Shallow Quantum Circuits: Algorithms, Complexity, and Fault Tolerance

Yunchao Liu

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2024-181
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2024/EECS-2024-181.html

August 16, 2024

Shallow Quantum Circuits: Algorithms, Complexity, and Fault Tolerance

By

Yunchao Liu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Umesh Vazirani, Chair
Professor Dorit Aharonov
Assistant Professor Avishay Tal
Assistant Professor John Wright
Professor Lin Lin

Summer 2024

Shallow Quantum Circuits: Algorithms, Complexity, and Fault Tolerance

Abstract

Shallow Quantum Circuits: Algorithms, Complexity, and Fault Tolerance

by

Yunchao Liu

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Umesh Vazirani, Chair

A fundamental goal of quantum computing is to demonstrate computational advantage over classical computers. Shallow quantum circuits play a central role in this effort as the field transits from the Noisy Intermediate Scale Quantum (NISQ) era to the Early Fault Tolerant era. This thesis describes three lines of research on the theoretical foundation of achieving quantum computational advantage using shallow quantum circuits.

- The first is complexity and applications of random circuit sampling (RCS), an experiment at the heart of recent "quantum supremacy" experiments. We describe recent progress in understanding the computational complexity of RCS and its applications in benchmarking noisy quantum devices.
- The second is learning algorithms. We give polynomial time algorithms for learning shallow quantum circuits and quantum states prepared by shallow circuits.
- The third is fault tolerance. We discuss the prospects of achieving quantum computational advantage using fault tolerance techniques within noisy shallow circuits, and describe two new techniques toward this goal, including fault tolerance against input noise and single-shot logical state preparation.

Contents

Acknowledgments

The journey of my PhD has been filled with the support, wisdom, and presence of mentors and friends, and I am grateful for their positive impact on me.

My advisor Umesh Vazirani is the person I learned most from over the last five years. Umesh has given me so much guidance and feedback on many different things, from finding good research problems to writing papers and giving talks. I can clearly feel my growth after each round of those back-and-forth conversations. Underlying all his guidance is his unique philosophy and approach to things, which I am also beginning to appreciate: the principle of focusing on the high-order bits, the deep philosophies behind simple stories, the gentle touch to push things toward the right direction, the great care about certain seemingly small things, and the mastery of cellphones... Above all, I think what Umesh did as an advisor is to help me discover myself: to realize what I really want to do in research.

Dorit Aharonov has had a great impact on this thesis through both her mentorship and her research work. This thesis builds upon the foundational work established in her PhD thesis "Noisy Quantum Computation". While reading her thesis, I realized that many of her thoughts, even though written 25 years ago, still feel deeply insightful today. The most important thing that Dorit taught me, through her deepest thoughts and strongest passion, is what the connection between physics and computation really means.

Math conversations with Zeph Landau have been an indispensable part of my PhD. Zeph has the magical power of distilling a math argument to its bare minimum and understanding its fundamental core, and every time he waves that magic, it gives me a feeling of beauty and joy. This is what led to our five papers together – and surely more will come. My conversations with Zeph also go much beyond math: they involve life, philosophy, and politics... And of course, there are cakes and waffles. Even more importantly, beyond all these things, what I learned from Zeph is an example of a truly kind person.

There are so many people who have mentored me and guided me over the years. Former postdocs Adam Bouland, Bill Fefferman, and Anurag Anshu helped me develop research skills during the early stage of my PhD. Former students Chinmay Nirkhe and Urmila Mahadev gave me invaluable advice throughout this journey. Berkeley faculty John Wright, Avishay Tal, and Lin Lin have taught me so much through their courses and by setting examples of great researchers. I spent wonderful times with Srinivasan Arunachalam and Kristan Temme at IBM, and they continue to be my mentors. I thank all my research collaborators, especially Yimu Bao, Thiago Bergamaschi, Xun Gao, and Jeongwan Haah, for being infinite sources of inspiration.

My research wouldn't have gone so smoothly without the hard work of the Simons leadership and staff. I would like to especially thank Sandy Irani for her contributions to the Simons community.

I would like to thank all members of the Berkeley theory group and Simons quantum community for bringing lots of fun into my life. Thank you to Seri Khoury for feeding me with his excellent barbeques and for educating me on the rules at UC Berkeley.

I am extremely fortunate to have enjoyed the friendships with so many great people. The Bonita community gave me unforgettable memories and fun, especially during covid. Former students of the Yao class have always been my closest friends and allies, and I especially enjoyed the presence of Binghui and Shunhua when they visited Simons. Hongxun and Xin always motivate me by reminding me how good the younger students have become. I would like to thank Fred and Lijie for all the experiences we have had together: all the food and drinks and philosophical discussions, all the thoughts and emotions, and ups and downs.

To all the mentors and friends: thank you.

Chapter 1

Quantum computational advantage: from NISQ to fault tolerance

A central goal of the theory of computation is to understand: *What are the fundamental capabilities and limitations of computers?* [1] To understand this question one needs to formulate a reasonable mathematical model of computation that can be realized in the physical world, and then study the computational power of this model. The Extended Church-Turing Thesis posits that any such realistic model of computation can be efficiently simulated on a probabilistic Turing machine. This suggests that Turing machine is the most powerful model in terms of understanding what computational problems can be efficiently solved in the physical world.

Over the last 40 years, quantum computation has emerged as a revolutionary model of computation based on quantum mechanics, not only in theory but also experimentally realized with increasing scale. It has been realized since the early days of quantum computing that this model potentially violates the Extended Church-Turing Thesis [2]: solving a computational problem that cannot be solved by classical computers in polynomial time. This was solidified by Shor's quantum algorithm for factoring [3].

To establish quantum computing as a physical reality, the next essential step is the *experimental violation* of the Extended Church-Turing Thesis, or an experimental demonstration of *quantum computational advantage*. However, demonstrating quantum computational advantage via factoring integers, for example, is still far from reality as it likely requires a large-scale fault tolerant quantum computer. On the other hand, rapid progress has been made on the hardware side, where noisy intermediate scale quantum (NISQ) devices and early fault tolerant devices of increasing scale and improving quality are being built (Fig. 1.1). These devices are already capable of implementing quantum circuits of nontrivial size.

This thesis aims to bridge the gap between theory and experiment: establishing the theoretical foundation for achieving quantum computational advantage using current or near-future quantum devices. This is part of an exciting ongoing effort in the field toward achieving deeper theoretical understanding and better experimental realization of quantum computational advantage in the next few years.

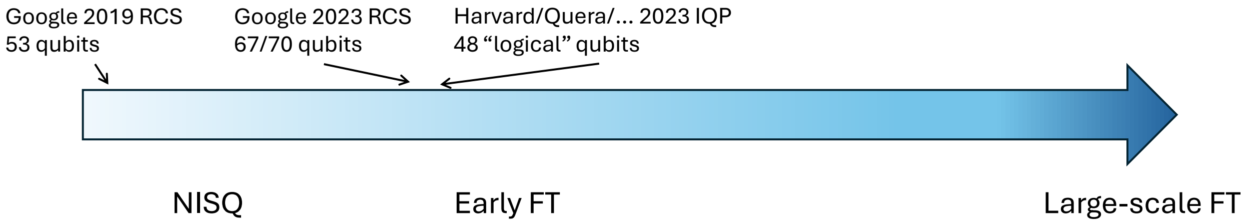


Figure 1.1: Progress on quantum computational advantage.

We will focus on three of the most important features of the current progress on quantum computational advantage (Fig. 1.1):

1. **Shallow quantum circuits**. Due to their limited coherence time, NISQ computers are naturally modeled as having small circuit depth. Despite their simplicity, shallow quantum circuits already demonstrate computational hardness: under plausible complexity assumptions, no efficient classical algorithm can approximately sample from the output distribution of a shallow quantum circuit. This is the starting point for the theoretical basis of quantum computational advantage.

2. **Transition from NISQ to fault tolerance**. Noise is a major obstacle to performing useful computation on a NISQ device which lacks error correction. A major effort today is to realize restricted forms of fault tolerance within the capability of current devices – the “early fault tolerant era”.

3. **Useful and practical quantum algorithms**. Besides the demonstration of computational hardness, an important direction is to develop useful quantum algorithms that solve problems of practical interest, which ideally can be implemented using shallow quantum circuits.

This thesis describes three lines of research on the theoretical foundation of achieving quantum computational advantage using shallow quantum circuits.

- Chapter 2 studies the complexity and applications of random circuit sampling (RCS). RCS is a basic primitive at the heart of recent “quantum supremacy” experiments, which is the leading effort toward demonstrating quantum computational advantage using NISQ devices. This chapter studies the complexity theoretic foundation of these experiments, as well as their applications in benchmarking noisy quantum devices.

- Chapter 3 studies learning algorithms. We give polynomial time algorithms for learning shallow quantum circuits and quantum states: given query access to an unknown black box shallow quantum circuit (or copies of an unknown quantum state prepared by a shallow circuit), learn a description of an equivalent shallow circuit (or description of a shallow circuit that prepares the state). An important genre of heuristic quantum algorithms for NISQ devices can be viewed as learning a good shallow quantum circuit for performing a certain task. Our learning algorithms may provide rigorous primitives for this genre of algorithms.

- Chapter 4 studies fault tolerance. We propose a direction for achieving quantum computational advantage in the early fault tolerant era – demonstrating computational hardness using fault tolerance techniques within noisy shallow circuits, and describe two new techniques toward this goal. The first is fault tolerance against input noise for shallow IQP circuits. As an application, we prove the existence of a family of local Hamiltonians for which the Gibbs state at constant temperature is easy to prepare on a quantum computer but hard to sample from classically. The second is single-shot logical state preparation. We prove that encoded logical states of arbitrary quantum LDPC codes can be prepared using a constant depth quantum circuit followed by a single round of measurement and classical feedforward.

Chapter 2

Complexity and applications of random circuit sampling

This chapter studies the computational complexity of random circuit sampling (RCS), as well as applications of RCS in benchmarking noisy quantum devices. In Section 2.1 we give an overview of the advances in understanding RCS and quantum supremacy experiments over the last five years, highlighting the exciting back-and-forth between theory and experiment. In Section 2.2 we study the complexity of ideal (noiseless) random circuit sampling and provide evidence for its classical hardness, based on joint work with Adam Bouland, Bill Fefferman and Zeph Landau [4]. In Section 2.3 we study the complexity of random circuit sampling in the regime of constant noise per gate and give a polynomial time classical algorithm for this problem, based on joint work with Dorit Aharonov, Xun Gao, Zeph Landau and Umesh Vazirani [5]. Finally, in Section 2.4 we discuss theoretical justifications of using random circuit sampling to characterize the noise models of a NISQ device, based on joint work with Matthew Otten, Roozbeh Bassirian, Liang Jiang and Bill Fefferman [6].

2.1 Overview: theoretical reflections on quantum supremacy

Motivation. This chapter studies random circuit sampling (RCS) – an experiment at the heart of the recent "quantum supremacy" experiments: demonstrating quantum computations that are hard to simulate classically. These experiments play a fundamental role in quantum computing in the NISQ era: from the experimental perspective, they achieve an important milestone toward building a practical quantum computer, and provide a benchmark for the power and fidelity of NISQ devices; from the theoretical perspective, they provide a fundamental test of quantum mechanics in the high complexity regime, and make progress toward an experimental violation of the Extended Church-Turing Thesis.

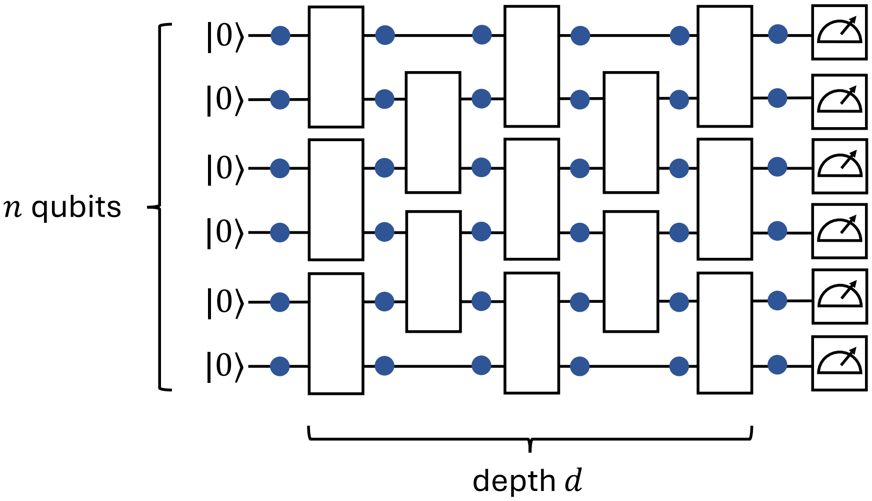


Figure 2.1: Random circuit sampling experiments. A random quantum circuit with n qubits and depth d is selected by choosing each gate (white box) randomly. Here we assume $d = \Omega(\log n)$, such that $p_C(x)$ satisfies a statistical property called *anti-concentration*[1]. The circuit is fixed and implemented on a noisy quantum device (noise is indicated by the blue dots) M times, obtaining M samples of n-bit strings.

Random circuit sampling experiments. In 2019, Google announced a quantum supremacy experiment [7] which performed random circuit sampling on a 53-qubit quantum device. A theoretical model of Google's random circuit sampling experiment is shown in Fig. 2.1. The experiment proceeds as follows:

1. Choose a random quantum circuit C with n qubits and depth d on a fixed circuit architecture (Fig. 2.1 shows a 1D geometry while the experiment uses a 2D geometry) by selecting each gate independently at random. The circuit acts on the input state $|0^n\rangle$ and is measured in the standard basis at the output. Ideally, the output samples are distributed as $p_C(x) = |\langle x|C|0^n\rangle|^2$, $x \in \{0,1\}^n$.

2. The circuit C is fixed and implemented on a noisy quantum device M times, obtaining M samples $x_1, x_2, \ldots, x_M \in \{0,1\}^n$. Due to noise in the device, the experiment actually implements a noisy circuit where each gate is subject to some amount of noise (blue dots in Fig. 2.1), and the experimental output may be far from $p_C(x)$.

3. After collecting the samples, a statistical test is used to test the consistency of the samples with $p_C(x)$. We will focus on the statistical test called linear cross entropy

[1] Formally, anti-concentration means that $\mathbb{E}_C\, 2^n \sum_{x\in\{0,1\}^n} p_C(x)^2 = O(1)$, that is, the distribution is not too concentrated.

benchmark (XEB), defined as

$$\mathrm{XEB} = \frac{2^n}{M} \sum_{i=1}^{M} p_C(x_i) - 1. \tag{2.1}$$

This can be viewed as a score between 0 and 1 (a higher score is better).

When the experiment is sampling from a completely uncorrelated distribution such as the uniform distribution, XEB $= 0$ in expectation. When the experiment is perfectly sampling from $p_C(x)$, XEB ≈ 1 in expectation (this follows from the Porter-Thomas property of $p_C(x)$). Google's 2019 experiment reported XEB $= 0.002$, which means that it only achieved a tiny correlation with $p_C(x)$. The claim of "quantum supremacy" relies on the conjecture that even this tiny correlation takes an enormous amount of computing resources classically to achieve (thousands of years on a supercomputer).

Overview. Since the 2019 experiment, much progress has been made over the last five years, including improved experiments and deeper theoretical understanding of how to interpret these experiments, as well as applications in quantum benchmarking. Here we give an overview of this progress. We refer to [8] for a more comprehensive survey of the literature.

From the theoretical perspective there are two basic questions about the "quantum supremacy" claim. The first is whether sampling from the ideal output distribution $p_C(x)$ of a random quantum circuit C is classically hard. Assuming this is true, there is the further question of whether the statistical tests used in the experiment can be "spoofed" classically, that is, whether an efficient classical algorithm can achieve the same performance as the noisy experiment, even without sampling from the ideal output distribution.

Separately, random circuit sampling and its associated statistical tests has become a standard approach to benchmark quantum devices: characterizing the fidelity and noise models in a NISQ device. This can be viewed as a practical application of quantum supremacy experiments.

Below we discuss the substantial progress that has been made on both fronts.

Complexity theoretic foundation. The key challenge to prove the classical hardness of ideal random circuit sampling is two-fold: the first is *average-case* hardness, meaning the hardness must hold with high probability over a random circuit C; the second is *robustness*, meaning it is hard to sample from any distribution that is close to $p_C(x)$. Putting together, the goal is to prove that under standard complexity assumptions (such as the non-collapsing of PH), no efficient classical algorithm can sample from $p_C(x)$ within inverse-polynomial total variation distance, with high probability over C. While this remains a conjecture, much progress has been made which provides positive theoretical evidence for this conjecture, including the results presented in Section 2.2. We refer to Section 2.2 for extended discussions on this subject.

Computational complexity of XEB and noisy RCS. The central question regarding the actual quantum supremacy experiments is the hardness of classically spoofing the XEB test.

To justify this statistical test, Aaronson and Gunn formulated a strong conjecture called XQUATH [9] which states that no efficient classical algorithm can achieve even a tiny correlation with p_C, which provided the complexity foundation of the XEB test and provided a way to heuristically argue that even the very small XEB achieved in actual experiments was a classically difficult computational task. The intuition behind this conjecture is the Feynman path integral: think of a quantum circuit as a walk on the computational basis with complex transition amplitudes. The ideal output of a quantum circuit is the exponential sum of all such trajectories (Feynman paths). Moreover, all these Feynman paths are a priori similar and contribute equally. This is so unstructured that the best thing an efficient classical algorithm can do is to calculate polynomially many such paths, which only achieves an exponentially small correlation.

Since then, there has been pushback on both the quantum supremacy claim as well as its theoretical foundation. On the practical side, there has been significant progress in tensor network based classical simulation algorithms [10, 11, 12, 13, 14, 15]. These algorithms can achieve the same XEB as Google's 2019 experiment using hundreds of GPUs in a few hours. However, these algorithms inherently run in exponential time and are not scalable. Indeed, subsequent RCS experiments by USTC [16, 17] with up to 60 qubits are estimated to take much more resources to spoof. This still left open the question of whether spoofing the XEB test is a computationally hard problem.

