sqrt{s})$

Table 5.2: Summary of results when the noise is depolarizing (Eq. (5.5)) with error parameter ϵ. These statements assume that the anti-concentration size is $s_{AC} = O(n\log(n))$ (which is known for the 1D architecture and conjectured generally), that the circuit size is larger than $\Omega(n\log(n))$, and that the quantity $\epsilon n \log(n)$ is small enough to be neglected. (It is believed this condition can be relaxed to $\epsilon < c/n$ for some constant c.)

samples from the noisy device, and the denominator is the expected score using samples from the ideal output distribution. The use of $\bar{F}$ as a fidelity benchmark has been further explored in Refs. [50, 151]. The denominator is also given by $q^n Z - 1$, where Z the collision probability studied in Chapter 4. The results of that chapter imply that the denominator becomes within a small constant factor of $(q^n - 1)/(q^n + 1) \approx 1$ after $s_{AC} = O(n\log(n))$ gates in the 1D or complete-graph architectures, and this fact is conjectured to hold widely for natural architectures.

Theorem 5.1. *Consider either the complete-graph architecture or a regularly connected, layered random quantum circuit architecture with n qudits of local Hilbert space dimension q and s gates, where the anti-concentration size is given by s_{AC}. Let r be the average infidelity of the local noise channels. Then there exists constants c and n_0 such that whenever $r \leq c/n$ and $n \geq n_0$, the following holds:*

$$\bar{F} \geq \exp\left(-2sr(1+q^{-1})\right) e^{-O(sr^2) - O(sq^{-2n}) - e^{O(s_{AC}/n)} e^{-\Omega(s/n)}} \tag{5.15}$$

$$\bar{F} \leq \exp\left(-2sr(1+q^{-1})\right) Q_1\,, \tag{5.16}$$

where

$$Q_1 = \exp\left(O(sr^2) + O(s_{AC}r) + e^{O(s_{AC}/n)} e^{-\Omega(s/n)} + O(nr\log(1/(nr)))\right)\,. \tag{5.17}$$

Note that the relationship $\epsilon = r(q+1)/q$ holds for the depolarizing channel as defined in Eq. (5.5), so, ignoring the $O(q^{-2n})$ corrections,

$$e^{-2s\epsilon - O(s\epsilon^2)} \leq \bar{F} \leq e^{-2s\epsilon + O(s\epsilon^2) + O(\epsilon s_{AC}) + O(n\epsilon\log(1/(n\epsilon)))}\,, \tag{5.18}$$

indicating that the fidelity decreases exponentially with the expected number of Pauli errors $2s\epsilon$, as long as the noise is sufficiently weak that the other terms can be ignored. In particular, three conditions must be met to approximate

Q_1 by 1 in Eq. (5.16): (1) $\epsilon^2 s \ll 1$, (2) anti-concentration has been reached, i.e. $s \geq s_{AC} + \Omega(n)$, and (3) $\epsilon \ll 1/s_{AC}$. Even if these conditions are not met, exponential decay is still observed, but our bounds on the decay constant are not tight. Note that if $s_{AC} = \Theta(n \log(n))$, condition (3) can be stated $1/\epsilon \geq \tilde{\Omega}(n)$ where the $\tilde{\Omega}$ suppresses log factors. We believe the analysis could be improved to remove these log factors.

5.3.2 Convergence to uniform

We show an upper bound on the expected total variation distance between the output of the noisy quantum device p_{noisy} and the uniform distribution. Our bound decays exponentially in the number of error locations, under certain circumstances. In particular, it decays exponentially in $(1-u)(1-q^{-2})s$ where u is the unitarity of the local noise channels.

Theorem 5.2. *Consider either the complete-graph architecture or a regularly connected, layered random quantum circuit architecture with n qudits of local Hilbert space dimension q and s gates, where the anti-concentration size is given by s_{AC}. Let u be the unitarity of the local noise channels (and define $v = 1 - u$). Then there exist constants c and n_0 such that as long as $v \leq c/n$ and $n \geq n_0$*

$$\mathop{\mathbb{E}}_{U} \left[\frac{1}{2} \| p_{noisy} - p_{unif} \|_1 \right] \leq \exp\left(-sv(1-q^{-2})\right) Q_2 \,, \tag{5.19}$$

where p_{unif} is the uniform distribution and

$$Q_2 = \exp\left(O(sv^2) + O(s_{AC} v) + e^{O(s_{AC}/n)} e^{-\Omega(s/n)} + O(nv \log(1/(nv))) \right) . \tag{5.20}$$

Note that Q_2 is small under a similar three conditions as in the fidelity decay result: (1) $s(1-u)^2 \ll 1$, (2) anti-concentration has been reached, and (3) $s_{AC}(1-u) \ll 1$.

For the depolarizing channel, $u = 1 - 2\epsilon(1-q^{-2})^{-1}$ up to first order in ϵ, so the distance to uniform decays like $e^{-2s\epsilon}$, which is identical to the rate of fidelity decay. On the other hand, the unitarity of the rotation channel is $u = 1$, so our upper bound does not decay with s, even though $\bar{F}$ does decay for the rotation channel. This is expected because the rotation channel is coherent; indeed, unlike the other two examples, it sends pure states to pure states. The ideal pure state and the noisy pure state will become less and less correlated as more noise channels act, which explains why $\bar{F}$ decays, but the output distribution for the noisy pure state will not converge to uniform.

5.3.3 Distance to white noise distribution

If the noise is sufficiently weak, we show a stronger statement. Not only does the output distribution decay to uniform, it does so in a very particular

way, preserving an uncorrupted signal from the ideal distribution. We show that p_{noisy} is close to p_{wn} by upper bounding the expected total variation distance between the two distributions.

Theorem 5.3. *Consider either the complete-graph architecture or a regularly connected, layered random quantum circuit architecture with n qudits of local Hilbert space dimension q and s gates, where the anti-concentration size is given by s_{AC}. Let r be the average infidelity and u the unitarity of the local noise channels (and define $v = 1 - u$). Let*

$$\delta = 2r(1 + q^{-1}) - (1 - u)(1 - q^{-2}) \,. \tag{5.21}$$

Then when we choose $F = \bar{F}$ as in Eq. (5.14), there exist constants c_1, c_2, and n_0 such that as long as $v \leq c_1/n$, $r \leq c_2/n$, and $n \geq n_0$,

$$\begin{aligned} \mathop{\mathbb{E}}_{U} \left[\frac{1}{2} \| p_{noisy} - p_{wn} \|_1 \right] \leq \bar{F}\sqrt{s} \left(\sqrt{\delta} + O(v) + O(r) \right) + O(\bar{F}\sqrt{s_{AC} v}) \\ + O(\bar{F}\sqrt{nv \log(1/nv)}) + \bar{F} e^{O(s_{AC}/n) - \Omega(s/n)} \,, \end{aligned} \tag{5.22}$$

whenever the right-hand side of Eq. (5.22) is less than $\bar{F}$.

We make a couple of comments. First, we emphasize how small the right-hand side of Eq. (5.22) is. The quantity $\bar{F}$ is decaying exponentially in the number of expected errors, as shown in Theorem 5.1. We showed in Theorem 5.2 that p_{noisy} converges to uniform at roughly the same rate. However, the distance between p_{noisy} and p_{wn} is much smaller than $\bar{F}$ if the parameters are sufficiently weak, demonstrating that the noisy and white-noise distribution are much closer even than either are to uniform.

Second, let us examine the quantity δ. For the depolarizing channel and the dephasing channel, the leading term in δ cancels out leaving $\delta = O(\epsilon^2)$, so the $\sqrt{\delta}$ term that appears is on the same order as the other terms. This is a signature of incoherent noise. The coherent rotation channel, which has $u = 1$ and $r = O(\theta^2)$, has $\delta = O(\theta^2)$, so $\sqrt{\delta}$ is large compared to the other terms in the expression. In this case, we would need $sr \ll 1$ for the approximation to be good, but if this is true, then $\bar{F} \approx 1$ and the white-noise approximation is trivial.

Relatedly, the parameter δ can be connected to the diamond distance D of the channel $\mathcal{N}$, which is the maximum amount action by $\mathcal{N}$ can change an input state (which might be entangled with an auxiliary system) as measured by the trace norm. If $\mathcal{N}$ is applied s times, the total deviation in trace norm from the ideal output can be as large as sD in the worst case. It was shown in Ref. [152] that $D = O(\sqrt{\delta})$, specifically

$$\frac{1}{2}\sqrt{\delta} \leq D \leq \frac{q^2}{2}\sqrt{\delta} \,. \tag{5.23}$$

It is also known that $r \leq O(D)$ and $1 - u \leq O(D)$. Thus, if we ignore the final three terms in Eq. (5.22), we can write our result as

$$\mathop{\mathbb{E}}_{U}\left[\frac{1}{2}\|p_{\text{noisy}} - p_{\text{wn}}\|_1\right] \leq O(FD\sqrt{s})\,. \tag{5.24}$$

This emphasizes that the fundamental result is an improved tradeoff between noise and circuit size; the strength of the signal decays exponentially, but the error due to noise grows quadratically slower for random quantum circuits with local noise than it does in the worst case.

5.4 Related work and implications

5.4.1 Quantum computational supremacy

A central motivation for our work has been recent quantum computational supremacy experiments [6, 7] that sampled from the output of noisy random quantum circuits on superconducting devices. In this context, the main claim is that no classical computer could have performed the same feat in any reasonable amount of time. As discussed in Chapter 1, while no efficient classical algorithms to simulate the quantum device performing this task are known, there is a lack of concrete theoretical evidence that no such algorithm exists.

Our work bolsters the theory behind these experiments in two ways, assuming that noise in the device is sufficiently well described by our local noise model. First, our fidelity decay result validates using the linear cross-entropy metric to benchmark the overall noise rate in the device, and quantify the amount of signal from the ideal computation that survives the noise. Second, convergence to the white-noise distribution has theoretical benefits with respect to a potential proof that the random circuit sampling task accomplished by the device is actually hard for classical computers.

Linear cross-entropy benchmarking

Quantum computational supremacy experiments are complicated by the fact that since (by definition) they cannot be replicated on a classical computer, it is non-trivial to classically verify that they actually performed the correct computational task. A partial solution to this issue has been the proposal of linear cross-entropy benchmarking, whereby a sample x is generated by the device according to the noisy output distribution p_{noisy}, and a classical supercomputer is used to compute $p_{\text{ideal}}(x)$.[4] When T samples $\{x_1, \ldots, x_T\}$ are chosen, the average

$$\mathcal{F} = \frac{1}{T}\sum_{i=1}^{T}(q^n p_{\text{ideal}}(x) - 1) \tag{5.25}$$

[4] This requires exponential time but can be tractable for circuit sizes up to $n = 50$ or so (in the case of a 2D architecture, it also depends on the depth of the circuit).

is calculated, which is an empirical measure of the circuit fidelity. We can see that the expected value of $\mathcal{F}$ is precisely $\sum_x p_{\text{noisy}}(x)(q^n p_{\text{ideal}}(x) - 1)$, which is the numerator of the quantity $\bar{F}$ defined in Eq. (5.14). Meanwhile, the denominator of $\bar{F}$ becomes close to 1, so long as the output is anti-concentrated. In Theorem 5.1, we show that if the depolarizing error rate ϵ satisfies $\epsilon \ll 1/(n \log(n))$ and as long as $\epsilon^2 s \ll 1$, then there are matching upper and lower bounds on the expected value of $\mathcal{F}$, which decays with the circuit size like $e^{-2\epsilon s}$. Thus, assuming our local noise model, we prove that one can infer ϵ given $\mathcal{F}$ and s. The inferred value of ϵ can then be compared to the noise strength estimated when testing each circuit component individually, thus providing one method of verification that the components are behaving as expected during the experiment.

Indeed, the idea of using random circuit sampling as an alternative to randomized benchmarking was formally proposed in Ref. [50], a work that has certain similarities to ours. In particular, like us, they find that a noise rate of $1/\epsilon \geq \Omega(n)$ appears necessary for controlled decay of the fidelity. (Our result can be expressed as requiring $1/\epsilon \geq \tilde{\Omega}(n)$, where the tilde hides log factors, and we believe those log factors are not necessary for our result.) Additionally, like us, they use the stat mech method to motivate their results. However, they only analytically study the fidelity decay up to first order in the error rate for a 1D architecture; that is, they compute the expected fidelity due to contributions with only one error location. Moreover, they propose using $\bar{F}$ in their algorithm but only analytically study decay of another quantity, namely the actual circuit fidelity $\mathbb{E}[\text{tr}[\rho_{\text{noisy}}\rho_{\text{ideal}}]]$ where ρ_{noisy} and ρ_{ideal} are the noisy and ideal output quantum states. Thus, our result might be regarded as a completion of their analysis, as we rigorously show a precise exponential decay of $\bar{F}$ without any need for approximation.

Note that as the fidelity decays, more samples must be generated to get a good estimate of the mean of $\mathcal{F}$. Since $p_{\text{ideal}}(x)$ for uniformly random x has standard deviation on the order of q^{-n} (assuming anti-concentration), the standard deviation of $\mathcal{F}$ is expected to decay with the number of samples like $1/\sqrt{T}$. Thus, resolving it with enough precision to differentiate it from 0 requires $T = \Omega(1/\mathcal{F}^2)$ samples.

We comment that while our analysis assumes that each noise location has the same value of ϵ, this is not essential to our method. We expect it could be shown that the expected value of $\mathcal{F}$ decays like $\exp(-\sum_i \epsilon_i)$ where i runs over all possible noise locations. Moreover, our analysis works for any kind of local noise, not just depolarizing noise; the only relevant parameter is the average infidelity of the noise channels.

Classical hardness of sampling from the noisy output distribution

To claim to have achieved quantum computational supremacy, the low-fidelity random circuit sampling experiments in Refs. [6, 7] must define a concrete

computational problem that their device solved, but a classical device could not also solve. Here there are a couple of options. One option is to simply rely directly on the linear cross-entropy benchmarking task and define the task to be generating a set of samples that scores at least $\mathcal{F} \geq 1/\operatorname{poly}(n)$. A related idea is the task of Heavy Output Generation (HOG) [23], which is to generate outputs x for which $p_{\text{ideal}}(x)$ is large (i.e. "heavy outputs") significantly more often than a uniform generator. The upshot of these definitions is that in the regime where $p_{\text{ideal}}(x)$ can be calculated classically with an exponential-time algorithm, it can be verified that the quantum device successfully performed the task. Their main drawback is that it is not clear whether running a (noisy) quantum computation is the only way to perform these tasks. Perhaps a classical algorithm can score well on the linear cross-entropy benchmark without performing an actual random circuit simulation; for example, this was the goal in Ref. [29].

Another option is to define the task specifically in terms of the white-noise distribution. Namely, one must produce samples from a distribution p_{noisy} for which $\frac{1}{2}\|p_{\text{noisy}} - p_{\text{wn}}\|_1 \leq \varepsilon F$ for some choice of F not too small (ideally at least inverse polynomial in n) and some small constant ε. In Chapter 1, we referred to this task as "white-noise RCS." A downside of this option is that even with unlimited computational power, an exponential number of samples from the device would be needed to definitively verify that the distribution is close to p_{wn} in total variation distance. Our work provides a partial solution here, as we show that a local error model allows a device to accomplish the white-noise RCS task, as long as the error rate is sufficiently weak compared to the circuit size. Thus, if the experimenters are sufficiently confident in the error model that describes their device, they can rely on our work to be confident they are performing the white-noise RCS task. The major upside of the white-noise RCS task is that one can give stronger evidence that it is classically hard to perform. For example, in the Supplementary Material of Ref. [6], it was shown that *exactly* (i.e. $\varepsilon = 0$) sampling from p_{wn} (a task they called "unbiased noise F-approximate random circuit sampling") is a hard computational task in the sense that an efficient classical algorithm for it would cause the collapse of the polynomial hierarchy (PH), and further that its computational cost should be at most a factor of F smaller than sampling exactly from p_{ideal}. In that spirit, we show in Theorem 5.4, in the appendix, that the more realistic task of sampling *approximately* from p_{wn} is essentially just as hard as sampling *approximately* from p_{ideal}, up to a factor polynomial in F in the classical computational cost. This is important because some mild progress has been made toward establishing that approximately sampling from p_{ideal} is hard for the polynomial hierarchy, through a series of work that reduce the task of computing $p_{\text{ideal}}(x)$ in the worst case to the task of computing $p_{\text{ideal}}(x)$ in the average case up to some small error [4, 25–27]. Weaknesses in this result as evidence for hardness of approximate sampling were discussed in more detail in Chapter 3, but it remains true that the white-noise-centered

definition of the computational task is the likeliest route to a more robust version of quantum computational supremacy that can be grounded in well-studied complexity theoretic principles.

5.4.2 Fast convergence to uniform

It is widely understood that incoherent, unbiased, and uncorrected noise in quantum circuits should typically lead the output of a quantum circuit to lose all correlation with the ideal circuit and become nearly uniform. It is further asserted that the decay to uniform should scale with the circuit size; however, rigorous results have only shown a decay in total variation distance to uniform with the circuit depth d, following the form $e^{-\Omega(\epsilon d)}$. In particular, Ref. [153] showed that any (even non-random) circuit with interspersed local depolarizing noise approaches uniform at least this quickly. Later, Ref. [31] showed the same is true for any Pauli noise model, at least for most circuits chosen from a particular random ensemble. However, in Ref. [26], a stronger convergence at the rate of $e^{-\Omega(\epsilon s)}$ in random quantum circuits like ours was desired in order to show a barrier on further improvements of their worst-to-average-case reduction for computing entries of p_{ideal}. To that end, they showed that exponential convergence in circuit size occurs in a toy model where each layer of unitary evolution enacts an exact global unitary 2-design as opposed to many disjoint local gates, and they conjectured the same is true in the local noise model we consider in this chapter. Thus, our result in Theorem 5.2 gets close to providing the missing ingredient for their claim; for their application, we would need to extend our result to show $e^{-\Omega(\epsilon s)}$ even in the regime where $\epsilon = O(1)$, independent of n. Our result applies only for $\epsilon = O(1/(n \log(n)))$, but we believe the extension to $\epsilon = O(1)$ might also be provable with our method.

5.4.3 Signal extraction in noisy experiments

One implication of our work is that, in the parameter regime where our results apply, the signal from the noiseless random circuit experiment can be extracted by taking many samples. To illustrate this, suppose we are interested in some classical function $f(x)$ for $x \in [q]^n$ that takes values between -1 and $+1$. Choosing x randomly from p_{ideal} induces a probability distribution over the resulting values of $f(x)$. To understand this distribution (e.g., estimate its mean or variance), samples x_i might be generated on a quantum device, but if the device is noisy, these samples will be drawn from p_{noisy} instead of p_{ideal}. However, if $p_{\text{noisy}} \approx p_{\text{wn}}$, then the sampled distribution over $f(x)$ will be a mixture of the ideal with weight F, and the distribution that arises from uniform choice of x with weight $1 - F$. Supposing the latter is well understood, inferences can be made about the former by repetition. For example, if $\sum_x p_{\text{ideal}}(x) f(x) = \mu = O(1)$ and $\sum_x f(x)/q^n = 0$, then the mean of f under samples from p_{wn} is $F\mu$. Meanwhile, the standard deviation of f can

be as large as $O(1)$, indicating that $O(1/F^2)$ samples from p_{wn} are required to compute the mean $F\mu$ up to $O(F)$ precision.

A concrete example of such a situation is the Quantum Approximate Optimization Algorithm (QAOA) [154], where samples x from the output of a parameterized quantum circuit are used to estimate the expectation of a classical cost function $C(x)$. The parameters can then be varied to optimize the expected value of the cost function. While our work is for Haar-random local quantum circuits, not QAOA circuits, it is plausible that generic QAOA circuits might exhibit a similar phenomenon. Indeed, in Refs. [155–157], numerical and analytic evidence was given for the conclusion that the expectation value of the cost function and its gradient with respect to the circuit parameters decay toward zero when local noise is inserted into a QAOA circuit. This behavior would be consistent with a stronger conclusion that the output is well-described by p_{wn}.

5.5 Summary of method and intuition

In this section, we present a heuristic argument about why the technical statements stated above should hold. Then we give an overview of how we actually show it using our method, which analyzes certain Markov processes derived from the quantum circuits, extending our work in Chapter 4.

5.5.1 Intuition behind error scrambling

Our result that p_{noisy} is very close to p_{wn} requires three conditions to be satisfied: (1) $\epsilon^2 s \ll 1$, (2) anti-concentration has been achieved, (3) $\epsilon s_{AC} \ll 1$. Here, we try to motivate why these conditions should be sufficient and speculate about whether they are also necessary. In particular, we believe condition (3) can be significantly relaxed.

For simplicity, lets restrict to qubits ($q = 2$). Let U denote the unitary enacted by the noiseless quantum circuit instance, so the ideal output state is the pure state $\rho_{\text{ideal}} = U|0^n\rangle\langle 0^n|U^\dagger$. If a location somewhere in the middle of the circuit experiences a Pauli error, then we could write the output state as $U_2 P U_1|0^n\rangle\langle 0^n|U_1^\dagger P^\dagger U_2^\dagger$, where P is a Pauli operator with support on only one qubit, and $U = U_2 U_1$ is a decomposition of the unitary into gates that act before and after the error location. If we like, we can commute P to act at the end of the circuit, giving $O_P U|0^n\rangle\langle 0^n|U^\dagger O_P^\dagger$ where $O_P = U_2 P U_2^\dagger$. Unlike P, the operator O_P will likely have support over many qubits. Indeed, this is what we mean by scrambling; the portion of the circuit acting after the error location scrambles the local noise P into more global noise O_P. We can handle error patterns E with multiple Pauli errors similarly, by commuting each to the end one at a time and forming an associated global noise operator O_E.

Next, we expand the output state ρ_{noisy} of the noisy circuit as a sum over all possible Pauli error patterns, weighted by the probability that each pattern occurs. Assuming that the local noise is depolarizing, the probability

of a pattern E depends only on the number of non-identity Pauli operators in the error pattern, denoted by $|E|$.

$$\rho_n = \sum_E \left(\frac{\epsilon}{3}\right)^{|E|} (1-\epsilon)^{2s-|E|} O_E \rho_{\text{ideal}} O_E^\dagger \,. \tag{5.26}$$

The classical probability distribution p_{noisy} is then given by $p_{\text{noisy}}(x) = \langle x|\rho_{\text{noisy}}|x\rangle$ for each measurement outcome x. Observe that for the error pattern with $|E| = 0$ (no errors), we have $\rho_E = \rho_{\text{ideal}}$. There can be other error patterns for which $O_E \rho_{\text{ideal}} O_E^\dagger = \rho_{\text{ideal}}$; for example, when a lone Pauli-Z error acts prior to any non-trivial gates, the state is unchanged since the initial state $|0^n\rangle$ is an eigenstate of all the Pauli-Z operators. However, these error patterns are rare, and for the sake of intuition, we ignore this possibility. In essence, the white-noise assumption is the claim that when we take the mixture over output states for all of the error patterns, we arrive at a state ρ_{err} that produces measurement outcomes that are very close to uniform. (Note that in general, ρ_{err} need not be close to maximally mixed to yield uniform measurements.) Letting $F = (1-\epsilon)^{2s}$, we may write

$$\rho_{\text{noisy}} = F\rho_{\text{ideal}} + F \sum_{E:|E|>0} \left(\frac{\epsilon/3}{1-\epsilon}\right)^{|E|} O_E \rho_{\text{ideal}} O_E^\dagger \tag{5.27}$$

$$= F\rho_{\text{ideal}} + (1-F)\frac{I}{2^n} + F \sum_{E:|E|>0} \left(\frac{\epsilon/3}{1-\epsilon}\right)^{|E|} \left(O_E \rho_{\text{ideal}} O_E^\dagger - \frac{I}{2^n}\right) . \tag{5.28}$$

This final term gives the deviations of the noisy output state ρ_{noisy} from a linear combination of the ideal state and the maximally mixed state.

This allows us to state more clearly the intuition for our result. Since the circuit is randomly chosen and scrambles the local error patterns, the operators O_E generally have large support and are essentially uncorrelated for different choices of error pattern E. Suppose we measure in the computational basis, and examine the probability of obtaining the outcome x. Let

$$p_E(x) = \langle x|O_E \rho_{\text{ideal}} O_E^\dagger|x\rangle \,. \tag{5.29}$$

We can calculate the squared deviation between this value and the white-noise value under expectation over instance U.

$$\mathop{\mathbb{E}}_U [(p_{\text{noisy}}(x) - p_{\text{wn}}(x))^2] \tag{5.30}$$

$$= \mathop{\mathbb{E}}_U \left[\left(\langle x|\rho_{\text{noisy}}|x\rangle - (F\langle x|\rho_{\text{ideal}}|x\rangle + (1-F)2^{-n}) \right)^2 \right] \tag{5.31}$$

$$= F^2 \sum_{\substack{E,E' \\ |E|,|E'|>0}} \left(\frac{\epsilon/3}{1-\epsilon}\right)^{|E|+|E'|} \mathop{\mathbb{E}}_U \left[\left(p_E(x) - 2^{-n}\right) \left(p_{E'}(x) - 2^{-n}\right) \right] . \tag{5.32}$$

Suppose we now make the approximation that the quantities $p_E(x)$ and $p_{E'}(x)$, when considered as functions of the random instance U, are independently distributed unless $E = E'$. Their mean is 2^{-n} and, assuming anti-concentration (condition (2)), their standard deviation is $O(2^{-n})$. Then we have

$$\mathop{\mathbb{E}}_{U}[(p_{\text{noisy}}(x) - p_{\text{wn}}(x))^2] \tag{5.33}$$

$$\approx F^2 \sum_{E:|E|>0} \left(\frac{\epsilon/3}{1-\epsilon}\right)^{2|E|} \mathop{\mathbb{E}}_{U}\left[\left(p_E(x) - 2^{-n}\right)^2\right] \tag{5.34}$$

$$= F^2 \sum_{E:|E|>0} \left(\frac{\epsilon/3}{1-\epsilon}\right)^{2|E|} O(2^{-2n}) \tag{5.35}$$

$$= F^2 \cdot O(2^{-2n}) \cdot \left((1+O(\epsilon^2))^{2s} - 1\right) \tag{5.36}$$

$$\approx O(F^2 2^{-2n} \epsilon^2 s) \tag{5.37}$$

when $\epsilon^2 s \ll 1$. This implies that the deviation of each entry in the probability distribution p_{noisy} from the white-noise distribution is on the order of $F2^{-n}\epsilon\sqrt{s}$, and since there are 2^n entries, we have

$$\mathop{\mathbb{E}}_{U}\left[\frac{1}{2}\|p_{\text{wn}} - p_{\text{noisy}}\|_1\right] \approx O(F\epsilon\sqrt{s})\,. \tag{5.38}$$

In other words, the total variation distance is much smaller than F when $\epsilon^2 s \ll 1$, giving an intuitive reason for condition (1). Moreover, without condition (2), the contribution of each term would be much larger than $O(q^{-2n})$, which illustrates why condition (2) is necessary.

The key step in this analysis was the assumption of independence between p_E and $p_{E'}$ when $E \neq E'$. This is only approximately true; indeed for a circuit that does not scramble errors, this will be a bad approximation because it might be common to have different error patterns E, E' that produce the same (or approximately the same) effective error $O_E = O_{E'}$. However, for random quantum circuits, this outcome is unlikely for the vast majority of error pairs. Our rigorous proof, later, might be regarded as a justification of this intuition above.

Condition (3) is more subtle to motivate. In our analysis, we require $\epsilon \ll 1/s_{AC}$ so that the chance an error occurs while the circuit is still anti-concentrating is small. This is helpful in the analysis because it allows us to essentially ignore the possibility that an error P occurs near the beginning or end of the circuit, where $O_P \rho_{\text{ideal}} O_P^\dagger = \rho_{\text{ideal}}$ is more likely to hold (or approximately hold). When $s_{AC} = \Theta(n\log(n))$, as is the case when the architecture is the 1D or complete-graph architecture, condition (3) reads $1/\epsilon \gg \Theta(n\log(n)) = \tilde{\Theta}(n)$. We believe this can be improved to read $1/\epsilon \geq n/c$ for some constant c that depends on the architecture (1D vs. complete-graph etc.). However, we do not believe that improvement beyond this point would be possible; there is a fundamental barrier that requires ϵ to scale as $O(1/n)$.

The reason for this is essentially that if the white-noise approximation is to hold, the errors need to be scrambled at least as fast as they appear. The fidelity F decreases like $(1-\epsilon)^{2s} = \exp(-2s\epsilon - O(s\epsilon^2))$, so each layer of $O(n)$ gates causes a decrease by a factor $\exp(-O(n\epsilon))$. Recall that we demand that the total variation distance between p_{noisy} and p_{wn} be much smaller than F, so as F decreases, this condition becomes increasingly stringent. Meanwhile, scrambling is fundamentally happening at the rate of increasing circuit depth, not size. One way to see this is simply that local Pauli errors P that appear at a certain circuit location are expected to be scrambled into larger operators that grow ballistically with the depth [41, 42]; each layer of $O(n)$ gates yields a constant amount of operator growth. Another way to see this is to consider a pair of error patterns E and E', where E consists of a single Pauli error on qudit j at layer d and E' consists of a single Pauli error on qudit j at layer $d+\Delta$. The correlation between $p_E(x)$ and $p_{E'}(x)$, as a function of the random instance U, which is roughly speaking the chance that the random circuit transforms the first error into something resembling the second error, will decay exponentially with Δ, the separation in depth between the two errors.[5] Yet a third way to see this fact is to notice that, after a circuit has initially reached anti-concentration, convergence of the collision probability Z to its limiting value Z_H occurs like $Z = Z_H + O(Z_H)\exp(-O(s/n))$. Each additional layer of $O(n)$ gates only decreases the deviation of Z from Z_H by a constant factor. The terms $\mathbb{E}_U[(p_E - 2^{-n})(p_{E'} - 2^{-n})]$ for $E \neq E'$ that were ignored above are expected to obey a similar kind of decay to the value 0 for most choices of (E, E'), but if F is decaying too fast, we are not be able to neglect these terms. Each layer of $O(n)$ gates must incur at most a constant-factor decay in fidelity to not exceed the rate of scrambling; equivalently, $n\epsilon < c$ must hold for some constant c.