In 2021, Gao *et al.* [18] developed an argument which cast doubt on the XQUATH conjecture of [9] and developed a practical spoofing algorithm. Practically, the algorithm uses a very small amount of resource, although it only achieved 10% of Google's XEB and fell short of spoofing the experiment. Theoretically, the computational hardness of the XEB test was called into question, although this work left unclear whether the hardness of the XEB test could be restored by formulating a new conjecture, and whether efficient classical algorithms exist for the other statistical tests for RCS output distributions such as the Heavy Output Generation (HOG) [19] and log XEB [20]. However, it reopens the question: is there high complexity in noisy RCS experiments?

Finally, in 2022, Aharonov *et al.* [5] gave a polynomial time classical algorithm to sample from the output distribution of a random quantum circuit with constant noise per gate, within inverse-polynomial total variation distance. This result is presented in detail in Section 2.3. This result implies the following: if the experiment collects M samples, then the classical algorithm can also produce M samples that are statistically indistinguishable from the experiment, in time $\mathrm{poly}(n, M)$. This means that no statistical test can distinguish between the experiment and the classical algorithm.

This classical algorithm is based on Pauli path integral – the Feynman path integral in the Pauli basis. Here we think of a quantum circuit as evolving a density matrix under unitary channels, which is a walk on the Pauli basis (I, X, Y, Z) with real transition amplitudes. The ideal output of a quantum circuit is the exponential sum of all Pauli paths. The key difference here is that: unlike the standard Feynman paths, different Pauli paths have very different contributions since the identity element I only goes to identity under unitary conjugation. The non-uniformity is further amplified in a noisy quantum circuit: as the identity element I is preserved by noise and non-identity elements X, Y, Z get contracted by noise, the contribution of a low-weight Pauli path is much higher than that of a high-weight Pauli path. Thus the algorithm calculates all low-weight Pauli paths, and the approximation error is bounded using anti-concentration.

Although this algorithm as it currently stands is not yet practical, it shows that there is no computational hardness in random circuit sampling in the asymptotic regime of constant noise per gate, thus ruling out an experimental violation of the Extended Church-Turing Thesis. This calls into question the basic definitions of quantum supremacy and the interpretation of the experimental results; a clarification is much needed.

Google's 2023 experiment. In 2023, Google performed improved random circuit sampling experiments with up to 70 qubits [21], which clearly exceeds far beyond classical simulability with reasonable resources estimated using current best-known practical simulation algorithms, due to increased system size and improved fidelity.

Moreover, they gave a major clarification on the interpretation of the computational hardness of the experiment. Instead of the asymptotic regime with constant noise per gate, Google claimed that their finite-size experiments are in a *low-noise regime*: the noise per gate scales as c/n where c is a small constant and n is the number of qubits. Thus the experiment is interpreted in a small, finite-size regime, where noise can still be viewed as a vanishing function in n, rather than the asymptotic regime where n goes to infinity and noise is a constant.

Within this low-noise regime, they gave theoretical arguments showing that the XEB is hard to spoof, which justifies the claimed difficulty of spoofing the XEB in practice. The argument is based on a new hypothesis: that low-weight Pauli paths is the optimal spoofing algorithm. Within the low-noise regime, the total noise on each layer of gates is a small constant denoted as λ, and the XEB of the experiment scales as $\text{XEB} \approx e^{-\lambda d}$ (see below). Furthermore, they conjecture that the XEB achieved by calculating all low-weight Pauli paths is at most $e^{-\Delta d}$, where Δ is some fixed constant. Therefore, the experiment cannot be spoofed as long as $\lambda < \Delta$, which is shown to be true in the experiment where Δ is estimated numerically. These ideas emerged from separate developments in benchmarking which we discuss next.

Cross-entropy benchmarking. Besides the above developments in understanding the computational hardness of quantum supremacy experiments, there have been significant developments in using XEB to characterize noisy quantum devices, where the significance of the low-noise regime has already been recognized.

In Google's 2019 experiment [7], it was suggested that the XEB is a proxy for fidelity. Let $|\psi\rangle = C\,|0^n\rangle$ denote the ideal output state of a RCS experiment (before measurement) and let ρ denote the mixed state of the noisy experiment. The fidelity is defined as $F = \langle\psi|\rho|\psi\rangle$, which is an important benchmark of the device and is hard to directly measure. Moreover, Google observed that their experimental XEB was consistent with an uncorrelated noise model defined by multiplying individual gate fidelities, and they claimed that these experimental results could be considered a way of verifying that the noise channel acting on a layer of gates is uncorrelated across each gate. This observation is remarkable in the sense that the fidelity of highly complex random circuits could be predicted by such a simple noise model, but also intriguing as little theoretical evidence has been shown that supports this observation.

Through subsequent theoretical developments (see Section 2.4), it was clarified that the XEB decays exponentially according to $e^{-\lambda d}$, where λ is the total mount of noise in each layer of gates, regardless of whether the noise is independent or correlated. This can be understood as the following: think of the noise on each layer as an arbitrarily correlated n-qubit Pauli channel (this is without loss of generality due to the twirling effects of random circuits), then λ is the probability that any nontrivial Pauli error happens on any qubit. Then, it holds in a RCS experiment that

$$\mathrm{XEB} \approx F \approx e^{-\lambda d} \approx \Pr[\text{no error happens in the circuit}]. \tag{2.2}$$

In particular, the main result of Section 2.4 provides theoretical evidence of $F \approx e^{-\lambda d}$, and Ref. [22] provides theoretical evidence of $\mathrm{XEB} \approx e^{-\lambda d}$. These results hold in the regime when λ is at most a small constant, which coincides with the low-noise regime where noise per gate scales as c/n. In addition, there have been theoretical and experimental works that apply XEB to benchmark analog quantum simulators [23]. These developments have significant consequences in quantum benchmarking:

1. It suggests that XEB is a proxy for fidelity, which gives a sample-efficient way to estimate fidelity.

2. It suggests that λ can be estimated by measuring the XEB, addressing a significant challenge in benchmarking non-Clifford gates (see Section 2.4 for more discussions).

3. Eq. (2.2) combined with experimental data on the noise of each gate suggests that the noise on each gate in Google's experiment is independently distributed (see Section 2.4 for more discussions), which is an important prerequisite for fault tolerance.

XEB phase transition. As discussed above, the XEB has been used as a benchmark of the computational hardness of RCS experiments, as well as a proxy of the fidelity of noisy devices. Google's 2023 experiment [21] provided a more detailed understanding of the XEB by identifying a *phase transition*, concluding that the low-noise regime of c/n noise per gate is precisely the threshold below which the XEB works for benchmarking. First, the expectation value of XEB (averaging over random circuits) is mapped to the partition function of a classical spin model. Second, under certain approximations, an analytical formula of this partition function is obtained, as a function of n, d, and ε (noise per gate). Third, by examining this expression, a phase transition in the parameter $\lambda = n\varepsilon$ is evident: when λ is below a certain constant threshold, XEB $\approx e^{-\lambda d}$ (low-noise regime); when λ is above this threshold, XEB $\gg e^{-\lambda d}$ (high-noise regime). Therefore, XEB is applicable for characterizing noisy devices (items 1-3 below Eq. (2.2)) only in the low-noise regime.

In addition, this phase transition also jusifies the computational hardness of XEB in the low-noise regime discussed earlier. Intuitively, in the low-noise regime the XEB captures global correlations within the system, while in the high-noise regime XEB only captures local correlations. Recall that efficient spoofing algorithms can be viewed as calculating low-weight Pauli paths, that is, classically simulating local correlations, and the hypothesis that the XEB achieved by low-weight Pauli paths is at most $e^{-\Delta d}$ for some fixed constant Δ, where Δ can be estimated numerically. In the low-noise regime, XEB $\approx e^{-\lambda d}$, which can exceed the total amount of local correlations (when $\lambda < \Delta$) because the experiment contains global correlations. This gives theoretical evidence that XEB is hard to spoof in the low-noise regime.

Discussion. Our current understanding of RCS and quantum supremacy experiments can be summarized as follows.

1. **Large practical classical simulation cost.** The lastest RCS experiment is estimated to take at least tens or hundreds of years to classically spoof using a supercomputer.

2. **No scalable computational hardness.** Noisy RCS experiments can be efficiently simulated classically in the asymptotic regime of constant noise per gate.

3. **Applications in benchmarking.** RCS and XEB have already become a standard approach for characterizing small-scale noisy quantum devices.

4. **Plausible computational hardness in the low-noise regime.** There are plausible arguments showing that XEB is hard to spoof in the low-noise regime. It remains an interesting open question to develop complexity theoretic evidence to put this claim on a more rigorous footing.

Reflecting on these exciting developments over the last five years, the scientific efforts have been very successful: compared with five year ago, we have now a much deeper theoretical understanding about the complexity of RCS experiments and NISQ computation in general, and have experimentally built much better devices. One of the reasons for this success is that there was a clear target in 2019: achieve quantum supremacy via random circuit sampling. It seemed clear at the time that this was the right target to achieve, and this provided strong motivations for both theory and experiment to focus on this target.

Looking forward, the field is now moving into the early fault tolerant era. An urgent question now is to formulate the next motivating target that drives our field. We refer to Chapter 4 for this discussion.

2.2 Complexity of ideal random circuit sampling

This section provides complexity theoretic evidence for the following conjecture on the classical hardness of ideal random circuit sampling.

Conjecture 2.1 (Hardness of approximate sampling)**.** *No polynomial-time randomized classical algorithm can do the following (unless the Polynomial Hierarchy collapses): on input a random quantum circuit C, obtain a sample from a distribution that is within $1/\text{poly}(n)$ total variation distance from the output distribution of C, with probability at least $1 - 1/\text{poly}(n)$ over the choice of C as well as the internal randomness of the algorithm.*

This conjecture provides the theoretical basis for quantum supremacy experiments (in the limit of very small noise), and from the complexity perspective it is interesting as it does not explicitly depend on $\mathsf{BQP} \neq \mathsf{BPP}$.

A line of research started from Aaronson and Arkhipov [24] in the context of Boson sampling developed a road map toward this conjecture. The crux is to show that the output probability of random quantum circuits are $\#\mathsf{P}$-hard to approximate on average. Here "approximate" corresponds to a $\frac{2^{-n}}{\text{poly}(n)}$ additive imprecision, which follows naturally from Stockmeyer's approximate counting theorem [25].

Conjecture 2.2. *It is $\#\mathsf{P}$-hard to compute $\mathsf{p}_{\mathbf{0}}(C) := |\langle 0^n|C|0^n\rangle|^2$ up to additive imprecision $2^{-n}/\text{poly}(n)$ on input a random circuit C, with probability at least $1 - \frac{1}{\text{poly}(n)}$ over the choice of C as well as the internal randomness of the algorithm.*

It is shown via Stockmeyer's approximate counting and anti-concentration that Conjecture 2.2 implies Conjecture 2.1 [24, 26]. Although Conjecture 2.2 has not been proven so far, necessary conditions of Conjecture 2.2 were established in the form of average-case hardness of $\mathsf{p}_{\mathbf{0}}(C)$ that is robust up to a smaller additive imprecision. An open direction is therefore to improve these results to match the $\frac{2^{-n}}{\text{poly}(n)}$ additive imprecision required by Conjecture 2.2. The main result of this section gives improved average case hardness results toward Conjecture 2.2, for both random circuit sampling and Boson sampling.

Theorem 2.1. *For any constant $\eta < 1/4$, it is $\#\mathsf{P}$-hard under a $\mathsf{BPP}^{\mathsf{NP}}$ reduction to compute the output probability of random quantum circuits C up to additive error $2^{-O(m \log m)}$, with probability at least $1 - \eta$ over the choice of C with m gates.*

We also give an analogous result for Boson sampling, showing that it is hard to compute the output probabilities of (noiseless) random n-photon $m = n^2$ mode linear optical networks to additive imprecision $e^{-6n \log n - O(n)}$. This nearly matches the desired robustness for Boson sampling ($O(e^{-n \log n})$) up to a constant factor in the exponent. To our knowledge, this is the first time that a quantum supremacy proposal exhibits a proven robustness of hardness which is polynomially related to the conjectured robustness in absolute terms.

Corollary 2.1. *For any constant $\eta < \frac{1}{4}$, it is #P-hard under a $\mathsf{BPP}^{\mathsf{NP}}$ reduction to compute the output probability of a n-photon $m = n^c$-mode Boson sampling experiment up to additive error $\delta = e^{-(c+4)n\log n - O(n)}$, with success probability $1 - \eta$.*

2.2.1 Average-case hardness

In this section we develop the average-case hardness of computing the output probability of random quantum circuits. An important ingredient of our proof is a robust polynomial interpolation technique, which is independently developed in Section 2.2.2.

2.2.1.1 Worst-to-average-case reduction

Our first result is to establish a worst-to-average-case reduction for computing the output probability of random quantum circuits in the noiseless setting. That is, if there exists an algorithm that can approximate the output probability of *most* random quantum circuits over an architecture, then there exists an algorithm that can approximate the output probability of *every* quantum circuit over the same architecture. Our average-case hardness result tolerates a constant failure probability (i.e. it is hard to compute the output probability for a large constant fraction of random circuits), and is robust up to a small additive error $2^{-O(m\log m)}$ for random circuits with m gates. Our result is therefore an improvement over previous results [26, 27] in both aspects.

The main idea of the worst-to-average-case reduction works as follows. Consider a circuit architecture over which computing the output probability up to an exponentially small additive error is #P-hard in the worst case. Then, to prove average-case hardness, it suffices to construct a reduction that belongs to some finite level of the Polynomial Hierarchy. That is, given an arbitrary circuit C_0, our goal is to compute $\mathsf{p_0}(C_0)$ using a procedure in the Polynomial Hierarchy, with access to an oracle $\mathcal{O}$ that can compute $\mathsf{p_0}(C)$ for a large fraction of random circuits C. For a circuit architecture $\mathcal{A}$, let $\mathcal{H}_{\mathcal{A}}$ be the distribution over circuits such that each gate is independently drawn from the Haar measure. Let $\{G_i\}_{i=1,\ldots,m}$ be the quantum gates in C_0. We create a new circuit C_1 by applying a "one-time pad" to C_0, where each gate G_i is replaced by

$$G_i \to H_i G_i, \tag{2.3}$$

where $\{H_i\}$ is independently drawn from the Haar measure over the unitary group and has the same dimension as G_i. By the invariance of Haar measure, C_1 is distributed the same as $\mathcal{H}_{\mathcal{A}}$. Therefore $\{H_i\}$ can be understood as the "random seed" used for the one-time pad.

By definition, $\mathcal{O}$ can compute an accurate approximation of $\mathsf{p_0}(C_1)$ with high probability. However, this number alone does not contain any information on our desired quantity $\mathsf{p_0}(C_0)$. The main insight that allows us to correlate average-case solutions to the worst-case quantity, which was originally developed to show the average-case hardness for permanents [28, 29], is to create many random instances $\{C_i\}$ by the following procedure: first sample a "random seed" $\{H_i\}$, then apply small and different perturbations to the random seed, and then apply each perturbed random seed to C_0. As a result, while each random instance is marginally distributed approximately according to $\mathcal{H}_A$, they are close to each other and are correlated in a way that reveals the worst-case quantity $\mathsf{p_0}\left(C_0\right)$.

More specifically, suppose there is a way of perturbing the circuit C_1 into a new circuit $C(\theta)$ ($\theta \in [0,1]$), such that $C(\theta) \approx C_1$ when $\theta \ll 1$, and $C(1) = C_0$. Moreover, suppose $\mathsf{p_0}\left(C(\theta)\right)$ for different values of θ are correlated in a way such that $\mathsf{p_0}\left(C(1)\right)$ can be inferred from $\mathsf{p_0}\left(C(\theta)\right)$ for small values of θ. Then, to compute $\mathsf{p_0}\left(C_0\right)$, it suffices to query $\mathcal{O}$ with $C(\theta_i)$ for many small θ_is.

One way to develop such a procedure is by perturbing the random seeds $\{H_i\}$ used in the construction of C_1, which we define as follows.

Definition 2.1 (θ-perturbed random circuit distribution)**.** *Suppose there is a transform on unitary matrices $H \mapsto H(\theta)$ parameterized by a real number $\theta \in [0,1]$, that satisfies $H(0) = H$ is unchanged, and $H(1) = I$ is the identity matrix. For any circuit C_0 with gates $\{G_i\}_{i=1,\dots,m}$ and a random seed $\{H_i\}_{i=1,\dots,m}$, the circuit $C(\theta)$ is defined by replacing G_i with*

$$G_i \mapsto H_i(\theta) G_i. \tag{2.4}$$

Denote the distribution induced by $C(\theta)$ as $\mathcal{H}_{\mathcal{A},\theta}$. Note that $\mathcal{H}_{\mathcal{A},0} = \mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{A},1} = \mathbf{1}_{C_0}$.

As discussed above, in order for such a perturbation to be useful for the worst-to-average-case reduction, we require that the following (informal) properties are satisfied.

Properties of the perturbed circuit distribution (informal).

1. When θ is small, $\mathcal{H}_{\mathcal{A},\theta}$ is close to $\mathcal{H}_{\mathcal{A}}$ in total variation distance. Therefore, when $\mathcal{O}$ is given an input $C(\theta) \sim \mathcal{H}_{\mathcal{A},\theta}$ from the perturbed distribution, it is guaranteed to return a correct approximation of $\mathsf{p_0}\left(C(\theta)\right)$ with high success probability.

2. An approximation of $\mathsf{p_0}\left(C(1)\right)$ can be inferred from approximations of $\mathsf{p_0}\left(C(\theta)\right)$ for small values of θ with a procedure in the Polynomial Hierarchy.