5.5.2 Noisy random quantum circuits as a stochastic process

In Chapter 4, we analyzed the collision probability $Z = \mathbb{E}_U[\sum_x p_{\text{ideal}}(x)^2]$ by mapping it to the expectation value of a Markov process, which could also be interpreted as the partition function of a classical stat mech system. In this chapter, we extend that analysis to account for the action of the single-qudit noise channels $\mathcal{N}$ that act after two-qudit gates.

This is a manifestation of the stat mech method for second-moment quantities, so the first step is to somehow express the distance between p_{noisy} and p_{wn} in terms of second-moment random quantum circuit quantities. The starting point for this is the general 1-norm to 2-norm bound: when p_1 and p_2 are

[5] This is particularly clear if the random circuits are Clifford circuits (for which our results also apply since random Clifford gates form an exact 2-design). Clifford circuits transform the error E at layer d more or less uniformly at random into one of the roughly 4^Δ possible Pauli operators at layer $d+\Delta$. The probability that this operator is E' is exponentially small in Δ.

vectors on a q^n-dimensional vector space, then

$$\|p_1 - p_2\|_1 \leq q^{n/2} \|p_1 - p_2\|_2 \,, \tag{5.39}$$

where $\|p_1 - p_2\|_2 = \sqrt{\sum_x (p_1(x) - p_2(x))^2}$ is a second-moment type quantity. Applying this identity with $p_1 = p_{\text{wn}}$ and $p_2 = p_{\text{noisy}}$ and invoking Jensen's inequality for the concave function $\sqrt{\cdot}$, we find

$$\mathop{\mathbb{E}}_{U} \left[\frac{1}{2} \|p_{\text{wn}} - p_{\text{noisy}}\|_1 \right] \leq q^{n/2} \mathop{\mathbb{E}}_{U} \left[\frac{1}{2} \|p_{\text{wn}} - p_{\text{noisy}}\|_2 \right] \leq \frac{1}{2} \sqrt{q^n \mathop{\mathbb{E}}_{U} \left[\|p_{\text{wn}} - p_{\text{noisy}}\|_2^2 \right]} \,. \tag{5.40}$$

Now we can expand

$$q^n \mathop{\mathbb{E}}_{U} \left[\|p_{\text{wn}} - p_{\text{noisy}}\|_2^2 \right] = q^n \mathop{\mathbb{E}}_{U} \left[\sum_x \left(\left(F p_{\text{ideal}}(x) + (1 - F) q^{-n} \right) - p_{\text{noisy}}(x) \right)^2 \right] \tag{5.41}$$

$$= (Z_2 - 1) - 2F(Z_1 - 1) + F^2 (Z_0 - 1) \,, \tag{5.42}$$

where

$$Z_0 = q^n \mathop{\mathbb{E}}_{U} \left[\sum_x p_{\text{ideal}}(x)^2 \right] = q^{2n} \mathop{\mathbb{E}}_{U} \left[p_{\text{ideal}}(0^n)^2 \right] \tag{5.43}$$

$$Z_1 = q^n \mathop{\mathbb{E}}_{U} \left[\sum_x p_{\text{noisy}}(x) p_{\text{ideal}}(x) \right] = q^{2n} \mathop{\mathbb{E}}_{U} \left[p_{\text{noisy}}(0^n) p_{\text{ideal}}(0^n) \right] \tag{5.44}$$

$$Z_2 = q^n \mathop{\mathbb{E}}_{U} \left[\sum_x p_{\text{noisy}}(x)^2 \right] = q^{2n} \mathop{\mathbb{E}}_{U} \left[p_{\text{noisy}}(0^n)^2 \right] \tag{5.45}$$

are second-moment quantities (the second equality holds since by symmetry each term in the sum has the same value under expectation), with Z_w containing w copies of the noisy output and $2 - w$ copies of the ideal output for each $w \in \{0, 1, 2\}$. Note that $Z_0 = q^n Z$ with Z the collision probability from Chapter 4. Further, note that F is a free parameter, and we may choose it so that it minimizes the right-hand side of Eq. (5.42), which occurs when

$$\bar{F} = \frac{Z_1 - 1}{Z_0 - 1} \,, \tag{5.46}$$

which matches the definition for $\bar{F}$ in Eq. (5.14). Plugging in $F = \bar{F}$ yields

$$\mathop{\mathbb{E}}_{U} \left[\frac{1}{2} \|p_{\text{wn}} - p_{\text{noisy}}\|_1 \right] \leq \frac{1}{2} \bar{F} \sqrt{(Z_0 - 1) \left(\frac{(Z_0 - 1)(Z_2 - 1)}{(Z_1 - 1)^2} - 1 \right)} \,. \tag{5.47}$$

Now, we have reduced an assessment of the expected total variation distance to an expression of second-moment quantities Z_0, Z_1, and Z_2. We bound these quantities by mapping them to stochastic processes. These stochastic

processes are the same as that in Chapter 4, except that the noise channels introduce slightly modified transition rules, as we now discuss.

The takeaway from Chapter 4 was that the quantity $Z = q^{-n} Z_0$ could be expressed as a weighted sum over *trajectories* $\gamma = (\vec{\gamma}^{(0)}, \ldots, \vec{\gamma}^{(s)})$, where each $\vec{\gamma} \in \{I, S\}^n$. In this chapter, the framework of the "biased random walk" from Appendix 4.B is most helpful. The biased random walk describes a process for generating a trajectory γ. For each time step t, if the tth gate acts on qudits i_t and j_t, then the transition from $\vec{\gamma}^{(t-1)}$ to $\vec{\gamma}^{(t)}$ can involve a bit flip at position i_t, at position j_t, or neither (but not at both), and no bit can flip at any other position. Moreover, $\gamma_{i_t}^{(t)} = \gamma_{j_t}^{(t)}$ must hold, so if $\gamma_{i_t}^{(t-1)} \neq \gamma_{j_t}^{(t-1)}$, then one of the two bits *must* be flipped. In this situation, when one bit is assigned I and one is assigned S, the S is flipped to I with probability $q^2/(q^2+1)$, and the I is flipped to S with probability $1/(q^2+1)$. Thus, there is a bias toward making more of the assignments I. The quantity Z_0 is given exactly by the expectation value of $q^{|\vec{\gamma}^{(s)}|}$ when trajectories γ are generated in this fashion, where $|\vec{\nu}|$ denotes the Hamming weight of the bit string $\vec{\nu}$, that is, the number of S assignments out of n.

With the biased random walk now defined, a vital observation is that the random walk has two fixed points, the I^n configuration and the S^n configuration, since whenever all the bits agree, none can be flipped. In Chapter 4, we could precisely compute the fraction of the probability mass that eventually reaches each of these fixed points if the circuit is infinitely long. Specifically, $q^n/(q^n+1)$ of the probability mass converges to I^n and $1/(q^n+1)$ converges to S^n; however, since the S^n fixed point receives a weighting of q^n and the I^n fixed point receives a weighting of 1, we find that $Z_0 \to 2q^n/(q^n+1)$.

Noise complicates this process. Suppose the configuration after the tth two-qudit gate is $\vec{\nu}$, and a noise channel $\mathcal{N}$ acts on qudit i_t. Since the noise channel is unital, if $\nu_{i_t} = I$, then the configuration is left unchanged. However, if $\nu_{i_t} = S$, then the action of the noise may cause a flip from S to I. For the calculation of Z_0, there is no noise, so this happens with probability 0. For the calculation of Z_1, where there is one copy of the noisy distribution and one copy of the ideal, we show that this happens with probability $rq/(q-1)$, where r is the average infidelity given in Eq. (5.3). For depolarizing noise, $rq/(q-1) = \frac{\epsilon}{1-q^{-2}} = \gamma$. For Z_2, where there are two copies of the noisy distribution, the probability of a transition is $1-u$, where u is the unitarity of the noise channel given in Eq. (5.4).

Since noise can flip an S to an I but not vice versa, I^n is the only fixed point of the noisy biased random walk; the S^n fixed point is only metastable. Eventually, the action of noise will flip one of the S bits to an I, and the trajectory might re-equilibrate to the I^n fixed point.

Now, we consider a toy example which captures the essence of our analysis. Suppose a circuit consists of alternating rounds of (1) a global Haar-

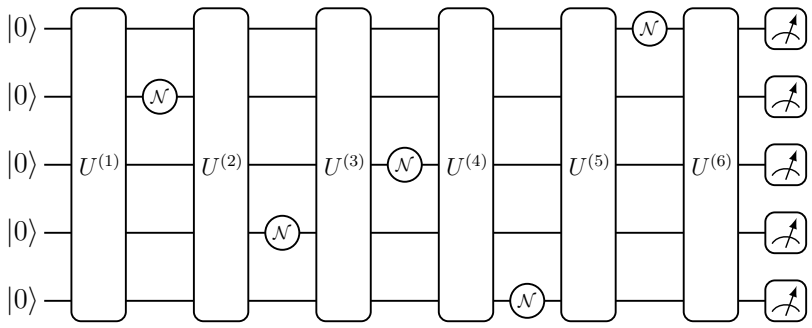


Figure 5.2: Toy example where global Haar-random gates $U^{(t)}$ act in between a depolarizing noise channel on a single qudit. In this model we can exactly compute quantities Z_0, Z_1, and Z_2 because the global Haar-random gates cause the probability mass in the stochastic process to fully re-equilibrate to one of the fixed points, I^n or S^n.

random transformation and (2) a depolarizing noise channel on a single qudit, as depicted in Figure 5.2. Step (1) can be approximately accomplished by performing a very large number of two-qudit gates. This model is similar to the toy model considered in Ref. [26] (the difference being that they considered single-qudit noise channels on all n qudits in step (2)), which they analyzed using the Pauli string method of Refs. [85, 131].

The initial global Haar-random transformation induces perfect equilibration to the two fixed points, with $q^n/(q^n+1)$ mass reaching the I^n fixed point and $1/(q^n+1)$ mass reaching the (metastable) S^n fixed point. This is already sufficient to compute $Z_0 - 1$, which is not sensitive to the noise.

$$Z_0 - 1 = \frac{q^n - 1}{q^n + 1}\,. \tag{5.48}$$

Now suppose we want to calculate Z_1, and that we are part of the $1/(q^n+1)$ fraction at the S^n fixed point. The single-qudit depolarizing noise channel will flip one of the S assignments to an I assignment with probability $\epsilon(1-q^{-2})^{-1}$. If this happens, there are $n-1$ S assignments and 1 I assignment. While it may seem that this new configuration is still close to the S^n fixed point, we must remember that the random walk is biased in the I direction. When we perform the global Haar-random transformation, we get perfect re-equilibration back to the two fixed points; with probability $\frac{1-q^{-2}}{1-q^{-2n}}$ we end at the I^n fixed point, and with probability $\frac{q^{-2}-q^{-2n}}{1-q^{-2n}}$ we end at the S^n fixed point. These probabilities were derived in Appendix 4.B. Now, the total mass that remains at the S^n fixed point is the $\frac{1}{q^n+1}(1-\frac{\epsilon}{1-q^{-2}})$ that never left and the $\frac{\epsilon}{1-q^{-2}}\frac{q^{-2}-q^{-2n}}{1-q^{-2n}}$ that left and returned, which comes out to $\frac{1}{q^n+1}(1-\frac{\epsilon}{1-q^{-2n}})$. After $2s$ single-qudit error channels have been applied, the probability mass remaining at the S^n

fixed point is precisely

$$\substack{\text{probability mass at } S^n \\ \text{after } 2s \text{ noise locations}} = \frac{1}{q^n+1}\left(1-\frac{\epsilon}{1-q^{-2n}}\right)^{2s} \approx \frac{1}{q^n+1}e^{-2\epsilon s}\,. \tag{5.49}$$

This mass receives weighting of q^n toward Z_1. Meanwhile the rest of the mass is at the I^n fixed point and receives weighting of 1. This tells us

$$Z_1 - 1 = \frac{q^n-1}{q^n+1}\left(1-\frac{\epsilon}{1-q^{-2n}}\right)^{2s}\,. \tag{5.50}$$

Calculating Z_2-1 is just as easy. Here transitions due to noise occur with probability $1-u$ where u is the unitarity of the noise channel. For depolarizing noise $1-u = 2\epsilon(1-q^{-2})^{-1} - O(\epsilon^2)$, so Z_2-1 is the same as Z_1-1 with the replacement $\epsilon \to 2\epsilon - O(\epsilon^2)$, giving

$$Z_2-1 = \frac{q^n-1}{q^n+1}\left(1-\frac{2\epsilon}{1-q^{-2n}}+O(\epsilon^2)\right)^{2s} = \frac{q^n-1}{q^n+1}\left(1-\frac{\epsilon}{1-q^{-2n}}\right)^{4s} e^{O(s\epsilon^2)}\,. \tag{5.51}$$

We can plug these calculations into Eq. (5.47) to find that

$$\mathop{\mathbb{E}}_{U}\left[\frac{1}{2}\|p_{\text{wn}} - p_{\text{noisy}}\|_1\right] \le \frac{1}{2}\bar{F}\sqrt{\frac{q^n-1}{q^n+1}\left(e^{O(\epsilon^2 s)}-1\right)} = O(\bar{F}\epsilon\sqrt{s})\,. \tag{5.52}$$

In the proofs of our theorems, the difficulty is that the probability mass does not fully equilibrate to a fixed point before the next error location acts. Nonetheless, we manage to calculate tight bounds on Z_1 and Z_2 by keeping track throughout the evolution of the amount of probability that the walk *would* re-equilibrate back to S^n and I^n if the rest of the gates were noiseless, which we refer to as S-destined and I-destined probability mass. We show that, as long as $\epsilon < c/n$ for some constant c, the S-destined probability mass is exponentially clustered near the S^n fixed point in the sense that the probability of being x bit flips away from S^n conditioned on being S-destined decays exponentially in x. Thus, nearly all the bits are still assigned S and the action of a noise channel reduces the S-destined mass by a factor of roughly $1-\epsilon$. If, on the other hand, a substantial number of bits were assigned I, the noise would cause a fewer number of flips and the fraction of the S-destined mass that stays S-destined after each noise channel would be larger than $1-\epsilon$, which ruins the analysis.

The reason $\epsilon < c/n$ is required is that we need errors to be rare enough that the S-destined mass *mostly* re-equilibrates back to S^n before new errors pop up; the errors must get scrambled at a faster rate than they appear. If a configuration has $n-1$ S assignments and 1 I assignment, it will take $O(n)$ gates before the single I-assigned qudit participates in a gate. Thus, if errors

occur at a slower rate than one per $O(n)$ gates, full re-equilibration will happen before a new error most of the time. It is not clear if this condition is truly necessary for the statement to hold. However, we need $\epsilon < c/n$ for another reason: we need the fidelity to decay more slowly than the anti-concentration rate. After all, even though the walk is I-biased, the I-destined mass does not make it to the I fixed point instantaneously. After s gates, there will be some residual contribution from the I-destined mass which decays by a constant factor every $O(n)$ gates. Thus, the fidelity must also decrease by at most a constant factor for every $O(n)$ gates, meaning $\epsilon < c/n$. In our analysis, we actually settle for something a bit weaker: we require that $\epsilon \ll 1/(n \log(n))$, which essentially means that very few errors occur *during* the initial anti-concentration period. However, this is done to make the analysis easier, and we do not believe this condition is necessary.

5.6 Outlook

In this chapter, we have presented a comprehensive picture of how the output distribution of typical random quantum circuits behaves under a weak incoherent local noise model. As more gates are applied, the output distribution decays toward the uniform distribution in total variation distance like $e^{-2\epsilon s}$ where ϵ is the local noise strength in a Pauli error model (for non-Pauli models, this can be expressed in terms of the average infidelity r) and s is the number of gates, confirming strong intuition that had not previously been rigorously proven. Moreover, we show that the convergence to uniform happens in a very special way: the residual non-uniform component of the noisy distribution is approximately in the direction of the ideal distribution. The random quantum circuits scramble the errors that occur locally during the evolution so that they can ultimately be treated as global white noise, allowing some signal of the ideal computation to be extracted even from a noisy device. While this property had previously been conjectured—it was an underlying assumption of quantum computational supremacy experiments [6, 7]—it had not received rigorous analytical study. Basic questions like how the error in the white-noise approximation scales with ϵ and s had not been investigated.

Our theorem statements are given for general, possibly coherent, noise channels. While we show that local coherent noise channels lead the output distribution to exhibit exponential decay in the linear cross-entropy benchmark for the fidelity, there is not generally also a decay toward the uniform distribution. As a result, the white-noise approximation is not good for coherent noise channels. Moreover, even for incoherent noise channels, our technical statements are only applicable if the Pauli noise strength ϵ (or for non-Pauli noise channels, the average infidelity) is beneath a threshold that shrinks with system size like $O(1/n)$ and if the circuit size is at least $\Omega(n \log(n))$. Furthermore, our bound on error in the white-noise approximation is only meaningful if $\epsilon \ll 1/(n \log(n))$. We believe the $\epsilon \ll 1/(n \log(n))$ requirement is merely a result of suboptimal analysis, but that the assumption $\epsilon < O(1/n)$ is funda-

mentally necessary for the approximation to be good; the errors are meaningfully transformed into white noise only if they can be scrambled faster than the fidelity $F = e^{-2\epsilon s}$ decays.

One implication of our result is to put low-fidelity random-circuit-based quantum computational supremacy experiments on stronger theoretical footing by showing that, as long as our local noise model is a reasonable approximation of noise in actual devices, the device produces samples from a well-understood output distribution, which can subsequently be argued is hard to classically sample. Indeed, in Appendix 5.C, we combine observations from previous work to show that the task of classically sampling from the white-noise distribution with fidelity F up to ηF error is essentially just as hard, in a certain complexity theoretic sense, as the task of classically sampling from the ideal distribution up to a $O(\eta)$ error. This is important because the latter task (and variants of it in other computational models [19, 22]), which we called "approximate RCS" in Chapter 1, has previously garnered significant theoretical scrutiny [4, 25, 26], although it is still not known whether it is hard in a formal complexity theoretic sense.

These results are good news for the utility of NISQ devices more broadly. In order to perform a larger and more interesting computation, noise rates must become smaller; our work shows that, in many applications, for circuits with s gates, noise rates need only decrease like $1/\sqrt{s}$, rather than $1/s$, as long as one is willing to repeat the experiment many times to extract the signal from the global white noise. A natural next question is when, besides the case of random quantum circuits, do we expect a similar white-noise phenomenon to occur? Our result shows that convergence to white-noise is a *generic* property, occurring for a large fraction of randomly chosen circuits. Heuristically, this is because random quantum circuits are known to be good scramblers. However, most interesting quantum circuits are non-generic in some way. An extreme example is quantum error-correcting circuits, which are specifically designed *not* to scramble errors (so that they can be corrected). The output of these circuits will not be close to the white-noise distribution. A fascinating follow-up question is whether other computations proposed for NISQ devices appear to scramble errors well enough that a similar approximation can be made. One leading candidate with relevance for many-body physics is circuits that simulate evolution by fixed chaotic Hamiltonians, since these systems are thought to scramble information efficiently. Indeed, random quantum circuits have been studied as a more analytically tractable model of such systems, precisely because of the similarity in their scrambling properties [41, 42, 87].

We conclude with a couple of comments on ways this analysis could be improved.

- We believe that a sharper analysis would allow the error term of the form $O(\epsilon s_{AC})$ to be omitted, as long as the error rate is a sufficiently small fraction of $1/n$. Can this be proven using our method?

- What are the actual constant prefactors of the error in the approximation? This would be relevant for knowing in practice how small one must make the error in order to assert that the white-noise approximation holds. We have not made significant attempt to optimize these factors; it might be possible to estimate them numerically.

- We showed that the classical output distribution over measurement outcomes is close to a mixture of the ideal output and the uniform output. A stronger statement would be that the output quantum state is a mixture of the ideal output state and the maximally mixed state. Our method struggles to prove this stronger claim owing to our usage of the 1-norm to 2-norm inequality and a difference that appears in the quantum case. The ideal output quantum state is a pure state, so both its 1-norm and its 2-norm are equal to 1. However, the ideal output classical distribution is expected to look like a Porter-Thomas distribution; its 1-norm is 1, but its 2-norm is $O(q^{-n/2})$. Thus, the 1-norm to 2-norm bound in Eq. (5.39) is tight in the classical case but not the quantum case, implying that we are unlikely to get tight bounds if we attempt to examine the output quantum state using an identical strategy.

APPENDIX TO CHAPTER 5

5.A Framework for noisy circuit analysis

We refer the reader back to Appendix 4.B, where the expected squared output probabilities of a noiseless quantum circuit are expressed as a sum over trajectories. Here we augment this by an analysis of the action of single-qudit noise channels.

5.A.1 Action of averaged noise channel on identity and swap

Since every single-qudit noise channel is followed by a Haar-random (either single-qudit or two-qudit) gate in the circuit diagram, we are free to add a single-qudit Haar-random gate immediately after every noise channel without changing the overall circuit ensemble (the Haar measure is invariant under multiplication by any unitary). Denote this single-qudit Haar-random matrix by V. There will be a difference in the analysis between the calculation of Z_0, Z_1 and Z_2, where Z_w contains w copies of the noisy output. Define

$$\mathcal{N}_0 = I \otimes I \tag{5.53}$$
$$\mathcal{N}_1 = I \otimes \mathcal{N} \tag{5.54}$$
$$\mathcal{N}_2 = \mathcal{N} \otimes \mathcal{N} \tag{5.55}$$

with I the single-qudit identity. Let ρ be a state on two copies of a single-qudit Hilbert space. Then for $w \in \{0, 1, 2\}$, let

$$N_w[\rho] = \mathop{\mathbb{E}}_{V} \left[V^{\otimes 2} \, \mathcal{N}_w(\rho) V^{\dagger \otimes 2} \right] \tag{5.56}$$

be the Haar-averaged noise channel.

We will only need to compute the action of N_w on input states $\rho = I$ (here I is the two-qudit identity) or $\rho = S$ since the random gates turn the initial state $|0^n\rangle\langle 0^n|$ into a linear combination of tensor products of I or S on each qudit. Note that since $\mathcal{N}$ is assumed to be unital, we have

$$N_w[I] = I \tag{5.57}$$

for all $w \in \{0, 1, 2\}$. However, computing the action on S is not as simple. Let

$$Y_w = \operatorname{tr}\left(S \mathcal{N}_w(S)\right) . \tag{5.58}$$

(Note that $Y_0 = q^2$ since $\mathcal{N}_0$ is the identity channel.) Then, use Eq. (4.45) and the fact that $\mathcal{N}$ is trace-preserving to show

$$N_w[S] = \frac{q - q^{-1} Y_w}{q^2 - 1} I + \frac{Y_w - 1}{q^2 - 1} S \, . \tag{5.59}$$

Now we relate the quantities Y_1 and Y_2 to the average infidelity and the unitarity, respectively. Recall that $\mathrm{tr}(AB) = \mathrm{tr}(S(A \otimes B))$. Using this trick and Eq. (4.45), the average infidelity from Eq. (5.3), can be evaluated as follows:

$$r = 1 - \int dV \, \mathrm{tr}\left[V|\psi\rangle\langle\psi|V^\dagger \mathcal{N}(V|\psi\rangle\langle\psi|V^\dagger)\right] \tag{5.60}$$

$$= 1 - \int dV \, \mathrm{tr}\left[S\left(V|\psi\rangle\langle\psi|V^\dagger \otimes \mathcal{N}(V|\psi\rangle\langle\psi|V^\dagger)\right)\right] \tag{5.61}$$

$$= 1 - \int dV \, \mathrm{tr}\left[S(I \otimes \mathcal{N})\left(\left(V|\psi\rangle\langle\psi|V^\dagger\right)^{\otimes 2}\right)\right] \tag{5.62}$$

$$= 1 - \mathrm{tr}\left[S\mathcal{N}_1\left(\frac{I+S}{q(q+1)}\right)\right] \tag{5.63}$$

$$= 1 - \frac{1 - q^{-1}Y_1}{q+1} = \frac{q - q^{-1}Y_1}{q+1}\,. \tag{5.64}$$

The unitarity from Eq. (5.4), can be evaluated in a similar way.

$$u = \frac{q}{q-1}\left(\int dV \, \mathrm{tr}\left[\mathcal{N}\left(V|\psi\rangle\langle\psi|V^\dagger\right)^2\right] - \frac{1}{q}\right) \tag{5.65}$$

$$= \frac{q}{q-1}\int dV \, \mathrm{tr}\left[S\left(\mathcal{N}\left(V|\psi\rangle\langle\psi|V^\dagger\right)\right)^{\otimes 2}\right] - \frac{1}{q-1} \tag{5.66}$$

$$= \frac{q}{q-1}\int dV \, \mathrm{tr}\left[S(\mathcal{N}\otimes\mathcal{N})\left(\left(V|\psi\rangle\langle\psi|V^\dagger\right)^{\otimes 2}\right)\right] - \frac{1}{q-1} \tag{5.67}$$

$$= \frac{q}{q-1}\,\mathrm{tr}\left[S\mathcal{N}_2\left(\frac{I+S}{q(q+1)}\right)\right] - \frac{1}{q-1} \tag{5.68}$$

$$= \frac{q+Y_2}{(q-1)(q+1)} - \frac{1}{q-1} \tag{5.69}$$

$$= \frac{Y_2 - 1}{q^2-1}\,. \tag{5.70}$$

Plugging these relations back into Eq. (5.59) gives us

$$N_0[S] = S \tag{5.71}$$

$$N_1[S] = \frac{r}{q-1}I + \left(1 - \frac{qr}{q-1}\right)S \tag{5.72}$$

$$N_2[S] = \frac{1-u}{q}I + uS\,. \tag{5.73}$$

For weak noise channels, r is close to 0 and u is close to 1. In this case, we see that the noise causes some small amount of leakage from the S state to the I state, but no leakage from the I state to the S state, introducing an asymmetry into the problem that did not exist in the noiseless analysis.

For $t = 1, \ldots, s$, let $N_w^{(t)} = I_{[n]\setminus\{i_t\}} \otimes N_{w,\{i_t\}}$ be the channel that acts with the averaged noise channel on site i_t and identity elsewhere, and let

$N_w^{'(t)} = I_{[n]\setminus\{j_t\}} \otimes N_{w,\{j_t\}}$ be the same for site j_t. For $t \leq 0$ and $t > s$, let $N_w^{(t)}$ be the identity channel. If ρ is a linear combination of tensor products of I and S, $N_w^{(t)}(\rho)$ and $N_w^{'(t)}(\rho)$ will be as well, with coefficients that transform linearly. For configurations $\vec{\gamma}, \vec{\nu} \in \{I, S\}^n$, let $N^{(t)}_{w,\vec{\nu}\vec{\gamma}}$ denote the matrix elements of this transformation, that is

$$N_w^{(t)}\left[\bigotimes_{j=0}^{n-1}\gamma_j\right] = \sum_{\vec{\nu}\in\{I,S\}^n} N^{(t)}_{w,\vec{\nu}\vec{\gamma}} \bigotimes_{j=0}^{n-1}\nu_j\,, \tag{5.74}$$

where for $1 \leq t \leq s$,

$$N^{(t)}_{0,\vec{\nu}\vec{\gamma}} = \begin{cases} 1 & \text{if } \vec{\gamma} = \vec{\nu} \\ 0 & \text{otherwise} \end{cases} \tag{5.75}$$

$$N^{(t)}_{1,\vec{\nu}\vec{\gamma}} = \begin{cases} 1 & \text{if } \gamma_{i_t} = \nu_{i_t} = I \text{ and } \vec{\gamma} = \vec{\nu} \\ 1 - \frac{qr}{q-1} & \text{if } \gamma_{i_t} = S \text{ and } \nu_{i_t} = S \text{ and } \vec{\gamma} = \vec{\nu} \\ \frac{r}{q-1} & \text{if } \gamma_{i_t} = S \text{ and } \nu_{i_t} = I \text{ and } \gamma_a = \nu_a \forall a \neq i_t \\ 0 & \text{otherwise} \end{cases} \tag{5.76}$$

$$N^{(t)}_{2,\vec{\nu}\vec{\gamma}} = \begin{cases} 1 & \text{if } \gamma_{i_t} = \nu_{i_t} = I \text{ and } \vec{\gamma} = \vec{\nu} \\ u & \text{if } \gamma_{i_t} = S \text{ and } \nu_{i_t} = S \text{ and } \vec{\gamma} = \vec{\nu} \\ \frac{1-u}{q} & \text{if } \gamma_{i_t} = S \text{ and } \nu_{i_t} = I \text{ and } \gamma_a = \nu_a \ \forall a \neq i_t \\ 0 & \text{otherwise}\,, \end{cases} \tag{5.77}$$

and $N_w^{'(t)}$ are given by the same equations, with j_t replacing i_t.