Finding such a good perturbation method is non-trivial. Two proposals were developed in previous work in the context of noiseless circuits. In [26], they considered the truncated Taylor series of $He^{-\theta \log H}$, which is given by

$$H(\theta) = H \cdot \left(\sum_{k=0}^{K} \frac{(-\theta \log H)^k}{k!} \right), \tag{2.5}$$

and showed that $\mathsf{p_0}\left(C(\theta)\right)$ is a degree $O(mK)$ polynomial in θ. Property 2 is then satisfied by using polynomial interpolation techniques. In this approach, $H(\theta)$ as defined by Eq. (2.5) is not unitary, and the resulting average-case hardness is for a circuit family that has a small deviation from $\mathcal{H}_{\mathcal{A}}$ given by the truncation error of the Taylor series. Later, Movassagh [27] developed a new perturbation method, called Cayley transform, that is able to stay within the unitary path.

Definition 2.2 (Cayley transform [27])**.** *The Cayley transform of a unitary matrix H parameterized by $\alpha \in [0,1]$ is a unitary matrix defined as*

$$H(\alpha) := \frac{\alpha I + (2-\alpha)H}{(2-\alpha)I + \alpha H}, \tag{2.6}$$

where A/B means $A \cdot B^{-1}$. Consider the diagonalization of a unitary matrix

$$H = \sum_j e^{i\varphi_j} |\psi_j\rangle\langle\psi_j| ,$$

the following is an equivalent form of the Cayley transform,

$$H(\alpha) = \sum_j \frac{1 + i(1-\alpha)\tan\frac{\varphi_j}{2}}{1 - i(1-\alpha)\tan\frac{\varphi_j}{2}} |\psi_j\rangle\langle\psi_j| . \tag{2.7}$$

Note that $H(0) = H$ and $H(1) = I$.

Note that we are using slightly different notations from [27]. A self-consistent presentation of the properties of the Cayley transform is given in [4, Appendix E].

While not having an evident polynomial structure as in the construction of [26], in [27] they applied Cayley transform to the gates $\{H_i\}$ parameterized by $\alpha = \theta$, and showed that the resulting $\mathsf{p_0}(C(\theta))$ is a degree $(O(m), O(m))$ rational function in θ. A similar polynomial interpolation technique can be applied to infer $\mathsf{p_0}(C(1))$ from $\mathsf{p_0}(C(\theta))$ with small values of θ.

In particular [27], building on [26], obtained the following average-case hardness results.

Theorem 2.2 ([26, 27])**.** *Let $\mathcal{A}$ be a circuit architecture so that computing $\mathsf{p_0}(C)$ to within additive error $2^{-O(m)}$ is #P-hard in the worst case. The following results hold:*

1. *It is #P-hard to compute $\mathsf{p_0}(C)$ exactly on input $C \sim \mathcal{H}_\mathcal{A}$, with success probability at least $1-\eta$ over the choice of C, for any constant $\eta < \frac{1}{4}$.*

2. *It is #P-hard to compute $\mathsf{p_0}(C)$ up to an additive imprecision $2^{-O(m^3)}$ on input $C \sim \mathcal{H}_\mathcal{A}$, with success probability at least $1 - O(1)/m$ over the choice of C.*

We improve these results by proving an average-case hardness that tolerates a larger additive imprecision $2^{-O(m\log m)}$, while only requiring a constant success probability over the choice of random circuits. These improvements follows from two new techniques: the first is a robust polynomial interpolation argument which tolerates a constant failure probability, and the second is an improved error bound for long distance polynomial extrapolation. Both results are developed in Section 2.2.2 and are used as black boxes here.

To establish our reduction, we first consider the θ-perturbed random circuit distribution instantiated by the Cayley transform.

Definition 2.3 (θ-Cayley perturbed random circuit distribution)**.** *For any circuit C_0 with gates $\{G_i\}_{i=1,\ldots,m}$ and a random seed $\{H_i\}_{i=1,\ldots,m}$, the circuit $C(\theta)$ is defined by replacing G_i with*

$$G_i \mapsto H_i(\theta) G_i, \tag{2.8}$$

where $H_i(\theta)$ denotes the Cayley transform of H_i parameterized by θ. Denote the distribution induced by $C(\theta)$ as $\mathcal{H}_{\mathcal{A},\theta}$. Note that $\mathcal{H}_{\mathcal{A},0} = \mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{A},1} = \mathbf{1}_{C_0}$.

As previously shown by [27], the output probability $\mathsf{p_0}\left(C(\theta)\right)$ is a low-degree rational function in θ.

Lemma 2.1 ([27])**.** *For any circuit C_0, let $C(\theta)$ be a circuit from the θ-Cayley perturbed random circuit distribution as in Definition 2.3. Then $\mathsf{p_0}\left(C(\theta)\right)$ is a degree $(O(m), O(m))$ rational function in θ.*

Proof. Here we give a self-consistent proof as the details are useful for our later developments.

First, notice that the output amplitude can be written as the Feynman path integral,

$$\langle 0^n | C(\theta) | 0^n \rangle = \sum_{y_1,\ldots,y_{m-1}\in\{0,1\}^n} \prod_{j=1}^{m} \langle y_j | (H_j(\theta) G_j \otimes I_{else}) | y_{j-1} \rangle \,, \tag{2.9}$$

where we have $y_0 = y_m = 0^n$, $H_j(\theta)G_j$ is a local gate and I_{else} denotes identity on all the other qubits. Let the diagonalization of H_j be $H_j = \sum_l e^{i\varphi_{jl}} |\psi_{jl}\rangle\langle\psi_{jl}|$. Then, consider an individual term in the above sum,

$$\begin{aligned}
&\prod_{j=1}^{m} \langle y_j | H_j(\theta) G_j | y_{j-1} \rangle \\
&= \prod_{j=1}^{m} \langle y_j | \sum_l \frac{\left(1 + i(1-\theta)\tan\frac{\varphi_{jl}}{2}\right) |\psi_{jl}\rangle\langle\psi_{jl}|}{1 - i(1-\theta)\tan\frac{\varphi_{jl}}{2}} G_j \, | y_{j-1} \rangle \\
&= \prod_{j=1}^{m} \langle y_j | \frac{\sum_l \left(1 + i(1-\theta)\tan\frac{\varphi_{jl}}{2}\right) |\psi_{jl}\rangle\langle\psi_{jl}| \prod_{t\neq l}\left(1 - i(1-\theta)\tan\frac{\varphi_{jt}}{2}\right)}{\prod_l \left(1 - i(1-\theta)\tan\frac{\varphi_{jl}}{2}\right)} G_j \, | y_{j-1} \rangle \,.
\end{aligned} \tag{2.10}$$

Let

$$\begin{aligned}
Q_0(\theta) &:= \prod_{j=1}^{m} \prod_l \left(1 - i(1-\theta)\tan\frac{\varphi_{jl}}{2}\right), \\
Q(\theta) &:= |Q_0(\theta)|^2.
\end{aligned} \tag{2.11}$$

Then Eq. (2.10) is a degree $(O(m), O(m))$ rational function in θ with $Q_0(\theta)$ being the denominator. Furthermore, notice that $Q_0(\theta)$ does not depend on the Feynman path $\{y_j\}$, and each term in the sum in Eq. (2.9) shares the same denominator $Q_0(\theta)$. Therefore, $\langle 0^n | C(\theta) | 0^n \rangle$ is a degree $(O(m), O(m))$ rational function in θ, and so is $\mathsf{p_0}\left(C(\theta)\right) = |\langle 0^n | C(\theta) | 0^n \rangle|^2$. $Q(\theta)$ is then the denominator of $\mathsf{p_0}\left(C(\theta)\right)$. Finally, note that if C_0 only consists of 2-qubit gates, then $\mathsf{p_0}\left(C(\theta)\right)$ is a degree $(8m, 8m)$ rational function in θ.

□

Lemma 2.1 formally supports Property 2 of the θ-Cayley perturbed random circuit distribution. Intuitively, this low-degree rational function structure allows us to apply polynomial interpolation techniques to obtain an approximation to the worst-case quantity, even though the worst-case point ($\theta = 1$) is far from the average-case data points ($0 < \theta \ll 1$). Details of this interpolation technique are presented in the proof of our main result (Theorem 2.3) and in Section 2.2.2. Before presenting the main result, it remains to establish Property 1 of the θ-Cayley perturbed random circuit distribution, which is given in the following lemma.

Lemma 2.2 ([27])**.** *Let $\mathcal{H}_{\mathcal{A},\theta}$ be the θ-Cayley perturbed random circuit distribution as in Definition 2.3, and $\mathcal{H}_{\mathcal{A}}$ be the distribution of Haar random circuits over $\mathcal{A}$. Then we have*

$$D_{\mathrm{TV}}(\mathcal{H}_{\mathcal{A},\theta}, \mathcal{H}_{\mathcal{A}}) = O(m\theta), \tag{2.12}$$

where $D_{\mathrm{TV}}(\cdot,\cdot)$ denotes the total variation distance between probability distributions and m is the number of gates in $\mathcal{A}$.

Proof. This total variation distance can be bounded by considering the distribution of each individual gates. In [27], it was shown that the total variation distance between the Cayley transformed random unitary $H(\theta)$ and Haar random unitary is $O(\theta)$. Therefore by additivity of the total variation distance we have $D_{\mathrm{TV}}(\mathcal{H}_{\mathcal{A},\theta}, \mathcal{H}_{\mathcal{A}}) = O(m\theta)$. □

Having established Property 1 and 2, we are now ready to state and prove our first result on the average-case hardness of random circuits.

Theorem 2.3. *Let $\mathcal{A}$ be a circuit architecture so that computing $\mathsf{p_0}(C)$ to within additive error $2^{-O(m)}$ is $\#\mathsf{P}$-hard in the worst case. Then the following problem is $\#\mathsf{P}$-hard under a $\mathsf{BPP}^{\mathsf{NP}}$ reduction: for any constant $\eta < \frac{1}{4}$, on input a random circuit $C \sim \mathcal{H}_{\mathcal{A}}$ with m gates, compute the output probability $\mathsf{p_0}(C)$ up to additive error $\delta = \exp(-O(m\log m))$, with probability at least $1-\eta$ over the choice of C.*

Remark 2.1. *Another way of stating Theorem 2.3 is the following: suppose there exists an algorithm that belongs to some finite level of PH that, on input a random circuit $C \sim \mathcal{H}_{\mathcal{A}}$ with m gates, computes the output probability $\mathsf{p_0}(C)$ up to additive error $\delta = \exp(-O(m\log m))$, with probability at least $1-\eta$ ($\eta < \frac{1}{4}$) over the choice of C as well as the randomness of the algorithm. Then PH collapses to a finite level.*

Proof. Let $\mathcal{O}$ be an algorithm that correctly approximates $\mathsf{p_0}(C)$ of a random circuit $C \sim \mathcal{H}_{\mathcal{A}}$ up to additive error δ, with success probability at least $1-\eta$ over the choice of C. In the following, we show that there exists a $\mathsf{BPP}^{\mathsf{NP}^{\mathcal{O}}}$ procedure that on input *any* circuit C_0, computes $\mathsf{p_0}(C_0)$ up to additive error $\delta' = \delta\exp(O(m\log m))$, with success probability at least $\frac{2}{3}$. The theorem statement then follows from the worst-case hardness of computing $\mathsf{p_0}(C_0)$ over $\mathcal{A}$.

Consider any circuit C_0 with m gates over the architecture $\mathcal{A}$. Create a new circuit C_1 as follows: for each gate G_i ($i = 1,\ldots,m$) in C_0, we replace G_i with H_iG_i, where $\{H_i\}$

is independently drawn from the Haar measure over the unitary group and has the same dimension as G_i. By the invariance of Haar measure, C_1 is distributed the same as $\mathcal{H}_{\mathcal{A}}$.

Fix the random unitary gates $\{H_i\}$. Next, we apply the θ-Cayley perturbation on C_0 using the random seed $\{H_i\}$ as in Definition 2.3 to get the perturbed circuit $C(\theta)$. By definition, we have $C(0) = C_1 \sim \mathcal{H}_{\mathcal{A}}$ and $C(1) = C_0$.

By Lemma 2.1, $\mathsf{p_0}(C(\theta))$ is a degree $(O(m), O(m))$ rational function in θ. Let $P(\theta)$, $Q(\theta)$ be the numerator and denominator of $\mathsf{p_0}(C(\theta))$, respectively, then $\mathsf{p_0}(C(\theta)) = \frac{P(\theta)}{Q(\theta)}$. Note that from the proof of Lemma 2.1, we have that

$$Q(\theta) = \prod_{j=1}^{m} \prod_{l} \left| 1 - i(1-\theta)\tan\frac{\varphi_{jl}}{2} \right|^2. \tag{2.13}$$

The goal is to recover $\mathsf{p_0}(C(1)) = \frac{P(1)}{Q(1)}$ from values of $\mathsf{p_0}(C(\theta))$ for small θ. To do this, first note that $Q(\theta)$ is a known polynomial whose value can be computed in time $O(m)$ for any θ. Therefore, the problem can be reduced to a polynomial interpolation for $P(\theta)$.

To analyze the error for the polynomial interpolation, it is useful to establish bounds for $Q(\theta)$. We can write $Q(\theta)$ as

$$Q(\theta) = \prod_{j=1}^{m} \prod_{l} \left| 1 - i(1-\theta)\tan\frac{\varphi_{jl}}{2} \right|^2 = \prod_{j=1}^{m} \prod_{l} \left(1 + (1-\theta)^2 \tan^2\frac{\varphi_{jl}}{2} \right), \tag{2.14}$$

where it is easy to see that $Q(\theta) \geq 1$. Next, call the set of Haar random gates $\{H_j\}_{j=1,\dots,m}$ β-good, if all eigenvalues φ_{jl} of all gates lie in the range $[-\pi+\beta, \pi-\beta]$. By Lemma 2.3, this happens with probability at least $1 - O(m\beta)$. Suppose we choose $\beta = O(m^{-1})$ such that $\{H_j\}$ is β-good with high constant probability. Then conditioned on $\{H_j\}$ being β-good, we have

$$\begin{aligned} Q(\theta) &= \prod_{j=1}^{m} \prod_{l} \left(1 + (1-\theta)^2 \tan^2\frac{\varphi_{jl}}{2} \right) \\ &\leq \prod_{j=1}^{m} \prod_{l} \left(1 + \tan^2\frac{\pi-\beta}{2} \right) \\ &\leq \prod_{j=1}^{m} \prod_{l} \left(1 + \frac{4}{\beta^2} \right) \\ &= \left(1 + O(m^2)\right)^{O(m)} \\ &= \exp\left(O(m\log m)\right). \end{aligned} \tag{2.15}$$

We also define the above upper bound of $Q(\theta)$ as K, where $K = \exp\left(O(m\log m)\right)$.

Next we apply the algorithm $\mathcal{O}$ on input $C(\theta)$. Notice that by definition, $\mathcal{O}$ works on inputs from the distribution $\mathcal{H}_{\mathcal{A}}$, while $C(\theta)$ is distributed according to $\mathcal{H}_{\mathcal{A},\theta}$ as in Definition 2.3. Therefore, the success probability of $\mathcal{O}$ depends on the distance between the two distributions,

$$\Pr_{C(\theta)\sim\mathcal{H}_{\mathcal{A},\theta}}\left[\left|\mathcal{O}(C(\theta)) - \mathsf{p_0}\left(C(\theta)\right)\right| \geq \delta\right] \leq \eta + D_{\mathrm{TV}}(\mathcal{H}_{\mathcal{A},\theta}, \mathcal{H}_{\mathcal{A}}), \tag{2.16}$$

where D_{TV} denotes total variation distance. By Lemma 2.2, we have $D_{\mathrm{TV}}(\mathcal{H}_{\mathcal{A},\theta}, \mathcal{H}_{\mathcal{A}}) = O(m\theta)$. Let $\Delta = O(m^{-1})$ and restrict θ in the interval $[0, \Delta]$, such that $D_{\mathrm{TV}}(\mathcal{H}_{\mathcal{A},\theta}, \mathcal{H}_{\mathcal{A}}) = O(m\Delta)$ is upper bounded by a small constant.

Conditioned on $\mathcal{O}$ being successful and $\{H_j\}$ being β-good, we have $\left|\mathcal{O}(C(\theta)) - \frac{P(\theta)}{Q(\theta)}\right| \leq \delta$. After seeing the output $\mathcal{O}(C(\theta))$ of $\mathcal{O}$, we multiply it with $Q(\theta)$ to get an estimate of $P(\theta)$, which satisfies

$$|\mathcal{O}(C(\theta))Q(\theta) - P(\theta)| \leq \delta Q(\theta) \leq \delta K. \tag{2.17}$$

By a simple union bound, the above equation holds with failure probability at most

$$\begin{aligned}\Pr\left[|\mathcal{O}(C(\theta))Q(\theta) - P(\theta)| \geq \delta K\right] &\leq \Pr\left[\left|\mathcal{O}(C(\theta)) - \mathsf{p_0}\left(C(\theta)\right)\right| \geq \delta \vee \{H_j\} \text{ not } \delta\text{-good}\right] \\ &\leq \eta + O(m\Delta) + O(m\beta) \leq \eta' < \frac{1}{4},\end{aligned} \tag{2.18}$$

for a suitable choice of constants in the O-notation for $\Delta = O(m^{-1})$ and $\beta = O(m^{-1})$, such that $\eta' < \frac{1}{4}$.