5.A.2 Mapping to partition function

Define

$$\mathcal{U}_0^{(t)} = \mathcal{U}^{(t)} \otimes \mathcal{U}^{(t)} \tag{5.78}$$

$$\mathcal{U}_1^{(t)} = \widetilde{\mathcal{U}}^{(t)} \otimes \mathcal{U}^{(t)} \tag{5.79}$$

$$\mathcal{U}_2^{(t)} = \widetilde{\mathcal{U}}^{(t)} \otimes \widetilde{\mathcal{U}}^{(t)}\,, \tag{5.80}$$

where $\mathcal{U}^{(t)}$ and $\widetilde{\mathcal{U}}^{(t)}$ are given in Eqs. (5.8) and (5.9). Then we may write, for $w \in \{0, 1, 2\}$

$$Z_w = q^{2n} \mathop{\mathbb{E}}_{U} \left[\mathrm{tr}\left[|0^n\rangle\langle 0^n|^{\otimes 2}\, \mathcal{U}_w^{(n+s)} \circ \cdots \circ \mathcal{U}_w^{(-n+1)}\left(|0^n\rangle\langle 0^n|^{\otimes 2}\right)\right]\right]. \tag{5.81}$$

Since each $U^{(t)}$ is chosen independently, we are free to perform the expectation value individually over each $\mathcal{U}_w^{(t)}$ channel. The noiseless channel $\mathcal{U}_0^{(t)} = \mathcal{U}^{(t)\otimes 2}$ averages to $M^{(t)}$, where $M^{(t)}$ is given in Eq. (4.49). The action of the noise may also be averaged, since, as discussed above, we may pull out a single-qudit Haar random gate to act after each noise location. Thus, the noiseless

single qudit gates at the end of the circuit may be dropped as they are being absorbed into the noise. Let

$$M_w^{(t)} = N_w^{\prime(t)} \circ N_w^{(t)} \circ M^{(t)} \tag{5.82}$$

so that

$$Z_w = q^{2n} \operatorname{tr}\left[|0^n\rangle\langle 0^n|^{\otimes 2} \; M_w^{(s)} \circ \cdots \circ M_w^{(-n+1)} \left(|0^n\rangle\langle 0^n|^{\otimes 2} \right) \right] . \tag{5.83}$$

Following Eq. (4.48), we have

$$M_w^{(0)} \circ \cdots \circ M_w^{(-n+1)} \left(|0^n\rangle\langle 0^n|^{\otimes 2} \right) = \frac{1}{q^n (q+1)^n} \sum_{\vec{\gamma} \in \{I,S\}^n} \bigotimes_{j=0}^{n-1} \gamma_j \, , \tag{5.84}$$

and thus

$$Z_w = \frac{q^n}{(q+1)^n} \sum_{\gamma \in \{I,S\}^{n \times (3s+1)}} \prod_{t=1}^{s} N^{\prime(t)}_{w,\vec{\gamma}^{(t)}\vec{\gamma}^{(t-1/3)}} N^{(t)}_{w,\vec{\gamma}^{(t-1/3)}\vec{\gamma}^{(t-2/3)}} M^{(t)}_{\vec{\gamma}^{(t-2/3)}\vec{\gamma}^{(t-1)}} \tag{5.85}$$

$$=: \frac{q^n}{(q+1)^n} \sum_{\gamma} \operatorname{weight}_w(\gamma) \, , \tag{5.86}$$

generalizing Eq. (4.59). We see, as was the case for Z in Appendix 4.B, Z_w is given by a weighted sum over trajectories, which can be interpreted as a partition function on the Ising-like stat mech model where each $\gamma_a^{(t)}$ is an Ising variable $\{+1, -1\}$. In the noisy case, these trajectories are of length $3s + 1$ (labeled as $t = 0, 1/3, 2/3, 1, \ldots, s$) instead of length $s + 1$, to account for the impact of the two single-qudit noise channels after each gate.

5.A.3 Modified biased random walk

For the noiseless case, we biased the random walk so that the weight of a trajectory only depended on its start and end points. The same will work for the noisy case. Inspection of Eqs. (5.76) and (5.77) reveals that a flip from S to I also comes with a loss of a factor of q. This is fundamentally a consequence of the trace-preserving property of the noise (since $\operatorname{tr}(I) = q^2$ but $\operatorname{tr}(S) = q$). The biased random walk from Eq. (4.63) has a built-in q^x weighting, where x is the number of S assignments in the final configuration. Thus, a flip from S to I, which decreases x by 1, fits naturally into the framework of the biased random walk.

We can say

$$Z_w = \frac{q^n}{(q+1)^n} \sum_{\vec{\gamma}^{(0)}} q^{-|\vec{\gamma}^{(0)}|} \mathop{\mathbb{E}}_{P_{b,w},\vec{\gamma}^{(0)}} \left[q^{|\vec{\gamma}^{(s)}|} \right] , \tag{5.87}$$

where $\mathbb{E}_{P_{b,w},\vec{\gamma}^{(0)}}$ denotes expectation over the dynamics of a biased random walk starting from configuration $\vec{\gamma}^{(0)}$. These dynamics are as follows. Each

step consists first of the same step from the biased random walk in Chapter 4: if a gate acts on qudits $\{i_t, j_t\}$, then one of the two bits is flipped in the case that they differ at those positions; the I is flipped to an S with probability $1/(q^2+1)$ and S to I with probability $q^2/(q^2+1)$. Then, if the assignment at position i_t and j_t is S, each is flipped to I independently with a certain probability. For Z_0 that probability is 0 (noiseless), for Z_1, it is $rq/(q+1)$, and for Z_2 it is $(1-u)$. We can see that any configuration γ has the same weight in Eq. (5.86) as it contributes to the expectation value above.

Defining Λ_b to be the initial configuration over configurations that weights a configuration $\vec{\nu}$ proportional to $q^{-|\vec{\nu}|}$, we may rewrite this as

$$Z_w = \mathop{\mathbb{E}}_{P_{b,w},\Lambda_b} \left[q^{|\vec{\gamma}^{(s)}|} \right] . \tag{5.88}$$

5.A.4 Matrix notation for the modified biased random walk

We can write this in slightly more precise notation, as follows. Consider a 2^n-dimensional vector space vector space, where basis states are labeled by configurations $|\vec{\nu}\rangle$ for each $\vec{\nu} \in \{I, S\}^n$. We will use the association $I \leftrightarrow 0$, $S \leftrightarrow 1$ and label these basis states explicitly by bit strings in $\{0,1\}^n$. Define the vectors

$$|\mathbf{1}\rangle = \sum_{\vec{\nu}} |\vec{\nu}\rangle \tag{5.89}$$

$$|\mathbf{q}\rangle = \sum_{\vec{\nu}} q^{|\vec{\nu}|} |\vec{\nu}\rangle \tag{5.90}$$

$$|\Lambda_b\rangle = \frac{1}{(q+1)^n} \sum_{\vec{\nu}} q^{n-|\vec{\nu}|} |\vec{\nu}\rangle . \tag{5.91}$$

Then we may define $2^n \times 2^n$ transition matrices $P^{(t)}$, which enacts the tth step of the noiseless biased walk, as well as matrices $Q_\sigma^{(t)}$ and $Q_\sigma^{\prime(t)}$ which enact the $S \to I$ transition with probability σ on qudits i_t and j_t, respectively. Explicitly, we let

$$P^{(t)} = I_{[n]\setminus\{i_t,j_t\}} \otimes P_{\{i_t,j_t\}} \tag{5.92}$$

$$Q_\sigma^{(t)} = I_{[n]\setminus\{i_t\}} \otimes \Big(|0\rangle\langle 0| + (1-\sigma)|1\rangle\langle 1| + \sigma |0\rangle\langle 1| \Big)_{\{i_t\}} \tag{5.93}$$

$$Q_\sigma^{\prime(t)} = I_{[n]\setminus\{j_t\}} \otimes \Big(|0\rangle\langle 0| + (1-\sigma)|1\rangle\langle 1| + \sigma |0\rangle\langle 1| \Big)_{\{j_t\}} , \tag{5.94}$$

where

$$D = |00\rangle\langle 00| + |11\rangle\langle 11| \tag{5.95}$$

$$T = \frac{q^2}{q^2+1}|00\rangle\langle 01| + \frac{q^2}{q^2+1}|00\rangle\langle 10| + \frac{1}{q^2+1}|11\rangle\langle 10| + \frac{1}{q^2+1}|11\rangle\langle 01| \tag{5.96}$$

$$P = D + T . \tag{5.97}$$

Note that P is a stochastic matrix. Then, define

$$\mathcal{Z}_\sigma = \langle \mathbf{q} | \left(\prod_{t=1}^{s} Q_\sigma'^{(t)} Q_\sigma^{(t)} P^{(t)} \right) |\Lambda_b\rangle \,, \tag{5.98}$$

where $|\Lambda_b\rangle$ is the initial distribution vector for the biased walk. Note that if the circuit diagram is generated randomly, as is the case for the complete-graph architecture, then $\mathcal{Z}_\sigma$ is the mean of the above expression over choice of circuit diagram.

The above equation for Z_w implies that

$$Z_0 = \mathcal{Z}_0 \tag{5.99}$$
$$Z_1 = \mathcal{Z}_{rq/(q-1)} \tag{5.100}$$
$$Z_2 = \mathcal{Z}_{1-u} \,. \tag{5.101}$$

5.B Detailed proofs

First, we state the definition of a layered architecture and two main lemmas, which are themselves dependent on more minor lemmas. Then, we prove our theorems based on the main lemmas. Afterward, we develop some more machinery and state the minor lemmas, deferring their proofs to Appendix 5.B.8.

5.B.1 Definitions and main lemmas

Our proofs apply to h-regularly connected (Definition 4.5), layered architectures. Here we define layered, which just means that you can always arrange the gates neatly into layers of $n/2$ non-overlapping gates.

Definition 5.1. *An architecture is layered if any sequence of gates $(A^{(1)}, \ldots, A^{(s)})$ it generates with non-zero probability has the property that for any integer $d \geq 0$, and any pair of gates in the same "layer"*

$$t_1, t_2 \in \{dn/2 + 1, dn/2 + 2, \ldots, (d+1)n/2\} \tag{5.102}$$

with $t_1 \neq t_2$, we have $A^{(t_1)} \cap A^{(t_2)} = \emptyset$. Thus, all n qudits are acted upon by exactly one gate out of every $n/2$ gates.

For layered architectures we can speak clearly about the depth $d = 2s/n$. Typically we always require s be a multiple of $n/2$ so that there are an integer number of layers. Regular lattice architectures in D spatial dimensions are typically layered, although adhering strictly to the definition might require applying periodic boundary conditions. We do not expect this condition is actually necessary for our results, but it is analytically convenient. The only place we need it is in Lemma 5.12.

Our theorems are corollaries of the following lemmas. Recall the definition of $\mathcal{Z}_\sigma$ from Eq. (5.98). Note that in these proofs, all constants are dependent

on q as well as h (the regularly connected parameter), but independent of n and the noise parameters.

Lemma 5.1. *If the random quantum circuit architecture is h-regularly connected and layered with anti-concentration depth d_{AC}, then there exist constants c_0, c_1, c_2, c_3, c_4, c_5, and n_0 that depend on h and q but not on n or σ, such that as long as $\sigma \leq c_5/n$ and $n \geq n_0$, for any value of the circuit depth d,*

$$\frac{q^n-1}{q^n+1}\left(1-f_\sigma\right)^d \leq \mathcal{Z}_\sigma - 1 \leq \frac{q^n-1}{q^n+1}\left(1-f_\sigma\right)^d e^{K_\sigma}\,, \tag{5.103}$$

where

$$f_\sigma = \frac{1-(1-\sigma(1-q^{-2}))^n}{1-q^{-2n}} \tag{5.104}$$

$$K_\sigma = c_0 nd\sigma^2 + c_1 n\sigma d_{AC} + c_2 e^{-c_3(d-d_{AC})+2\sigma dn} + c_4 n\sigma \log(1/(n\sigma))\,. \tag{5.105}$$

Proof. The lower bound is an immediate consequence of two lemmas that appear later, Lemma 5.11 and Lemma 5.12. The upper bound is also an immediate consequence, with the constant c_1 absorbing an $O(n\sigma)$ term since $d_{AC} = 2s_{AC}/n \geq \Omega(\log(n))$ by the results of Chapter 4. □

We show the analogous statement for the complete-graph architecture.

Lemma 5.2. *If the random quantum circuit architecture is the complete-graph architecture, then there exist constants c_0', c_1', c_2', c_3', c_4', c_5', and n_0 that depend on q but not on n or σ, such that as long as $\sigma \leq c_5'/n$ and $n \geq n_0$, for any value of the circuit size s,*

$$\frac{q^n-1}{q^n+1}\left(1-f'_\sigma\right)^s \leq \mathcal{Z}_\sigma - 1 \leq \frac{q^n-1}{q^n+1}\left(1-f'_\sigma\right)^s e^{K'_\sigma}\,, \tag{5.106}$$

where

$$f'_\sigma = \frac{1-(1-\sigma(1-q^{-2}))^2}{1-q^{-2n}} \tag{5.107}$$

$$K'_\sigma = c_0' s\sigma^2 + c_1' \sigma s_{AC} + c_2' e^{-c_3'(s-s_{AC})/n+4\sigma s} + c_4' n\sigma \log(1/(n\sigma))\,, \tag{5.108}$$

and $s_{AC} = \Theta(n\log(n))$ is the anti-concentration size for the complete-graph architecture.

Proof. The proof is the same as Lemma 5.1 except using Lemma 5.13 in place of Lemma 5.12. □

Note that in the regime $\sigma \leq O(1/n)$, we can bound $1 - \sigma(1 - q^{-2}) \geq e^{-\sigma(1-q^{-2})} e^{-O(\sigma^2)}$ and the following holds:

$$e^{-n\sigma(1-q^{-2})} e^{-O(n\sigma^2)-O(q^{-2n})} \leq 1 - f_\sigma \leq e^{-n\sigma(1-q^{-2})} \tag{5.109}$$

$$e^{-2\sigma(1-q^{-2})} e^{-O(\sigma^2)-O(q^{-2n})} \leq 1 - f'_\sigma \leq e^{-2\sigma(1-q^{-2})}\,. \tag{5.110}$$

The upper bound in Eqs. (5.109) and (5.110) actually holds generally for all σ.

5.B.2 Proofs of main theorems from main lemmas

Proof of Theorem 5.1: fidelity decay

Proof. The quantity $\bar{F}$ is precisely $(Z_1 - 1)/(Z_0 - 1) = (\mathcal{Z}_\sigma - 1)/(\mathcal{Z}_0 - 1)$ with $\sigma = rq/(q-1)$. The statements are then direct consequences of Lemma 5.1 for layered architectures and Lemma 5.2 for the complete-graph architecture, combined with the observation in Eqs. (5.109) and (5.110). Note also that $nd = 2s$. □

Proof of Theorem 5.2: convergence to the uniform distribution

Proof. We can use the 1-norm to 2-norm inequality in Eq. (5.39), along with Jensen's inequality for the concave $\sqrt{\cdot}$ function to say

$$\mathop{\mathbb{E}}_{U}\left[\frac{1}{2}\|p_{\text{noisy}} - p_{\text{unif}}\|_1\right] \leq \frac{1}{2}\sqrt{q^n \mathop{\mathbb{E}}_{U}\left[\sum_x \left(p_{\text{noisy}}(x) - q^{-n}\right)^2\right]} \tag{5.111}$$

$$= \frac{1}{2}\sqrt{q^{2n} \mathop{\mathbb{E}}_{U}\left[p_{\text{noisy}}(0^n)^2\right] - 1} = \frac{1}{2}\sqrt{Z_2 - 1} \tag{5.112}$$

$$= \frac{1}{2}\sqrt{\mathcal{Z}_v - 1} \tag{5.113}$$

Then, the theorem follows from the upper bound in Lemma 5.1 for layered architectures and Lemma 5.2 for the complete-graph architecture, with $\sigma = v$, combined with the observation in Eqs. (5.109) and (5.110). Note also that $nd = 2s$. □

Proof of Theorem 5.3: approximation by white noise

Proof. Following Section 5.5.2, we first use the 1-norm to 2-norm bound and Jensen's inequality, and then we optimize the value of F. The bound on the distance between p_{wn} and p_{noisy} is minimized when we choose $F = \bar{F} = (Z_1 - 1)/(Z_0 - 1)$. When this value is chosen, the bound can be expressed as

$$\mathop{\mathbb{E}}_{U}\left[\frac{1}{2}\|p_{\text{noisy}} - p_{\text{wn}}\|_1\right] \leq \frac{1}{2}\bar{F}\sqrt{\frac{(Z_2 - 1)(Z_0 - 1)^2}{(Z_1 - 1)^2} - (Z_0 - 1)} \tag{5.114}$$

Note that after the anti-concentration size has been surpassed, the quantity $Z_0 - 1$ rapidly approaches $\frac{q^n-1}{q^n+1} \approx 1$ from above. To evaluate Z_0, Z_1 and Z_2 we use the correspondence $Z_0 = \mathcal{Z}_0$, $Z_1 = \mathcal{Z}_{rq/(q-1)}$ and $Z_2 = \mathcal{Z}_v$. The bounds from Lemma 5.1 for layered architectures and Lemma 5.2 for the complete-graph architecture then allow us to upper bound $(Z_2 - 1)(Z_0 - 1)^2/(Z_1 - 1)^2$, arriving at

$$\begin{aligned}
&\frac{(Z_2 - 1)(Z_0 - 1)^2}{(Z_1 - 1)^2} \\
&\leq \frac{q^n - 1}{q^n + 1} e^{2s\left(2r(1+q^{-1}) - v(1-q^{-2})\right)} e^{O(sr^2) + O(sq^{-2n}) + e^{O(s_{AC}/n)} e^{-\Omega(s/n)}} Q_2 \qquad (5.115) \\
&= \frac{q^n - 1}{q^n + 1} e^{2s\delta} e^{O(sr^2 + sq^{-2n} + sv^2 + s_{AC}v - nv\log(nv)) + e^{O(s_{AC}/n)} e^{-\Omega(s/n)}}, \qquad (5.116)
\end{aligned}$$

where Q_2 is given in Eq. (5.20), and δ is given in Eq. (5.21). Now, working back from Eq. (5.114), and noting that $e^x - 1 < 2x$ for all $x \leq 1$, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{U} \left[\frac{1}{2} \| p_{\text{noisy}} - p_{\text{wn}} \|_1 \right] \\
&\leq \frac{\bar{F}}{2} \sqrt{4s\delta + O(sr^2 + sq^{-2n} + sv^2 + s_{AC}v - nv\log(nv)) + e^{O(s_{AC}/n)} e^{-\Omega(s/n)}} \qquad (5.117) \\
&= \bar{F}\sqrt{s} \left(\sqrt{\delta} + O(v) + O(r) \right) + O(\bar{F}\sqrt{s_{AC}v}) \\
&\qquad + O(\bar{F}\sqrt{nv\log(1/nv)}) + \bar{F} e^{O(s_{AC}/n) - \Omega(s/n)} \qquad (5.118)
\end{aligned}$$

when the quantity under the square root is less than 1 (and using $\sqrt{A+B} \leq \sqrt{A} + \sqrt{B}$). □

5.B.3 Machinery for proof

We now develop some more notation, and we precisely state our lemmas. We defer the proofs of these lemmas to Appendix 5.B.8. As we state them, we attempt to give some commentary about the meaning and purpose of the different lemmas.

Coupling a noiseless and noisy copy of the dynamics

We have a fairly good understanding of the noiseless random walk from Chapter 4. Our strategy here is to try to examine how introducing noise perturbs that walk. To that end, we consider *two* copies of the random walk, where one is noiseless and one is noisy, but they are correlated so that we can isolate the impact of the noise.

Recall that we have reduced the calculation of $\mathcal{Z}_\sigma$ to the expectation value of a random variable (the configuration) that evolves according to the stochastic transition matrix $P^{(t)}$ (representing the noiseless gate) followed by transition matrices $Q_\sigma^{(t)}$ and $Q_\sigma^{\prime(t)}$, which represent the impact of noise.

Let X denote the 2^n-dimensional vector space for the first "noiseless" copy and Y for the second "noisy" copy. To define the dynamics formally, recall the definition of D and T from Eqs. (5.95) and (5.96), and define the following matrix that acts on four bits.

$$\begin{aligned} R =& D \otimes D + D \otimes T + T \otimes D + T \otimes T \left(|01,10\rangle\langle 01,10| + |10,01\rangle\langle 10,01| \right) \\ &+ \frac{q^2}{q^2+1} |00,00\rangle\langle 01,01| + \frac{1}{q^2+1} |11,11\rangle\langle 01,01| \\ &+ \frac{q^2}{q^2+1} |00,00\rangle\langle 10,10| + \frac{1}{q^2+1} |11,11\rangle\langle 10,10| \end{aligned} . \tag{5.119}$$

The matrix R is stochastic. It should be understood as a correlated bit flip where, if the first and third bits are equal and the second and fourth bits are equal, they are sent to a state where that is still true. However, its marginal on either the first two bits or the last two bits is precisely P from Eq. (5.97). Refer to the ith bit of the first random variable as X_i and the ith bit of the second random variable as Y_i. Then define

$$R_\sigma^{(t)} = \left(I \otimes Q_\sigma^{\prime(t)} Q_\sigma^{(t)} \right) \left(I_{XY \setminus \{X_{i_t}, X_{j_t}, Y_{i_t}, Y_{j_t}\}} \otimes R_{\{X_{i_t}, X_{j_t}, Y_{i_t}, Y_{j_t}\}} \right) . \tag{5.120}$$

In words, what $R_\sigma^{(t)}$ does is first generate a correlated noiseless transition among the bits involved in the gate $\{X_{i_t} X_{j_t}, Y_{i_t} Y_{j_t}\}$ for both the first "noiseless" X copy and the second "noisy" Y copy, and then apply the noise transitions only to the Y copy. Since the marginal dynamics of the matrix R restricted either to the first two bits or to the last two bits is the matrix P, the marginal dynamics of $R_\sigma^{(t)}$ are $P^{(t)}$ on the X copy and $Q_\sigma^{\prime(t)} Q_\sigma^{(t)} P^{(t)}$ on the Y copy.

An additional property of $R_\sigma^{(t)}$ is that it preserves a certain subspace of the $2^n \times 2^n$ Hilbert space. If we define the projector $\pi_i = (|00\rangle\langle 00| + |11\rangle\langle 11| + |10\rangle\langle 10|)_{\{X_i Y_i\}}$, then the support of $\bigotimes_{i=0}^{n-1} \pi_i$ is not coupled with its orthogonal complement by the matrix $R_\sigma^{(t)}$. Let us refer to this subspace as the *accessible subspace*. This corresponds to the fact that the noise can send $1 \to 0$, but not vice versa.

We define the initial state to be the correlated version of $|\Lambda_b\rangle$

$$|\Lambda_b \Lambda_b\rangle = \frac{1}{(q+1)^n} \sum_{\vec{\nu}} q^{n-|\vec{\nu}|} |\vec{\nu}\rangle_X \otimes |\vec{\nu}\rangle_Y , \tag{5.121}$$

which lies in this subspace, so evolution by $R_\sigma^{(t)}$ is guaranteed to remain within the subspace for the entire evolution.

In terms of $R_\sigma^{(t)}$, we can rewrite Eq. (5.98) as

$$\mathcal{Z}_\sigma = \langle \mathbf{1}, \mathbf{q}| \prod_{t=1}^{s} R_\sigma^{(t)} |\Lambda_b \Lambda_b\rangle , \tag{5.122}$$

where $|a, b\rangle$ is shorthand for $|a\rangle_X \otimes |b\rangle_Y$.

Note also that since the marginal dynamics of the X copy is the noiseless dynamics, we have

$$\mathcal{Z}_0 = \langle \mathbf{q}, \mathbf{1}| \prod_{t=1}^{s} R_\sigma^{(t)} |\Lambda_b \Lambda_b\rangle \tag{5.123}$$

for any σ.

In our proof, we find it convenient to let

$$|v^{(t)}\rangle = \prod_{t'=1}^{t} R_\sigma^{(t')} |\Lambda_b \Lambda_b\rangle \,. \tag{5.124}$$

Note that for circuit architectures where the circuit diagram is chosen randomly, such as the complete-graph architecture, $|v^{(t)}\rangle$ is defined as the above expression averaged over all circuit diagrams.

Also let W refer to a third copy of the Hilbert space and define a mapping from the ith bits of X and Y to the ith bit of W, as follows:

$$\Delta_i = |1\rangle_{W_i}\langle 11|_{X_i Y_i} + |1\rangle_{W_i}\langle 00|_{X_i Y_i} + |0\rangle_{W_i}\langle 01|_{X_i Y_i} + |0\rangle_{W_i}\langle 10|_{X_i Y_i} \tag{5.125}$$

It maps a bit pair to $|1\rangle$ if they agree and $|0\rangle$ if they disagree. Let

$$\Delta = \bigotimes_{i=0}^{n-1} \Delta_i \tag{5.126}$$

be the map from $X \otimes Y$ to W. Note that $\Delta|\Lambda_b \Lambda_b\rangle = |1^n\rangle$.

I-destined and S-destined probability mass

We view $|v^{(t)}\rangle$ as the probability vector for the correlated biased random walk. Suppose starting at timestep $t+1$, we begin running noiseless dynamics on *both* copies, i.e. we apply $R_0^{(t)}$, and we continue for an infinite number of gates. Then we will get full convergence to the fixed points $|0^n\rangle \otimes |0^n\rangle$, $|1^n\rangle \otimes |1^n\rangle$ and $|1^n\rangle \otimes |0^n\rangle$. The fourth fixed point $|0^n\rangle \otimes |1^n\rangle$ is not in the accessible subspace. We know precisely the probability of each of these outcomes, owing to our derivation of the functions P_I and P_S for the biased random walk.

We define the diagonal matrices

$$L_I = \sum_{\vec{\nu}} \frac{1 - q^{-2n+2|\vec{\nu}|}}{1 - q^{-2n}} |\vec{\nu}\rangle\langle\vec{\nu}| \tag{5.127}$$

$$L_S = \sum_{\vec{\nu}} \frac{q^{-2n+2|\vec{\nu}|} - q^{-2n}}{1 - q^{-2n}} |\vec{\nu}\rangle\langle\vec{\nu}| \,, \tag{5.128}$$

and note that $L_I + L_S$ is the identity matrix. The coefficient of $|\vec{\nu}\rangle\langle\vec{\nu}|$ in L_I gives the probability that a configuration that starts at $|\vec{\nu}\rangle$ ends at the $I^n \equiv 0^n$ fixed point if it undergoes completely noiseless dynamics, and the coefficient in L_S gives the probability of ending at the $S^n \equiv 1^n$ fixed point.

Then define

$$L_{II} = L_I \otimes I \tag{5.129}$$
$$L_{SS} = I \otimes L_S \tag{5.130}$$
$$L_{SI} = I \otimes L_I - L_I \otimes I\,, \tag{5.131}$$

which are the analogous matrices for the joint dynamics to end at $|0^n\rangle \otimes |0^n\rangle$, $|1^n\rangle \otimes |1^n\rangle$, and $|1^n\rangle \otimes |0^n\rangle$, respectively.

Now we may define

$$P_I^{(t)} = L_I P^{(t)} L_I^{-1} \tag{5.132}$$
$$P_S^{(t)} = L_S P^{(t)} L_S^{-1} \tag{5.133}$$

and

$$R_{II}^{(t)} = L_{II} R_0^{(t)} L_{II}^{-1} \tag{5.134}$$
$$R_{SS}^{(t)} = L_{SS} R_0^{(t)} L_{SS}^{-1} \tag{5.135}$$
$$R_{SI}^{(t)} = L_{SI} R_0^{(t)} L_{SI}^{-1}\,, \tag{5.136}$$

where in each case O^{-1} denotes the Moore-Penrose pseudo-inverse of O. We interpret these matrices as the transition operators for probability mass that has been conditioned to end up at a certain fixed point. For example, $P_S^{(t)}$ is the transition operator for a single copy conditioned on eventually ending up at the $S^n \equiv 1^n$ fixed point. Even though the walk is generally biased toward I, it will be biased toward S when you condition on ending at the 1^n fixed point. The following lemma asserts that these are indeed stochastic matrices.

Lemma 5.3. *The matrices $P_I^{(t)}$, $P_S^{(t)}$, $R_{II}^{(t)}$, $R_{SS}^{(t)}$, $R_{SI}^{(t)}$, restricted to their support, are stochastic matrices.*

The next lemma asserts that if the $X \otimes Y$ system undergoes dynamics under $R_{SI}^{(t)}$, then the W system undergoes dynamics under $P_I^{(t)}$. This makes sense, since conditioning on X to go to $S^n \equiv 1^n$ and Y to go to $I^n \equiv 0^n$ should be equivalent to conditioning the W system to go to 0^n.

Lemma 5.4. *Within the accessible subspace, the following holds:*

$$\Delta R_{SI}^{(t)} = P_I^{(t)} \Delta\,. \tag{5.137}$$

We now introduce some more notation. For any vector $|x\rangle$ on a single copy of the vector space, let

$$|x_I\rangle = L_I|x\rangle \tag{5.138}$$
$$|x_S\rangle = L_S|x\rangle\,, \tag{5.139}$$

and for any vector $|v\rangle$ on two copies of the vector space, let

$$|v_{II}\rangle = L_{II}|v\rangle \tag{5.140}$$
$$|v_{SS}\rangle = L_{SS}|v\rangle \tag{5.141}$$
$$|v_{SI}\rangle = L_{SI}|v\rangle\,. \tag{5.142}$$

Thus, if $|x\rangle$ represents a probability distribution over the 2^n basis states on a single copy of the Hilbert space, then the vector $|x_I\rangle$ is the portion of $|x\rangle$ that is destined to end at the fixed point 0^n, and $|x_S\rangle$ is the portion destined to end at 1^n (if all future gates are noiseless).