To compute $P(1)$, we apply the algorithm $\mathcal{O}$ to a set of circuits $\{C(\theta_i)\}$, where θ_i ($i = 1, \ldots, O(m^2)$) is a set of equally spaced points in the interval $[0, \Delta]$, and different perturbed circuits $C(\theta_i)$ shares the same random seed $\{H_j\}$. By Eq. (2.18), we obtain a set of points $\{(\theta_i, y_i)\}$ such that

$$\Pr\left[|y_i - P(\theta_i)| \geq \delta K\right] \leq \eta' < \frac{1}{4}. \tag{2.19}$$

The problem is then reduced to a polynomial interpolation for $P(\theta)$, a degree $O(m)$ polynomial, with the noisy data $\{(\theta_i, y_i)\}$. Using our robust Berlekamp-Welch theorem which we develop in the next section (see Theorem 2.4), on input $\{(\theta_i, y_i)\}$ we can compute a number $p \approx P(1)$ with access to an NP oracle, where the error can be bounded by

$$\begin{aligned}|p - P(1)| &\leq \delta K \exp\left(O(m \log \Delta^{-1}) + O(m)\right) \\ &= \delta \cdot \exp(O(m \log m)).\end{aligned} \tag{2.20}$$

Finally, notice that

$$\mathsf{p_0}\left(C_0\right) = \mathsf{p_0}\left(C(1)\right) = \frac{P(1)}{Q(1)} = P(1) \tag{2.21}$$

as $Q(1) = 1$, therefore by choosing $\delta = \exp\left(-O(m \log m)\right)$ with a sufficiently large constant, we can compute the worst-case output probability $\mathsf{p_0}\left(C_0\right)$ up to additive error $\exp\left(-O(m \log m)\right)$. The overall procedure is in $\mathsf{BPP}^{\mathsf{NP}^{\mathcal{O}}}$. If $\mathcal{O}$ is an algorithm that belongs to some finite level of PH, then this procedure also belongs to a finite level of PH, and therefore by the worst-case $\#\mathsf{P}$ hardness the PH collapses to a finite level. □

Lemma 2.3. *The Haar distribution μ_H over the unitary group $\mathbb{U}(N)$ satisfies*

$$\Pr_{\mu_H}\left[\varphi_j \in [-\pi+\delta, \pi-\delta],\ \forall j\right] \geq 1 - \frac{N\delta}{\pi}, \tag{2.22}$$

where φ_j $(j = 1 \ldots N)$ denotes the eigenvalues of the random unitary distributed according to μ_H.

Proof. The proof is a simple union bound:

$$\begin{aligned}\Pr_{\mu_H}\left[\varphi_j \in [-\pi+\delta, \pi-\delta],\ \forall j\right] &= 1 - \Pr_{\mu_H}\left[\exists j : \varphi_j \notin [-\pi+\delta, \pi-\delta]\right] \\ &\geq 1 - \sum_{j=1}^{N} \Pr_{\mu_H}\left[\varphi_j \notin [-\pi+\delta, \pi-\delta]\right] \\ &= 1 - N \Pr_{\mu_H}\left[\varphi_1 \notin [-\pi+\delta, \pi-\delta]\right] \\ &= 1 - \frac{N\delta}{\pi}.\end{aligned} \tag{2.23}$$

□

In the above proof we used a result developed in Section 2.2.2 as a black box, which we refer to as robust Berlekamp-Welch algorithm. This algorithm allows us to do polynomial interpolation on noisy data, while tolerating a constant fraction of data points that can be arbitrarily wrong. Similar to the Learning with Errors problem for solving noisy linear equations, the polynomial interpolation problem with both noise and corruption seems unlikely to be solved in polynomial time. However, recall that given the #P hardness of computing $\mathsf{p_0}(C)$ in the worst case, in our worst-to-average-case reduction we are allowed to use any finite level of PH. We show how to do such a polynomial interpolation in Section 2.2.2 having access to a NP oracle. As a result, we are able to tolerate a constant failure probability in our average-case hardness.

Furthermore, our proof techniques can also be naturally applied to BosonSampling, where a fundamental question is to prove robust average-case hardness results for computing permanents of matrices with i.i.d. Gaussian entries [24, 30]. Consider a BosonSampling experiment with n photons and $m = n^c$ $(c > 2)$ output modes. Then assuming no collision, the output probability corresponding to a output pattern $S = (s_1, \ldots, s_m)$ $(s_i \in \{0, 1\}, \sum_i s_i = n)$ is

$$\Pr[S] = \left|\mathrm{Per}(A_S)\right|^2, \tag{2.24}$$

where A is a $m \times n$ submatrix of a $m \times m$ Haar random unitary matrix, and A_S is the $n \times n$ submatrix of A whose rows are selected according to S. When c is a large enough constant, A_S is distributed close in total variation distance to matrices with i.i.d. complex Gaussian entries of mean 0 and variance $\frac{1}{m}$. We then obtain the following result using similar proof techniques:

Corollary 2.2 (Restatement of Corollary 2.1)**.** *For any constant $\eta < \frac{1}{4}$, it is $\#\mathsf{P}$-hard under a $\mathsf{BPP}^{\mathsf{NP}}$ reduction to compute the output probability of a n-photon $m = n^c$-mode BosonSampling experiment up to additive error $\delta = e^{-(c+4)n\log n - O(n)}$, with success probability $1-\eta$.*

Remark 2.2. *As shown in [24], to prove the hardness of approximate sampling for BosonSampling experiments, it suffices to improve the constant in our result from $c+4$ to $c-1$. This is because on average the output probability is roughly*

$$\frac{1}{\binom{m}{n}} \approx \frac{n!}{m^n} \approx e^{-(c-1)n\log n}. \tag{2.25}$$

However, the barrier result shown by [24] implies that this improvement cannot be obtained using similar techniques based on polynomial interpolation. See [4, Appendix B] for a detailed discussion, where we also show that this barrier result does not rule out improving the constant to $c+1$.

Proof. Let $\mathcal{G}^{n\times n}$ denote the distribution over $n\times n$ matrices where each entry is independently distributed according to $\mathcal{CN}(0,1)$, the standard complex Gaussian distribution. Suppose c is a large enough constant[2], then an algorithm for approximating the output probability of a BosonSampling experiment can be used to approximate

$$p_X := \frac{|\mathrm{Per}(X)|^2}{m^n}, \quad X \sim \mathcal{G}^{n\times n} \tag{2.26}$$

with a small loss in the success probability.

Let $\mathcal{O}$ be an algorithm that, on input a random matrix $X \sim \mathcal{G}^{n\times n}$, correctly approximates p_X up to additive error δ, with success probability at least $1-\eta$ over the choice of X. In the following, we show that there exists a $\mathsf{BPP}^{\mathsf{NP}^{\mathcal{O}}}$ procedure that on input *any* 0/1 matrix X_0, computes $|\mathrm{Per}(X_0)|^2$ up to additive error $\delta' = \delta\exp\left((c+4)n\log n + O(n)\right)$, with success probability at least $\frac{2}{3}$. The theorem statement then follows from the worst-case $\#\mathsf{P}$ hardness of computing $\mathrm{Per}(X_0)$.

Let $X_1 \sim \mathcal{G}^{n\times n}$ and define

$$X(\theta) := (1-\theta)X_1 + \theta X_0. \tag{2.27}$$

Then $|\mathrm{Per}(X(\theta))|^2$ is a degree $d = 2n$ polynomial in θ. By a result of [24], the total variation distance between the distribution of $X(\theta)$ and $\mathcal{G}^{n\times n}$ is $O(n^2\theta)$. This can be shown by calculating the distance for one entry, which is $O(\theta)$, and then multiply by n^2 as the entries are independently distributed.

Consider $O(n^2)$ uniformly spaced points $\{\theta_i\}$ in $[0,\Delta]$ with $\Delta = O(n^{-2})$. For a suitable choice of constants, we can guarantee that for each data point θ_i,

$$\Pr\left[\left|\mathcal{O}(X(\theta_i)) - \frac{|\mathrm{Per}(X(\theta_i))|^2}{m^n}\right| \geq \delta\right] \leq \eta + O(n^2\Delta) \leq \eta' \tag{2.28}$$

[2] The proof in [24] requires c to be at least 5, and it was conjectured that $c > 2$ suffices.

for some constant $\eta' < \frac{1}{4}$. The procedure then follows by sending the points $\{(\theta_i, \mathcal{O}(X(\theta_i)))\}$ to the robust Berlekamp-Welch algorithm (Theorem 2.4) to obtain the desired approximation to the worst-case quantity $|\mathrm{Per}(X_0)|^2 = |\mathrm{Per}(X(1))|^2$. Overall we have a $\mathsf{BPP}^{\mathsf{NP}^{\mathcal{O}}}$ procedure for computing $|\mathrm{Per}(X_0)|^2$ up to additive error

$$\delta \cdot m^n \cdot \exp\left(d \log \Delta^{-1} + O(d)\right) = \delta \cdot \exp\left((c+4)n \log n + O(n)\right), \tag{2.29}$$

which concludes the proof. □

2.2.2 Robust Berlekamp-Welch

We prove the following theorem for robust polynomial interpolation. All polynomials considered here have real coefficients, and our results can be generalized to the complex case, as for polynomials with complex coefficients we can deal with the real and imaginary parts separately. Below our theorem is stated with success probability $\frac{2}{3}$, which can be amplified to $1 - 1/\mathsf{exp}(d)$ by taking the median of poly(d) independent experiments.

Theorem 2.4 (Robust Berlekamp-Welch)**.** *For a degree d polynomial $P(x)$, suppose there is a set of data points $D = \{(x_i, y_i)\}$ such that $|D| = 100d^2$ and $\{x_i\}$ is equally spaced in the interval $[0, \Delta]$ $(\Delta < 1)$. Furthermore, assume that each point (x_i, y_i) satisfies*

$$\Pr\left[|y_i - P(x_i)| \geq \delta\right] \leq \eta, \tag{2.30}$$

where $\eta < \frac{1}{4}$ is a constant, then there exists a P^{NP} algorithm which, takes D as input, returns a number p such that $|p - P(1)| \leq \delta e^{d \log \Delta^{-1} + O(d)}$, with success probability at least $\frac{2}{3}$.

Remark 2.3. *Note that our result can be improved to using only $O(d)$ data points, by applying a result of Rakhmanov [31]. However, the number of data points we use does not affect our main results, as long as it is polynomial in d. Here we give a simpler proof using $O(d^2)$ data points.*

Proof. For simplicity, our proof is presented by specifying $\eta = \frac{1}{300}$, which can be naturally generalized to any $\eta < \frac{1}{4}$. Say a point (x_i, y_i) is correct if $|y_i - P(x_i)| \leq \delta$. By Eq. (2.30), the expected number of wrong points is less than $\frac{1}{300}$ fraction. By Markov's inequality, with probability at least $\frac{2}{3}$, 0.99 fraction of points in D are correct. In the following we show that conditioned on 0.99 fraction of points being correct (denote the set of correct points as F), there exists a deterministic P^{NP} algorithm that computes a number p that satisfies $|p - P(1)| \leq \delta e^{d \log \Delta^{-1} + O(d)}$. This implies the statement of the theorem.

Consider the following computational problem:

Problem 2.1. *Given $100d^2$ points D, decide if there exists at least 0.99 fraction of them (denoted as F') and a degree d polynomial $Q(x)$ that satisfy $|y_i - Q(x_i)| \leq \delta, \forall (x_i, y_i) \in F'$.*

Problem 2.1 is in NP because a certificate (F', Q) can be efficiently verified by checking each point in F'.

When giving D in the statement of the theorem as input, this problem has a satisfying certificate (F, P). Having access to a NP oracle, as this problem is guaranteed to have a solution, we can find a certificate (F', Q) by performing binary search in the certificate space, which may be different from (F, P). However, in the following we show that any solution will satisfy the requirement of the theorem, and the algorithm simply outputs $Q(1)$.

Let $R(x) = P(x) - Q(x)$. As $|F \cap F'| \geq 0.98|D|$, there are at least 0.98 fraction of points in D that satisfies

$$|R(x_i)| \leq |P(x_i) - y_i| + |Q(x_i) - y_i| = 2\delta. \tag{2.31}$$

Recall that D is a set of equally spaced points in $[0, \Delta]$. Lemma 2.5 says that in such condition $R(x)$ can be uniformly bounded in the interval $[0, \Delta]$:

$$|R(x)| \leq \delta e^{O(d)}, \quad \forall x \in [0, \Delta]. \tag{2.32}$$

Finally, as R is uniformly bounded in $[0, \Delta]$, we can use Lemma 2.7 to bound the error at $x = 1$,

$$|R(1)| = |P(1) - Q(1)| \leq \delta e^{O(d)} e^{d \log \Delta^{-1} + d \log 8} = \delta e^{d \log \Delta^{-1} + O(d)}, \tag{2.33}$$

which concludes the proof. □

The above proof needs several additional results which we develop in the following. First, recall the following result of Paturi, which gives a bound of how fast a polynomial can grow when it is uniformly bounded in a small interval.

Lemma 2.4 (Paturi [32])**.** *Let P be a degree d polynomial which satisfies $|P(x)| \leq \varepsilon$ for $x \in [0, \Delta]$ ($\Delta < 1$). Then we have $|P(1)| \leq \varepsilon e^{4d\Delta^{-1}}$.*

Proof. This is implied by Fact 2 and Corollary 2 of [32]. □

The first missing step in the proof of Theorem 2.4, which is in Eq. (2.32), is to show the following: if a degree d polynomial is bounded by δ for a constant fraction of a set of uniformly spaced points in $[0, \Delta]$, then the maximum value of the polynomial in the interval $[0, \Delta]$ is at most $\delta e^{O(d)}$. We show this in the following lemma, which is the main technical part of the full proof.

Lemma 2.5. *Let P be a degree d polynomial which satisfies $|P(x)| \leq 1$ for any $x \in A$, where $A \subseteq B$ and B is a set of equally spaced points in $[0, 1]$ such that $|B| = 100d^2$ and $|A| \geq 0.98|B|$. Then P satisfies $|P(x)| \leq 2^{O(d)}, \forall x \in [0, 1]$.*

Proof. Denote A as good points at which P is bounded, and denote $B - A$ as bad points. We first prove the bound under a special case, where the good points are exactly the first 0.98 fraction. More specifically, for the set of uniformly spaced points

$$B = \{0, \delta, 2\delta, \ldots, (100d^2 - 1)\delta\} \subseteq [0, 1] \tag{2.34}$$

where $\delta = \frac{1}{100d^2}$, suppose the polynomial is bounded at the first 0.98 fraction of points,

$$|P(x)| \leq 1, \quad \forall x \in A = \{0, \delta, \dots, (98d^2 - 1)\delta\}. \tag{2.35}$$

When a polynomial is bounded at consecutive uniformly spaced points in an interval, it can also be uniformly bounded at all points in that interval. Using Lemma 2.6 we get

$$|P(x)| \leq O(1), \quad \forall x \in [0, (98d^2 - 1)\delta]. \tag{2.36}$$

As $P(x)$ is uniformly bounded on a constant subinterval of $[0, 1]$, we can then use Lemma 2.4 to obtain the desired bound $|P(x)| \leq 2^{O(d)}, \forall x \in [0, 1]$. Also note that both inequalities are saturated by Chebyshev polynomial, so the bound is tight.

We proceed by proving the bound for the general case where A can be arbitrarily distributed in B. Let $M = \max_{x\in[0,1]} |P(x)|$ and suppose the maximum is achieved at $x^* \in [\frac{1}{2}, 1]$. Consider the roots of P

$$P(x) = L \prod_{k=1}^{d} (x - r_k), \quad L > 0, \tag{2.37}$$

where r_k can have real and imaginary parts. We construct a new polynomial $Q(x) = L \prod_{k=1}^{d} (x - r'_k)$ according to the following rule:

$$r'_k = \begin{cases} r_k, & \mathrm{Re}(r_k) \leq x^*, \\ 2x^* - \mathrm{Re}(r_k) + i\,\mathrm{Im}(r_k), & \mathrm{Re}(r_k) > x^*. \end{cases} \tag{2.38}$$

Let $M' = \max_{x\in[0,1]} |Q(x)|$. We show that this construction is useful because $Q(x)$ has the following properties:

1. $\deg(Q) = \deg(P) = d$.
   Proof. By construction.

2. $M' \geq M$.
   Proof. $M' \geq |Q(x^*)| = |P(x^*)| = M$.

3. $|Q(x)| \leq |P(x)|, \forall x \in [0, x^*]$.
   Proof. This follows because $|x - r'_k| \leq |x - r_k|, \forall k, \forall x \in [0, x^*]$.

4. $M' = |Q(1)|$.
   Proof. $|Q(x)|$ is monotonic increasing at $[x^*, 1]$ because $\forall x^* \leq x < y \leq 1$, we have $|x - r'_k| \leq |y - r'_k|$. Also, $\forall x \in [0, x^*], |Q(x)| \leq |P(x)| \leq M = |Q(x^*)|$. This suggests that the maximum occurs at $x = 1$.

According to property 2, we proceed by proving an upper bound for M'. Let

$$B' = \{0, \delta, 2\delta, \dots, (50d^2 - 1)\delta\} \subseteq [0, \frac{1}{2}].$$

Among these points, at least 0.96 fraction of them are good for $Q(x)$, denoted by A'. More specifically,

$$|Q(x)| \le |P(x)| \le 1, \forall x \in A' \subseteq B'. \tag{2.39}$$

If A' is a consecutive set starting from 0 (i.e. $A' = \{0, \delta, 2\delta, \ldots, (|A'|-1)\delta\}$), then we can obtain the desired bound by repeating the argument at the beginning of this proof. When this is not the case, there exists $y \in B'$ such that y is bad and $y+\delta$ is good. We proceed by converting $Q(x)$ to a new polynomial to eliminate this kind of points, while maintaining the good properties of Q.

Let y be the maximum element in B' such that y is bad and $y+\delta$ is good (assuming such y exists). By definition, all good points larger than y is consecutive, denoted as

$$A' \cap [y, \frac{1}{2}] = \{y+\delta, y+2\delta, \ldots, y+T\delta\} \tag{2.40}$$

for some $T \ge 1$. We perform the following operation on $Q(x) = L\prod_{k=1}^{d}(x-r_k)$, resulting in a new polynomial $R(x) = L\prod_{k=1}^{d}(x-r'_k)$:

$$r'_k = \begin{cases} r_k, & \mathrm{Re}(r_k) \le y, \\ \mathrm{Re}(r_k) - \delta + i\,\mathrm{Im}(r_k), & \mathrm{Re}(r_k) > y. \end{cases} \tag{2.41}$$

We prove the following properties for R.