The amount of probability mass for which the noisy copy is destined for the 1^n fixed point cannot decay too quickly with the number of noise locations (note that if the noisy copy ends at 1^n, the noiseless copy must also end at 1^n).

Lemma 5.5. *The S-destined probability mass obeys the following inequality, for any $t' \geq t$.*

$$\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t')}\rangle \geq (1-\sigma)^{2(t'-t)}\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t)}\rangle\,. \tag{5.143}$$

Proof idea. Recall that the inner product with $\langle \mathbf{1}, \mathbf{1}|$ gives the sum of the entries of the vector. We interpret $|v_{SS}^{(t)}\rangle$ as the probability vector of mass destined to reach the 1^n fixed point on both copies. Each time a noise location acts, it can affect at most a σ fraction of the mass, so even after two noise locations act, at least a $(1-\sigma)^2$ fraction of the mass that was S-destined before will still be S-destined. □

Decomposing the I-destined probability mass

The final piece of machinery we need is an accounting of which error leads to each piece of I-destined probability mass. To do this, for each $t \geq 1$ define

$$|v_{SI}^{(t,t)}\rangle = |v_{SI}^{(t)}\rangle - (I \otimes Q_\sigma'^{(t)} Q_\sigma^{(t)}) R_{SI}^{(t)} |v_{SI}^{(t-1)}\rangle \tag{5.144}$$
$$= \left(L_{SI}\left(I \otimes Q_\sigma'^{(t)} Q_\sigma^{(t)}\right) - \left(I \otimes Q_\sigma'^{(t)} Q_\sigma^{(t)}\right) L_{SI}\right) R_0^{(t)} |v^{(t-1)}\rangle\,, \tag{5.145}$$

and define the evolution rule

$$|v_{SI}^{(t'+1,t)}\rangle = Q_\sigma'^{(t'+1)} Q_\sigma^{(t'+1)} R_{SI}^{(t'+1)} |v_{SI}^{(t',t)}\rangle\,. \tag{5.146}$$

The vector $|v_{SI}^{(t',t)}\rangle$ represents the probability mass that would have gone to the 1^n fixed point, but the noise at time step t caused it to be redirected to the 0^n fixed point, and we have subsequently evolved it forward to timestep t'.

Importantly, we can verify from the definition that

$$\sum_{t=1}^{t'} |v_{SI}^{(t',t)}\rangle = |v_{SI}^{(t')}\rangle\,, \tag{5.147}$$

indicating that all of the mass at time step t' is accounted for as having originated at some previous time step t.

Lemma 5.6. *For all t and $t' \geq t$,*

$$\langle \mathbf{1}, \mathbf{1} | v_{SI}^{(t',t)} \rangle \leq (1 - (1-\sigma)^2) \langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t-1)} \rangle\,. \tag{5.148}$$

Proof idea. The vector $|v_{SI}^{(t,t)}\rangle$ represents the mass that satisfies two conditions: (1) it was destined for the $|1^n\rangle \otimes |1^n\rangle$ fixed point at time step $t-1$, and (2) the noise at time step t caused it to be destined for the $|1^n\rangle \otimes |0^n\rangle$ fixed point at time step t. At most $\langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t-1)} \rangle$ mass qualifies under condition (1). Among that mass, each of the two noise location can only impact a σ fraction of the mass, so the fraction of mass that can be re-directed is at most $(1 - (1-\sigma)^2)$. □

5.B.4 Consequences of anti-concentration

In all of our rigorous proofs, we assume that we have a random quantum circuit architecture that is h-regularly connected for some constant $h = O(1)$, and has anti-concentration size equal to s_{AC}. Recall that this means that Z_0 becomes twice its limiting value at s_{AC}. When this is the case, we have the following lemmas. All constants are dependent on q and h, but not on n or any noise parameters.

Lemma 5.7. *Suppose the random quantum circuit architecture is regularly connected. There exist constants χ_1 and χ_2 such that for all $t \geq s_{AC}$*

$$\langle \mathbf{q}, \mathbf{1} | v^{(t)} \rangle \leq \frac{2q^n}{q^n + 1} + \eta_t\,, \tag{5.149}$$

where

$$\eta_t = \chi_2 \exp\left(-\frac{\chi_1}{n} (t - s_{AC}) \right)\,. \tag{5.150}$$

Proof idea. The left-hand side is precisely Z_0 for a circuit with size t. The regularly connected property indicates that for any configuration not at a fixed point, there will be a gate that couples an I with an S roughly once every $O(n)$ gates. When this happens, the difference between Z_0 and its infinite-size limit is reduced by a constant factor, leading to the scaling in the lemma. □

Lemma 5.8. *Suppose the random quantum circuit architecture is regularly connected. There exist constants χ_3 and χ_4 such that for all t*

$$\langle 1^n, \mathbf{1} | v^{(t)} \rangle \geq \frac{1 - \eta_t'}{q^n + 1}\,, \tag{5.151}$$

where

$$\eta'_t = \chi_4 \exp\left(-\frac{\chi_3}{n}(t - s_{AC})\right) . \tag{5.152}$$

Proof idea. Anti-concentration happens because most of the probability mass makes it to one of the fixed points. This lemma states that after the anti-concentration size, most of the mass destined for the 1^n fixed point has already reached it. The fraction that has not yet reached is η'_t, which decays exponentially with t/n. We show that if this were not the case, then the bound in Lemma 5.7 could not hold. □

Lemma 5.9. *Suppose the random quantum circuit architecture is regularly connected. There exist constants χ_5 and χ_6 such that for any non-negative vector $|v\rangle$ that is normalized (i.e. $\langle \mathbf{1}, \mathbf{1}|v\rangle = 1$), the following holds for any t_0 and any $t_1 \geq t_0$:*

$$\begin{aligned} &\langle \mathbf{q}|\Delta \prod_{t=t_0+1}^{t_1} \left((I \otimes Q'^{(t)}_\sigma Q^{(t)}_\sigma) R^{(t)}_{SI}\right) |v\rangle - 1 \\ &\leq \left(\langle \mathbf{q}|\Delta|v\rangle - 1\right) \chi_6 \exp\left(-\frac{\chi_5(t_1 - t_0)}{n}\right) . \end{aligned} \tag{5.153}$$

Proof idea. Recall from Lemma 5.4 that if $|v\rangle$ evolves by $R^{(t)}_{SI}$, then $\Delta|v\rangle$ evolves by $P^{(t)}_I$. The transition matrix $P^{(t)}_I$ is the matrix that conditions on sending the vector to the 0^n fixed point, so it is even more I-biased than the transition matrix $P^{(t)}$. Thus, each time a bit is flipped, the Hamming weight is likely to decrease, and the inner product with $\langle \mathbf{q}| - \langle \mathbf{1}|$ will be reduced by a constant factor. This will happen once every $O(n)$ gates if the architecture is regularly connected. The insertion of the $Q^{(t)}_\sigma$ operators will only make the Hamming weight smaller since they can only flip $1 \to 0$. □

5.B.5 Exponential clustering of S-destined probability mass

A key step in our analysis is that the S-destined mass stays close to the 1^n fixed point, as long as $\sigma = O(1/n)$. In fact, the probability of deviating from the fixed point by x bit flips decays exponentially in x. Intuitively, this is because the S-destined mass is biased to move upward in Hamming weight, and when σ is small enough, this upward pressure will be greater than the downward pressure coming from the noise itself.

We prove this for the W system, which captures the difference between the (noiseless) X and (noisy) Y systems. We cannot directly analyze the Y system because at time step 0, the statement is definitively not true. It takes s_{AC} gates for the S-destined mass in the Y system to initially converge. Meanwhile, the W system begins at the 1^n fixed point. This is the main reason we introduced the W system in the first place.

Define the projector

$$\Pi_w = \sum_{\vec{\nu}:|\vec{\nu}|=w} |\vec{\nu}\rangle\langle\vec{\nu}| \,. \tag{5.154}$$

Lemma 5.10. *There exist constants χ_7, χ_8, χ_9, and n_0 such that as long as $\sigma \leq \chi_7/n$ and $n \geq n_0$, the following holds for any t and any integer w with $1 \leq w < n$:*

$$\frac{\langle \mathbf{1}|\Pi_w \Delta|v_{SS}^{(t)}\rangle}{\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle} \leq n\sigma\xi_w \,, \tag{5.155}$$

where

$$\xi_w = \chi_9(n-w)q^{-(n-w)}e^{-\chi_8(n-w)} \,. \tag{5.156}$$

Proof idea. The S-destined portion of the mass within the W system starts at the 1^n fixed point. When noise acts at time step t, some of the mass moves to Hamming weight $n-1$ but continues to be S-destined, and some of it is "redirected" to become I-destined, which is captured in the $|v_{SI}^{(t,t)}\rangle$ vector. The total amount of redirected mass cannot be too large, as we see in Lemma 5.6. Moreover, the redirected mass must steadily move downward in Hamming weight (after all, it is I-destined), which we quantify with Lemma 5.9. This is important because for each value of the Hamming weight w, the amount of S-destined mass divided by the amount of I-destined mass at that Hamming weight is precisely $\frac{q^{-2n+2w}-q^{-2n}}{1-q^{-2n+2w}} \approx q^{-2(n-w)}$, so as the I-destined mass moves down in Hamming weight, the S-destined mass that corresponds to it decreases exponentially. After accounting for each bit of I-destined mass by summing over all $|v_{SI}^{(t',t)}\rangle$, we can prove the lemma. □

5.B.6 Relating $\mathcal{Z}_\sigma$ to the S-destined weight

The following lemma states that keeping track of the amount of S-destined mass is sufficient to get good upper and lower bounds on the quantity $\mathcal{Z}_\sigma$.

Lemma 5.11. *The following lower bound always holds*

$$\mathcal{Z}_\sigma - 1 \geq (q^n - 1)\,\langle \mathbf{1},\mathbf{1}|v_{SS}^{(s)}\rangle \tag{5.157}$$

Moreover, there exist constants χ_{10}, χ_{11}, χ_{12}, χ_{13}, and n_0 such that as long as $\sigma \leq \chi_{13}/n$ and $n \geq n_0$, the following upper bound holds.

$$\mathcal{Z}_\sigma - 1 \leq (q^n - 1)\,\langle \mathbf{1},\mathbf{1}|v_{SS}^{(s)}\rangle \exp\left(1 + \chi_{10}n\sigma + \chi_{12}e^{-\frac{\chi_{11}}{n}(s-s_{AC})+4s\sigma}\right) \tag{5.158}$$

Proof idea. For each w, we know the ratio of the I-destined and S-destined mass at Hamming weight w: for each portion of S-destined probability mass, there is roughly $q^{2(n-w)}$ I-destined probability mass. This decreases with w like q^{-2w}. The contribution of mass at Hamming weight w to $\mathcal{Z}_\sigma$ increases,

but at the slower rate of q^w. Thus, for a fixed amount of S-destined mass, $\mathcal{Z}_\sigma$ is minimized when all of it is at the 1^n fixed point, leading to our lower bound. On the other hand, we know that the S-destined mass is exponentially clustered near the 1^n fixed point (Lemma 5.10), so this lower bound cannot be too loose, which we leverage into an upper bound. □

5.B.7 Bounding the S-destined mass

Now, all that remains is to compute the amount of S-destined mass. Here we show upper and lower bounds on this quantity for layered architectures and for the complete-graph architecture.

Lemma 5.12. *Suppose the random quantum circuit architecture is regularly connected and layered. Let d_{AC} be its anti-concentration depth. Then, for any d,*

$$\langle \mathbf{1}, \mathbf{1} | v_{SS}^{(dn/2)} \rangle \geq \frac{\left(1 - \frac{1-(1-\sigma(1-q^{-2}))^n}{1-q^{-2n}}\right)^d}{q^n + 1} . \tag{5.159}$$

Moreover, there exist constants a_0, a_1, a_2, a_3, and n_0 such that, as long as $\sigma \leq a_3/n$ and $n \geq n_0$,

$$\langle \mathbf{1}, \mathbf{1} | v_{SS}^{(dn/2)} \rangle \leq \frac{\left(1 - \frac{1-(1-\sigma(1-q^{-2}))^n}{1-q^{-2n}}\right)^d}{q^n + 1} e^{a_0 \sigma^2 dn + a_1 \sigma n d_{AC} + a_2 n \sigma \log(1/(n\sigma))} , \tag{5.160}$$

where d_{AC} is the anti-concentration depth.

Lemma 5.13. *Suppose the random quantum circuit architecture is the complete-graph architecture. Let s_{AC} be its anti-concentration size. Then, for any s,*

$$\langle \mathbf{1}, \mathbf{1} | v_{SS}^{(s)} \rangle \geq \frac{\left(1 - \frac{1-(1-\sigma(1-q^{-2}))^2}{1-q^{-2n}}\right)^s}{q^n + 1} . \tag{5.161}$$

Moreover, there exist constants b_0, b_1, b_2, b_3, and n_0 such that, as long as $\sigma \leq b_3/n$ and $n \geq n_0$,

$$\langle \mathbf{1}, \mathbf{1} | v_{SS}^{(s)} \rangle \leq \frac{\left(1 - \frac{1-(1-\sigma(1-q^{-2}))^2}{1-q^{-2n}}\right)^s}{q^n + 1} e^{b_0 \sigma^2 s + b_1 \sigma s_{AC} + b_2 n \sigma \log(1/(n\sigma))} \tag{5.162}$$

Proof idea for Lemma 5.12 and Lemma 5.13. When a portion of S-destined mass is at the 1^n fixed point, and noise acts to move it to Hamming weight $n-1$, we have a good understanding of what fraction remains S-destined. Specifically, there is a $\frac{q^{-2}-q^{-2n}}{1-q^{-2n}}$ chance that it re-equilibrates to 1^n. We also know the chance that it will make the transition in the first place; the transition from $1 \to 0$ happens with probability precisely σ. This scenario gives the maximum amount of lost S-destined mass, and gives rise to our lower bound. However, if the portion of S-destined mass is not at the 1^n fixed point, then

this is complicated in two ways. First, the probability of re-equilibrating back to 1^n is a slightly different expression, and, more importantly, the noise will not cause a transition as often, as there is a chance it acts on a bit that is already 0. If the configuration has Hamming weight w and the noise acts on a random bit, the chance of a transition is $\frac{n-w}{n}\sigma$ so a smaller amount of S-destined mass is lost at each step. Luckily, we know that the S-destined mass is exponentially clustered near $w = n$ (Lemma 5.10), so the corrections are small, which gives rise to the upper bound.

We utilize the layered architecture property to be able to say that *every* qudit is acted upon by noise after each layer, and thus, from the perspective of the amount of S-destined mass, all that matters is the Hamming weight of the configuration prior to the noise. The same is true for the complete-graph case because the gates are chosen randomly and each qudit is equally likely to participate. However, we do not believe this property is necessary for our result to be true. □

5.B.8 Deferred proofs of lemmas

Proof of Lemma 5.3

Proof. We demonstrate this for $P_I^{(t)}$ and leave the others to be verified in a similar fashion. First of all, since $P^{(t)}$ is a stochastic matrix, its matrix elements are non-negative. Since L_I and L_I^{-1} are diagonal matrices with non-negative entries, $P_I^{(t)} = L_I P^{(t)} L_I^{-1}$ also has non-negative matrix elements. The support of P_I is the entire vector space except for the span of $|1^n\rangle$. Consider another basis state $|\vec{\nu}\rangle$. Since gate t acts on qudits $\{i_t, j_t\}$, if $\nu_{i_t} = \nu_{j_t}$, then it is a $+1$ eigenvector of $|P^{(t)}\rangle$ and

$$\langle \mathbf{1}|P_I^{(t)}|\vec{\nu}\rangle = \sum_{\vec{\mu}} \langle \vec{\mu}|L_I P^{(t)} L_I^{-1}|\vec{\nu}\rangle \tag{5.163}$$

$$= \sum_{\vec{\mu}} \frac{1-q^{-2n+2|\vec{\mu}|}}{1-q^{-2n+2|\vec{\nu}|}} \langle \vec{\mu}|P^{(t)}|\vec{\nu}\rangle \tag{5.164}$$

$$= \sum_{\vec{\mu}} \frac{1-q^{-2n+2|\vec{\mu}|}}{1-q^{-2n+2|\vec{\nu}|}} \langle \vec{\mu}|\vec{\nu}\rangle = 1\,. \tag{5.165}$$

If $\nu_{i_t} \neq \nu_{j_t}$, then $P^{(t)}$ sends $|\vec{\nu}\rangle$ to a basis state with Hamming weight reduced by 1 with probability $q^2/(q^2+1)$, and to Hamming weight increased by 1 with probability $1/(q^2+1)$, so

$$\langle \mathbf{1}|P_I^{(t)}|\vec{\nu}\rangle = \sum_{\vec{\mu}} \frac{1-q^{-2n+2|\vec{\mu}|}}{1-q^{-2n+2|\vec{\nu}|}} \langle \vec{\mu}|P^{(t)}|\vec{\nu}\rangle \tag{5.166}$$

$$= \left(\frac{q^2}{q^2+1} \frac{1-q^{-2n+2|\vec{\nu}|-2}}{1-q^{-2n+2|\vec{\nu}|}} + \frac{1}{q^2+1} \frac{1-q^{-2n+2|\vec{\nu}|+2}}{1-q^{-2n+2|\vec{\nu}|}} \right) = 1\,. \tag{5.167}$$

This demonstrates that $P_I^{(t)}$ is a stochastic matrix when restricted to its support. □

Proof of Lemma 5.4

Proof. We consider the action of both sides of the equation on an input state $|\vec{\nu}, \vec{\mu}\rangle$. Let a and b be the number of 1 entries in $\vec{\nu}$ and $\vec{\mu}$, excluding the positions $\{i_t, j_t\}$, respectively, and let c be the number of entries on which $\vec{\nu}$ and $\vec{\mu}$ agree. Since we are restricting to the accessible subspace, we have $c = n-2-a+b$. Since Δ is a tensor product across all bits $i \in \{0, \ldots, n-1\}$, and both $P_I^{(t)}$ and $R_{SI}^{(t)}$ modify only bits i_t and j_t, it is sufficient to consider the transitions among just bits i_t and j_t. First, define

$$c_0 = \frac{1 - q^{-2n+2c}}{1 - q^{-2n+2c+2}} \frac{q^2}{q^2+1} \tag{5.168}$$

$$c_1 = \frac{1 - q^{-2n+2c+4}}{1 - q^{-2n+2c+2}} \frac{1}{q^2+1} \,. \tag{5.169}$$

Let the four bits below be ordered $X_{i_t} X_{j_t}$, $Y_{i_t} Y_{j_t}$. The right-hand side has the following effect, where the first arrow is application of Δ and the second is application of $P_I^{(t)}$.

$$\begin{aligned}
&|11, 11\rangle \to |11\rangle \to |11\rangle \\
&|11, 10\rangle \to |10\rangle \to c_0|00\rangle + c_1|11\rangle \\
&|11, 01\rangle \to |01\rangle \to c_0|00\rangle + c_1|11\rangle \\
&|11, 00\rangle \to |00\rangle \to |00\rangle \\
&|10, 10\rangle \to |11\rangle \to |11\rangle \\
&|10, 00\rangle \to |01\rangle \to c_0|00\rangle + c_1|11\rangle \\
&|01, 01\rangle \to |11\rangle \to |11\rangle \\
&|01, 00\rangle \to |10\rangle \to c_0|00\rangle + c_1|11\rangle \\
&|00, 00\rangle \to |11\rangle \to |11\rangle \,.
\end{aligned}$$

Now, we can do the same for the left-hand side. For example, consider the input state $|11, 10\rangle$. Action by $R_{SI}^{(t)}$ sends it to

$$|11, 10\rangle \to \frac{q^{-2n+2a+4} - q^{-2n+2b}}{q^{-2n+2a+4} - q^{-2n+2b+2}} \frac{q^2}{q^2+1} |11, 00\rangle \tag{5.170}$$

$$+ \frac{q^{-2n+2a+4} - q^{-2n+2b+4}}{q^{-2n+2a+4} - q^{-2n+2b+2}} \frac{1}{q^2+1} |11, 11\rangle \tag{5.171}$$

$$= c_0|11, 00\rangle + c_1|11, 11\rangle \,, \tag{5.172}$$

where the last line follows by recalling the relation $c = n - 2 - a + b$. Action by Δ then yields the state $c_0|00\rangle + c_1|11\rangle$. We can now list this calculation for

each input state, where the first arrow is action by $R_{SI}^{(t)}$ and the second by Δ.

$$|11,11\rangle \to |11,11\rangle \to |11\rangle \tag{5.173}$$
$$|11,10\rangle \to c_0|11,00\rangle + c_1|11,11\rangle \to c_0|00\rangle + c_1|11\rangle \tag{5.174}$$
$$|11,01\rangle \to c_0|11,00\rangle + c_1|11,11\rangle \to c_0|00\rangle + c_1|11\rangle \tag{5.175}$$
$$|11,00\rangle \to |11,00\rangle \to |00\rangle \tag{5.176}$$
$$|10,10\rangle \to \frac{q^{-2n+2a} - q^{-2n+2b}}{q^{-2n+2a+2} - q^{-2n+2b+2}} \frac{q^2}{q^2+1} |00,00\rangle \tag{5.177}$$
$$+ \frac{q^{-2n+2a+4} - q^{-2n+2b+4}}{q^{-2n+2a+2} - q^{-2n+2b+2}} \frac{1}{q^2+1} |11,11\rangle \tag{5.178}$$
$$\to |11\rangle \tag{5.179}$$
$$|10,00\rangle \to c_1|00,00\rangle + c_0|11,00\rangle \to c_0|00\rangle + c_1|11\rangle \tag{5.180}$$
$$|01,01\rangle \to \frac{q^{-2n+2a} - q^{-2n+2b}}{q^{-2n+2a+2} - q^{-2n+2b+2}} \frac{q^2}{q^2+1} |00,00\rangle \tag{5.181}$$
$$+ \frac{q^{-2n+2a+4} - q^{-2n+2b+4}}{q^{-2n+2a+2} - q^{-2n+2b+2}} \frac{1}{q^2+1} |11,11\rangle \tag{5.182}$$
$$\to |11\rangle \tag{5.183}$$
$$|01,00\rangle \to c_1|00,00\rangle + c_0|11,00\rangle \to c_0|00\rangle + c_1|11\rangle \tag{5.184}$$
$$\tag{5.185}$$
$$|00,00\rangle \to |00,00\rangle \to |11\rangle\,, \tag{5.186}$$

which verifies that the left-hand and right-hand sides are equal. □

Proof of Lemma 5.5

Proof.

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle = \langle \mathbf{1},\mathbf{1}|L_{SS}R_\sigma^{(t)}|v^{(t-1)}\rangle = \langle \mathbf{1},\mathbf{1}|L_{SS}R_\sigma^{(t)}L_{SS}^{-1}|v_{SS}^{(t-1)}\rangle \tag{5.187}$$
$$= \sum_{\substack{\vec{\mu}\\ \vec{\nu}\neq 0^n}} \langle \mathbf{1},\mathbf{1}|L_{SS}|\mathbf{1},\vec{\mu}\rangle\langle \mathbf{1},\vec{\mu}|R_\sigma^{(t)}|\mathbf{1},\vec{\nu}\rangle\langle \mathbf{1},\vec{\nu}|L_{SS}^{-1}|v_{SS}^{(t-1)}\rangle \tag{5.188}$$
$$= \sum_{\substack{\vec{\mu}\\ \vec{\nu}\neq 0^n}} \frac{q^{-2n+2|\vec{\mu}|} - q^{-2n}}{q^{-2n+2|\vec{\nu}|} - q^{-2n}} \langle \mathbf{1},\vec{\mu}|R_\sigma^{(t)}|\mathbf{1},\vec{\nu}\rangle\langle \mathbf{1},\vec{\nu}|v_{SS}^{(t-1)}\rangle \tag{5.189}$$
$$= \sum_{\substack{\vec{\mu}\\ \vec{\nu}\neq 0^n}} \frac{q^{-2n+2|\vec{\mu}|} - q^{-2n}}{q^{-2n+2|\vec{\nu}|} - q^{-2n}} \langle \vec{\mu}|Q_\sigma'^{(t)}Q_\sigma^{(t)}P^{(t)}|\vec{\nu}\rangle\langle \mathbf{1},\vec{\nu}|v_{SS}^{(t-1)}\rangle \tag{5.190}$$
$$= \sum_{\substack{\vec{\mu}\\ \vec{\nu},\vec{\zeta}\neq 0^n}} E_{\vec{\mu}\vec{\zeta}}G_{\vec{\zeta}\vec{\nu}}\langle \mathbf{1},\vec{\nu}|v_{SS}^{(t-1)}\rangle \tag{5.191}$$

where

$$E_{\vec{\mu}\vec{\zeta}} = \frac{q^{-2n+2|\vec{\mu}|} - q^{-2n}}{q^{-2n+2|\vec{\zeta}|} - q^{-2n}} \langle \vec{\mu} | Q_\sigma^{'(t)} Q_\sigma^{(t)} | \vec{\zeta} \rangle \tag{5.192}$$

$$G_{\vec{\zeta}\vec{\nu}} = \frac{q^{-2n+2|\vec{\zeta}|} - q^{-2n}}{q^{-2n+2|\vec{\nu}|} - q^{-2n}} \langle \vec{\zeta} | P^{(t)} | \vec{\nu} \rangle = \langle \vec{\zeta} | P_S^{(t)} | \vec{\nu} \rangle \tag{5.193}$$

However, note that $E_{\vec{\zeta}\vec{\zeta}} \geq (1-\sigma)^2$ (with equality when $\zeta_{i_t} = \zeta_{j_t} = 1$), and all $E_{\vec{\mu}\vec{\zeta}}$ are non-negative. Moreover, note that

$$\sum_{\vec{\zeta}} G_{\vec{\zeta}\vec{\nu}} = 1 \,, \tag{5.194}$$

owing to the fact that $P_S^{(t)}$ is stochastic. Thus $\langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t)} \rangle \geq (1-\sigma)^2 \langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t-1)} \rangle$, and by recursion, the statement holds. □

Proof of Lemma 5.6

Proof. Recall that $L_{SI} = I \otimes L_I - L_I \otimes I$, but the second term commutes with $I \otimes Q_\sigma^{'(t)} Q_\sigma^{(t)}$, thus we may ignore it in the following calculation.

$$\langle \mathbf{1}, \mathbf{1} | v_{SI}^{(t,t)} \rangle = \sum_{\vec{\mu},\vec{\nu}} \langle \vec{\mu} | L_I Q_\sigma^{'(t)} Q_\sigma^{(t)} - Q_\sigma^{'(t)} Q_\sigma^{(t)} L_I | \vec{\nu} \rangle \langle \mathbf{1}, \vec{\nu} | R_0^{(t)} | v^{(t-1)} \rangle \tag{5.195}$$

$$= \sum_{\vec{\mu},\vec{\nu}} \frac{q^{-2n+2|\vec{\nu}|} - q^{-2n+2|\vec{\mu}|}}{1 - q^{-2n}} \langle \vec{\mu} | Q_\sigma^{'(t)} Q_\sigma^{(t)} | \vec{\nu} \rangle \langle \mathbf{1}, \vec{\nu} | R_0^{(t)} | v^{(t-1)} \rangle \tag{5.196}$$

If $\vec{\mu} = \vec{\nu}$ the factor gives 0. For each $\vec{\nu}$ there are at most three possible $\vec{\mu} \neq \vec{\nu}$ for which the matrix element $\langle \vec{\mu} | Q_\sigma^{'(t)} Q_\sigma^{(t)} | \vec{\nu} \rangle \neq 0$, corresponding to a single error on either qudit or an error on both at once. In those cases, the matrix element is $\sigma(1-\sigma)$ (for single error) or σ^2 (for double error). The double error is only possible if $|\vec{\nu}| \geq 2$, but note that we may assume $|\vec{\nu}| \neq 1$ since action by $R_0^{(t)}$ will leave the two bits it acts on equal, and cannot lead to a configuration with Hamming weight 1. We have

$$\sum_{\vec{\mu}} \frac{q^{-2n+2|\vec{\nu}|} - q^{-2n+2|\vec{\mu}|}}{1 - q^{-2n}} \langle \vec{\mu} | Q_\sigma^{'(t)} Q_\sigma^{(t)} | \vec{\nu} \rangle$$

$$\leq 2\sigma(1-\sigma) \frac{q^{-2n+2|\vec{\nu}|} - q^{-2n+2|\vec{\nu}|-2}}{1 - q^{-2n}} + \sigma^2 \frac{q^{-2n+2|\vec{\nu}|} - q^{-2n+2|\vec{\nu}|-4}}{1 - q^{-2n}} \tag{5.197}$$

$$= \left(\frac{q^{-2n+2|\vec{\nu}|} - q^{-2n}}{1 - q^{-2n}} \right) \frac{2\sigma(1-\sigma)(1-q^{-2}) + \sigma^2(1-q^{-4})}{1 - q^{-2|\vec{\nu}|}} \tag{5.198}$$

$$\leq \left(\frac{q^{-2n+2|\vec{\nu}|} - q^{-2n}}{1 - q^{-2n}} \right) \left(2\sigma - \sigma^2 \right) \,. \tag{5.199}$$

This lets us say

$$\langle \mathbf{1}, \mathbf{1} | v_{SI}^{(t,t)} \rangle \leq \sum_{\vec{\nu}} \left(\frac{q^{-2n+2|\vec{\nu}|} - q^{-2n}}{1 - q^{-2n}} \right) \left(2\sigma - \sigma^2 \right) \langle \mathbf{1}, \vec{\nu} | R_0^{(t)} | v^{(t-1)} \rangle \tag{5.200}$$

$$= \sum_{\vec{\nu}} \left(2\sigma - \sigma^2 \right) \langle \mathbf{1}, \vec{\nu} | L_{SS} R_0^{(t)} | v^{(t-1)} \rangle \tag{5.201}$$

$$= \sum_{\vec{\nu}} \left(2\sigma - \sigma^2 \right) \langle \mathbf{1}, \vec{\nu} | R_{SS}^{(t)} L_{SS} | v^{(t-1)} \rangle \tag{5.202}$$

$$= \left(2\sigma - \sigma^2 \right) \sum_{\vec{\nu}} \langle \mathbf{1}, \vec{\nu} | R_{SS}^{(t)} | v_{SS}^{(t-1)} \rangle \tag{5.203}$$

$$= (1 - (1 - \sigma)^2) \langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t-1)} \rangle \,, \tag{5.204}$$

where the last equality follows because R_{SS} is stochastic.