5. $|R(x-\delta)| \le 1, \forall x \in A' \cap [y, \frac{1}{2}]$.
   Proof. $\forall x \in A' \cap [y, \frac{1}{2}]$, notice that $|x-\delta-r'_k| = |x-r_k|$ when $\mathrm{Re}(r_k) > y$, and $|x-\delta-r'_k| \le |x-r_k|$ when $\mathrm{Re}(r_k) \le y$. Therefore $|R(x-\delta)| \le |Q(x)| \le 1$.

6. $|R(x)| \le 1, \forall x \in A' \cap [0, y]$.
   Proof. $\forall x \in A' \cap [0, y]$, notice that $|x-r'_k| = |x-r_k|$ when $\mathrm{Re}(r_k) \le y$, and $|x-r'_k| \le |x-r_k|$ when $\mathrm{Re}(r_k) > y$. Therefore $|R(x)| \le |Q(x)| \le 1$.

From property 5-6 we conclude that $R(x)$ has the same number of good points as $Q(x)$ (if some point outside of $(A' \cap [0,y]) \cup \{y, y+\delta, \ldots, y+(T-1)\delta\}$ is good for R, we still label it as bad). As y is good for R, we also have $\max\{z \in B' : z+\delta \in A'\} < y$. Finally we show that R has a larger maximum.

7. Let $M'' = \max_{x\in[0,1]} |R(x)|$. Then $M'' \ge M'$.
   Proof. As $|1-r'_k| \ge |1-r_k|$, we have $M'' \ge |R(1)| \ge |Q(1)| = M'$.

Next, we repeat the above process until $\{z \in B' : z+\delta \in A'\}$ is empty, and denote the resulting polynomial as $\tilde{R}$, for which property 5-7 still holds. As $\tilde{R}$ is bounded by 1 at $\{0, \delta, 2\delta, \ldots, (48d^2-1)\delta\}$, we obtain

$$M' \le \max_{x\in[0,1]} |\tilde{R}(x)| \le 2^{O(d)} \tag{2.42}$$

by repeating the argument at the beginning of the proof, which concludes the full proof. □

A missing ingredient in the above proof is to show that when a polynomial in bounded on a set of $O(d^2)$ equally spaced points in an interval, then it can be uniformly bounded on that interval. As discussed in Remark 2.3, the number of points needed can be improved to $O(d)$ by using a powerful result of Rakhmanov [31].

Lemma 2.6. *Let P be a degree d polynomial which satisfies $|P(x)| \leq c$ for equally spaced points $x \in \{0, \frac{a}{N}, \frac{2a}{N}, \ldots, a\}$ in the interval $[0,a]$, where $\frac{d^2}{N} \leq 1 - \varepsilon$. Then P is uniformly bounded as $|P(x)| \leq \frac{c}{\varepsilon}, \forall x \in [0,a]$.*

Proof. To prove a uniform bound for $P(x)$, we use a Markov's inequality [33] to bound the maximum derivative,

$$\max_{x\in[0,a]} |P'(x)| \leq \frac{2d^2}{a} \max_{x\in[0,a]} |P(x)|. \tag{2.43}$$

Let $M = \max_{x\in[0,a]} |P(x)|$ which is achieved at $t\frac{a}{N} \leq x \leq (t+1)\frac{a}{N}$. Then

$$\frac{2d^2}{a} M \geq \max_{x\in[0,a]} |P'(x)| \geq (M - c)\frac{2N}{a}, \tag{2.44}$$

which gives $M \leq \frac{c}{\varepsilon}$. □

We are now ready to fill in the final missing part in the proof of Theorem 2.4 which is in Eq. (2.33). A first idea is to directly apply Paturi's Lemma 2.4 which gives the final error bound $\delta e^{O(d\Delta^{-1})}$. However, this is clearly over pessimistic: for degree d polynomials, the error growth should not be much bigger than $(1/\Delta)^d$ when Δ is very small. Based on this intuition, our following result gives an improvement for the error bound of long distance polynomial extrapolation.

Lemma 2.7. *Let P be a degree d polynomial that satisfies $|P(x)| \leq \delta$, $\forall x \in [0, \Delta]$ ($\Delta < 1$). Then we have*

$$|P(1)| \leq \delta e^{d \log \Delta^{-1} + d \log 8}. \tag{2.45}$$

Proof. Let $Q(x) = P\left(\frac{x+1}{2}\Delta\right)$. Then $Q(x)$ is a degree d polynomial that satisfies $|Q(x)| \leq \delta$, $\forall x \in [-1, 1]$. Next we use a well-known fact about polynomials, namely that the coefficients of bounded polynomials are at most exponential in the degree. Let $Q(x) = \sum_{i=0}^{d} a_i x^i$. Lemma 2.8 implies that

$$\sum_{i=0}^{d} |a_i| \leq 4^d \delta. \tag{2.46}$$

Then we have

$$\begin{aligned} |P(1)| &= \left|Q\left(\frac{2}{\Delta}-1\right)\right| \\ &\le \sum_{i=0}^{d} |a_i| \left(\frac{2}{\Delta}-1\right)^i \\ &\le \sum_{i=0}^{d} |a_i| \left(\frac{2}{\Delta}\right)^d \\ &\le \delta \frac{8^d}{\Delta^d} = \delta e^{d\log\Delta^{-1}+d\log 8}. \end{aligned} \tag{2.47}$$

□

Lemma 2.8 (Lemma 4.1 in [34]). *Let $P(x) = \sum_{i=0}^{d} a_i x^i$ be a polynomial. Then*

$$\sum_{i=0}^{d} |a_i| \le 4^d \max_{x\in[-1,1]} |P(x)|. \tag{2.48}$$

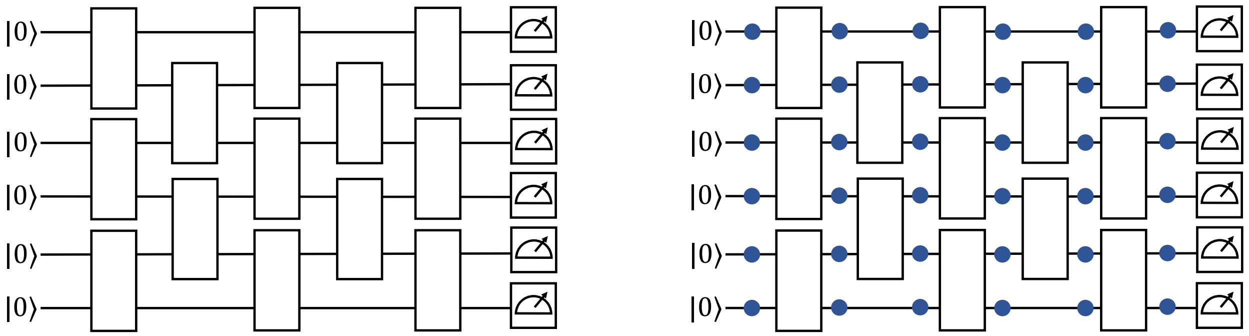
2.3 Complexity of noisy random circuit sampling

In this section we study the classical complexity of RCS in the presence of a constant rate of noise per gate. Specifically we consider a simple noise model shown in Fig. 2.2 (b) where a (arbitrarily small) constant amount of depolarizing noise is applied to each qubit at each time step, which is a theoretical model for the actual RCS experiments. Our main result shows that sampling from the output distribution of a noisy random circuit can be approximately simulated by an efficient classical algorithm within small total variation distance.

Theorem 2.5 (Main result). *Assuming anti-concentration, there is a classical algorithm that, on input a random circuit C on any fixed architecture, outputs a sample from a distribution that is ε-close to the noisy output distribution $\tilde{p}(C)$ in total variation distance with success probability at least $1-\delta$ over the choice of C, in time* $\mathrm{poly}(n, 1/\varepsilon, 1/\delta)$.

To put this in perspective, consider a RCS quantum supremacy experiment that collects M samples. We claim that Theorem 2.5 implies that there is a classical algorithm running in time bounded by polynomial in M, that outputs M samples that are indistinguishable, i.e. no statistical test can distinguish the output of the algorithm from the output of the experiment with probability greater than $1/2+\mu$, for any constant $\mu > 0$. This is because to achieve statistical indistinguishability it suffices to choose $\varepsilon = \mu/M$, which by the main result above gives a running time $\mathrm{poly}(n, M/\mu)$. Thus the running time of our algorithm is at most a polynomial in the running time of the experiment.

Corollary 2.3. *Assuming anti-concentration, no statistical test applied to M samples can distinguish between the output of a noisy random circuit and the above classical algorithm*



(a) Ideal RCS

(b) Noisy RCS

Figure 2.2: Random circuit sampling, each white box is an independent Haar random 2-qubit gate. (a) Ideal RCS generates an output distribution $p(C)$ that satisfies anti-concentration when $d = \Omega(\log n)$. (b) Noisy RCS, where an arbitrarily small constant amount of depolarizing noise is applied to each qubit at each step, which generates a noisy output distribution $\tilde{p}(C)$. Here the 1D architecture is for illustration; the result applies to general architectures (Definition 2.6).

with running time $\mathrm{poly}(n, M)$*. In particular, if* $M = \mathrm{poly}(n)$*, the classical algorithm runs in* $\mathrm{poly}(n)$ *time.*

We note that the implications of our result are complexity theoretic and do not directly address the soundness of finite-size quantum supremacy experiments.

Also note that anti-concentration is a central assumption for both the RCS experiments and our algorithm, which is believed to hold for general architectures above $\Omega(\log n)$ depth [35]. At the same time, the output distribution of noisy random circuits is $2^{-\Theta(d)}$ close to uniform in total variation distance [36, 37, 38]. This means that any quantum supremacy experiment must collect $M = 2^{\Omega(d)}$ samples. Thus $d = \Theta(\log n)$ was recognized as the sweet spot for scalable experimental demonstration of quantum computational advantage [38], depth $O(\log n)$ to guarantee polynomial number of samples and $\Omega(\log n)$ to guarantee anti-concentration. In this regime both the sample complexity of the experiment and running time of our classical algorithm scale polynomially in n.

Our approach builds upon the work of Gao and Duan in 2018 [37]. They developed the idea of performing a Fourier transform on quantum circuits and an algorithm for simulating noisy random circuits via a truncation in Fourier domain and calculating low-degree Fourier coefficients. They used the resulting algorithm to efficiently estimate local observables for random analogs of fault-tolerance circuits, thus showing that structure is necessary for quantum fault-tolerance. While not explicitly mentioned in [37], their approach in fact produces a quasi-polynomial time algorithm for sampling from the output distribution to within inverse polynomial total variation distance[3]. This raises the challenge of giving a polynomial time

[3] This fact was unknown at the time, as it was believed that anti-concentration requires large circuit

algorithm for the sampling problem.

We start by reformulating the Fourier transform defined by [37] as Feynman path integral in the Pauli basis, and the simulation algorithm as calculating those Feynman paths with lowest Hamming weight. The Pauli basis framework was also used by [18] to give an alternative argument for achieving a $2^{-O(d)}$ XEB (see Section 2.3.7 for a more formal treatment). The advantage of using the Pauli basis for Feynman path integral is that most low-Hamming-weight Feynman paths have 0 contribution to the path integral. This view helps design an enumeration algorithm that calculates the contributions of only non-trivial paths in polynomial time. From the perspective of Fourier analysis, prior algorithms of [40, 37] based on low-degree Fourier approximation mainly rely on noise sensitivity and have running time $n^{O(\log 1/\varepsilon)}$ where ε is the desired approximation error which results in quasi-polynomial running time for our purpose, but our algorithm has running time $2^{O(\log 1/\varepsilon)} = \mathrm{poly}(1/\varepsilon)$ due to the additional property of Fourier sparsity.

Our algorithm is not practical in its current form due to a large exponent in the running time, and we leave as an interesting future direction to develop practical implementations using our framework. See Section 2.3.3 for discussions regarding finite-size noisy RCS experiments.

2.3.1 Description of algorithm

Let ρ be an n qubit density matrix. We can write $\rho = \sum_{s\in\mathsf{P}_n} \alpha_s \cdot s$ where P_n are the normalized n-qubit Pauli operators, and $\alpha_s = \mathrm{Tr}(s\rho)$ is real. We keep track of the coefficients in the Pauli basis after unitary evolution $\rho \mapsto U\rho U^\dagger$, which evolve according to the rule $\mathrm{Tr}\big(sU\rho U^\dagger\big) = \sum_{t\in\mathsf{P}_n} \mathrm{Tr}\big(sUtU^\dagger\big)\,\mathrm{Tr}(t\rho)$. Comparing with the transition rule $\langle x|U|\psi\rangle = \sum_y \langle x|U|y\rangle\langle y|\psi\rangle$ we can see that while $\langle x|U|y\rangle$ is the transition amplitude from $|y\rangle$ to $|x\rangle$, $\mathrm{Tr}\big(sUtU^\dagger\big)$ plays the role of transition amplitude from t to s.

Consider a quantum circuit $C = U_d U_{d-1} \cdots U_1$ where U_i is a layer of 2-qubit gates and d is circuit depth. A **Pauli path** is a sequence $s = (s_0, \dots, s_d) \in \mathsf{P}_n^{d+1}$. The Feynman path integral in the Pauli basis (in short, Pauli path integral) is written as sum of product of transition amplitudes,

$$p(C,x) = \sum_{s_0,\dots,s_d\in\mathsf{P}_n} \mathrm{Tr}(|x\rangle\langle x|\, s_d)\,\mathrm{Tr}\Big(s_d U_d s_{d-1} U_d^\dagger\Big) \cdots \mathrm{Tr}\Big(s_1 U_1 s_0 U_1^\dagger\Big)\,\mathrm{Tr}(s_0\,|0^n\rangle\langle 0^n|). \quad (2.49)$$

Note that LHS is the probability $p(C,x) = |\langle x|C|0^n\rangle|^2$ instead of amplitude. Denote the contribution of a Pauli path $s = (s_0, \dots, s_d) \in \mathsf{P}_n^{d+1}$ to the path integral as $f(C,s,x)$, which gives $p(C,x) = \sum_{s\in\mathsf{P}_n^{d+1}} f(C,s,x)$.

Our algorithm for simulating noisy random circuits is based on a simple but powerful fact, used in [41, 37]. Consider the single-qubit depolarizing noise with strength γ, $\mathcal{E}(\rho) := (1-\gamma)\rho + \gamma\frac{I}{2}\mathrm{Tr}(\rho)$. Then the contribution of a Pauli path of a noisy quantum circuit subject

depth. Recent developments [39, 35] suggest otherwise.

to this noise, decays exponentially with the Hamming weight of the Pauli path:

$$\tilde{p}(C,x) = \sum_{s\in \mathsf{P}_n^{d+1}} (1-\gamma)^{|s|} f(C,s,x), \tag{2.50}$$

where $\tilde{p}(C,x)$ is the output probability of the noisy circuit and $|s|$ is the **Hamming weight** of s (the number of non-identity Pauli in s). We would like to approximate the value $\tilde{p}(C,x)$ by summing only over the low-weight Pauli paths,

$$\tilde{p}(C,x) \approx \sum_{s\in \mathsf{P}_n^{d+1}:|s|\le \ell} (1-\gamma)^{|s|} f(C,s,x), \tag{2.51}$$

and claim that the total variation distance achieved by the approximation is $2^{-\Omega(\ell)}$ on average. This is not immediate since the $f(C,s,x)$ can be both positive and negative. We invoke two properties of random circuits: the first is **orthogonality**, which says that on average over random circuits the product of the contributions from two different Pauli paths equals 0, i.e. $\mathbb{E}_C[f(C,s,x)f(C,s',x)] = 0$ when $s \neq s'$; the second is **anti-concentration**, which says that the sum of squares of the output probability of a random circuit is small, i.e. $\mathbb{E}_C \sum_x p(C,x)^2 = O(1)\cdot 2^{-n}$. Roughly speaking, orthogonality allows us to upper bound the total variation distance by a sum of squares quantity, which is then upper bounded using anti-concentration.

The next step is to develop an algorithm to calculate the RHS of Eq. (2.51). Note that a straightforward sum over all paths up to weight ℓ gives a running time of $O(nd)^{O(\ell)}$ leading to a quasi-polynomial time algorithm as in [37]. Here we develop a **counting argument** and efficient enumeration method for all Pauli paths of weight at most ℓ which takes only $2^{O(\ell)}$ time. The idea is **sparsity** of the low-weight paths, meaning that for most Pauli paths in P_n^{d+1}, its contribution $f(C,s,x)$ is 0; therefore we design a combinatorial algorithm that only enumerates those paths that have non-zero contributions. Finally, the sampling algorithm follows from a general sampling-to-computing reduction of [40].

At a high level, the hardness assumptions in [19, 9] may be intuitively viewed as asserting that Feynman path integral in the computational basis is essentially the best classical algorithm for RCS, and achieving non-trivial correlation requires following exponentially many paths. Instead, the Pauli path integral approach has the virtue that low weight paths have the most significant contribution.

2.3.2 Prior work regarding the computational complexity of RCS

To put the above results in context, let us recall the background regarding complexity theoretic evidence that classical computers cannot efficiently sample from the output of a random quantum circuit (this section focuses on asymptotic hardness; see next section for discussions regarding finite-size experiments). There are two main genres of results along those lines, which we review below (see [8] for a more comprehensive survey).

The first is in the form of a worst-case to average-case reduction, showing that if an efficient classical algorithm can sample from the output distribution of ideal RCS within small total variation distance, then the Polynomial Hierarchy collapses [26, 27, 4, 42, 43]. The eventual goal of this program was to show classical hardness for sampling within constant total variation distance, which would require showing average-case hardness of computing the output probability of ideal RCS within additive error $O(2^{-n})$. While the earliest average-case hardness results could only tolerate very small additive error, it was hoped that over time the reductions could be made more robust. This has indeed been the case, with an improvement from a large polynomial in the exponent [26] to $2^{-O(m)}$ [43], but this line of work has hit an obstacle that may prove difficult to overcome (see e.g. [4, Section 3] and [38, Section II A]). Moreover, these results do not address the actual RCS experiments which are highly noisy[4].