The fact that this is also true for $|v^{(t',t)}\rangle$ with $t' > t$ follows from the fact that $|v^{(t',t)}\rangle$ is related to $|v^{(t,t)}\rangle$ by a sequence of stochastic matrices, which preserves the left-hand side of the lemma statement. □

Proof of Lemma 5.7

Proof. This proof is similar to that of Theorem 4.3 from Chapter 4. Define $Z^{(t')} = \langle \mathbf{q}, \mathbf{1} | v^{(t')} \rangle$. If the anti-concentration size is s_{AC}, this means that

$$Z^{(s_{AC})} \leq 2q^n Z_H = \frac{4q^n}{q^n + 1} \,. \tag{5.205}$$

Note that $Z^{(t')}$ is monotonically non-increasing with t' (i.e., collision probability only decreases as more gates are applied). Recall that for architectures where the circuit diagram is random, $|v^{(t')}\rangle$ represents an average over choice of circuit diagram. The h-regularly connected property says that, no matter what the circuit diagram has looked like up to time step t', given any partition of the qudits into two parts, there is at least a $1/2$ probability that the next hn gates in the circuit diagram will include at least one gate that couples qudits from opposite parts. Conditioned on coupling the two parts, the portion of the collision probability associated with configurations not already at a fixed point will decrease by a factor $2q/(q^2+1)$, as was seen in Eq. (4.80). Thus for all t',

$$Z^{(t'+rn)} - \frac{2q^n}{q^n + 1} \leq \left(\frac{1}{2} + \frac{1}{2} \frac{2q}{q^2 + 1} \right) \left(Z^{(t')} - \frac{2q^n}{q^n + 1} \right) \tag{5.206}$$

$$= \frac{(q+1)^2}{2(q^2+1)} \left(Z^{(t')} - \frac{2q^n}{q^n + 1} \right) . \tag{5.207}$$

Applying the above recursively, we have

$$Z^{(s_{AC} + zhn)} - \frac{2q^n}{q^n + 1} \leq \left(\frac{(q+1)^2}{2(q^2+1)} \right)^z \frac{2q^n}{q^n + 1} \leq 2 \left(\frac{(q+1)^2}{2(q^2+1)} \right)^z . \tag{5.208}$$

Now we ensure something similar holds for every value of t and not just $t = s_{AC} + zhn$ for integers z. Let t_0 be the maximum integer for which $t_0 \leq t$, and $t_0 = s_{AC} + z_0 hn$ for some integer z_0. So $t - t_0 \leq hn$ and $z_0 \geq (t - s_{AC})/(hn) - 1$. Moreover, by monotonicity, we have $Z^{(t)} \leq Z^{(t_0)}$. Together, this implies

$$Z^{(t)} \leq \frac{2q^n}{q^n+1} + 2\left(\frac{(q+1)^2}{2(q^2+1)}\right)^{z_0} = \frac{2q^n}{q^n+1} + 2\left(\frac{(q+1)^2}{2(q^2+1)}\right)^{\frac{t-s_{AC}}{hn}-1} \tag{5.209}$$

$$= \frac{2q^n}{q^n+1} + \chi_2 e^{-\chi_1(t-s_{AC})/n}\,, \tag{5.210}$$

where $\chi_2 = 4(q^2+1)/(q+1)^2$ and $\chi_1 = \frac{1}{h}\log(2(q^2+1)/(q+1)^2)$. □

Proof of Lemma 5.8

Proof. We have

$$\frac{\langle \mathbf{q}, \mathbf{1}|v^{(t)}\rangle - 1}{q^n - 1}$$

$$= \sum_{\vec{\nu}} \frac{q^{|\vec{\nu}|}-1}{q^n-1}\langle \vec{\nu}, \mathbf{1}|v^{(t)}\rangle \tag{5.211}$$

$$= \langle 1^n, \mathbf{1}|v^{(t)}\rangle + \sum_{\vec{\nu}\neq 0^n, 1^n} \frac{q^{|\vec{\nu}|}-1}{q^n-1}\langle \vec{\nu}, \mathbf{1}|v^{(t)}\rangle \tag{5.212}$$

$$= \langle 1^n, \mathbf{1}|v^{(t)}\rangle + \sum_{\vec{\nu}\neq 0^n, 1^n} \frac{q^{|\vec{\nu}|}-1}{q^n-1}\langle \vec{\nu}, \mathbf{1}|(L_S^{-1}L_S \otimes I)|v^{(t)}\rangle \tag{5.213}$$

$$= \langle 1^n, \mathbf{1}|v^{(t)}\rangle + \sum_{\vec{\nu}\neq 0^n, 1^n} \frac{(1-q^{-2n})\left(q^{|\vec{\nu}|}-1\right)}{(q^{-2n+2|\vec{\nu}|}-q^{-2n})(q^n-1)}\langle \vec{\nu}, \mathbf{1}|L_S \otimes I|v^{(t)}\rangle \tag{5.214}$$

$$\geq \langle 1^n, \mathbf{1}|v^{(t)}\rangle + \frac{(1-q^{-2n})\,(q^{n-1}-1)}{(q^{-2}-q^{-2n})(q^n-1)} \sum_{\vec{\nu}\neq 0^n, 1^n} \langle \vec{\nu}, \mathbf{1}|L_S \otimes I|v^{(t)}\rangle \tag{5.215}$$

$$= \langle 1^n, \mathbf{1}|v^{(t)}\rangle + \frac{q\,(1+q^{-n})}{1+q^{-n+1}} \sum_{\vec{\nu}\neq 0^n, 1^n} \langle \vec{\nu}, \mathbf{1}|L_S \otimes I|v^{(t)}\rangle \tag{5.216}$$

$$= -\left(\frac{q\,(1+q^{-n})}{1+q^{-n+1}} - 1\right)\langle 1^n, \mathbf{1}|v^{(t)}\rangle + \frac{q\,(1+q^{-n})}{1+q^{-n+1}} \sum_{\vec{\nu}\neq 0^n} \langle \vec{\nu}, \mathbf{1}|L_S \otimes I|v^{(t)}\rangle \tag{5.217}$$

$$= -\frac{q-1}{1+q^{-n+1}}\langle 1^n, \mathbf{1}|v^{(t)}\rangle + \frac{q\,(1+q^{-n})}{1+q^{-n+1}}\langle \mathbf{1}, \mathbf{1}|L_S \otimes I|v^{(t)}\rangle \tag{5.218}$$

$$= -\frac{q-1}{1+q^{-n+1}}\langle 1^n, \mathbf{1}|v^{(t)}\rangle + \frac{q\,(1+q^{-n})}{1+q^{-n+1}}\frac{1}{q^n+1}\,, \tag{5.219}$$

where the last line follows because the total amount of S-destined mass for the noiseless copy is exactly $1/(q^n+1)$. From Lemma 5.7, we have

$$\frac{\langle \mathbf{q}, \mathbf{1}|v^{(t)}\rangle - 1}{q^n - 1} \leq \frac{1}{q^n+1} + \frac{\eta_t}{q^n-1}\,. \tag{5.220}$$

Combining the above, we have

$$\langle 1^n, \mathbf{1}|v^{(t)}\rangle \frac{q-1}{1+q^{-n+1}} \geq \frac{1}{q^n+1}\left(\frac{q\,(1+q^{-n})}{1+q^{-n+1}} - 1\right) - \frac{\eta_t}{q^n-1}\,, \tag{5.221}$$

and hence

$$\langle 1^n, \mathbf{1}|v^{(t)}\rangle \geq \frac{1-\eta_t'}{q^n+1}\,, \tag{5.222}$$

where

$$\eta_t' = \eta_t \frac{(q^n+1)(1+q^{-n+1})}{(q-1)(q^n-1)} \leq 6\eta_t = 6\chi_2 e^{-\frac{\chi_1}{n}(t-s_{AC})}\,. \tag{5.223}$$

The inequality above is true for all $n \geq 1$ and $q \geq 2$. We choose $\chi_4 = 6\chi_2$ and $\chi_3 = \chi_1$, and the lemma is proved. □

Proof of Lemma 5.9

Proof. The gate at time step t acts on bits i_t and j_t. Suppose for some configuration $\vec{\nu}$ these bits disagree, i.e. $\nu_{i_t} \neq \nu_{j_t}$. Consider a state $|\vec{\eta}, \vec{\eta}'\rangle$ for which $\Delta|\vec{\eta}, \vec{\eta}'\rangle = |\vec{\nu}\rangle$. Then consider the quantity

$$\langle \mathbf{q}|\Delta R_{SI}^{(t)}|\vec{\eta}, \vec{\eta}'\rangle - 1 = \langle \mathbf{q}|P_I^{(t)}\Delta|\vec{\eta}, \vec{\eta}'\rangle - 1 = \langle \mathbf{q}|P_I^{(t)}|\vec{\nu}\rangle - 1 \tag{5.224}$$

$$= \sum_{\vec{\mu}} (q^{|\vec{\mu}|} - 1)\langle \vec{\mu}|L_I P^{(t)} L_I^{-1}|\vec{\nu}\rangle \tag{5.225}$$

$$= \sum_{\vec{\mu}} \frac{(q^{|\vec{\mu}|}-1)(1-q^{-2n+2|\vec{\mu}|})}{1-q^{-2n+2|\vec{\nu}|}} \langle \vec{\mu}|P^{(t)}|\vec{\nu}\rangle\,. \tag{5.226}$$

The action of $P^{(t)}$ on $|\vec{\nu}\rangle$ will force a bit flip, so there are only two possible $\vec{\mu}$ that lead to a non-zero contribution, one for which $|\vec{\mu}| = |\vec{\nu}| + 1$ and one for which $|\vec{\mu}| = |\vec{\nu}| - 1$. The matrix element (probability) of the former is $1/(q^2+1)$ and the matrix element for the latter is $q^2/(q^2+1)$. Thus, we have

$$\langle \mathbf{q}|P_I^{(t)}|\vec{\nu}\rangle - 1 = \frac{q^2(q^{|\vec{\nu}|-1}-1)(1-q^{-2n+2|\vec{\nu}|-2})}{(q^2+1)(1-q^{-2n+2|\vec{\nu}|})} + \frac{(q^{|\vec{\nu}|+1}-1)(1-q^{-2n+2|\vec{\nu}|+2})}{(q^2+1)(1-q^{-2n+2|\vec{\nu}|})} \tag{5.227}$$

$$= \frac{2q}{q^2+1} \frac{q^{|\vec{\nu}|} - \frac{q+q^{-1}}{2} - q^{-2n+2|\vec{\nu}|}\left(q^{|\vec{\nu}|}\frac{q^2+q^{-2}}{2} - \frac{q+q^{-1}}{2}\right)}{1-q^{-2n+2|\vec{\nu}|}} \tag{5.228}$$

$$\leq \frac{2q}{q^2+1}(q^{|\vec{\nu}|}-1) = \frac{2q}{q^2+1}\left(\langle \mathbf{q}|\vec{\nu}\rangle - 1\right)\,. \tag{5.229}$$

The above is true for all $\vec{\nu}$, and demonstrates that each time disagreeing bits are coupled, the total contribution under inner product with $(\langle \mathbf{q}| - \langle \mathbf{1}|)\Delta$ decreases by a constant factor.

Now consider the sequence $\prod_{t=t_0+1}^{t_1} \left(I \otimes Q_\sigma^{'(t)} Q_\sigma^{(t)}\right) R_{SI}^{(t)}$ acting on $|\vec{\eta}, \vec{\eta}'\rangle$. Since the architecture is h-regularly connected, for any t there is at least a $1/2$ chance that there will be some pair $(i_{t'}, j_{t'})$ with $t < t' \leq t + hn$ for which $\nu_{i_{t'}} \neq \nu_{j_{t'}}$ (assuming that $\vec{\nu}$ is not a fixed point). The first time this happens, it will lead to a decrease in inner product with $(\langle \mathbf{q}| - \langle \mathbf{1}|)\Delta$ by the factor $2q/(q^2+1)$. The only way this would not happen is if one of the bits $\nu_{i_{t'}}$ or $\nu_{j_{t'}}$ was flipped already by action by one of the operators $Q^{(t'')}$. However, since the $Q_\sigma^{(t)}$ operators act only on the noisy Y copy, they can only flip a bit of $\vec{\eta}'$ from a 1 to a 0, which would also induce a bit flip in $\vec{\nu}$ from a 1 to a 0. In this case, the Hamming weight decreases by 1 and the inner product with $(\langle \mathbf{q}| - \langle \mathbf{1}|)\Delta$ would decrease by a factor of $\frac{q^{|\vec{\nu}|-1}-1}{q^{|\vec{\nu}|}-1}$ which is less than $2q/(q^2+1)$.

Thus, if z_0 is the largest integer such that $t_0 + z_0 hn \leq t_1$, then

$$\begin{aligned}
&\langle \mathbf{q}|\Delta \prod_{t=t_0+1}^{t_1} \left((I \otimes Q_\sigma^{'(t)} Q_\sigma^{(t)}) R_{SI}^{(t)}\right) |v\rangle - 1 \\
&\leq \left(\frac{1}{2} + \frac{1}{2}\frac{2q}{q^2+1}\right)^{z_0} (\langle \mathbf{q}|\Delta|v\rangle - 1) && (5.230)\\
&\leq \left(\frac{1}{2} + \frac{1}{2}\frac{2q}{q^2+1}\right)^{\frac{t_1-t_0}{hn}-1} (\langle \mathbf{q}|\Delta|v\rangle - 1) && (5.231)\\
&= \chi_6 \exp\left(-\frac{\chi_5(t_1-t_0)}{n}\right) (\langle \mathbf{q}|\Delta|v\rangle - 1) && (5.232)
\end{aligned}$$

for appropriate choice of χ_5 and χ_6. □

Proof of Lemma 5.10

Proof. When probability mass is redirected from S-destined at time step $t-1$ to I-destined at time step t', it may begin with Hamming weight as large as $n-1$. But since it is I-destined, it will quickly move down in Hamming weight. We wish to quantify this phenomenon. First of all,

$$\begin{aligned}
\langle \mathbf{1}|\Pi_w \Delta|v_{SI}^{(t,t')}\rangle &= \sum_{\vec{\mu}:|\vec{\mu}|=w} \langle \vec{\mu}|\Delta|v_{SI}^{(t,t')}\rangle = \frac{\sum_{\vec{\mu}:|\vec{\mu}|=w}(q^{|\vec{\mu}|}-1)\langle \vec{\mu}|\Delta|v_{SI}^{(t,t')}\rangle}{q^w - 1} && (5.233)\\
&\leq \frac{\langle \mathbf{q}|\Delta|v_{SI}^{(t,t')}\rangle - \langle \mathbf{1}|v_{SI}^{(t,t')}\rangle}{q^w - 1}\,. && (5.234)
\end{aligned}$$

Now, note that $|v_{SI}^{(t,t')}\rangle = \prod_{t''=t'+1}^{t} \left((I \otimes Q_\sigma^{\prime(t'')} Q_\sigma^{(t'')}) R_{SI}^{(t'')} \right) |v_{SI}^{(t',t')}\rangle$, so we can invoke Lemma 5.9.

$$\langle \mathbf{1}|\Pi_w \Delta|v_{SI}^{(t,t')}\rangle \leq \frac{\langle \mathbf{q}|\Delta|v_{SI}^{(t',t')}\rangle - \langle \mathbf{1}|v_{SI}^{(t,t')}\rangle}{q^w - 1} \chi_6 \exp\left(-\frac{\chi_5(t-t')}{n}\right) \tag{5.235}$$

$$\leq \frac{q^n - 1}{q^w - 1} \langle \mathbf{1}, \mathbf{1}|v_{SI}^{(t',t')}\rangle \chi_6 \exp\left(-\frac{\chi_5(t-t')}{n}\right), \tag{5.236}$$

where the second line follows because q^n is the maximum entry in $\langle \mathbf{q}|$, and the quantity $\langle \mathbf{1}|v_{SI}^{(t,t')}\rangle$ does not change as t increases (it evolves by stochastic transformations).

We now invoke Lemma 5.6 (in the first line) and Lemma 5.5 (in the second line) to say

$$\langle \mathbf{1}|\Pi_w \Delta|v_{SI}^{(t,t')}\rangle \leq \frac{q^n - 1}{q^w - 1}(2\sigma - \sigma^2)\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t'-1)}\rangle \chi_6 e^{-\frac{\chi_5(t-t')}{n}} \tag{5.237}$$

$$\leq \frac{q^n - 1}{q^w - 1}(2\sigma - \sigma^2)(1-\sigma)^{-2(t-t'+1)}\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t)}\rangle \chi_6 e^{-\frac{\chi_5(t-t')}{n}} \tag{5.238}$$

$$\leq \sigma(4\chi_6 q^{n-w}) \exp\left(-\frac{\chi_5(t-t')}{n} + 2(t-t'+1)\log\left(\frac{1}{1-\sigma}\right)\right) \langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t)}\rangle, \tag{5.239}$$

where the extra factor of 2 comes from a very crude bound $(q^n - 1)/(q^w - 1) \leq 2q^{n-w}$. As long as χ_5/n is greater than $2\log(1/(1-\sigma))$, the above is exponentially decaying in t. This will be the case whenever $\sigma \leq 1 - \exp(-\chi_5/2n))$. There is an n_0 and χ_7 such that $\sigma \leq \chi_7/n$ whenever $n \geq n_0$ is a weaker condition. Alternatively, we could make a simpler bound by invoking Lemma 5.6 and Lemma 5.5, but not Lemma 5.9.

$$\langle \mathbf{1}|\Pi_w \Delta|v_{SI}^{(t,t')}\rangle \leq \langle \mathbf{1}, \mathbf{1}|v_{SI}^{(t,t')}\rangle \leq 2\sigma\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t'-1)}\rangle \tag{5.240}$$

$$\leq 2\sigma(1-\sigma)^{-2(t-t'+1)}\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t)}\rangle \tag{5.241}$$

Both Eq. (5.239) and Eq. (5.241) will be useful.

Now, we connect $|v_{SS}^{(t)}\rangle$ to $|v_{SI}^{(t,t')}\rangle$. First we note that

$$\begin{aligned}
&\langle \mathbf{1}|\Pi_w \Delta|v_{SS}^{(t)}\rangle \\
&= \langle \mathbf{1}|\Pi_w \Delta L_{SS}|v^{(t)}\rangle = \sum_{\vec{\mu},\vec{\nu}} \langle \mathbf{1}|\Pi_w \Delta|\vec{\mu},\vec{\nu}\rangle\langle \vec{\mu},\vec{\nu}|L_{SS}|v^{(t)}\rangle && (5.242)\\
&= \sum_{\substack{\vec{\mu},\vec{\nu}\\ |\vec{\mu}|=|\vec{\nu}|+n-w}} \langle \vec{\mu},\vec{\nu}|L_{SS}|v^{(t)}\rangle = \sum_{\substack{\vec{\mu},\vec{\nu}\\ |\vec{\mu}|=|\vec{\nu}|+n-w}} \frac{q^{-2n+2|\vec{\nu}|}-q^{-2n}}{1-q^{-2n}}\langle \vec{\mu},\vec{\nu}|v^{(t)}\rangle && (5.243)\\
&= \sum_{\substack{\vec{\mu},\vec{\nu}\\ |\vec{\mu}|=|\vec{\nu}|+n-w}} \frac{q^{2|\vec{\nu}|}-1}{q^{2|\vec{\mu}|}-q^{2|\vec{\nu}|}} \frac{q^{-2n+2|\vec{\mu}|}-q^{-2n+2|\vec{\nu}|}}{1-q^{-2n}}\langle \vec{\mu},\vec{\nu}|v^{(t)}\rangle && (5.244)\\
&= \sum_{\substack{\vec{\mu},\vec{\nu}\\ |\vec{\mu}|=|\vec{\nu}|+n-w}} q^{-2(n-w)} \frac{1-q^{-2|\vec{\nu}|}}{1-q^{-2(n-w)}}\langle \vec{\mu},\vec{\nu}|L_{SI}|v^{(t)}\rangle && (5.245)\\
&\leq \frac{q^{-2(n-w)}}{1-q^{-2}} \sum_{\substack{\vec{\mu},\vec{\nu}\\ |\vec{\mu}|=|\vec{\nu}|+n-w}} \langle \vec{\mu},\vec{\nu}|v_{SI}^{(t)}\rangle = \frac{q^{-2(n-w)}}{1-q^{-2}}\langle \mathbf{1}|\Pi_w \Delta|v_{SI}^{(t)}\rangle\,. && (5.246)
\end{aligned}$$

This allows us to use Eq. (5.147) and assert

$$\langle \mathbf{1}|\Pi_w \Delta|v_{SS}^{(t)}\rangle = \frac{q^{-2(n-w)}}{1-q^{-2}} \sum_{t'=1}^{t} \langle \mathbf{1}|\Pi_w \Delta|v_{SI}^{(t,t')}\rangle\,. \tag{5.247}$$

Let $t_w = t - \lceil n(n-w)\log(q)/\chi_5 \rceil$. For $t' > t_w$, we will bound $|v_{SI}^{(t,t')}\rangle$ with Eq. (5.241), and for $t' \leq t_w$, we will use Eq. (5.239). Let us examine these sums separately. For the $t' > t_w$ portion, we make the substitution $a = t' - t_w - 1$, and we have

$$\begin{aligned}
&\sum_{t'=t_w+1}^{t} \langle \mathbf{1}|\Pi_w \Delta|v_{SI}^{(t,t')}\rangle \leq \sum_{t'=t_w+1}^{t} 2\sigma(1-\sigma)^{-2(t-t'+1)}\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle && (5.248)\\
&= \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle 2\sigma(1-\sigma)^{-2(t-t_w)} \sum_{a=0}^{t-t_w-1} (1-\sigma)^{2a} && (5.249)\\
&= \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle 2\sigma(1-\sigma)^{-2(t-t_w)} \frac{1-(1-\sigma)^{2(t-t_w-1)}}{2\sigma-\sigma^2} && (5.250)\\
&\leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle (1-\sigma)^{-2(t-t_w)}\,(4\sigma(t-t_w)) && (5.251)\\
&\leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle (1-\sigma)^{-2\lceil n(n-w)\log(q)/\chi_5 \rceil}\,(4\sigma\lceil n(n-w)\log(q)/\chi_5 \rceil) && (5.252)\\
&\leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle q^{-2n(n-w)\log(1-\sigma)/\chi_5}\chi_5' n\sigma(n-w) && (5.253)
\end{aligned}$$

for some constant χ_5' slightly larger than $4\log(q)/\chi_5$ to account for dropping the ceiling in the last line. Note that in the third-to-last line, the extra factor of 2 comes from the bound $2\sigma/(2\sigma-\sigma^2) \leq 2$.

For the $t \leq t_w$ portion, we use the substitution $a = t_w - t'$ and find (assuming that $\chi_5/n \geq 2\log(1/(1-\sigma))$)

$$\sum_{t'=1}^{t_w} \langle \mathbf{1}|\Pi_w \Delta|v_{SI}^{(t,t')}\rangle \leq \sum_{t'=1}^{t_w} \sigma(4\chi_6 q^{n-w}) e^{-\frac{\chi_5(t-t')}{n} + 2(t-t'+1)\log\left(\frac{1}{1-\sigma}\right)} \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \tag{5.254}$$

$$= \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \sigma(4\chi_6 q^{n-w}) \sum_{a=0}^{t_w-1} e^{-\frac{\chi_5(t-t_w+a)}{n} + 2(t-t_w+a+1)\log\left(\frac{1}{1-\sigma}\right)} \tag{5.255}$$

$$\leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \sigma(4\chi_6 q^{n-w}) \sum_{a=0}^{\infty} e^{-\frac{\chi_5(t-t_w+a)}{n} + 2(t-t_w+a+1)\log\left(\frac{1}{1-\sigma}\right)} \tag{5.256}$$

$$= \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \sigma(4\chi_6 q^{n-w}) \frac{\exp\left(-\lceil n(n-w)\log(q)/\chi_5\rceil(\frac{\chi_5}{n} + 2\log(1-\sigma))\right)}{(1-e^{-\chi_5/n - 2\log(1-\sigma)})(1-\sigma)^2} \tag{5.257}$$

$$\leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \sigma(4\chi_6) \frac{\exp\left(-2\lceil n(n-w)\log(q)/\chi_5\rceil \log(1-\sigma)\right)}{(1-e^{-\chi_5/n - 2\log(1-\sigma)})(1-\sigma)^2} \tag{5.258}$$

$$\leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \sigma \chi_6' q^{-2n(n-w)\log(1-\sigma)/\chi_5} \tag{5.259}$$

for some constant χ_6'. Plugging the bounds on the two parts of the sum into Eq. (5.247), we find

$$\frac{\langle \mathbf{1}|\Pi_w \Delta|v_{SS}^{(t)}\rangle}{\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle} \leq \frac{q^{-2(n-w)}}{1-q^{-2}} n\sigma q^{-2n(n-w)\log(1-\sigma)/\chi_5} \left(\chi_5'(n-w) + \frac{\chi_6'}{n}\right) \tag{5.260}$$

$$\leq \chi_9 q^{-2(n-w)} n\sigma(n-w) q^{c'(n-w)} \tag{5.261}$$

$$= \chi_9 n\sigma(n-w) q^{-(n-w)} q^{-(1-c')(n-w)} \tag{5.262}$$

for some constants χ_9 and c' which is less than 1 whenever $\sigma \leq \chi_7/n$ and $n \geq n_0$ hold. Thus we may define $\chi_8 = (1-c')\log(q)$ and the lemma is proved. □

Proof of Lemma 5.11

Proof. Recall that $\mathcal{Z}_\sigma = \langle \mathbf{1},\mathbf{q}|v^{(s)}\rangle$. and that $|v_{SS}^{(t)}\rangle = L_{SS}|v^{(t)}\rangle$. The matrix L_{SS}^{-1} is defined to be the Moore-Penrose pseudo-inverse of L_{SS}, and note that the null space of L_{SS} is the space spanned by $|\vec{\nu},0^n\rangle$ for all $\vec{\nu}$. The projector onto this subspace is $|\mathbf{1},0^n\rangle\langle\mathbf{1},0^n|$. Thus,

$$|v^{(s)}\rangle = I|v^{(s)}\rangle = (|\mathbf{1},0^n\rangle\langle\mathbf{1},0^n| + L_{SS}^{-1}L_{SS})|v^{(s)}\rangle \tag{5.263}$$

$$= |\mathbf{1},0^n\rangle\langle\mathbf{1},0^n|v^{(s)}\rangle + L_{SS}^{-1}|v_{SS}^{(s)}\rangle\,. \tag{5.264}$$

The lower bound is shown as follows:

$$\mathcal{Z}_\sigma - 1 = \sum_{\vec{\nu}} \left(q^{|\vec{\nu}|} - 1\right) \langle \mathbf{1}, \vec{\nu} | v^{(s)} \rangle = \sum_{\vec{\nu} \neq 0^n} \left(q^{|\vec{\nu}|} - 1\right) \langle \mathbf{1}, \vec{\nu} | v^{(s)} \rangle \tag{5.265}$$

$$= \sum_{\vec{\nu} \neq 0^n} \left(q^{|\vec{\nu}|} - 1\right) \langle \mathbf{1}, \vec{\nu}| \left(|\mathbf{1}, 0^n\rangle \langle \mathbf{1}, 0^n | v^{(s)} \rangle + L_{SS}^{-1} | v_{SS}^{(s)} \rangle \right) \tag{5.266}$$

$$= \sum_{\vec{\nu} \neq 0^n} \left(q^{|\vec{\nu}|} - 1\right) \langle \mathbf{1}, \vec{\nu} | L_{SS}^{-1} | v_{SS}^{(s)} \rangle \tag{5.267}$$

$$= \sum_{\vec{\nu} \neq 0^n} \left(q^{|\vec{\nu}|} - 1\right) \langle \vec{\nu} | \frac{1 - q^{-2n}}{q^{-2n+2|\vec{\nu}|} - q^{-2n}} | v_S^{(s)} \rangle \tag{5.268}$$

$$= \sum_{\vec{\nu} \neq 0^n} \left(q^n - 1\right) \left(\frac{1 + q^n}{1 + q^{|\vec{\nu}|}} \right) \langle \mathbf{1}, \vec{\nu} | v_{SS}^{(s)} \rangle \tag{5.269}$$

$$\geq \sum_{\vec{\nu} \neq 0^n} \left(q^n - 1\right) \langle \mathbf{1}, \vec{\nu} | v_{SS}^{(s)} \rangle \tag{5.270}$$

$$= (q^n - 1) \langle \mathbf{1}, \mathbf{1} | v_{SS}^{(s)} \rangle \,. \tag{5.271}$$