The second genre is based on complexity theoretic assumptions about the difficulty of distinguishing heavy and light outputs of the random circuit [19, 9]. These assumptions essentially say that even a tiny correlation (order 2^{-n}) with the output distribution of ideal RCS is hard to achieve classically. While these assumptions are quite strong, they have the virtue of yielding robust bounds. Indeed a specific conjecture in this genre called XQUATH [9] has provided robust complexity theoretic foundation of the linear cross entropy benchmark (XEB) used in recent experiments [7, 16, 17]. This provided a way to heuristically argue that even the very small XEB achieved in actual 50-70 qubit experiments was a classically difficult computational task. However, the strong parameters in the assumption (correlation of order 2^{-n}) was called into question by the result of [18], although it remained unclear if the hardness of the XEB test can be restored by changing the parameters in XQUATH. In addition, it was unclear whether the hardness of the other statistical tests such as HOG or log XEB was impacted. Our results address these questions by showing that no statistical tests, like the XEB, HOG and log XEB, can distinguish between noisy RCS and our classical algorithm.

2.3.3 Concluding remarks: what our results do not address

We importantly note several points left unaddressed by our results.

- **Practical speed-ups.** We note that our results do not address RCS based quantum supremacy in its non-asymptotic, practical form. In particular, much progress has been made in developing practical spoofing algorithms for achieving a similar numerical value as the XEB in current 53-60 qubit RCS experiments. Practical tensor network algorithms [10, 11, 12, 13, 14, 15] can achieve this goal using hundreds of GPUs in a few hours, but these algorithms have exponential scaling and become impractical if the system size increases by a few qubits. A numerical implementation of the algorithm in [18] achieved roughly 10% of Google's XEB using 1 GPU in 1 second, though it

[4] The related work of [22] argued that the hardness of approximate sampling for noisy RCS can be reduced to ideal RCS, but the argument required a local noise model that decreases as $\widetilde{O}(1/n)$, which is not scalable.

remains unclear whether this algorithm can achieve Google's XEB (using much less than hundreds of GPUs). Our algorithm is not practical in its current form, as there is a large constant (of order $1/\gamma$ where γ is the error per gate) in the degree of the polynomial of the running time. An interesting future direction is to develop practical implementations using our framework and ideas from [18] that achieves similar XEB as in the experiments [7, 16, 17] using a small amount of resource.

- **Sublogarithmic depth.** Our algorithm assumes anti-concentration and therefore works for random circuits with depth at least $\Omega(\log n)$.[5] The issue with sub-logarithmic depth random circuits (with Haar random 2-qubit gates) is that there is no evidence for hardness of sampling even for ideal RCS, as all existing results for average-case hardness (the first genre discussed above) are only relevant for sampling when anti-concentration holds. In addition, [44] gives evidence that 2D ideal RCS can be efficiently simulated when depth is smaller than some fixed constant. The complexity of ideal and noisy RCS remains unclear at depth between constant and $o(\log n)$. Separately, existing quantum supremacy experiments rely on the assumption that the ideal circuit is close to Porter-Thomas for benchmarking; closeness to Porter-Thomas is even stronger than anti-concentration.

  Notwithstanding the above discussion, it remains possible that a different approach based on RCS of sublogarithmic depth circuits, which does not rely on anti-concentration, could lead to a scalable experimental violation of the extended Church-Turing thesis.

- **Less random gate sets.** Besides anti-concentration, our algorithm also requires randomness in the gate set. The simplest distribution over the gate set to think of is that of Haar random 2-qubit gates. However, the gate sets used in actual experiments [7, 16, 17] are not Haar random 2-qubit gates, but gates with more limited randomness. While we do not know if our results hold for the exact gate sets used in those recent experiments, we show in Section 2.3.6 that our algorithm works for a gate set which is closely related to the gate sets used in those experiments; more generally, the required condition for our results is in fact much weaker than Haar random 2-qubit gates (see Definition 2.6).

Overview of remainder of this section. In Section 2.3.4 we give formal definitions of the Pauli path integral and derive useful properties of this framework. In Section 2.3.5 we give the proof of our main result, and discuss Google and USTC's gate set in Section 2.3.6. Section 2.3.7 contains a formal proof for refuting XQUATH using the Pauli basis framework. As an application of the Pauli basis framework, we provide simple proofs for existing results about random circuits, including a lower bound on the depth for anti-concentration previously shown by [35] (Corollary 2.4), and an improved lower bound on the convergence to uniform for noisy random circuits previously shown by [38] (Section 2.3.8).

[5] It was shown [35, 38] that anti-concentration requires at least $\Omega(\log n)$ depth for random circuits with Haar random 2-qubit gates.

	(a) Vector basis	(b) Operator basis	(c) Operator basis
State	$\lvert\psi\rangle = \sum_{x\in\{0,1\}^n} \langle x\vert\psi\rangle \lvert x\rangle$	$\rho = \sum_{s\in\mathsf{P}_n} \mathrm{Tr}(s\rho)s$	$\lvert\rho\rangle\rangle = \sum_{s\in\mathsf{P}_n} \langle\langle s\vert\rho\rangle\rangle \lvert s\rangle\rangle$
Evolution	$\lvert\psi\rangle \mapsto U\lvert\psi\rangle$	$\rho \mapsto U\rho U^\dagger$	$\lvert\rho\rangle\rangle \mapsto \mathcal{U}\lvert\rho\rangle\rangle$
Path integral	$\langle x\vert U\vert\psi\rangle$ $= \sum_{y\in\{0,1\}^n} \langle x\vert U\vert y\rangle \langle y\vert\psi\rangle$	$\mathrm{Tr}\big(sU\rho U^\dagger\big)$ $= \sum_{t\in\mathsf{P}_n} \mathrm{Tr}\big(sUtU^\dagger\big)\,\mathrm{Tr}(t\rho)$	$\langle\langle s\vert\mathcal{U}\vert\rho\rangle\rangle$ $= \sum_{t\in\mathsf{P}_n} \langle\langle s\vert\mathcal{U}\vert t\rangle\rangle\langle\langle t\vert\rho\rangle\rangle$

Table 2.1: The Feynman path integral can be viewed as decomposing the state into basis states at each step of time evolution. (a) The standard decomposition with computational basis states. (b) Decomposition using the Pauli operator basis, where states are represented as density matrices and time evolution is represented as unitary channels. (c) The same decomposition using the Pauli operator basis, presented with operator ket notation.

2.3.4 The Pauli basis framework

We first give formal definitions of the Pauli path integral discussed in Section 2.3.1 and then derive useful properties of this framework.

Let $C = U_d U_{d-1} \cdots U_1$ be a quantum circuit acting on n qubits, where U_i is a layer of 2-qubit gates and d is circuit depth. The Feynman path integral in the computational basis is written as

$$\langle 0^n\vert C\vert 0^n\rangle = \sum_{x_1,\ldots,x_{d-1}\in\{0,1\}^n} \langle 0^n\vert U_d\vert x_{d-1}\rangle\, \langle x_{d-1}\vert U_{d-1}\vert x_{d-2}\rangle \cdots \langle x_1\vert U_1\vert 0^n\rangle\,. \tag{2.52}$$

The main difference when switching to the Pauli basis is that instead of thinking about a quantum circuit as applying unitary matrices to vectors, we think of it as unitary channels applied to density matrices, $\mathcal{C} = \mathcal{U}_d\mathcal{U}_{d-1}\cdots\mathcal{U}_1$ where each $\mathcal{U}_i(\cdot) := U_i(\cdot)U_i^\dagger$ is a unitary channel. Similar to decomposing a pure state vector into a superposition of computational basis states, we consider the normalized Pauli operators

$$\mathsf{P}_n := \left\{I/\sqrt{2}, X/\sqrt{2}, Y/\sqrt{2}, Z/\sqrt{2}\right\}^{\otimes n} \tag{2.53}$$

as an operator basis and decompose a density matrix into a linear combination of Pauli operators (Table 2.1). In Table 2.1 we present the operator basis as a direct analogy of vector basis by switching to the operator ket notation (Table 2.1 (c)).

Definition 2.4 (Pauli path integral). *Let $C = U_d U_{d-1} \cdots U_1$ be a quantum circuit acting on n qubits, where U_i is a layer of 2-qubit gates and d is circuit depth, and let $p(C,x) :=$*

$|\langle x|C|0^n\rangle|^2$ be the output probability distribution. The Pauli path integral is written as

$$\begin{aligned} p(C,x) &= \sum_{s_0,\dots,s_d\in\mathsf{P}_n} \mathrm{Tr}(|x\rangle\langle x|\, s_d)\,\mathrm{Tr}\Big(s_d U_d s_{d-1} U_d^\dagger\Big)\cdots\mathrm{Tr}\Big(s_1 U_1 s_0 U_1^\dagger\Big)\,\mathrm{Tr}(s_0\,|0^n\rangle\langle 0^n|) \\ &= \sum_{s_0,\dots,s_d\in\mathsf{P}_n} \langle\!\langle x|s_d\rangle\!\rangle\langle\!\langle s_d|\mathcal{U}_d|s_{d-1}\rangle\!\rangle\cdots\langle\!\langle s_1|\mathcal{U}_1|s_0\rangle\!\rangle\langle\!\langle s_0|0^n\rangle\!\rangle. \end{aligned} \tag{2.54}$$

*Here each term on RHS corresponds to a **Pauli path** $s=(s_0,\dots,s_d)\in\mathsf{P}_n^{d+1}$. We also define the **Fourier coefficient** of a quantum circuit C with output x and Pauli path s as*

$$f(C,s,x) := \langle\!\langle x|s_d\rangle\!\rangle\langle\!\langle s_d|\mathcal{U}_d|s_{d-1}\rangle\!\rangle\cdots\langle\!\langle s_1|\mathcal{U}_1|s_0\rangle\!\rangle\langle\!\langle s_0|0^n\rangle\!\rangle \tag{2.55}$$

and the output probability is written as

$$p(C,x) = \sum_{s\in\mathsf{P}_n^{d+1}} f(C,s,x). \tag{2.56}$$

Eq. (2.54) follows from repeatedly applying the rules shown in Table 2.1. The above definition can also be extended to noisy quantum circuits. Let $\mathcal{E}(\rho) := (1-\gamma)\rho + \gamma\frac{I}{2}\mathrm{Tr}(\rho)$ be the single-qubit depolarizing noise with strength γ. It has the property that $\mathcal{E}(I) = I$ and $\mathcal{E}(P) = (1-\gamma)P$ when $P\in\{X,Y,Z\}$.

Definition 2.5 (Pauli path integral for noisy quantum circuits)**.** *For a quantum circuit $C = U_d U_{d-1}\cdots U_1$, let $\tilde{C}$ be a noisy quantum circuit where each qubit in C is subject to γ depolarizing noise in each layer (Fig. 2.2 (b)). Let*

$$\tilde{p}(C,x) := \langle\!\langle x|\mathcal{E}^{\otimes n}\mathcal{U}_d\mathcal{E}^{\otimes n}\cdots\mathcal{U}_1\mathcal{E}^{\otimes n}|0^n\rangle\!\rangle \tag{2.57}$$

be the output probability distribution of the noisy circuit $\tilde{C}$. The Pauli path integral for $\tilde{C}$ is defined as

$$\tilde{p}(C,x) = \sum_{s\in\mathsf{P}_n^{d+1}} \tilde{f}(C,s,x) \tag{2.58}$$

where

$$\tilde{f}(C,s,x) := \langle\!\langle x|\mathcal{E}^{\otimes n}|s_d\rangle\!\rangle\langle\!\langle s_d|\mathcal{U}_d\mathcal{E}^{\otimes n}|s_{d-1}\rangle\!\rangle\cdots\langle\!\langle s_1|\mathcal{U}_1\mathcal{E}^{\otimes n}|s_0\rangle\!\rangle\langle\!\langle s_0|0^n\rangle\!\rangle. \tag{2.59}$$

*Let $|s|$ be the **Hamming weight** of s (the number of non-identity Pauli in s). The definition of depolarizing noise implies that*

$$\tilde{f}(C,s,x) = (1-\gamma)^{|s|} f(C,s,x). \tag{2.60}$$

Our algorithm described in Section 2.3.1 is summarized in Algorithm 1 (“legal” Pauli path is defined in Definition 2.9). Next we develop properties of the Pauli basis that are useful later.

Algorithm 1 Simulating noisy random circuits by low-degree Fourier approximation

```
Input: quantum circuit $C$, truncation parameter $\ell$, $x \in \{0,1\}^n$
Output: an approximation of $\tilde{p}(C,x)$
$q \leftarrow 0$
for all legal Pauli path $s$ with $|s| \le \ell$ do
    calculate $f(C,s,x)$
    $q \leftarrow q + (1-\gamma)^{|s|} f(C,s,x)$
end for
Return $q$
```

First, note that the Fourier coefficients $f(C,s,x)$ can be further decomposed into products of transition amplitudes of 2-qubit gates $\langle\!\langle q|\mathcal{U}|p\rangle\!\rangle = \mathrm{Tr}\big(qUpU^\dagger\big)$ where $U \in \mathbb{U}(4)$, $p,q \in \mathsf{P}_2$, so any Fourier coefficient can be computed in time $O(nd)$. The Fourier coefficients satisfy $f(C,s,x) \in \mathbb{R}$ and $|f(C,s,x)| \le \frac{1}{2^n}$. This is because for any $x \in \{0,1\}^n$ and $s \in \mathsf{P}_n$ we have

$$\langle\!\langle x|s\rangle\!\rangle = \mathrm{Tr}(|x\rangle\!\langle x|\, s) \in \left\{0, -\frac{1}{\sqrt{2^n}}, \frac{1}{\sqrt{2^n}}\right\}. \tag{2.61}$$

In addition, the output x only affects the sign of the Fourier coefficient, as

$$f(C,s,x)^2 = f(C,s,0^n)^2, \quad \forall x \in \{0,1\}^n. \tag{2.62}$$

The rest of the properties we develop in this section crucially rely on the randomness of the gate set. We first recall the properties of Haar random 2-qubit gates.

Lemma 2.9 (Properties of Haar random 2-qubit gates [45])**.** *Let $U \in \mathbb{U}(4)$ be a Haar random 2-qubit gate, and $p,q,r,s \in \mathsf{P}_2$. Then*

$$\mathop{\mathbb{E}}_{U\sim\mathbb{U}(4)} \langle\!\langle p|\mathcal{U}|q\rangle\!\rangle \langle\!\langle r|\mathcal{U}|s\rangle\!\rangle = 0 \quad \textit{if } p \neq r \textit{ or } q \neq s. \tag{2.63}$$

We also have

$$\mathop{\mathbb{E}}_{U\sim\mathbb{U}(4)} \langle\!\langle p|\mathcal{U}|q\rangle\!\rangle^2 = \begin{cases} 1, & p = q = I^{\otimes 2}/2, \\ 0, & p = I^{\otimes 2}/2, q \neq I^{\otimes 2}/2, \\ 0, & p \neq I^{\otimes 2}/2, q = I^{\otimes 2}/2, \\ \frac{1}{15}, & \textit{else.} \end{cases} \tag{2.64}$$

Eq. (2.63) is a key property which we refer to as gate-set orthogonality. It says that if we consider the Pauli basis decomposition and average over two copies of a random unitary, then the randomness forces the input and output Paulis to be the same across the two copies. Next we show that this property does not require full randomness over $\mathbb{U}(4)$; randomness over Pauli operators already suffices.

Lemma 2.10 (Gate-set orthogonality)**.** *Let $\mathcal{D}$ be any distribution over $\mathbb{U}(4)$ that is invariant under right-multiplication of random Pauli, i.e. for any measurable function F,*

$$\mathop{\mathbb{E}}_{U\sim\mathcal{D}}[F(U)] = \mathop{\mathbb{E}}_{U\sim\mathcal{D}}\mathop{\mathbb{E}}_{V\sim\{I,X,Y,Z\}^2}[F(UV)]. \tag{2.65}$$

Then for any $P, Q \in \{I, X, Y, Z\}^2$ such that $P \neq Q$, we have

$$\mathop{\mathbb{E}}_{U\sim\mathcal{D}}\left[UPU^\dagger \otimes UQU^\dagger\right] = 0. \tag{2.66}$$

Proof. Due to invariance under right-multiplication of random Pauli and linearity, it suffices to prove that

$$\mathop{\mathbb{E}}_{V\sim\{I,X,Y,Z\}^2}\left[VPV^\dagger \otimes VQV^\dagger\right] = 0 \quad \text{if } P \neq Q. \tag{2.67}$$

Let $\langle P, Q\rangle := 1[P, Q \text{ anticommute}]$. Then

$$\begin{aligned}\mathop{\mathbb{E}}_{V\sim\{I,X,Y,Z\}^2}\left[VPV^\dagger \otimes VQV^\dagger\right] &= \frac{1}{16}\sum_{V\in\{I,X,Y,Z\}^2} VPV^\dagger \otimes VQV^\dagger \\ &= \frac{1}{16}\sum_{V\in\{I,X,Y,Z\}^2} (-1)^{\langle V,P\rangle+\langle V,Q\rangle} P\otimes Q \\ &= \frac{1}{16}\sum_{V\in\{I,X,Y,Z\}^2} (-1)^{\langle V,PQ\rangle} P\otimes Q \\ &= 0,\end{aligned} \tag{2.68}$$

where the last line follows from the fact that PQ is not identity, and therefore commutes with half Paulis and anticommutes with the other half. □

Our main result holds for any gate set and architecture that satisfies gate-set orthogonality and anti-concentration. We discuss these two properties separately and start with orthogonality.

Definition 2.6 (Gate set and architecture of random circuits)**.** *We consider random quantum circuits defined over a fixed architecture described as follows. In each layer, each qubit experiences a 2-qubit gate (so the number of qubits n is even, and there are $n/2$ 2-qubit gates per layer). The 2-qubit gates can be applied to any pair of qubits, without geometric locality. Each 2-qubit gate is independently drawn from some distribution that is invariant under right-multiplication of random Pauli. The final layer is drawn from a distribution that is invariant under both left- and right-multiplication of random Pauli.*

Note that the requirement that each qubit experiences a 2-qubit gate in each layer is for convenience; more general architectures can be handled by a suitable redefinition of circuit depth (this was also noted in [38]).