Now, we will show the upper bound.

$$\mathcal{Z}_\sigma - 1 = \sum_{\vec{\nu}} \left(q^{|\vec{\nu}|} - 1\right) \langle \mathbf{1}, \vec{\nu} | v^{(s)} \rangle \tag{5.272}$$

$$= \sum_{\vec{\nu}} \left(q^{|\vec{\nu}|} - 1\right) \left(\langle 1^n, \vec{\nu}| + \sum_{\vec{\mu} \neq 1^n} \langle \vec{\mu}, \vec{\nu}| \right) |v^{(s)}\rangle \tag{5.273}$$

$$= \sum_{\vec{\nu}} \left(\left(q^{|\vec{\nu}|} - 1\right) \langle 1^n, \vec{\nu} | v^{(s)} \rangle + \sum_{\vec{\mu} \neq 1^n} \left(q^{|\vec{\nu}|} - 1\right) \langle \vec{\mu}, \vec{\nu} | v^{(s)} \rangle \right) \tag{5.274}$$

$$\leq \sum_{\vec{\nu}} \left(\left(q^{|\vec{\nu}|} - 1\right) \langle 1^n, \vec{\nu} | v^{(s)} \rangle + \sum_{\vec{\mu} \neq 1^n} \left(q^{|\vec{\mu}|} - 1\right) \langle \vec{\mu}, \vec{\nu} | v^{(s)} \rangle \right) \tag{5.275}$$

$$= \sum_{\vec{\nu}} \left(\left(q^{|\vec{\nu}|} - 1\right) \langle 1^n, \vec{\nu} | v^{(s)} \rangle \right) + Z_0 - 1 - (q^n - 1) \langle 1^n, \mathbf{1} | v^{(s)} \rangle \,, \tag{5.276}$$

where we have used $Z_0 = \sum_{\vec{\nu}} \sum_{\vec{\mu}} q^{|\vec{\mu}|} \langle \vec{\mu}, \vec{\nu} | v^{(s)} \rangle$. Now we invoke Lemma 5.8, to say

$$\mathcal{Z}_\sigma - 1 \leq \sum_{\vec{\nu}} \left(q^{|\vec{\nu}|} - 1\right) \langle 1^n, \vec{\nu} | v^{(s)} \rangle + Z_0 - 1 - \frac{q^n - 1}{q^n + 1} (1 - \eta_s') \tag{5.277}$$

$$= \sum_{\vec{\nu}} \left(q^{|\vec{\nu}|} - 1\right) \langle 1^n, \vec{\nu} | v^{(s)} \rangle + \left(Z_0 - \frac{2q^n}{q^n + 1} \right) + \eta_s' \left(\frac{q^n - 1}{q^n + 1} \right) \tag{5.278}$$

$$\leq \sum_{\vec{\nu}} \left(q^{|\vec{\nu}|} - 1\right) \langle 1^n, \vec{\nu} | v^{(s)} \rangle + \left(Z_0 - \frac{2q^n}{q^n + 1} \right) + \eta_s' \,. \tag{5.279}$$

Now we invoke Lemma 5.7 to bound $Z_0 - 2q^n/(q^n+1)$ in the first step below, and continue on. Denote $\eta_s'' = \eta_s + \eta_s'$.

$$\mathcal{Z}_\sigma - 1 \leq \sum_{\vec{\nu}} \left(q^{|\vec{\nu}|} - 1\right) \langle 1^n, \vec{\nu}| v^{(s)}\rangle + \eta_s' + \eta_s \tag{5.280}$$

$$= \sum_{\vec{\nu}} \left(q^{|\vec{\nu}|} - 1\right) \langle 1^n, \vec{\nu}| L_{SS}^{-1} L_{SS} |v^{(s)}\rangle + \eta_s'' \tag{5.281}$$

$$= \sum_{\vec{\nu}} \left(\frac{(q^{|\vec{\nu}|}-1)(1-q^{-2n})}{q^{-2n+2|\vec{\nu}|} - q^{-2n}}\right) \langle 1^n, \vec{\nu}| v_{SS}^{(s)}\rangle + \eta_s'' \tag{5.282}$$

$$= \sum_{\vec{\nu}} (q^n - 1)\left(\frac{q^n+1}{q^{|\vec{\nu}|}+1}\right) \langle 1^n, \vec{\nu}| v_{SS}^{(s)}\rangle + \eta_s'' \tag{5.283}$$

$$\leq \eta_s'' + (q^n-1)\sum_{\vec{\nu}} q^{n-|\vec{\nu}|} \langle 1^n, \vec{\nu}| v_{SS}^{(s)}\rangle \tag{5.284}$$

$$= \eta_s'' + (q^n-1)\sum_{\vec{\nu}} q^{n-|\vec{\nu}|} \langle \vec{\nu}| \Delta |v_{SS}^{(s)}\rangle \tag{5.285}$$

$$= \eta_s'' + (q^n-1)\langle 1^n, 1^n| v_{SS}^{(s)}\rangle + (q^n-1)\sum_{w=1}^{n-1} q^{n-w} \langle \mathbf{1}| \Pi_w \Delta |v_{SS}^{(s)}\rangle \tag{5.286}$$

$$\leq \eta_s'' + (q^n-1)\langle 1^n, 1^n| v_{SS}^{(s)}\rangle + (q^n-1)\sum_{w=1}^{n-1} q^{n-w} n\sigma\xi_w \langle \mathbf{1}, \mathbf{1}| v_{SS}^{(s)}\rangle \tag{5.287}$$

$$\leq \eta_s'' + (q^n-1)\langle \mathbf{1}, \mathbf{1}| v_{SS}^{(s)}\rangle \left(1 + \chi_9 n\sigma \sum_{w=1}^{n-1} (n-w) e^{-\chi_8(n-w)}\right), \tag{5.288}$$

where in the second-to-last line, we have invoked Lemma 5.10, which requires $\sigma \leq \chi_7/n$ and $n \geq n_0$ (leading to our requirements in this lemma that $\sigma \leq \chi_{13}/n$ and $n \geq n_0$). Now, we make the choice of $\chi_{10} = \chi_9 \sum_{w=1}^{n-1}(n-w)e^{-\chi_8(n-w)} \leq \chi_9 \sum_{w=1}^{\infty} we^{-\chi_8 w} = O(1)$, which yields the following. (In line 2, we invoke Lemma 5.5.)

$$\mathcal{Z}_\sigma - 1 \leq (q^n-1)\langle \mathbf{1}, \mathbf{1}| v_{SS}^{(s)}\rangle \left(1 + \chi_{10} n\sigma + \frac{\eta_s''}{(q^n-1)\langle \mathbf{1}, \mathbf{1}| v_{SS}^{(s)}\rangle}\right) \tag{5.289}$$

$$\leq (q^n-1)\langle \mathbf{1}, \mathbf{1}| v_{SS}^{(s)}\rangle \left(1 + \chi_{10} n\sigma + \frac{q^n+1}{q^n-1}\eta_s''(1-\sigma)^{-2s}\right) \tag{5.290}$$

$$\leq (q^n-1)\langle \mathbf{1}, \mathbf{1}| v_{SS}^{(s)}\rangle \left(1 + \chi_{10} n\sigma + 3\eta_s''(1-\sigma)^{-2s}\right) \tag{5.291}$$

$$\leq (q^n-1)\langle \mathbf{1}, \mathbf{1}| v_{SS}^{(s)}\rangle \left(1 + \chi_{10} n\sigma + \chi_{12} e^{-\frac{\chi_{11}}{n}(s-s_{AC})+4s\sigma}\right) \tag{5.292}$$

$$\leq (q^n-1)\langle \mathbf{1}, \mathbf{1}| v_{SS}^{(s)}\rangle \exp\left(1 + \chi_{10} n\sigma + \chi_{12} e^{-\frac{\chi_{11}}{n}(s-s_{AC})+4s\sigma}\right), \tag{5.293}$$

where the third-to-last line is true for all $q \geq 2$ and $n \geq 1$, and the second-to-last line plugs in the equations for η_s and η_s', chooses constants χ_{11} and χ_{12} appropriately, and asserts $(1-\sigma)^{2s} \leq e^{-4\sigma s}$, which is true whenever $\sigma \leq 0.79$, so it is certainly true under the assumption $\sigma \leq \chi_7/n$ for sufficiently large n. □

Proof of Lemma 5.12

Proof. Recall that $\langle \mathbf{1},\mathbf{1}|v_{SS}^{(0)}\rangle = 1/(q^n+1)$. Let $t_0 = dn/2$.

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t_0+n/2)}\rangle = \langle \mathbf{1},\mathbf{1}|L_{SS}\prod_{t=t_0+1}^{t_0+n/2} R_\sigma^{(t)}|v^{(t_0)}\rangle \tag{5.294}$$

$$= \langle \mathbf{1},\mathbf{1}|L_{SS}\prod_{t=t_0+1}^{t_0+n/2} R_\sigma^{(t)}L_{SS}^{-1}|v_{SS}^{(t_0)}\rangle \tag{5.295}$$

$$= \langle \mathbf{1},\mathbf{1}|L_{SS}\prod_{t=t_0+1}^{t_0+n/2} (I\otimes Q_\sigma^{'(t)}Q_\sigma^{(t)})\prod_{t=t_0+1}^{t_0+n/2} (I\otimes R_0^{(t)})L_{SS}^{-1}|v_{SS}^{(t_0)}\rangle \tag{5.296}$$

$$= \langle \mathbf{1},\mathbf{1}|L_{SS}\prod_{t=t_0+1}^{t_0+n/2} (I\otimes Q_\sigma^{'(t)}Q_\sigma^{(t)})L_{SS}^{-1}\prod_{t=t_0+1}^{t_0+n/2} R_{SS}^{(t)}|v_{SS}^{(t_0)}\rangle \tag{5.297}$$

$$= \langle \mathbf{1},\mathbf{1}|(I\otimes L_S)\prod_{t=t_0+1}^{t_0+n/2} (I\otimes Q_\sigma^{'(t)}Q_\sigma^{(t)})(I\otimes L_S^{-1})\prod_{t=t_0+1}^{t_0+n/2} R_{SS}^{(t)}|v_{SS}^{(t_0)}\rangle \tag{5.298}$$

$$= \langle \mathbf{1}|L_S\prod_{t=t_0+1}^{t_0+n/2} (Q_\sigma^{'(t)}Q_\sigma^{(t)})L_S^{-1}(\langle \mathbf{1}|\otimes I)\prod_{t=t_0+1}^{t_0+n/2} R_{SS}^{(t)}|v_{SS}^{(t_0)}\rangle \tag{5.299}$$

$$= \sum_{\vec{\nu}\neq 0^n}\langle \mathbf{1}|L_S\prod_{t=t_0+1}^{t_0+n/2} (Q_\sigma^{'(t)}Q_\sigma^{(t)})L_S^{-1}|\vec{\nu}\rangle\langle \mathbf{1},\vec{\nu}|\prod_{t=t_0+1}^{t_0+n/2} R_{SS}^{(t)}|v_{SS}^{(t_0)}\rangle\,. \tag{5.300}$$

We now examine the quantity

$$\langle \mathbf{1}|L_S\prod_{t=t_0+1}^{t_0+n/2} (Q_\sigma^{'(t)}Q_\sigma^{(t)})L_S^{-1}|\vec{\nu}\rangle = \sum_{\vec{\mu}}\langle \vec{\mu}|L_S\prod_{t=t_0+1}^{t_0+n/2} (Q_\sigma^{'(t)}Q_\sigma^{(t)})L_S^{-1}|\vec{\nu}\rangle \tag{5.301}$$

$$= \sum_{\vec{\mu}}\frac{q^{-2n+2|\vec{\mu}|}-q^{-2n}}{q^{-2n+2|\vec{\nu}|}-q^{-2n}}\langle \vec{\mu}|\prod_{t=t_0+1}^{t_0+n/2} (Q_\sigma^{'(t)}Q_\sigma^{(t)})|\vec{\nu}\rangle\,. \tag{5.302}$$

Note that, because of the layered property, all n qudits are acted upon by one of the $Q_\sigma^{(t)}$ or $Q_\sigma^{'(t)}$. This can cause some 1s to flip to 0s (with probability σ). For $\vec{\mu}$ to have non-zero contribution it must have $\mu_i \leq \nu_i$ for all i, a condition we denote by $\vec{\mu}\leq\vec{\nu}$, and in this case we have

$$\langle \vec{\mu}|\prod_{t=t_0+1}^{t_0+n/2} (Q_\sigma^{'(t)}Q_\sigma^{(t)})|\vec{\nu}\rangle = (1-\sigma)^{|\vec{\mu}|}\sigma^{|\vec{\nu}|-|\vec{\mu}|}\,. \tag{5.303}$$

Note also the following sum formula, which holds for any z.

$$\sum_{\vec{\mu}\leq\vec{\nu}} q^{z|\vec{\mu}|}(1-\sigma)^{|\vec{\mu}|}\sigma^{|\vec{\nu}|-|\vec{\mu}|} = \sum_{x=0}^{|\vec{\nu}|}\binom{|\vec{\nu}|}{x} q^{zx}(1-\sigma)^{x}\sigma^{|\vec{\nu}|-x} = (\sigma+q^{z}(1-\sigma))^{|\vec{\nu}|}\,. \tag{5.304}$$

We find

$$\langle \mathbf{1}|L_S \prod_{t=t_0+1}^{t_0+n/2} (Q_\sigma^{\prime(t)}Q_\sigma^{(t)})L_S^{-1}|\vec{\nu}\rangle \tag{5.305}$$

$$= \frac{1}{q^{2|\vec{\nu}|}-1}\sum_{\vec{\mu}\leq\vec{\nu}}(q^{2|\vec{\mu}|}-1)(1-\sigma)^{|\vec{\mu}|}\sigma^{|\vec{\nu}|-|\vec{\mu}|} \tag{5.306}$$

$$= \frac{(\sigma+q^2(1-\sigma))^{|\vec{\nu}|}-1}{q^{2|\vec{\nu}|}-1} = \frac{(1-\sigma')^{|\vec{\nu}|}-q^{-2|\vec{\nu}|}}{1-q^{-2|\vec{\nu}|}}\,, \tag{5.307}$$

where $\sigma' = \sigma(1-q^{-2})$. Denote this final expression by

$$E_w = \frac{(1-\sigma')^w - q^{-2w}}{1-q^{-2w}}\,. \tag{5.308}$$

Now we claim that, for any $|\vec{\nu}| \neq 0$,

$$E_n \leq E_{|\vec{\nu}|}\,. \tag{5.309}$$

We can prove the statement above by noting that it holds for $|\vec{\nu}| = n$ and observing that the derivative with respect to $|\vec{\nu}|$ is always negative (in this verification, note that $(1-\sigma') \geq 1/q$ holds for all $\sigma \leq 1$).

Collecting these observations, we have

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t_0+n/2)}\rangle \geq \sum_{\vec{\nu}\neq 0^n} E_n \langle \mathbf{1},\vec{\nu}| \prod_{t=t_0+1}^{t_0+n/2} R_{SS}^{(t)}|v_{SS}^{(t_0)}\rangle \tag{5.310}$$

$$= E_n \langle \mathbf{1},\mathbf{1}| \prod_{t=t_0+1}^{t_0+n/2} R_{SS}^{(t)}|v_{SS}^{(t_0)}\rangle \tag{5.311}$$

$$= E_n \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t_0)}\rangle\,. \tag{5.312}$$

Hence, the lower bound in the lemma statement follows by recursively applying the above conclusion for increasing d.

To show the upper bound, we return to Eq. (5.300). Note that $E_w \leq 1$. We can restate what we know and divide the mass into whether or not the noiseless copy has reached the S^n fixed point, and if it has, what value w for the Hamming weight the noisy copy ends up at.

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t_0+n/2)}\rangle = A_{not} + \sum_{w=1}^{n} E_w A_w\,, \tag{5.313}$$

where

$$A_{not} = \sum_{\vec{\nu}\neq 0^n, \vec{\mu}\neq 1^n} E_{|\vec{\nu}|} \langle \mathbf{1}, \vec{\nu}| \prod_{t=t_0+1}^{t_0+n/2} R_{SS}^{(t)} \left(|\vec{\mu}\rangle\langle\vec{\mu}| \otimes I \right) |v_{SS}^{(t_0)}\rangle \tag{5.314}$$

$$A_w = \sum_{\vec{\nu}:|\vec{\nu}|=w} \langle \mathbf{1}, \vec{\nu}| \prod_{t=t_0+1}^{t_0+n/2} R_{SS}^{(t)} \left(|1^n\rangle\langle 1^n| \otimes I \right) |v_{SS}^{(t_0)}\rangle \,. \tag{5.315}$$

Since $E_{|\vec{\nu}|} \leq 1$, we may directly apply Lemma 5.8 and bound $A_{not} \leq \eta'_{t_0}/(q^n + 1)$.

To bound A_w, we will need to use Lemma 5.10. Applying the layer of $R_{SS}^{(t)}$ from $t = t_0 + 1$ to $t = t_0 + n/2$ can at most double the number of 0 bits, since each qudit participates in at most one gate. So, in order to land at a configuration with Hamming weight w, you have to start with Hamming weight at most $\lfloor \frac{n+w}{2} \rfloor$. In other words,

$$A_w \leq \sum_{w'=1}^{\lfloor \frac{n+w}{2} \rfloor} \sum_{\vec{\mu}:|\vec{\mu}|=w'} \langle 1^n, \vec{\mu}| v_{SS}^{(t_0)} \rangle \,. \tag{5.316}$$

When $w < n$, the right-hand side of the above is then bounded with Lemma 5.10, which requires $\sigma \leq \chi_7/n$ and $n \geq n_0$ (and thus the upper bound portion of lemma inherits these requirements).

$$A_w \leq \sum_{w'=1}^{\lfloor \frac{n+w}{2} \rfloor} \sum_{\vec{\mu}:|\vec{\mu}|=w'} \langle \vec{\mu}|\Delta|v_{SS}^{(t_0)}\rangle = \sum_{w'=1}^{\lfloor \frac{n+w}{2} \rfloor} \langle \mathbf{1}|\Pi_{w'}\Delta|v_{SS}^{(t_0)}\rangle \tag{5.317}$$

$$\leq \sum_{w'=1}^{\lfloor \frac{n+w}{2} \rfloor} n\sigma\chi_9(n-w')q^{-(n-w')}e^{-\chi_8(n-w')}\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t_0)}\rangle \tag{5.318}$$

$$\leq n\sigma\chi_9\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t_0)}\rangle \sum_{a=\lceil \frac{n-w}{2} \rceil}^{\infty} a e^{-a(\chi_8+\log(q))} \,, \tag{5.319}$$

where we have used the substitution $a = n - w'$. For any c, there is a constant c'' such that $\sum_{a=a_0}^{\infty} ae^{-ca}$ is bounded by $c''e^{-ca_0}$. Thus, there is a constant c'' such that the sum in the expression above is bounded by

$$c''e^{-(\chi_8+\log(q))\lceil \frac{n-w}{2} \rceil} \leq c''e^{-(\chi_8+\log(q))\frac{n-w}{2}} \,. \tag{5.320}$$

Note also that by construction $\sum_{w=1}^{n} A_w \leq \langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t_0)}\rangle$. Thus,

$$\sum_{w=1}^{n} E_w A_w = \sum_{w=1}^{n} (E_n + E_w - E_n) A_w \leq E_n \langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t_0)}\rangle + \sum_{w=1}^{n-1} (E_w - E_n) A_w \,, \tag{5.321}$$

which we can insert into Eq. (5.313), along with the bounds on A_w, giving

$$\langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t_0+n/2)} \rangle \leq \langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t_0)} \rangle \left(E_n + n\sigma \sum_{w=1}^{n-1} (E_w - E_n) f e^{-f'(n-w)} \right) + \frac{\eta'_{t_0}}{q^n + 1} \tag{5.322}$$

for appropriate constants f and f'. We also have

$$\frac{E_w}{E_n} = \frac{1 - q^{-2n}}{1 - q^{-2w}} \frac{(1-\sigma')^w - q^{-2w}}{(1-\sigma')^n - q^{-2n}} \leq (1-\sigma')^{-(n-w)}, \tag{5.323}$$

which can be verified by observing that the quantity

$$\frac{E_w}{E_n}(1-\sigma')^{n-w} = \frac{(1-q^{-2n})(1-(q\sqrt{1-\sigma'})^{-2w})}{(1-q^{-2w})(1-(q\sqrt{1-\sigma'})^{-2n})} \tag{5.324}$$

achieves its maximum with respect to σ' when $\sigma' = 0$, where it equals 1. The quantity in parentheses in Eq. (5.322) is now at most

$$\begin{aligned} &\left(E_n + n\sigma E_n \sum_{w=1}^{n-1} f e^{-f'(n-w)} (e^{-\log(1-\sigma')(n-w)} - 1) \right) \\ &\leq \left(E_n + n\sigma E_n \sum_{w=1}^{n-1} f e^{-f'(n-w)} \tau\sigma(n-w) \right) && (5.325) \\ &\leq E_n \left(1 + f'' n\sigma^2\right), && (5.326) \end{aligned}$$

where in the second line, we bound $e^{-x\log(1-\sigma)} - 1$ by $\tau\sigma x$ for some constant τ, which holds for x sufficiently small, as is the case when $\sigma \leq O(1/n)$ with n sufficiently large; in the third line, we choose the appropriate constant f'' as the exponentially decaying sum is bounded by a constant. This gives us the recursion relation

$$\langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t_0+n/2)} \rangle \leq \langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t_0)} \rangle E_n (1 + f'' n\sigma^2) + \frac{\eta'_{t_0}}{q^n + 1}. \tag{5.327}$$

For the first few layers, before anti-concentration has been reached and η'_{t_0} has become small, we will just use the simpler naive bound $\langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t_0+n/2)} \rangle \leq \langle \mathbf{1}, \mathbf{1} | v_{SS}^{(t_0)} \rangle$. Suppose the anti-concentration depth is $d_{AC} = 2s_{AC}/n$. Then we have

$$\begin{aligned} \frac{\eta'_{dn/2}}{q^n + 1} &\leq \frac{\chi_4}{q^n + 1} e^{-\chi_3(d - d_{AC})/2} \leq \frac{\chi'_4}{q^n + 1} E_n e^{-\chi_3(d-d_{AC})/2 - n\log(1-\sigma)} && (5.328) \\ &\leq \chi'_4 E_n \langle \mathbf{1}, \mathbf{1} | v_{SS}^{(dn/2)} \rangle e^{-\chi_3(d-d_{AC})/2 - n\log(1-\sigma) - dn\log(1-\sigma)} && (5.329) \\ &\leq E_n \langle \mathbf{1}, \mathbf{1} | v_{SS}^{(dn/2)} \rangle n\sigma e^{-\chi'_3(d-d^*)}, && (5.330) \end{aligned}$$

where in line 1, we refer back to the definition of E_n and choose χ_4' slightly larger than χ_4, in line 2, we use Lemma 5.5, and in line 3 we choose

$$d^* = d_{AC}\chi_3/2\chi_3' + f''' + \log(1/n\sigma)/\chi_3' \tag{5.331}$$

for some constant f''' that is $O(1)$ whenever $-n\log(1-\sigma)$ is $O(1)$. Note that this also requires $n\log(1-\sigma) \leq \chi_3$. We can choose the constant a_3 such that the condition $\sigma \leq a_3/n$ implies these requirements hold. Note we also must choose a weaker exponential decay constant χ_3'. Thus our recursion relation is

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t_0+n/2)}\rangle \leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t_0)}\rangle E_n(1 + f''n\sigma^2 + n\sigma e^{-\chi_3'(d-d^*)}) \,. \tag{5.332}$$

Iterating this equation starting at $d = d^*$, we get

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(dn/2)}\rangle \leq \frac{E_n^{d-d^*}}{q^n+1} \prod_{d'=d^*+1}^{d} (1 + f''n\sigma^2 + n\sigma e^{-\chi_3'(d'-d^*)}) \tag{5.333}$$

$$\leq \frac{E_n^{d-d^*}}{q^n+1} \exp\left(\sum_{d'=d^*+1}^{d} (f''n\sigma^2 + n\sigma e^{-\chi_3'(d'-d^*)}) \right) \tag{5.334}$$

$$\leq \frac{E_n^{d-d^*}}{q^n+1} \exp\left((d-d^*)(f''n\sigma^2) + n\sigma\chi_3''\right) \tag{5.335}$$

for some choice of χ_3'' (the exponentially decaying sum is bounded). Now, we note from the definition of E_n that as long as $\sigma \leq O(1/n)$, there is a constant g (slightly larger than 1) such that $E_n \geq \exp(-gn\sigma')$, allowing us to say

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(dn/2)}\rangle \leq \frac{E_n^{d}}{q^n+1} \exp\left(f''n\sigma^2 d + gn\sigma' d^* + n\sigma\chi_3''\right) \,, \tag{5.336}$$

which, recalling the definition of d^* in Eq. (5.331), implies the lemma statement for appropriate choices of a_0, a_1, and a_2. □

Proof of Lemma 5.13

Proof. In the layered case (proof of Lemma 5.12), we considered the action of all $n/2$ gates in a layer at once. For complete-graph, we can treat each gate individually. Following the layered derivation to Eq. (5.300), for complete-graph we have

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle = \sum_{\vec{\nu}\neq 0^n} \langle \mathbf{1}|L_S Q_\sigma'^{(t)} Q_\sigma^{(t)} L_S^{-1}|\vec{\nu}\rangle \langle \mathbf{1},\vec{\nu}|R_{SS}^{(t)}|v_{SS}^{(t-1)}\rangle \,.$$

Here the tth gate acts on two qudits i_t and j_t, but in forming $|v^{(t)}\rangle$ from $|v^{(t-1)}\rangle$, we take the average over all possible choices. After action by $R_{SS}^{(t)}$ the values assigned at position i_t and j_t must be set equal. If they are assigned 1, then errors can send the new configuration to one of four possible configurations,

corresponding to errors on none, one, or both qudits. If they are assigned 0 then no errors are possible. If we assume that $\nu_{i_t} = \nu_{j_t} = 1$, then zero errors occurs with probability $(1-\sigma)^2$, one error with probability $2\sigma(1-\sigma)$, and two errors with probability σ^2. Thus, we have

$$\begin{aligned}
&\langle \mathbf{1}| L_S Q_\sigma^{'(t)} Q_\sigma^{(t)} L_S^{-1} |\vec{\nu}\rangle \\
&= (1-\sigma)^2 + 2\sigma(1-\sigma)\frac{q^{-2n+2|\vec{\nu}|-2} - q^{-2n}}{q^{-2n+2|\vec{\nu}|} - q^{-2n}} + \sigma^2 \frac{q^{-2n+2|\vec{\nu}|-4} - q^{-2n}}{q^{-2n+2|\vec{\nu}|} - q^{-2n}} \qquad (5.337)\\
&= \frac{(1-\sigma')^2 - q^{-2|\vec{\nu}|}}{1 - q^{-2|\vec{\nu}|}}\,, \qquad (5.338)
\end{aligned}$$

where $\sigma' = \sigma(1-q^{-2})$. Define the final expression as

$$J_w = \frac{(1-\sigma')^2 - q^{-2w}}{1-q^{-2w}}\,. \qquad (5.339)$$

The quantity J_w is monotonically increasing in w and satisfies $J_w \leq J_n$ for all w. Meanwhile, if $\nu_{i_t} = \nu_{j_t} = 0$, then $\langle \mathbf{1}| L_S Q_\sigma^{'(t)} Q_\sigma^{(t)} L_S^{-1} |\vec{\nu}\rangle = 1$.

Recall the marginal dynamics of $R_{SS}^{(t)}$ on the noisy copy are simply $P_S^{(t)}$. Suppose the noisy copy starts at a configuration $|\vec{\eta}\rangle$. If $|\vec{\eta}| = w$, then let $\phi_{SS,w}$ be the probability that the qudits i_t and j_t are both assigned S, $\phi_{IS,w}$ be the probability one is assigned S and one is assigned I, and $\phi_{II,w}$ be the probability both are assigned I.

$$\phi_{SS,w} = \frac{w(w-1)}{n(n-1)} \qquad (5.340)$$

$$\phi_{IS,w} = \frac{2w(n-w)}{n(n-1)} \qquad (5.341)$$

$$\phi_{II,w} = \frac{(n-w)(n-w-1)}{n(n-1)}\,. \qquad (5.342)$$

Note that $\phi_{SS,w} + \phi_{IS,w} + \phi_{II,w} = 1$. In the case where one is I and one is S, the I is flipped to S by $P_S^{(t)}$ with probability $P_{\uparrow,w}$ and the S is flipped to I with probability $P_{\downarrow,w}$, where

$$P_{\uparrow,w} = \frac{1}{q^2+1}\frac{q^{-2n+2w+2} - q^{-2n}}{q^{-2n+2w} - q^{-2n}} \qquad (5.343)$$

$$P_{\downarrow,w} = 1 - P_{\uparrow,w}\,, \qquad (5.344)$$

which increases or decreases the Hamming weight of w by 1.