Examples of gate sets that satisfy Definition 2.6 include Haar random 2-qubit gates as well as a fixed 2-qubit gate surrounded by Haar random single qubit gates. A fixed 2-qubit gate surrounded by random Pauli gates also satisfies Definition 2.6 but may violate anti-concentration (see Remark 2.5). Any ensemble of random circuits that satisfies Definition 2.6 has the following crucial property that we frequently use.

Lemma 2.11 (Orthogonality of Fourier coefficients)**.** *Let C be a random circuit drawn from some distribution $\mathcal{D}$ that satisfies Definition 2.6. Then for any Pauli paths $s \neq s' \in \mathsf{P}_n^{d+1}$ and for any $x \in \{0,1\}^n$ we have*

$$\mathop{\mathbb{E}}_{C\sim\mathcal{D}}[f(C,s,x)f(C,s',x)] = 0. \tag{2.69}$$

Proof. As $s \neq s'$, there exists a 2-qubit gate U that contributes transition amplitude $\langle\langle q_1|\mathcal{U}|p_1\rangle\rangle$ to $f(C,s,x)$ and contributes $\langle\langle q_2|\mathcal{U}|p_2\rangle\rangle$ to $f(C,s',x)$, such that $p_1 \neq p_2 \in \mathsf{P}_2$. Lemma 2.10 implies that

$$\mathop{\mathbb{E}}_{U}[\langle\langle q_1|\mathcal{U}|p_1\rangle\rangle\langle\langle q_2|\mathcal{U}|p_2\rangle\rangle] = 0. \tag{2.70}$$

Due to the independence between different gates, we can separately calculate the expectation over each gate in Eq. (2.69). Therefore the above equation implies that the overall expectation in Eq. (2.69) equals 0. One special case is that the difference between s and s' happens at the last step s_d. For this case we use the left-invariance under random Pauli of the final layer of gates. □

Next we discuss anti-concentration, which is formally defined as follows.

Definition 2.7 (Anti-concentration)**.** *A distribution over quantum circuits $\mathcal{D}$ satisfies anti-concentration if*

$$\mathop{\mathbb{E}}_{C\sim\mathcal{D}} 2^n \sum_{x\in\{0,1\}^n} p(C,x)^2 = O(1). \tag{2.71}$$

Remark 2.4. *The following is known about anti-concentration:*

- *[39, 35] showed that anti-concentration is satisfied for 1D random circuits with Haar random 2-qubit gates as long as circuit depth is above some constant times $\log n$.*
- *[35] also showed that $\Theta(n\log n)$ 2-qubit gates are necessary and sufficient for anti-concentration for a stochastic all-to-all connected architecture with Haar random 2-qubit gates.*
- *[35, 38] showed that at least $\Omega(\log n)$ depth is necessary for anti-concentration, for any architecture with Haar random 2-qubit gates. We also give a simple proof of this fact using the Pauli basis framework in Corollary 2.4.*

- *[35] remarked that, as anti-concentration is proven for two architectures which are two opposite extremes of geometric locality, they conjecture $\Theta(n \log n)$ size (which is $\Theta(\log n)$ depth in our case) to be necessary and sufficient for anti-concentration for any reasonably well-connected architecture.*

Remark 2.5. *The results discussed in Remark 2.4 only concern Haar random 2-qubit gates. We expect the same results to hold for a fixed 2-qubit gate surrounded by Haar random single qubit gates. It is worth mentioning that while a fixed 2-qubit gate surrounded by random Pauli gates satisfies Definition 2.6, we do not expect it to satisfy anti-concentration, due to the fact that it does not generate the entire Clifford group when, for example, the 2-qubit gate is a CNOT gate.*

The reason for requiring anti-concentration for our results is because it is closely related to the Fourier weights of random circuits, which is then related to the error of the simulation algorithm.

Definition 2.8 (Fourier weight)**.** *The **Fourier weight** of a random circuit C at degree k is defined as*

$$W_k = 2^{2n} \mathop{\mathbb{E}}_{C} \sum_{s \in \mathsf{P}_n^{d+1} : |s| = k} f(C, s, 0^n)^2. \tag{2.72}$$

Here the 2^{2n} factor is a normalization factor that comes from Eq. (2.61). A crucial property for our arguments is that anti-concentration implies that the total Fourier weight is upper bounded by a constant.

Lemma 2.12 (Total Fourier weight)**.** *Let $\mathcal{D}$ be a distribution over quantum circuits that satisfies anti-concentration and Definition 2.6. The Fourier weights $\{W_k\}$ satisfy*

1. $W_0 = 1$,

2. $W_k = 0$, $\forall 0 < k \leq d$,

3. $\sum_{k \geq d+1} W_k = O(1)$.

Proof. $W_0 = 1$ corresponds to the unique all-identity path. Let s be a Pauli path of Hamming weight $k = |s| \in (0, d]$. Then there exists a 2-qubit gate U that contributes a transition amplitude $\langle\langle q|\mathcal{U}|p\rangle\rangle$ to $f(C, s, 0^n)$, where either p is identity and q is non-identity, or vice versa. In either case we have $\langle\langle q|\mathcal{U}|p\rangle\rangle = 0$. This implies that $W_k = 0$.

To bound the total weight, we start with anti-concentration.

$$\begin{aligned} O(1) &= \mathop{\mathbb{E}}_{C\sim\mathcal{D}} 2^n \sum_{x\in\{0,1\}^n} p(C,x)^2 \\ &= \mathop{\mathbb{E}}_{C\sim\mathcal{D}} 2^n \sum_{x\in\{0,1\}^n} \left(\sum_{s\in\mathsf{P}_n^{d+1}} f(C,s,x)\right)^2 \\ &= \mathop{\mathbb{E}}_{C\sim\mathcal{D}} 2^n \sum_{x\in\{0,1\}^n} \sum_{s,s'\in\mathsf{P}_n^{d+1}} f(C,s,x)f(C,s',x) \\ &= \mathop{\mathbb{E}}_{C\sim\mathcal{D}} 2^n \sum_{x\in\{0,1\}^n} \sum_{s\in\mathsf{P}_n^{d+1}} f(C,s,x)^2 \\ &= 2^{2n} \mathop{\mathbb{E}}_{C\sim\mathcal{D}} \sum_{s\in\mathsf{P}_n^{d+1}} f(C,s,0^n)^2 \\ &= 2^{2n} \mathop{\mathbb{E}}_{C\sim\mathcal{D}} \sum_{k\geq 0} \sum_{s\in\mathsf{P}_n^{d+1}:|s|=k} f(C,s,0^n)^2 \\ &= 1 + \sum_{k\geq d+1} W_k. \end{aligned} \tag{2.73}$$

Here, the first line follows from anti-concentration; the second line follows from the Pauli path integral; the fourth line follows from orthogonality (Lemma 2.11); the fifth line follows from Eq. (2.62). □

Finally we give a detailed clarification regarding the assumptions we make about the architecture and gate set for our main result.

Remark 2.6. *For our main result Theorem 2.5, we assume Definition 2.6 and anti-concentration as defined in Definition 2.7.*

- *If the gate set is Haar random 2-qubit gates, no further assumption is needed.*
- *If not, then we further assume that the circuit depth is at least $\Omega(\log n)$. This is because our algorithm requires $\Omega(\log n)$ depth to be efficient, and we cannot rule out the possibility that there is an ensemble of random circuits below log depth that satisfies both Definition 2.6 and 2.7.*

2.3.5 Simulating noisy random circuit sampling

Given a random circuit C and an output x, let $p(C,x) = |\langle x|C|0^n\rangle|^2$ be the ideal output distribution and let $\tilde{p}(C,x)$ be the output distribution of the noisy circuit where C is subject to local depolarizing noise of rate γ. This section shows the following:

Theorem 2.6 (Restatement of Theorem 2.5)**.** *Let $\mathcal{D}$ be a distribution over quantum circuits that satisfies anti-concentration and Definition 2.6 (also see Remark 2.6). There is a classical algorithm that, on input $C \sim \mathcal{D}$, outputs a sample from a distribution that is ε-close to $\tilde{p}(C,x)$ in total variation distance with success probability at least $1-\delta$ over the choice of C, in time* $\mathrm{poly}(n, 1/\varepsilon, 1/\delta)$.

Our goal is to compute a function $\bar{q}(C,x)$ that achieves small L_1 distance

$$\Delta := \|\tilde{p} - \bar{q}\|_1 := \sum_{x\in\{0,1\}^n} |\tilde{p}(C,x) - \bar{q}(C,x)| \tag{2.74}$$

with high probability. Here $\{\bar{q}(C,x)\}_x$ is not necessarily a distribution, and $\bar{q}(C,x)$ is not necessarily positive (the bar notation indicates that $\bar{q}$ is a quasi-probability distribution). The main result is derived in three steps:

1. We use a general sampling-to-computing reduction shown by [40] which says that given the ability to compute $\bar{q}(C,x)$ as well as its marginals, we can sample from a distribution that is $O(\Delta)$-close to $\tilde{p}(C,x)$ with a polynomial overhead. This is discussed in Section 2.3.5.3. It remains to develop an efficient algorithm to compute $\bar{q}(C,x)$ and its marginals.

2. The algorithm is to approximate $\tilde{p}(C,x)$ by summing its low-degree Fourier coefficients, defined as
$$\bar{q}(C,x) := \sum_{s:|s|\le\ell} \tilde{f}(C,s,x) = \sum_{s:|s|\le\ell} (1-\gamma)^{|s|} f(C,s,x), \tag{2.75}$$
where ℓ is to be determined. In Section 2.3.5.1 we upper bound the total variation distance Δ achieved by this approximation. It shows that choosing $\ell = O(\log 1/\varepsilon)$ suffices to achieve ε total variation distance.

3. It remains to bound the running time of the algorithm. In Section 2.3.5.2 which is the main technical part, we show that each $\bar{q}(C,x)$ can be computed in time $2^{O(\ell)}$. This completes the argument.

2.3.5.1 Bounds for the total variation distance

We show that the expected total variation distance square is upper bounded by an exponential decay of the Fourier weights.

$$\begin{aligned}
\mathop{\mathbb{E}}_C\left[\Delta^2\right] &\leq 2^n \mathop{\mathbb{E}}_C \sum_{x\in\{0,1\}^n} \left(\tilde{p}(C,x)-\bar{q}(C,x)\right)^2 \\
&= 2^n \mathop{\mathbb{E}}_C \sum_{x\in\{0,1\}^n} \left(\sum_{s:|s|>\ell}(1-\gamma)^{|s|} f(C,s,x)\right)^2 \\
&= 2^n \mathop{\mathbb{E}}_C \sum_{x\in\{0,1\}^n} \sum_{s:|s|>\ell}(1-\gamma)^{2|s|} f(C,s,x)^2 \\
&= 2^{2n} \mathop{\mathbb{E}}_C \sum_{s:|s|>\ell}(1-\gamma)^{2|s|} f(C,s,0^n)^2 \\
&= \sum_{k>\ell}(1-\gamma)^{2k} W_k .
\end{aligned} \tag{2.76}$$

Here, the first line follows from Cauchy–Schwarz; the second line is by definition of $\bar{q}$; the third line follows from orthogonality (Lemma 2.11); the fourth line follows from Eq. (2.62); the fifth line is by definition of Fourier weight.

A simple upper bound can be derived assuming anti-concentration (using item 3 from Lemma 2.12),

$$\mathop{\mathbb{E}}_C\left[\Delta^2\right] \leq \sum_{k>\ell}(1-\gamma)^{2k} W_k \leq \sum_{k>\ell}(1-\gamma)^{2\ell} W_k \leq O(1)\cdot e^{-2\gamma\ell}. \tag{2.77}$$

By choosing $\ell = O(\log 1/\varepsilon)$ (roughly $\ell \approx \frac{1}{\gamma}\cdot\log 1/\varepsilon$) we can guarantee that $\Delta \leq \varepsilon$ with high probability.

2.3.5.2 Counting and enumerating legal Pauli paths

For a given truncation parameter ℓ, the running time of the algorithm depends on the number of Pauli paths with Hamming weight at most ℓ, as well as the efficiency for finding and enumerating these paths. A simple argument for bounding the number of paths is as follows. There are $n(d+1)$ locations in the circuit to insert Pauli paths. The total number of ways to insert ℓ non-identity Pauli into the Pauli path is at most $\binom{n(d+1)}{\ell}$, and the choice of X, Y, Z for each non-identity gives a 3^ℓ factor. Therefore the total number of paths with Hamming weight at most ℓ is at most

$$\ell \cdot \binom{n(d+1)}{\ell}\cdot 3^\ell \leq (nd)^{O(\ell)}. \tag{2.78}$$

In this section we show that this bound is a significant overestimate and can be improved to $2^{O(\ell)}$. The key point here is that only the “legal” paths matters, and therefore we design an algorithm that only counts and enumerates legal paths.

Definition 2.9 (Legal Pauli path)**.** *For a given circuit architecture, a Pauli path $s = (s_0, s_1, \ldots, s_d)$ is legal if the following two conditions are satisfied:*

1. *For all 2-qubit gates in the circuit, its input and output Paulis are either both* `II`*, or both not* `II`*.*

2. s_0 *and* s_d *contains only* `I` *and* `Z`*.*

The reason for considering legal Pauli paths is that the illegal ones are irrelevant, as they contribute 0 to the Pauli path integral.

Lemma 2.13. *Any illegal Pauli path s gives $f(C, s, x) = 0$ for any C and x.*

Proof. Let s be an illegal Pauli path. Then there are two cases: either the first or the second condition of Definition 2.9 is violated. If the second condition is violated, then $f(C, s, x) = 0$ because the inner product between computational basis states with s_0 or s_d equals 0, due to the fact that

$$\langle\langle x|s\rangle\rangle = \mathrm{Tr}(|x\rangle\langle x| \cdot s) = 0, \quad \forall x \in \{0,1\}^n, s \notin \{I/\sqrt{2}, Z/\sqrt{2}\}^{\otimes n}. \tag{2.79}$$

If the first condition is violated, then there is a 2-qubit gate U whose input Pauli is `II` and the output is not `II`, or vice versa. Then $f(C, s, x) = 0$ because the transition amplitude contributed by U equals 0 due to the fact that unitary channel is trace preserving, i.e.

$$\langle\langle p|\mathcal{U}|q\rangle\rangle = \mathrm{Tr}\big(pUqU^\dagger\big) = 0 \quad \begin{array}{l}\text{if } p = I \otimes I/2, q \neq I \otimes I/2, \\ \text{or } p \neq I \otimes I/2, q = I \otimes I/2.\end{array} \tag{2.80}$$

□

Next we develop arguments to count legal paths. The number of legal Pauli paths up to a given Hamming weight is a combinatorial property that only depends on the circuit architecture, independent of the gate set.

We first give a simple example that counts the number of legal paths with weight $d + 1$. Lemma 2.12 says that $d+1$ is the smallest non-zero Hamming weight with legal paths. The result below is interesting by itself, as we will show later that this result gives a simple lower bound on the depth for anti-concentration (Corollary 2.4).

Lemma 2.14. *The number of legal Pauli paths with Hamming weight $d+1$ equals $n \cdot 2^d \cdot 3^{d-1}$.*

Proof. As the Pauli path $s = (s_0, s_1, \ldots, s_d)$ has Hamming weight $d+1$, it has to be the case that $|s_i| = 1$ for $i = 0, \ldots, d$. We first choose the location of the non-identity in the first layer s_0, which has n choices. Suppose this non-identity Pauli is at the input of some 2-qubit gate U. Then the output of U can be either `IR` or `RI` (We use `R` to represent a non-identity), which gives two choices. Repeating this argument for each layer, we know that the number of configurations of locations of non-identities is $n \cdot 2^d$. Finally, the 3^{d-1} factor comes from the fact that the non-identity Pauli at the first and last layer has to be Z, while each of the other $d - 1$ layers has three choices among X, Y, Z. □

Next we show that anti-concentration implies the desired $2^{O(\ell)}$ upper bound for the number of legal paths. This bound is clearly tight up to the constant in the exponent, as even the choice of X, Y, Z for a single path of weight ℓ gives a 3^{ℓ} factor. The problem with the result below is that it does not give an algorithm to find and enumerate the legal paths. This is addressed later.