Note the following equalities and inequalities:

$$P_{\downarrow,w} = \frac{1}{q^2+1}\frac{1-q^{-2w+2}}{1-q^{-2w}} \geq \frac{1}{q^2+1} - q^{-2w} \tag{5.345}$$

$$1 - J_w = \frac{1-(1-\sigma')^2}{1-q^{-2w}} = \frac{2\sigma' - \sigma'^2}{1-q^{-2w}} \tag{5.346}$$

$$J_n - J_w = \frac{(1-(1-\sigma')^2)(q^{-2w}-q^{-2n})}{(1-q^{-2n})(1-q^{-2w})} \leq \frac{q^{-2w}(2\sigma'-\sigma'^2)}{1-q^{-2w}} \tag{5.347}$$

$$\phi_{II,w} + \phi_{IS,w}P_{\downarrow,w} \geq \begin{cases} \frac{n-w}{n-1}\left(\frac{1}{q^2+1} - q^{-2w}\right) \geq \frac{n-w}{n-1}\frac{1}{q^2+1} - q^{-2w} & \text{if } w \geq n/2 \\ \frac{1}{4} & \text{if } w < n/2 \end{cases}, \tag{5.348}$$

where the last inequality follows because, when $w \geq n/2$, $\phi_{IS,w} \geq \frac{n-w}{n-1}$, and when $w < n/2$, $\phi_{II,w} \geq \frac{1}{4}$.

We may now define G_w by the following equation, where $|\vec{\eta}| = w$,

$$G_w = \sum_{\vec{\nu} \neq I^n} \langle \mathbf{1}|L_S Q_\sigma^{\prime(t)} Q_\sigma^{(t)} L_S^{-1}|\vec{\nu}\rangle\langle\vec{\nu}|P_S^{(t)}|\vec{\eta}\rangle \tag{5.349}$$

$$= \phi_{SS,w}J_w + \phi_{IS,w}(P_{\uparrow,w}J_{w+1} + P_{\downarrow,w}) + \phi_{II,w}\,. \tag{5.350}$$

We want to lower bound this quantity. If $n = 2$, then $G_1 = G_2 = J_2$. If $n > 2$, we have

$$G_w \geq \phi_{SS,w}J_w + \phi_{IS,w}(P_{\uparrow,w}J_w + P_{\downarrow,w}) + \phi_{II,w} \tag{5.351}$$

$$= J_n + (1-J_w)(\phi_{II,w} + P_{\downarrow,w}\phi_{IS,w}) - (J_n - J_w) \tag{5.352}$$

$$\geq J_n + \frac{2\sigma'-\sigma'^2}{1-q^{-2w}} \begin{cases} \frac{n-w}{n-1}\frac{1}{q^2+1} - 2q^{-2w} & \text{if } w \geq n/2 \\ \frac{1}{4} - q^{-2w} & \text{if } w < n/2 \end{cases}. \tag{5.353}$$

By inspection of the final equation, we see that $G_w \geq J_n$ for every combination $n > 2$, $w \geq 1$ (since $q > 2$) except when $w = n$, but for $w = n$, $G_w = J_n$ by definition, so $G_w \geq J_n$ also holds.

This immediately gives us

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle = \sum_{w=1}^{n} G_w \sum_{\vec{\eta}:|\vec{\eta}|=w} \langle \mathbf{1},\vec{\eta}|v_{SS}^{(t-1)}\rangle \geq J_n \sum_{w=1}^{n}\sum_{\vec{\eta}:|\vec{\eta}|=w} \langle \mathbf{1},\vec{\eta}|v_{SS}^{(t-1)}\rangle \tag{5.354}$$

$$= J_n \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t-1)}\rangle\,, \tag{5.355}$$

which proves the lower bound by recursion on increasing t and the fact that $\langle \mathbf{1},\mathbf{1}|v_{SS}^{(0)}\rangle = 1/(q^n+1)$.

To show the upper bound, we first observe

$$G_w \leq J_n + (1-J_n)(\phi_{II,w} + P_{\downarrow,w}\phi_{IS,w})\,. \tag{5.356}$$

We have the inequalities

$$1 - J_n = \frac{2\sigma' - \sigma'^2}{1 - q^{-2n}} \leq 2\sigma \tag{5.357}$$

$$\phi_{II,w} + P_{\downarrow,w}\phi_{IS,w} \leq \phi_{II,w} + \frac{1}{2}\phi_{IS,w} = \frac{n-w}{n}\,. \tag{5.358}$$

Moreover, there exists a constant b such that $J_n \geq 1/b$ as long as $n \geq 2$ and $\sigma \leq 0.5$. and thus

$$G_w \leq J_n(1 + 2b\sigma\frac{n-w}{n})\,. \tag{5.359}$$

Similar to the proof of Lemma 5.12, we can split the initial weight into parts for which the noiseless copy has reached the S^n fixed point, and a part that has not.

$$\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t)}\rangle = A_{not} + \sum_{w=1}^{n} G_w A_w\,, \tag{5.360}$$

where

$$A_{not} = \sum_{\vec{\eta},\vec{\mu}\neq S^n} G_{|\vec{\eta}|}\langle \vec{\mu}, \vec{\eta}|v_{SS}^{(t-1)}\rangle \tag{5.361}$$

$$A_w = \sum_{\vec{\eta}:|\vec{\eta}|=w} \langle S^n, \vec{\eta}|v_{SS}^{(t-1)}\rangle \tag{5.362}$$

Since $G_{|\vec{\eta}|} \leq 1$ by definition, we may directly apply Lemma 5.8 and bound $A_{not} \leq \eta'_{t-1}/(q^n+1)$.

When $w < n$, we also have

$$A_w \leq \sum_{\vec{\eta}:|\vec{\eta}|=w} \langle \vec{\eta}|\Delta|v_{SS}^{(t-1)}\rangle \leq n\sigma(n-w)q^{-(n-w)}\chi_9 e^{-\chi_8(n-w)}\langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t-1)}\rangle \tag{5.363}$$

by Lemma 5.10. This requires $\sigma \leq \chi_7/n$ and $n \geq n_0$, so the upper bound inherits these requirements. Meanwhile by definition $\sum_{w=1}^{n} A_w \leq \langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t-1)}\rangle$.

Thus we have

$$\sum_{w=1}^{n} G_w A_w = G_n \sum_{w=1}^{n} A_w + \sum_{w=1}^{n}(G_w - G_n)A_w \tag{5.364}$$

$$\leq \langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t-1)}\rangle \left(G_n + \sum_{w=1}^{n-1}(G_w - G_n)n\sigma(n-w)q^{-(n-w)}\chi_9 e^{-\chi_8(n-w)} \right) \tag{5.365}$$

$$\leq \langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t-1)}\rangle J_n \left(1 + \sum_{w=1}^{n-1} 2b\sigma\frac{n-w}{n} n\sigma(n-w)q^{-(n-w)}\chi_9 e^{-\chi_8(n-w)} \right) \tag{5.366}$$

$$\leq \langle \mathbf{1}, \mathbf{1}|v_{SS}^{(t-1)}\rangle J_n(1 + f\sigma^2) \tag{5.367}$$

for some constant f, since $\sum_{a=1}^{\infty} a^2 e^{-ca}$ is bounded by a constant.

This gives us the recursion relation

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t-1)}\rangle J_n(1+f\sigma^2) + \frac{\eta'_{t-1}}{q^n+1} \tag{5.368}$$

However, for the first roughly s_{AC} gates, we will use the naive recursion relation $\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t-1)}\rangle$. We will begin to use Eq. (5.368) once η'_{t-1} is small. We have

$$\frac{\eta'_{t-1}}{q^n+1} \leq \frac{\chi_4}{q^n+1} e^{-\chi_3(t-1-s_{AC})/n} \leq \frac{\chi_4}{q^n+1} J_n e^{-\chi_3(t-1-s_{AC})/n - 2\log(1-\sigma)} \tag{5.369}$$

$$\leq \chi_4 J_n \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t-1)}\rangle e^{-\chi_3(t-1-s_{AC})/n - 2\log(1-\sigma) - 2(t-1)\log(1-\sigma)} \tag{5.370}$$

$$\leq J_n n\sigma \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t-1)}\rangle e^{-\chi'_3(t-s^*)/n}\,, \tag{5.371}$$

where in the first line we used the fact that $J_n \geq (1-\sigma)^2$, in the second line we invoked Lemma 5.5, and in the third line we have defined

$$s^* = s_{AC} + n\log(1/n\sigma)/\chi'_3 + f'' + n\log(\chi_4)/\chi'_3 \tag{5.372}$$

for an appropriate constant f'' and a weaker exponential decay coefficient χ'_3. This requires $-2\log(1-\sigma) < \chi_3/n$, which will hold as long as $\sigma \leq b_3/n$ for a properly chosen constant b_3. This gives us

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \leq \langle \mathbf{1},\mathbf{1}|v_{SS}^{(t-1)}\rangle J_n(1+f\sigma^2 + n\sigma e^{-\chi'_3(s-s^*)/n})\,. \tag{5.373}$$

Iterating this equation starting at $t = s^*$, and recalling that $\langle \mathbf{1},\mathbf{1}|v_{SS}^{(s^*)}\rangle \leq 1/(q^n+1)$,

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \leq \frac{J_n^{t-s^*}}{q^n+1} \prod_{t'=s^*+1}^{t} \left(1+f\sigma^2 + n\sigma e^{-\chi'_3(t'-s^*)/n}\right) \tag{5.374}$$

$$\leq \frac{J_n^{t-s^*}}{q^n+1} \exp\left(\sum_{t'=s^*+1}^{t} \left(f\sigma^2 + n\sigma e^{-\chi'_3(t'-s^*)/n}\right)\right) \tag{5.375}$$

$$\leq \frac{J_n^{t-s^*}}{q^n+1} e^{ft\sigma^2 + \chi''_3 n\sigma} \tag{5.376}$$

for some choice of $\chi''_3 = O(1)$ (the exponentially decaying sum is bounded). Now, we note that $J_n \geq \exp(-g\sigma')$ for a constant g slightly larger than 2 (when σ is beneath some constant), allowing us to say

$$\langle \mathbf{1},\mathbf{1}|v_{SS}^{(t)}\rangle \leq \frac{J_n^{t}}{q^n+1} e^{ft\sigma^2 + \chi''_3 n\sigma + g\sigma' s^*}\,, \tag{5.377}$$

which, recalling the definition of s^* in Eq. (5.372), implies the lemma statement for appropriate choices of b_0, b_1, and b_2. Note that the $O(n\sigma)$ term can be collected with the $O(s_{AC}\sigma)$ term since $s_{AC} \geq \Omega(n\log(n))$. □

5.C Complexity theory of the white-noise sampling problem

Recent experiments on superconducting qubit devices [6, 7] have claimed that the output distribution p_{noisy} sampled by their device would be intractable to sample on a classical computer. This claim is motivated by progress in complexity theory on showing that sampling the outputs of quantum computations is hard, but ultimately these claims rely on conjecture.

As discussed in Chapter 1, the argument that quantum computations should be hard to simulate classically begins with the observation that an efficient classical algorithm for sampling p_{ideal} exactly with probability 1 over choice of U (i.e. in the worst case) would lead to a contradiction of the widely believed conjecture that the polynomial hierarchy (PH) does not collapse [82]. The main problem with this result in practice is that noisy quantum devices cannot sample exactly from p_{ideal}. It has been conjectured that the task of *approximately* sampling p_{ideal} with high probability over circuit instance cannot be efficiently classically performed, assuming that the PH does not collapse. Here "approximate" means that the sampled distribution p_{noisy} is close to p_{ideal} in total variation distance. This is the approximate Random Circuit Sampling (RCS) problem.

In the following, when we say a task is PH-hard, we mean that there is a level of the polynomial hierarchy for which granting access to an oracle that performs the task would imply that that level contains the entire PH. Thus a polynomial time algorithm for the task would imply that the PH is contained within one of its levels and collapses.

Conjecture 5.1 (Approximate RCS is PH-hard)**.** *There exists a choice of* $\varepsilon = O(1)$ *and* $\delta \geq 1/\operatorname{poly}(n)$ *such that the task of sampling from a distribution* p_{noisy} *for which* $\frac{1}{2}\|p_{noisy} - p_{ideal}\|_1 \leq \varepsilon$ *for at least a* $1-\delta$ *fraction of random quantum circuit instances is PH-hard.*

This conjecture mirrors similar conjectures for random linear optical networks and random "instantaneous" quantum (IQP) circuits in Refs. [19, 22]. There is weak evidence for these conjectures in the form of worst-to-average case reductions for *computing* the entries of p_{ideal} with very small error tolerance [4, 19, 25–27, 137], but as noted in Chapter 1, these results are multiple steps away from proving Conjecture 5.1.

However, another issue with applying the conjecture in practice is that actual devices are unlikely to be able to sample from a distribution with such small total variation distance from ideal, as doing so requires error rates to be exceedingly small. Sampling from a distribution p_{noisy} that is close in total variation distance to p_{wn} (for some non-negligible choice of F) is potentially much more tractable in the near term; indeed, the experiments from Refs. [6, 7] claim to have performed this task (although note that their random circuits were not Haar random, but rather chosen from some other discrete random ensemble). We refer to this task as *white-noise RCS*.

Conjecture 5.2 (White-noise RCS is PH-hard)**.** *There exists a choice of* $\varepsilon = O(1)$ *and* $\delta \geq 1/\operatorname{poly}(n)$ *such that whenever the fidelity* F *satisfies* $F \geq 1/\operatorname{poly}(n)$*, the task of sampling from a distribution* p_{noisy} *for which* $\frac{1}{2}\|p_{wn} - p_{noisy}\|_1 \leq \varepsilon F$ *for at least a* $1-\delta$ *fraction of random quantum circuit instances is PH-hard.*

Note that exact worst-case white-noise sampling is PH-hard for the same reason that exact worst-case sampling is PH-hard (as long as F is at least inverse polynomial). A fine-grained version of this statement, which further claims that the exact worst-case white-noise task can be at most a factor of F easier for classical computers than the exact worst-case noiseless task, appears in the Supplementary Material of Ref. [6]. However, allowing error of size εF was not explicitly considered. Here we show that this is not an issue, and that approximate white-noise RCS and approximate RCS are essentially equivalent in this context, up to a linear factor in F, whenever the underlying random quantum circuits have the anti-concentration property.

Theorem 5.4. *Consider a random quantum circuit architecture that has the anti-concentration property. That is, there is a constant* z *such that* $\mathbb{E}_U[\sum_x p_{ideal}(x)^2] \leq zq^{-n}$*. Define an oracle* $\mathcal{O}$ *as follows. On input* (U, b)*, where* U *is a description of a* n*-qudit circuit with* $\operatorname{poly}(n)$ *gates drawn randomly from the architecture, and* b *is a string of* $\operatorname{poly}(n)$ *uniformly random bits,* $\mathcal{O}$ *produces an output* x *from a distribution* p_{noisy} *for which* $\frac{1}{2}\|p_{noisy} - p_{wn}\|_1 \leq \varepsilon F$ *holds for a certain constant* F *on at least* $1-\delta$ *fraction of random circuit instances* U*.*

Then, given access to $\mathcal{O}$ *and an NP oracle, there is an algorithm with runtime* $F^{-1} * \operatorname{poly}(n)$ *that produces samples from a distribution* p *for which* $\frac{1}{2}\|p - p_{ideal}\|_1 \leq \varepsilon'$ *on at least* $1-\delta'$ *fraction of circuit instances, with*

$$\varepsilon' = 4\varepsilon + 1/\operatorname{poly}(n) \tag{5.378}$$

$$\delta' = \delta + 1/\operatorname{poly}(n)\,. \tag{5.379}$$

Corollary 5.1. *For a random quantum circuit architecture with the anti-concentration property, Conjecture 5.1 is true if and only if Conjecture 5.2 is true.*

Proof of Corollary 5.1. It is straightforward to show that Conjecture 5.2 implies Conjecture 5.1 simply by reduction from the white-noise RCS task to the approximate RCS task: suppose one could efficiently classically produce samples from a distribution p_{noisy} for which $\frac{1}{2}\|p_{\text{noisy}} - p_{\text{ideal}}\|_1 \leq \varepsilon$. Then, for any choice of F, one can design another algorithm that samples from a distribution p'_{noisy} by producing a uniformly random output with probability $1-F$ and an output drawn from p_{noisy} with probability F. Then we have $\frac{1}{2}\|p'_{\text{noisy}} - p_{\text{wn}}\|_1 \leq \varepsilon F$. Thus whenever approximate RCS can be performed

efficiently, white-noise RCS can also be performed efficiently with the same (ε, δ) parameters, and if the latter is PH-hard then the former is also PH-hard.

The fact that Conjecture 5.1 implies Conjecture 5.2 is a direct implication of Theorem 5.4. Given a target (ε', δ') pair for which approximate RCS is hard, we can choose $\varepsilon = O(1)$ and $\delta \geq 1/\operatorname{poly}(n)$ such that if a white-noise sampler exists with those parameters, there is also an approximate sampler with parameters (ε', δ') that runs in $\operatorname{poly}(n)$ time and requires access to an NP oracle. However, since NP lies within the PH, this would still imply a collapse of the PH to one of its levels. □

The part of the proof that shows Conjecture 5.2 implies Conjecture 5.1 also illustrates why a linear factor of F is optimal. To simulate a white-noise output, one need only produce an output from p_{ideal} an F fraction of the time, so producing T samples requires only FT queries to a sampler for p_{ideal}. If sampling from p_{ideal} is a hard classical task, sampling from p_{wn} is thus at least a factor of F easier. Theorem 5.4 shows that, in a sense, it is also *at most* a factor of F easier.

This observation essentially puts the low-fidelity and high-fidelity noise regimes on the same theoretical footing when it comes to hardness of sampling, as long as the fidelity is at least inverse polynomial in n. One might object that $F \geq 1/\operatorname{poly}(n)$ is unrealistic in an asymptotic sense, and in many cases, this may be true. One way to achieve $F \geq 1/\operatorname{poly}(n)$ is to run circuits with Pauli error rate $\epsilon = \Theta(1/n)$ and circuit size $s = \Theta(n \log(n))$, which conveniently, is precisely the size required to achieve the anti-concentration property, as shown in Chapter 4. Moreover, when the fidelity is inverse exponential in n (but larger than 2^{-n}), there is still a sense in which the low-fidelity regime can be at most a factor of F easier for a classical computer than the high-noise regime.

Proof of Theorem 5.4. The idea behind our reduction is to combine approximate rejection sampling with the ability to efficiently estimate $p_{\text{noisy}}(x)$ up to $1/\text{poly}(n)$ relative error for any fixed instance U using an NP oracle (Stockmeyer's approximate counting algorithm [158]). To be precise, for any ν, any μ, and any x, there is a randomized algorithm (with access to NP oracle) that produces a number, denoted p' such that with probability at least $1 - \mu$,

$$|p_{\text{noisy}}(x) - p'| \leq 2\nu p_{\text{noisy}}(x), \tag{5.380}$$

and the algorithm runs in time $\nu^{-1} * \operatorname{poly}(n, \log(1/\mu))$. For the linear dependence on ν^{-1}, see the Supplementary Material of Ref. [6] or the lecture notes in Ref. [159]. For a fixed ν and μ, we may take $\mu' = q^{-n}\mu$ and note that $\log(1/\mu') = \operatorname{poly}(n) + \log(1/\mu)$. Now fix a set of random bits ω to feed into the randomized algorithm above. If we feed the same bits ω for every choice

of x with parameters ν and μ', then we have a fixed set of outputs $p'_{\text{noisy}}(x)$ for each possible x, and by the union bound, these values satisfy

$$|p_{\text{noisy}}(x) - p'_{\text{noisy}}(x)| \leq 2\nu p_{\text{noisy}}(x) \tag{5.381}$$

for every x simultaneously with probability at least $1 - \mu$ over the choice of ω. On any particular x, the algorithm still runs in time $\nu^{-1}\,\text{poly}(n, \log(1/\mu))$. When this is the case,

$$\frac{1}{2}\|p_{\text{noisy}}(x) - p'_{\text{noisy}}(x)\|_1 \leq \nu\,. \tag{5.382}$$

Also, let

$$\overline{p_{\text{ideal}}}(x) = \frac{p_{\text{noisy}}(x) - (1-F)q^{-n}}{F} \tag{5.383}$$

and

$$\overline{p_{\text{ideal}}}'(x) = \begin{cases} \frac{p'_{\text{noisy}}(x) - (1-F)q^{-n}}{F} & \text{if } p'_{\text{noisy}}(x) > (1-F)q^{-n} \\ 0 & \text{otherwise} \end{cases}, \tag{5.384}$$

so that, as long as the instance U is among the $1 - \delta$ fraction for which $\mathcal{O}$ succeeds, the following hold:

$$\frac{1}{2}\|\overline{p_{\text{ideal}}} - p_{\text{ideal}}\|_1 \leq \varepsilon \tag{5.385}$$

$$\frac{1}{2}\|\overline{p_{\text{ideal}}} - \overline{p_{\text{ideal}}}'\|_1 \leq \nu/F\,, \tag{5.386}$$

and by the triangle inequality

$$\frac{1}{2}\|p_{\text{ideal}} - \overline{p_{\text{ideal}}}'\|_1 \leq \nu/F + \varepsilon\,. \tag{5.387}$$

Note that in general, the function $\overline{p_{\text{ideal}}}'$ as defined does not describe a probability distribution since it is not necessarily normalized.

Now let $k > 1$ and consider the following approximate rejection sampling algorithm, similar to that in the Supplementary Information of Ref. [160].

1. Choose a set of random bits ω, which implicitly determines a function p'_{noisy}.
2. Choose an x uniformly at random, and use the estimation algorithm with bits ω to produce $p'_{\text{noisy}}(x)$.
3. Generate a random real number $0 \leq \eta \leq 1$
4. If $\overline{p_{\text{ideal}}}'(x) \leq 2kq^{-n}$ and if $\eta \leq \overline{p_{\text{ideal}}}'(x)q^n/(2k)$, output x (accept); otherwise, return to step 2 (reject).

.

Following the observations in Ref. [160], we first analyze the output distribution p_ω of our algorithm for a certain choice of ω in step 1. We see that p_ω is precisely the distribution $\overline{p_{\text{ideal}}}'$ conditioned on $x \in W$ where W is the set of x for which $\overline{p_{\text{ideal}}}'(x) \leq 2kq^{-n}$. Define

$$\mathcal{M} = \sum_x \overline{p_{\text{ideal}}}'(x) \tag{5.388}$$

$$\mathcal{N} = \sum_{x \in W} \overline{p_{\text{ideal}}}'(x)\,. \tag{5.389}$$

Then,

$$p_\omega(x) = \begin{cases} \mathcal{N}^{-1}\overline{p_{\text{ideal}}}'(x) & \text{if } x \in W \\ 0 & \text{otherwise} \end{cases}. \tag{5.390}$$

Hence,

$$\frac{1}{2}\|p_\omega - \overline{p_{\text{ideal}}}'(x)\|_1 = \frac{1}{2}\sum_{x \in W} |\mathcal{N}^{-1}\overline{p_{\text{ideal}}}'(x) - \overline{p_{\text{ideal}}}'(x)| + \frac{1}{2}\sum_{x \notin W} \overline{p_{\text{ideal}}}'(x) \tag{5.391}$$

$$= \frac{1}{2}|1 - \mathcal{N}| + \frac{1}{2}(\mathcal{M} - \mathcal{N}) \tag{5.392}$$

$$\leq \frac{1}{2}|1 - \mathcal{M}| + (\mathcal{M} - \mathcal{N})\,. \tag{5.393}$$

Note that $|1 - \mathcal{M}| \leq 2\nu/F$ by Eq. (5.386) and the fact that the values of $\overline{p_{\text{ideal}}}$ sum to 1 (although note some can in principle be negative). To handle the quantity $\mathcal{M} - \mathcal{N} = \sum_{x \notin W} \overline{p_{\text{ideal}}}'(x)$, we invoke Lemma 5.14, with $p_1 = \overline{p_{\text{ideal}}}'$, $p_2 = p_{\text{ideal}}$ and $T = 2kq^{-n}$. It shows that

$$\mathcal{M} - \mathcal{N} \leq 4\varepsilon + 4\nu/F + \sum_{x : p_{\text{ideal}}(x) > kq^{-n}} p_{\text{ideal}}(x)\,, \tag{5.394}$$

and thus

$$\frac{1}{2}\|p_\omega - \overline{p_{\text{ideal}}}'(x)\|_1 \leq 5\nu/F + 4\varepsilon + \sum_{x : p_{\text{ideal}}(x) > kq^{-n}} p_{\text{ideal}}(x)\,. \tag{5.395}$$

This is progress because the right-hand side only has dependence on the ideal distribution p_{ideal}, and not the approximate distribution output by the estimator.

Now, recall that we assume that $\mathbb{E}_U[\sum_x p_{\text{ideal}}(x)^2] \leq zq^{-n}$. By Markov's inequality, for any z', $\sum_x p_{\text{ideal}}(x)^2 \leq z'q^{-n}$ for at least $1 - z/z'$ fraction of

instances U. Suppose we have such an instance. Then

$$\sum_{x:p_{\text{ideal}}(x)>kq^{-n}} p_{\text{ideal}}(x) = \sum_{x:p_{\text{ideal}}(x)>kq^{-n}} \frac{p_{\text{ideal}}(x)^2}{p_{\text{ideal}}(x)} \tag{5.396}$$

$$\leq \sum_{x:p_{\text{ideal}}(x)>kq^{-n}} \frac{p_{\text{ideal}}(x)^2}{kq^{-n}} \tag{5.397}$$

$$\leq z'/k\,. \tag{5.398}$$

We conclude that the algorithm produces outputs from a distribution p_ω for which

$$\frac{1}{2}\|p_\omega - p_{\text{ideal}}\|_1 \leq 5\nu/F + 4\varepsilon + z'/k \tag{5.399}$$

(with probability at least $1-\mu$ over its internal randomness) and succeeds on at least $1-\delta'$ fraction of circuit instances, where

$$\delta' = \delta + z/z'\,. \tag{5.400}$$

The δ' fraction of failed instances arise either because the underlying white-noise sampler also fails on those instances or because the output distribution is not sufficiently anti-concentrated. Either way, whether an instance is among this δ' fraction is independent of the choice of ω. Thus, we may note that in the μ chance that the total variation distance bound is not satisfied for the random choice of ω, it will be at most its maximal value of 1, and thus, for any of the $1-\delta'$ successful instances, the overall total variation distance of the sampler is at most ε', where

$$\varepsilon' = 5\nu/F + 4\varepsilon + z'/k + \mu\,. \tag{5.401}$$

Now, we analyze the algorithm's runtime. Each random choice of x and subsequent calculation of $\overline{p_{\text{ideal}}}'(x)$ takes at most $\nu^{-1}\,\text{poly}(n, \log(1/\mu))$ time, but sometimes this step must be repeated. Each time the algorithm returns to step 2, it will end up accepting on step 4 with probability $\mathcal{N}/2k$. By the above analysis,

$$|\mathcal{N}-1| \leq |\mathcal{M}-1| + (\mathcal{M}-\mathcal{N}) \leq 4\varepsilon + 6\nu/F + z'/k\,. \tag{5.402}$$

Thus, as long $4\varepsilon + 6\nu/F + z'/k \leq 1/2$, then the acceptance probability will be at least $1/4k$, and the expected number of repetitions required to produce an output is at most $4k$.

Recall that $z = O(1)$. Then we may choose $z' = \text{poly}(n)$ sufficiently large, $k = \text{poly}(n)$ even larger, $\nu^{-1} = F^{-1} * \text{poly}(n)$ sufficiently large, and $\mu^{-1} = \text{poly}(n)$ sufficiently large that the algorithm runs in expected[6] time

[6]To make the runtime bounded, we could impose a cap on the number of times the algorithm returns to step 2 of $4k \cdot \text{polylog}(n)$ which, if hit, results in a uniformly random output. This would increase the total variation distance ε' by only $1/\text{poly}(n)$ and can thus be ignored.

$F^{-1}\,\mathrm{poly}(n)$ and solves the approximate RCS task with parameters $\varepsilon' = 4\varepsilon + 1/\,\mathrm{poly}(n)$ and $\delta' = \delta + 1/\,\mathrm{poly}(n)$. It is likely the factor of 4 could be optimized. □

Lemma 5.14. *Suppose p_1 and p_2 are two real functions on $[q]^n$ for which*

$$\frac{1}{2}\|p_1 - p_2\|_1 \leq \varepsilon\,. \tag{5.403}$$

Let $\mathbf{1}(\cdot)$ be the indicator function. Then for any threshold $T > 0$, we have

$$\sum_x p_1(x)\mathbf{1}(p_1(x) > T) \leq 4\varepsilon + \sum_x p_2(x)\mathbf{1}(p_2(x) > T/2)\,. \tag{5.404}$$

Proof. Let A_1 be the subset of $[q]^n$ for which $p_1(x) > T$, A_2 be the subset for which $p_2(x) > T$, and A_3 be the subset for which $p_2(x) > T/2$. For a subset X let $\overline{X}$ denote its complement.

$$\begin{aligned}
\sum_x p_1(x)\mathbf{1}(p_1(x) > T) &= \sum_{x\in A_1} p_1(x) && (5.405)\\
&= \sum_{x\in A_1} (p_1(x) - p_2(x)) + \sum_{x\in A_1} p_2(x) && (5.406)\\
&\leq 2\varepsilon + \sum_{x\in A_1} p_2(x) && (5.407)\\
&= 2\varepsilon + \sum_{x\in A_1\cap\overline{A}_3} p_2(x) + \sum_{x\in A_1\cap A_3} p_2(x) && (5.408)\\
&\leq 2\varepsilon + \sum_{x\in A_1\cap\overline{A}_3} p_2(x) + \sum_{x\in A_3} p_2(x) && (5.409)\\
&\leq 2\varepsilon + (T/2)|A_1\cap\overline{A}_3| + \sum_{x\in A_3} p_2(x) && (5.410)\\
&\leq 2\varepsilon + (T/2)\frac{2\varepsilon}{T/2} + \sum_{x\in A_3} p_2(x) && (5.411)\\
&= 4\varepsilon + \sum_x p_2(x)\mathbf{1}(p_2(x) > T/2)\,, && (5.412)
\end{aligned}$$

where the second-to-last line follows because any element of $A_1 \cap \overline{A}_3$ must contribute at least $T/2$ toward the 2ε total allowed deviation between the two functions. □

Bibliography

[1] "National quantum initiative supplement to the president's FY 2021 budget," tech. rep., Subcommittee on quantum information science, Committee on science, National Science & Technology Council, 2021. https://www.quantum.gov/wp-content/uploads/2021/01/NQI-Annual-Report-FY2021.pdf.

[2] "Press release: PsiQuantum closes $450 million funding round to build the world's first commercially viable quantum computer," 2021. https://psiquantum.com/news/psiquantum-closes-450-million-funding-round-to-build-the-worlds-first-commercially-viable-quantum-computer.

[3] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum* **2** (2018) 79, arXiv:1801.00862.

[4] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, "On the complexity and verification of quantum random circuit sampling," *Nature Physics* **15** (2019) 159, arXiv:1803.04402.

[5] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, "Characterizing quantum supremacy in near-term devices," *Nature Physics* **14** (2018) 595, arXiv:1608.00263.

[6] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature* **574** (2019) 505–510.

[7] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, *et al.*, "Strong quantum computational advantage using a superconducting quantum processor," arXiv:2106.14734.

[8] R. P. Feynman, "Simulating physics with computers," *International Journal of Theoretical Physics* **21** (1982) 467–488.

[9] G. Vidal, "Efficient classical simulation of slightly entangled quantum computations," *Physical Review Letters* **91** (2003) 147902, arXiv:quant-ph/0301063.

[10] G. Vidal, "Efficient simulation of one-dimensional quantum many-body systems," *Physical Review Letters* **93** (2004) 040502, arXiv:quant-ph/0310089.

[11] T. J. Osborne, "Efficient approximation of the dynamics of one-dimensional quantum spin systems," *Physical Review Letters* **97** (2006) 157202.

[12] J. C. Bridgeman and C. T. Chubb, "Hand-waving and interpretive dance: An introductory course on tensor networks," *Journal of Physics A: Mathematical and Theoretical* **50** (2017) 223001, arXiv:1603.03039.

[13] D. Gottesman, "The Heisenberg representation of quantum computers," *Proc. XXII International Colloquium on Group Theoretical Methods in Physics* (1998) 32–43, arXiv:quant-ph/9807006.

[14] S. Aaronson and D. Gottesman, "Improved simulation of stabilizer circuits," *Physical Review A* **70** (2004) 052328, arXiv:quant-ph/0406196.

[15] D. Deutsch and R. Jozsa, "Rapid solution of problems by quantum computation," *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* **439** (1992) 553–558.

[16] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, "Quantum algorithms revisited," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **454** (1998) 339–354, arXiv:quant-ph/9708016.

[17] S. Fenner, F. Green, S. Homer, and R. Pruim, "Determining acceptance possibility for a quantum computation is hard for the polynomial hierarchy," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **455** (1999) 3953–3966, arXiv:quant-ph/9812056.

[18] S. Toda and M. Ogiwara, "Counting classes are at least as hard as the polynomial-time hierarchy," *SIAM Journal on Computing* **21** (1992) 316–328.

[19] S. Aaronson and A. Arkhipov, "The computational complexity of linear optics," *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (2011) 333–342, arXiv:1011.3245.

[20] K. Fujii, H. Kobayashi, T. Morimae, H. Nishimura, S. Tamate, and S. Tani, "Power of quantum computation with few clean qubits," *43rd International Colloquium on Automata, Languages, and Programming* **55** (2016) 13:1–13:14, arXiv:1509.07276.

[21] K. Fujii, H. Kobayashi, T. Morimae, H. Nishimura, S. Tamate, and S. Tani, "Impossibility of classically simulating one-clean-qubit model with multiplicative error," *Physical Review Letters* **120** (2018) 200502, arXiv:1409.6777.

[22] M. J. Bremner, A. Montanaro, and D. J. Shepherd, "Average-case complexity versus approximate simulation of commuting quantum computations," *Physical Review Letters* **117** (2016) 080501, arXiv:1504.07999.

[23] S. Aaronson and L. Chen, "Complexity-theoretic foundations of quantum supremacy experiments," *Proceedings of the 32nd Computational Complexity Conference* (2017) 22:1–22:67, arXiv:1612.05903.

[24] D. Aharonov and M. Ben-Or, "Fault-tolerant quantum computation with constant error rate," *SIAM Journal on Computing* (2008) 1207–1282, arXiv:quant-ph/9906129.

[25] R. Movassagh, "Quantum supremacy and random circuits," arXiv:1909.06210.

[26] A. Bouland, B. Fefferman, Z. Landau, and Y. Liu, "Noise and the frontier of quantum supremacy," arXiv:2102.01738.

[27] Y. Kondo, R. Mori, and R. Movassagh, "Improved robustness of quantum supremacy for random circuit sampling," arXiv:2102.01960.

[28] A. W. Harrow and A. Montanaro, "Quantum computational supremacy," *Nature* **549** (2017) 203, arXiv:1809.07442.

[29] B. Barak, C.-N. Chou, and X. Gao, "Spoofing linear cross-entropy benchmarking in shallow quantum circuits," arXiv:2005.02421.

[30] M. J. Bremner, A. Montanaro, and D. J. Shepherd, "Achieving quantum supremacy with sparse and noisy commuting quantum computations," *Quantum* **1** (2017) 8, arXiv:1610.01808.

[31] X. Gao and L. Duan, "Efficient classical simulation of noisy quantum computation," arXiv:1810.03176.

[32] R. Durrett, *Probability: Theory and examples*. Cambridge University Press, 2019.

[33] B. Collins, "Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability," *International Mathematics Research Notices* (2003) 953–982, arXiv:math-ph/0205010.

[34] B. Collins and P. Śniady, "Integration with respect to the Haar measure on unitary, orthogonal and symplectic group," *Communications in Mathematical Physics* **264** (2006) 773–795, arXiv:math-ph/0402073.

[35] Y. Gu, *Moments of random matrices and Weingarten functions*. PhD thesis, Queen's University, 2013.

[36] T. Zhou and A. Nahum, "Emergent statistical mechanics of entanglement in random unitary circuits," *Physical Review B* **99** (2019) 174205, arXiv:1804.09737.

[37] N. Hunter-Jones, "Unitary designs from statistical mechanics in random quantum circuits," arXiv:1905.12053.

[38] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter, and Z. Yang, "Holographic duality from random tensor networks," *Journal of High Energy Physics* (2016) 9, arXiv:1601.01694.

[39] R. Vasseur, A. C. Potter, Y.-Z. You, and A. W. W. Ludwig, "Entanglement transitions from holographic random tensor networks," *Physical Review B* **100** (2019) 134203, arXiv:1807.07082.

[40] J. Lopez-Piqueres, B. Ware, and R. Vasseur, "Mean-field entanglement transitions in random tree tensor networks," *Physical Review B* **102** (2020) 064202, arXiv:2003.01138.

[41] A. Nahum, S. Vijay, and J. Haah, "Operator spreading in random unitary circuits," *Physical Review X* **8** (2018) 021014, arXiv:1705.08975.

[42] C. W. von Keyserlingk, T. Rakovszky, F. Pollmann, and S. L. Sondhi, "Operator hydrodynamics, OTOCs, and entanglement growth in systems without conservation laws," *Physical Review X* **8** (2018) 021013, arXiv:1705.08910.

[43] B. Bertini and L. Piroli, "Scrambling in random unitary circuits: Exact results," *Physical Review B* **102** (2020) 064305, arXiv:2004.13697.

[44] C.-M. Jian, Y.-Z. You, R. Vasseur, and A. W. W. Ludwig, "Measurement-induced criticality in random quantum circuits," *Physical Review B* **101** (2020) 104302, arXiv:1908.08051.

[45] Y. Bao, S. Choi, and E. Altman, "Theory of the phase transition in random unitary circuits with measurements," *Physical Review B* **101** (2020) 104301, arXiv:1908.04305.

[46] Y. Li, X. Chen, and M. P. A. Fisher, "Measurement-driven entanglement transition in hybrid quantum circuits," *Physical Review B* **100** (2019) 134306, arXiv:1901.08092.

[47] J. Napp, R. L. La Placa, A. M. Dalzell, F. G. S. L. Brandao, and A. W. Harrow, "Efficient classical simulation of random shallow 2D quantum circuits," arXiv:2001.00021.

[48] Y. Li and M. P. A. Fisher, "Statistical mechanics of quantum error correcting codes," *Physical Review B* **103** (2021) 104306, arXiv:2007.03822.

[49] M. J. Gullans, S. Krastanov, D. A. Huse, L. Jiang, and S. T. Flammia, "Quantum coding with low-depth random circuits," *Physical Review X* **11** (2021) 031066, arXiv:2010.09775.

[50] Y. Liu, M. Otten, R. Bassirianjahromi, L. Jiang, and B. Fefferman, "Benchmarking near-term quantum computers via random circuit sampling," arXiv:2105.05232.

[51] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, *et al.*, "Quantum computational advantage using photons," *Science* **370** (2020) 1460–1463.

[52] I. L. Markov and Y. Shi, "Simulating quantum computation by contracting tensor networks," *SIAM Journal on Computing* **38** (2008) 963–981, arXiv:quant-ph/0511069.

[53] B. M. Terhal and D. P. DiVincenzo, "Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games," *Quantum Information & Computation* **4** (2004) 134–145, arXiv:quant-ph/0205133.

[54] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, "Validating quantum computers using randomized model circuits," *Physical Review A* **100** (2019) 032328, arXiv:1811.12926.

[55] D. J. Rosenbaum, "Optimal quantum circuits for nearest-neighbor architectures," *8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)* **22** (2013) 294–307, arXiv:1205.0036.

[56] C. Neill, P. Roushan, K. Kechedzhi, S. Boixo, S. V. Isakov, V. Smelyanskiy, A. Megrant, B. Chiaro, A. Dunsworth, K. Arya, *et al.*, "A blueprint for demonstrating quantum supremacy with superconducting qubits," *Science* **360** (2018) 195–199, arXiv:1709.06678.

[57] Y. Li, X. Chen, and M. P. A. Fisher, "Quantum Zeno effect and the many-body entanglement transition," *Physical Review B* **98** (2018) 205136, arXiv:1808.06134.

[58] A. Chan, R. M. Nandkishore, M. Pretko, and G. Smith, "Unitary-projective entanglement dynamics," *Physical Review B* **99** (2019) 224307, arXiv:1808.05949.

[59] B. Skinner, J. Ruhman, and A. Nahum, "Measurement-induced phase transitions in the dynamics of entanglement," *Physical Review X* **9** (2019) 031009, arXiv:1808.05953.

[60] M. Szyniszewski, A. Romito, and H. Schomerus, "Entanglement transition from variable-strength weak measurements," *Physical Review B* **100** (2019) 064204, arXiv:1903.05452.

[61] S. Choi, Y. Bao, X.-L. Qi, and E. Altman, "Quantum error correction in scrambling dynamics and measurement-induced phase transition," *Physical Review Letters* **125** (2020) 030505, arXiv:1903.05124.

[62] M. J. Gullans and D. A. Huse, "Dynamical purification phase transition induced by quantum measurements," *Physical Review X* **10** (2020) 041020, arXiv:1905.05195.

[63] M. J. Gullans and D. A. Huse, "Scalable probes of measurement-induced criticality," *Physical Review Letters* **125** (2020) 070606, arXiv:1910.00020.

[64] A. Zabalo, M. J. Gullans, J. H. Wilson, S. Gopalakrishnan, D. A. Huse, and J. H. Pixley, "Critical properties of the measurement-induced transition in random quantum circuits," *Physical Review B* **101** (2020) 060301, arXiv:1911.00008.

[65] Q. Tang and W. Zhu, "Measurement-induced phase transition: A case study in the nonintegrable model by density-matrix renormalization group calculations," *Physical Review Research* **2** (2020) 013022, arXiv:1908.11253.

[66] A. Nahum and B. Skinner, "Entanglement and dynamics of diffusion-annihilation processes with Majorana defects," *Physical Review Research* **2** (2020) 023288, arXiv:1911.11169.

[67] K. Agarwal and N. Bao, "Toy model for decoherence in the black hole information problem," *Physical Review D* **102** (2020) 086017, arXiv:1912.09491.

[68] R. Fan, S. Vijay, A. Vishwanath, and Y.-Z. You, "Self-organized error correction in random unitary circuits with measurement," *Physical Review B* **103** (2021) 174309, arXiv:2002.12385.

[69] Y. Li, X. Chen, A. W. W. Ludwig, and M. P. A. Fisher, "Conformal invariance and quantum nonlocality in critical hybrid circuits," *Physical Review B* **104** (2021) 104305, arXiv:2003.12721.

[70] A. Lavasani, Y. Alavirad, and M. Barkeshli, "Measurement-induced topological entanglement transitions in symmetric random quantum circuits," *Nature Physics* **17** (2021) 342–347, arXiv:2004.07243.

[71] S. Sang and T. H. Hsieh, "Measurement-protected quantum phases," *Physical Review Research* **3** (2021) 023200, arXiv:2004.09509.

[72] M. Ippoliti, M. J. Gullans, S. Gopalakrishnan, D. A. Huse, and V. Khemani, "Entanglement phase transitions in measurement-only dynamics," *Physical Review X* **11** (2021) 011030, arXiv:2004.09560.

[73] Y. Fuji and Y. Ashida, "Measurement-induced quantum criticality under continuous monitoring," *Physical Review B* **102** (2020) 054302, arXiv:2004.11957.

[74] M. Szyniszewski, A. Romito, and H. Schomerus, "Universality of entanglement transitions from stroboscopic to continuous measurements," *Physical Review Letters* **125** (2020) 210602, arXiv:2005.01863.

[75] S. Vijay, "Measurement-driven phase transition within a volume-law entangled phase," arXiv:2005.03052.

[76] O. Lunt and A. Pal, "Measurement-induced entanglement transitions in many-body localized systems," *Physical Review Research* **2** (2020) 043072, arXiv:2005.13603.

[77] X. Turkeshi, R. Fazio, and M. Dalmonte, "Measurement-induced criticality in $(2+1)$-dimensional hybrid quantum circuits," *Physical Review B* **102** (2020) 014315, arXiv:2007.02970.

[78] L. Fidkowski, J. Haah, and M. B. Hastings, "How dynamical quantum memories forget," *Quantum* **5** (2021) 382, arXiv:2008.10611.

[79] A. Nahum, S. Roy, B. Skinner, and J. Ruhman, "Measurement and entanglement phase transitions in all-to-all quantum circuits, on quantum trees, and in Landau-Ginsburg theory," *PRX Quantum* **2** (2021) 010352, arXiv:2009.11311.

[80] M. Ippoliti and V. Khemani, "Postselection-free entanglement dynamics via spacetime duality," *Physical Review Letters* **126** (2021) 060501, arXiv:2010.15840.

[81] S. Aaronson, "Quantum computing, postselection, and probabilistic polynomial-time," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **461** (2005) 3473–3482, arXiv:quant-ph/0412187.

[82] M. J. Bremner, R. Jozsa, and D. J. Shepherd, "Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **467** (2010) 459–472, arXiv:1005.1407.

[83] D. N. Page, "Average entropy of a subsystem," *Physical Review Letters* **71** (1993) 1291–1294, arXiv:gr-qc/9305007.

[84] P. Hayden, D. W. Leung, and A. Winter, "Aspects of generic entanglement," *Communications in Mathematical Physics* **265** (2006) 95–117, arXiv:quant-ph/0407049.

[85] O. C. Dahlsten, R. Oliveira, and M. B. Plenio, "The emergence of typical entanglement in two-party random processes," *Journal of Physics A: Mathematical and Theoretical* **40** (2007) 8081, arXiv:quant-ph/0701125.

[86] A. W. Harrow and S. Mehraban, "Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates," arXiv:1809.06957.

[87] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, "Quantum entanglement growth under random unitary dynamics," *Physical Review X* **7** (2017) 031016, arXiv:1608.06950.

[88] M. B. Hastings, "The asymptotics of quantum max-flow min-cut," *Communications in Mathematical Physics* **351** (2017) 387–418, arXiv:1603.03717.

[89] M.-H. Yung and X. Gao, "Can chaotic quantum circuits maintain quantum supremacy under noise?" arXiv:1706.08913.

[90] G. Kalai and G. Kindler, "Gaussian noise sensitivity and bosonsampling," arXiv:1409.3093.

[91] M. Oszmaniec and D. J. Brod, "Classical simulation of photonic linear optics with lost particles," *New Journal of Physics* **20** (2018) 092002, arXiv:1801.06166.

[92] L. Eldar and S. Mehraban, "Approximating the permanent of a random matrix with vanishing mean," *2018 IEEE 59th Annual Symposium on Foundations of Computer Science* (2018) 23–34, arXiv:1711.09457.

[93] F. Pan, P. Zhou, S. Li, and P. Zhang, "Contracting arbitrary tensor networks: General approximate algorithm and applications in graphical models and quantum circuit simulations," *Physical Review Letters* **125** (2020) 060503, arXiv:1912.03014.

[94] S. Bravyi, D. Gosset, and R. Movassagh, "Classical algorithms for quantum mean values," *Nature Physics* **17** (2021) 337–341, arXiv:1909.11485.

[95] D. Shepherd and M. J. Bremner, "Temporally unstructured quantum computation," *Proceedings of the Royal Society A* **465** (2009) 1413–1439, arXiv:0809.0847.

[96] P. W. Shor, "Fault-tolerant quantum computation," *Proceedings of 37th Conference on Foundations of Computer Science* (1996) 56–65, arXiv:quant-ph/9605011.

[97] D. Aharonov and M. Ben-Or, "Polynomial simulations of decohered quantum computers," *Proceedings of 37th Conference on Foundations of Computer Science* (1996) 46–55, arXiv:quant-ph/9611029.

[98] A. W. Harrow and M. A. Nielsen, "Robustness of quantum gates in the presence of noise," *Physical Review A* **68** (2003) 012308, arXiv:quant-ph/0301108.

[99] A. A. Razborov, "An upper bound on the threshold quantum decoherence rate," *Quantum Information & Computation* **4** (2004) 222–228, arXiv:quant-ph/0310136.

[100] S. Virmani, S. F. Huelga, and M. B. Plenio, "Classical simulability, entanglement breaking, and quantum computation thresholds," *Physical Review A* **71** (2005) 042328, arXiv:quant-ph/0408076.

[101] H. Buhrman, R. Cleve, M. Laurent, N. Linden, A. Schrijver, and F. Unger, "New limits on fault-tolerant quantum computation," *2006 47th Annual IEEE Symposium on Foundations of Computer Science* (2006) 411–419, arXiv:quant-ph/0604141.

[102] J. Kempe, O. Regev, F. Unger, and R. De Wolf, "Upper bounds on the noise threshold for fault-tolerant quantum computing," *International Colloquium on Automata, Languages, and Programming* (2008) 845–856, arXiv:0802.1464.

[103] R. Raussendorf, S. Bravyi, and J. Harrington, "Long-range quantum entanglement in noisy cluster states," *Physical Review A* **71** (2005) 062313, arXiv:quant-ph/0407255.

[104] S. D. Barrett, S. D. Bartlett, A. C. Doherty, D. Jennings, and T. Rudolph, "Transitions in the computational power of thermal states for measurement-based quantum computation," *Physical Review A* **80** (2009) 062328, arXiv:0807.4797.

[105] D. E. Browne, M. B. Elliott, S. T. Flammia, S. T. Merkel, A. Miyake, and A. J. Short, "Phase transition of computational power in the resource states for one-way quantum computation," *New Journal of Physics* **10** (2008) 023010, arXiv:0709.1729.

[106] A. Deshpande, B. Fefferman, M. C. Tran, M. Foss-Feig, and A. V. Gorshkov, "Dynamical phase transitions in sampling complexity," *Physical Review Letters* **121** (2018) 030501, arXiv:1703.05332.

[107] G. Muraleedharan, A. Miyake, and I. H. Deutsch, "Quantum computational supremacy in the sampling of bosonic random walkers on a one-dimensional lattice," *New Journal of Physics* **21** (2019) 055003, arXiv:1805.01858.

[108] A. Broadbent, J. Fitzsimons, and E. Kashefi, "Universal blind quantum computation," *2009 50th Annual IEEE Symposium on Foundations of Computer Science* (2009) 517–526, arXiv:0807.4154.

[109] R. Raussendorf and H. J. Briegel, "A one-way quantum computer," *Physical Review Letters* **86** (2001) 5188–5191.

[110] I. H. Kim, "Holographic quantum simulation," arXiv:1702.02093.

[111] I. H. Kim, "Noise-resilient preparation of quantum many-body ground states," arXiv:1703.00032.

[112] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, "Entropy scaling and simulability by matrix product states," *Physical Review Letters* **100** (2008) 030504, arXiv:0705.0292.

[113] R. Orús, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states," *Annals of Physics* **349** (2014) 117–158, arXiv:1306.2164.

[114] M. Horodecki, J. Oppenheim, and A. Winter, "Quantum state merging and negative information," *Communications in Mathematical Physics* **269** (2007) 107–136, arXiv:quant-ph/0512247.

[115] B. Villalonga, D. Lyakh, S. Boixo, H. Neven, T. S. Humble, R. Biswas, E. G. Rieffel, A. Ho, and S. Mandrà, "Establishing the quantum supremacy frontier with a 281 pflop/s simulation," *Quantum Science and Technology* **5** (2020) 034003, arXiv:1905.00444.

[116] Y. Huang, "Dynamics of Rényi entanglement entropy in diffusive qudit systems," *IOP SciNotes* **1** (2020) 035205, arXiv:1902.00977.

[117] R. M. F. Houtappel, "Order-disorder in hexagonal lattices," *Physica* **16** (1950) 425–455.

[118] J. Stephenson, "Ising-model spin correlations on the triangular lattice. IV. anisotropic ferromagnetic and antiferromagnetic lattices," *Journal of Mathematical Physics* **11** (1970) 420–431.

[119] F. G. Brandão and M. J. Kastoryano, "Finite correlation length implies efficient preparation of quantum thermal states," *Communications in Mathematical Physics* **365** (2019) 1–16, arXiv:1609.07877.

[120] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.

[121] E. A. Rakhmanov, "Bounds for polynomials with a unit discrete norm," *Annals of Mathematics* **165** (2007) 55–88.

[122] R. Paturi, "On the degree of polynomials that approximate symmetric Boolean functions (preliminary version)," *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing* (1992) 468–474.

[123] F. Verstraete and J. I. Cirac, "Matrix product states represent ground states faithfully," *Physical Review B* **73** (2006) 094423, arXiv:cond-mat/0505140.

[124] R. Ayoub, *Introduction to the analytic theory of numbers*. American Mathematical Society, 1963.

[125] A. M. Dalzell, N. Hunter-Jones, and F. G. Brandão, "Random quantum circuits anti-concentrate in log depth," arXiv:2011.12277.

[126] P. Hayden and J. Preskill, "Black holes as mirrors: Quantum information in random subsystems," *Journal of High Energy Physics* **9** (2007) 120, arXiv:0708.4025.

[127] R. Oliveira, O. C. O. Dahlsten, and M. B. Plenio, "Generic entanglement can be generated efficiently," *Physical Review Letters* **98** (2007) 130502, arXiv:quant-ph/0605126.

[128] W. Brown and O. Fawzi, "Scrambling speed of random quantum circuits," arXiv:1210.6644.

[129] W. Brown and O. Fawzi, "Decoupling with random quantum circuits," *Comm. Math. Phys.* **340** (2015) 867, arXiv:1307.0632.

[130] W. Brown and O. Fawzi, "Short random circuits define good quantum error correcting codes," *IEEE International Symposium on Information Theory - Proceedings* (2013) 346–350, arXiv:1312.7646.

[131] A. W. Harrow and R. A. Low, "Random quantum circuits are approximate 2-designs," *Communications in Mathematical Physics* **291** (2009) 257, arXiv:0802.1919.

[132] W. G. Brown and L. Viola, "Convergence rates for arbitrary statistical moments of random quantum circuits," *Physical Review Letters* **104** (2010) 250501, arXiv:0910.0913.

[133] F. G. Brandao, A. W. Harrow, and M. Horodecki, "Local random quantum circuits are approximate polynomial-designs," *Communications in Mathematical Physics* **346** (2016) 397–434, arXiv:1208.0692.

[134] T. Morimae, "Hardness of classically sampling the one-clean-qubit model with constant total variation distance error," *Physical Review A* **96** (2017) 040302, arXiv:1704.03640.

[135] A. Bouland, J. F. Fitzsimons, and D. E. Koh, "Complexity classification of conjugated Clifford circuits," *Proceedings of the 33rd Computational Complexity Conference* (2018) 21:1–21:25, arXiv:1709.01805.

[136] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert, "Anticoncentration theorems for schemes showing a quantum speedup," *Quantum* **2** (2018) 65, arXiv:1706.03786.

[137] A. M. Dalzell, A. W. Harrow, D. E. Koh, and R. L. La Placa, "How many qubits are needed for quantum computational supremacy?" *Quantum* **4** (2020) 264, arXiv:1805.05224.

[138] T. Morimae and S. Tamaki, "Additive-error fine-grained quantum supremacy," *Quantum* **4** (2020) 329, arXiv:1912.06336.

[139] J. Haferkamp, D. Hangleiter, A. Bouland, B. Fefferman, J. Eisert, and J. Bermejo-Vega, "Closing gaps of a quantum advantage with short-time hamiltonian dynamics," *Physical Review Letters* **125** (2020) 250501, arXiv:1908.08069.

[140] E. Onorati, O. Buerschaper, M. Kliesch, W. Brown, A. H. Werner, and J. Eisert, "Mixing properties of stochastic quantum Hamiltonians," *Commun. Math. Phys.* **355** (2017) 905, arXiv:1606.01914.

[141] H. Gharibyan, M. Hanada, S. H. Shenker, and M. Tezuka, "Onset of random matrix behavior in scrambling systems," *Journal of High Energy Physics* **7** (2018) 124, arXiv:1803.08050.

[142] N. Hunter-Jones, "Operator growth in random quantum circuits with symmetry," arXiv:1812.08219.

[143] R. Movassagh, "Efficient unitary paths and quantum computational supremacy: A proof of average-case hardness of random circuit sampling," arXiv:1810.04681.

[144] R. Goodman and N. R. Wallach, *Representations and Invariants of the Classical Groups*. Cambridge University Press, 2000.

[145] M. Christandl, *The structure of bipartite quantum states-insights from group theory and cryptography*. PhD thesis, University of Cambridge, 2006. arXiv:quant-ph/0604183.

[146] F. G. S. L. Brandão and M. Horodecki, "Exponential quantum speed-ups are generic," *Quantum Information & Computation* **13** (2013) 0901–0924, arXiv:1010.3654.

[147] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, "Random quantum circuits transform local noise into global white noise," arXiv:2111.14907.

[148] Q. Zhu, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, *et al.*, "Quantum computational advantage via 60-qubit 24-cycle random circuit sampling," arXiv:2109.03494.

[149] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, "Estimating the coherence of noise," *New Journal of Physics* **17** (2015) 113020, arXiv:1503.07865.

[150] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, "Bounding the average gate fidelity of composite channels using the unitarity," *New Journal of Physics* **21** (2019) 053016, arXiv:1610.05296.

[151] Y. Rinott, T. Shoham, and G. Kalai, "Statistical aspects of the quantum supremacy demonstration," arXiv:2008.05177.

[152] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia, "Comparing experiments to the fault-tolerance threshold," *Physical Review Letters* **117** (2016) 170502, arXiv:1510.05653.

[153] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, "Limitations of noisy reversible computation," arXiv:quant-ph/9611028.

[154] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," arXiv:1411.4028.

[155] C. Xue, Z.-Y. Chen, Y.-C. Wu, and G.-P. Guo, "Effects of quantum noise on quantum approximate optimization algorithm," *Chinese Physics Letters* **38** (2021) 030302, arXiv:1909.02196.

[156] J. Marshall, F. Wudarski, S. Hadfield, and T. Hogg, "Characterizing local noise in QAOA circuits," *IOP SciNotes* **1** (2020) 025208, arXiv:2002.11682.

[157] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, "Noise-induced barren plateaus in variational quantum algorithms," arXiv:2007.14384.

[158] L. Stockmeyer, "Complexity of approximate counting," *8th Conference on the Theory of Quantum Computation, Communication and Cryptography* (1983) 118–126.

[159] L. Trevisan, "Lecture notes on computational complexity," 2002. http://theory.stanford.edu/~trevisan/notes/complexitynotes02.pdf.

[160] A. Neville, C. Sparrow, R. Clifford, E. Johnston, P. M. Birchall, A. Montanaro, and A. Laing, "Classical boson sampling algorithms with superior performance to near-term experiments," *Nature Physics* **13** (2017) 1153–1157.