Lemma 2.15. *Consider any circuit architecture which satisfies anti-concentration with Haar random 2-qubit gates. For any $\ell \geq d+1$, the total number of legal Pauli paths with Hamming weight at most ℓ is upper bounded by $2^{O(\ell)}$.*

Proof. We have shown in Lemma 2.12 that anti-concentration implies that $\sum_{k \geq d+1} W_k = O(1)$. Below we give a lower bound on the Fourier weight up to degree ℓ. Consider any legal Pauli path s with Hamming weight at most ℓ. We will calculate its contribution to the Fourier weight $2^{2n}\,\mathbb{E}_C\, f(C, s, 0^n)^2$ as follows.

$$\begin{aligned} 2^{2n} \mathop{\mathbb{E}}_{C} \left[f(C, s, 0^n)^2 \right] &= 2^{2n} \mathop{\mathbb{E}}_{C} \left[(\langle\!\langle x | s_d \rangle\!\rangle \langle\!\langle s_d | \mathcal{U}_d | s_{d-1} \rangle\!\rangle \cdots \langle\!\langle s_1 | \mathcal{U}_1 | s_0 \rangle\!\rangle \langle\!\langle s_0 | 0^n \rangle\!\rangle)^2 \right] \\ &= \mathop{\mathbb{E}}_{C} \left[\langle\!\langle s_d | \mathcal{U}_d | s_{d-1} \rangle\!\rangle^2 \cdots \langle\!\langle s_1 | \mathcal{U}_1 | s_0 \rangle\!\rangle^2 \right] \\ &= \mathop{\mathbb{E}}_{\mathcal{U}_d} \left[\langle\!\langle s_d | \mathcal{U}_d | s_{d-1} \rangle\!\rangle^2 \right] \cdots \mathop{\mathbb{E}}_{\mathcal{U}_1} \left[\langle\!\langle s_1 | \mathcal{U}_1 | s_0 \rangle\!\rangle^2 \right] \\ &= \left(\frac{1}{15} \right)^{G(s)} \end{aligned} \tag{2.81}$$

Here the second line follows from the fact that $|\langle\!\langle x | s_d \rangle\!\rangle| = |\langle\!\langle s_0 | 0^n \rangle\!\rangle| = \frac{1}{\sqrt{2^n}}$, the third line is due to the independence between different random gates, and the fourth line is due to Lemma 2.9, where we define

$$G(s) := \text{the number of 2-qubit gates whose input and output are not } \mathtt{II} \text{ in } s. \tag{2.82}$$

The above calculation says that any 2-qubit gate whose input and output are not II contributes a $\frac{1}{15}$ factor to the Fourier weight. A simple bound on $G(s)$ is

$$\frac{|s|}{4} \leq G(s) \leq |s|, \tag{2.83}$$

where LHS is because each gate corresponds to at most 4 non-identity Paulis, and RHS is because each gate has at least 1 input non-identity Pauli. This implies that

$$2^{2n} \mathop{\mathbb{E}}_{C} \left[f(C, s, 0^n)^2 \right] \geq \left(\frac{1}{15} \right)^{|s|}. \tag{2.84}$$

Using this we have

$$\begin{aligned}O(1) &= \sum_{k=d+1}^{\ell} W_k \\ &= \sum_{k=d+1}^{\ell} 2^{2n} \mathop{\mathbb{E}}_{C} \sum_{s\in \mathsf{P}_n^{d+1}:|s|=k} f(C,s,0^n)^2 \\ &\geq \sum_{k=d+1}^{\ell} \sum_{s\in \mathsf{P}_n^{d+1}:|s|=k} \left(\frac{1}{15}\right)^{|s|} 1[s \text{ is legal}] \\ &\geq \left(\frac{1}{15}\right)^{\ell} (\text{Number of legal paths of weight at most } \ell)\,,\end{aligned} \tag{2.85}$$

which means that the number of legal paths of weight at most ℓ is at most $O(1) \cdot 15^{\ell}$. □

We have remarked earlier that the number of legal paths is a combinatorial property that only depends on the circuit architecture, independent of the gate set. We introduce Haar random 2-qubit gates in Lemma 2.15 as a proof technique for bounding the Fourier weights. We further show that the above results imply a lower bound on the depth for anti-concentration, which has been shown by [35, 38] using different techniques.

Corollary 2.4. *Consider any circuit architecture which satisfies anti-concentration with Haar random 2-qubit gates, then the circuit depth satisfies $d = \Omega(\log n)$.*

Proof. Consider $\ell = d+1$, using Lemma 2.14 and Lemma 2.15 we have

$$n \cdot 2^d \cdot 3^{d-1} \leq O(1) \cdot 15^{d+1}, \tag{2.86}$$

which implies that $d = \Omega(\log n)$. □

Next we present the main result of this section, an algorithm for efficiently enumerating low-weight legal Pauli paths.

Lemma 2.16. *For any $\ell \geq d+1$, the number of legal Pauli paths with Hamming weight at most ℓ is at most $n^{\ell/d} \cdot 2^{O(\ell)}$ (the circuit architecture does not need to satisfy anti-concentration). Furthermore there is an efficient algorithm to enumerate the legal paths in time $n^{\ell/d} \cdot 2^{O(\ell)}$ and memory $\tilde{O}(nd)$.*

The proof of Lemma 2.16 is deferred to the end of this section. Next we discuss its relationship with the above results.

First, it appears that Lemma 2.16 is not tight as it has an additional $n^{\ell/d}$ factor compared with Lemma 2.15. In fact this is not the case, due to the fact that Lemma 2.15 assumes anti-concentration, which by Corollary 2.4 means that Lemma 2.15 only holds when $d = \Omega(\log n)$. Note that in this case

$$n^{\ell/d} = e^{\frac{\ell}{d}\cdot \log n} = 2^{O(\ell)}, \tag{2.87}$$

so in the anti-concentration regime Lemma 2.16 gives the same asymptotic result as in Lemma 2.15, which is tight up to the constant in the exponent.

Second, when $\ell = O(d)$, Lemma 2.16 gives $\text{poly}(n) \cdot 2^{O(d)}$. Therefore compared with Lemma 2.14 we conclude that Lemma 2.16 with $\ell = O(d)$ is tight up to the constant in the exponent, regardless of whether anti-concentration holds.

Proof of Lemma 2.16. We prove Lemma 2.16 in the rest of this section. We will enumerate legal Pauli paths $s = (s_0, s_1, \dots, s_d)$ using the following method.

1. For each $d+1 \leq k \leq \ell$, choose the Hamming weight $w_0, \dots, w_d$ for each layer, such that $w_0 + \cdots + w_d = k$.

2. Choose the configuration (positions of identities and non-identities) for each layer.

3. Choose $X/Y/Z$ for each non-identity.

The following is a detailed counting argument and enumeration method for the legal Pauli paths. Consider a fixed total Hamming weight $d+1 \leq k \leq \ell$.

1. Choose the Hamming weight $w_0, \dots, w_d$ for each layer, such that the total weight is k. The number of choices equals the number of solutions to the equation $w_0 + w_2 + \cdots + w_d = k$ ($w_i \geq 1$), which equals to $\binom{k-1}{d} \leq 2^{k-1}$. The enumeration of such solutions can be achieved using a combinations enumerator which efficiently enumerates all combinations of choosing d objects from $k-1$ objects, with memory cost $\widetilde{O}(d)$. Note that not all solutions correspond to legal Pauli paths; the illegal ones will be rejected later.

2. For each Hamming weight configuration $w_0, \dots, w_d$, let t be the index of the layer with smallest Hamming weight (if there are tiebreaks, choose the smallest t). As the total weight is k, we know that $w_t \leq k/d$. Next we enumerate the configuration (locations of non-identities) of this layer. The number of choices is $\binom{n}{w_t} \leq n^{k/d}$ and can be enumerated using a combinations enumerator. We can store a configuration of a layer using n bits.

3. We choose the configurations for the other layers in a way that evolves the t-th layer both forwards and backwards. For example, consider choosing the configuration for the $t+1$-th layer, conditioned on a given configuration for the t-th layer. Consider the layer of 2-qubit gates that connects the t-th layer of the Pauli path with the $t+1$-th layer of the Pauli path. Those 2-qubit gates that have input `II` have to have output `II`. The number of 2-qubit gates whose input is not `II` is at most w_t. For each of these gates, its output can be `IR`, `RI`, or `RR` (We use `R` to represent a non-identity). So there are at most 3^{w_t} configurations for the $t+1$-th layer. Not all of these configurations satisfy the constraint that the $t+1$-th layer has Hamming weight w_{t+1}. So within these (at most) 3^{w_t} configurations, we reject those that do not have weight w_{t+1}. Repeating

this procedure for the next layer, we have that the number of configurations for the $t+2$-th layer is at most $3^{w_{t+1}}$, conditioned on a given configuration for the $t+1$-th layer. Using the same argument but evolve backward from the t-th layer, the number of configurations for the $t-1$-th layer is at most 3^{w_t}, and the number of configurations for the $t-2$-th layer is at most $3^{w_{t-1}}$ and so on.

4. Repeat the above argument for $t+1, t+2, \ldots, d$ as well as $t-1, t-2, \ldots, 0$. The total number of configurations for the entire Pauli path (conditioned on a given partition $w_0, \ldots, w_d$ and a given configuration for the t-th layer) is at most $3^{\sum_i w_i} = 3^k$. The memory cost for enumerating a configuration for the entire circuit is at most $\tilde{O}(nd)$.

5. Replace each `R` with X, Y, Z (except for the first and last layer, where `R` is only replaced with Z), giving another 3^k factor.

Taking into account all factors in the above steps, the total number of legal paths of Hamming weight at most ℓ (and the total running time of the enumeration algorithm) is at most

$$\sum_{k=d+1}^{\ell} 2^{k-1} \cdot n^{k/d} \cdot 3^k \cdot 3^k \leq \ell \cdot n^{\ell/d} \cdot 18^{\ell} = n^{\ell/d} \cdot 2^{O(\ell)}. \tag{2.88}$$

2.3.5.3 Putting everything together

Summarizing the main results of the previous section, we have the following.

Lemma 2.17. *Consider the same assumptions as our main result (Remark 2.6) and fix a truncation parameter ℓ. There is an algorithm that computes the function $\bar{q}(C,x) = \sum_{s:|s|\leq\ell}(1-\gamma)^{|s|} f(C,s,x)$ and its marginals in time $nd \cdot 2^{O(\ell)}$. Here by marginal we mean $\sum_{i\in T}\sum_{x_i\in\{0,1\}} \bar{q}(C,x_1,\ldots,x_n)$ for any $T \subseteq [n]$.*

Proof. As circuit depth $d = \Omega(\log n)$, Lemma 2.16 says that for any $x \in \{0,1\}^n$, $\bar{q}(C,x)$ can be computed in time $nd \cdot 2^{O(\ell)}$ using the enumeration algorithm, as there are $2^{O(\ell)}$ paths and each path takes $O(nd)$ time to compute. To compute a certain marginal

$$\sum_{i\in T}\sum_{x_i\in\{0,1\}} \bar{q}(C,x_1,\ldots,x_n),$$

note that we cannot straightforwardly compute each $\bar{q}(C,x_1,\ldots,x_n)$ and sum them up because it has an additional factor $2^{|T|}$. However, the marginal can be easily computed by exchanging the summation order,

$$\begin{aligned}\sum_{i\in T}\sum_{x_i\in\{0,1\}} \bar{q}(C,x_1,\ldots,x_n) &= \sum_{i\in T}\sum_{x_i\in\{0,1\}}\sum_{s:|s|\leq\ell}(1-\gamma)^{|s|} f(C,s,x_1,\ldots,x_n) \\ &= \sum_{s:|s|\leq\ell}(1-\gamma)^{|s|}\left(\sum_{i\in T}\sum_{x_i\in\{0,1\}} f(C,s,x_1,\ldots,x_n)\right).\end{aligned} \tag{2.89}$$

The statement follows from the fact that the summation in the bracket can be computed in time $O(nd)$. This is because

$$\begin{aligned}\sum_{i\in T}\sum_{x_i\in\{0,1\}} f(C,s,x_1,\ldots,x_n) &= \sum_{i\in T}\sum_{x_i\in\{0,1\}} \langle\langle x|s_d\rangle\rangle\langle\langle s_d|\mathcal{U}_d|s_{d-1}\rangle\rangle\cdots\langle\langle s_1|\mathcal{U}_1|s_0\rangle\rangle\langle\langle s_0|0^n\rangle\rangle \\ &= \langle\langle x'|s_d\rangle\rangle\langle\langle s_d|\mathcal{U}_d|s_{d-1}\rangle\rangle\cdots\langle\langle s_1|\mathcal{U}_1|s_0\rangle\rangle\langle\langle s_0|0^n\rangle\rangle,\end{aligned} \tag{2.90}$$

where

$$\langle\langle x'|s_d\rangle\rangle = \mathrm{Tr}\left(s_d\cdot\bigotimes_{j\notin T}|x_j\rangle\langle x_j|\bigotimes_{i\in T} I_i\right). \tag{2.91}$$

□

Lemma 2.17 allows us to use the standard reduction of sampling from a probability distribution via computing its marginals. An issue here is that $\bar{q}(C,x)$ is not necessarily a distribution; it is only guaranteed to be close to $\tilde{p}(C,x)$ in L_1 norm. We use the following result of [40] which allows us to sample from a distribution that is close to $\tilde{p}(C,x)$.

Lemma 2.18 (Lemma 10 in [40]). *Let p be a probability distribution on $\{0,1\}^n$. Assume there is an oracle that computes a function $\bar{q}:\{0,1\}^n\to\mathbb{R}$ as well as its marginals, such that $\|p-\bar{q}\|_1\le\delta$. Then there is an algorithm that samples from a probability distribution q using $O(n)$ calls to the oracle, such that $\|p-q\|_1\le 4\delta/(1-\delta)$.*

Proof of Main result. In Section 2.3.5.1 we have shown that $\mathbb{E}_C\left[\Delta^2\right]\le O(1)\cdot e^{-2\gamma\ell}$. By Markov's inequality,

$$\Pr\left[\Delta\ge\frac{1}{\sqrt{\delta}}\sqrt{\mathbb{E}\left[\Delta^2\right]}\right] = \Pr\left[\Delta^2\ge\frac{1}{\delta}\,\mathbb{E}\left[\Delta^2\right]\right]\le\delta. \tag{2.92}$$

Therefore, with probability at least $1-\delta$ over random circuit C, we have

$$\Delta\le\frac{1}{\sqrt{\delta}}\sqrt{\mathbb{E}\left[\Delta^2\right]}\le\frac{O(1)}{\sqrt{\delta}}e^{-\gamma\ell}. \tag{2.93}$$

Using Lemma 2.17 and Lemma 2.18, for those circuits that satisfy Eq. (2.93) we can sample from a probability distribution that is $O(1)\cdot\Delta$-close to $\tilde{p}(C,x)$ in total variation distance. Let ε be the desired total variation distance, then

$$\frac{O(1)}{\sqrt{\delta}}e^{-\gamma\ell}\le\varepsilon \text{ is satisfied when } \ell\ge\frac{1}{\gamma}\log\frac{O(1)}{\varepsilon\cdot\sqrt{\delta}}. \tag{2.94}$$

Obtaining one sample requires $O(n)$ calls to the algorithm in Lemma 2.17. Assuming circuit depth is $d\le\mathrm{poly}(n)$, the total running time for obtaining one sample is $n\cdot nd\cdot 2^{O(\ell)} = \mathrm{poly}(n)\cdot\left(O(1)/(\varepsilon\cdot\sqrt{\delta})\right)^{O(1/\gamma)} = \mathrm{poly}(n,1/\varepsilon,1/\delta)$.

2.3.5.4 Statistical indistinguishability

Next we show that our main result implies statistical indistinguishability. We first recall the basic notions and then give a proof of Corollary 2.3.

Given two known probability distributions p, q over the same finite alphabet ($\{0,1\}^n$ in our case), and given M samples from either p or q, we would like to tell which is the case with high success probability. That is, two known distributions p and q are statistically distinguishable if there is an algorithm $\mathcal{A}$ (with unbounded running time) that, on input $x_1, \ldots, x_M \sim \mathcal{D}$,

- if $\mathcal{D} = p$, $\mathcal{A}$ returns "$\mathcal{D} = p$" with probability at least $\frac{2}{3}$;
- if $\mathcal{D} = q$, $\mathcal{A}$ returns "$\mathcal{D} = q$" with probability at least $\frac{2}{3}$.

Two known distributions p and q are statistically indistinguishable with M samples if there is no algorithm $\mathcal{A}$ that satisfies the above condition. We use the following well-known fact that closeness in total variation distance implies statistical indistinguishability.

Lemma 2.19. *Two known distributions p and q are statistically indistinguishable with M samples if*

$$\frac{1}{2}\|p - q\|_1 < \frac{1}{3M}. \tag{2.95}$$

In the context of random circuit sampling, statistical distinguishability is similarly defined with an additional averaging over the random circuit.

Definition 2.10 (Statistical distinguishability)**.** *For a random circuit C, let $\tilde{p}(C, x)$ be the noisy RCS output distribution and let $q(C, x)$ be a classical mock-up distribution (the output distribution of a classical simulation algorithm). $\tilde{p}(C, x)$ is statistically distinguishable from $q(C, x)$ with M samples if there is an algorithm $\mathcal{A}$ with input C as well as $x_1, \ldots, x_M \in \{0,1\}^n$ and output one of $\{$noisy RCS, mock-up$\}$ (with unbounded running time) such that*

- $\mathbb{E}_C \Pr_{x_1,\ldots,x_M \sim \tilde{p}(C)}[\mathcal{A}(C, x_1, \ldots, x_M) = \textit{noisy RCS}] \geq \frac{2}{3}$,
- $\mathbb{E}_C \Pr_{x_1,\ldots,x_M \sim q(C)}[\mathcal{A}(C, x_1, \ldots, x_M) = \textit{noisy RCS}] \leq \frac{1}{3}$.

Proof of Corollary 2.3. In order to prove statistical indistinguishability it suffices to show that

$$\mathop{\mathbb{E}}_{C}\left[\left\|\tilde{p}(C)^{\otimes M}, q(C)^{\otimes M}\right\|_1\right] < \frac{1}{3}. \tag{2.96}$$

Our main result says that $\|\tilde{p}(C) - q(C)\|_1 \leq \varepsilon$ with probability at least $1 - \delta$ over C. Call those C that satisfy $\|\tilde{p}(C) - q(C)\|_1 \leq \varepsilon$ good, and the rest bad. We have

$$\begin{aligned}
\mathop{\mathbb{E}}_{C}\left[\left\|\tilde{p}(C)^{\otimes M}, q(C)^{\otimes M}\right\|_1\right] &\leq \mathop{\mathbb{E}}_{C}\left[\left\|\tilde{p}(C)^{\otimes M}, q(C)^{\otimes M}\right\|_1 \middle| C \text{ is good}\right] + \Pr[C \text{ is bad}] \\
&\leq \mathop{\mathbb{E}}_{C}\left[M \cdot \|\tilde{p}(C), q(C)\|_1 \,|C \text{ is good}\right] + \delta \qquad (2.97) \\
&\leq M\varepsilon + \delta,
\end{aligned}$$

where the first line follows from the law of total expectation and the second line follows from subadditivity of total variation distance with respect to tensor product. Therefore, statistical indistinguishability is guaranteed by choosing $\varepsilon = 0.01/M$ and $\delta = 0.01$, which gives running time $\mathrm{poly}(n, M)$ in our algorithm.

2.3.6 Generalizing to an approximation of Google and USTC's gate sets

In this section we discuss the role of gate sets in our main result. Assuming anti-concentration holds and at least $\Omega(\log n)$ depth, then in fact the only place in the proof of our main result where the gate set is relevant is in the third line of Eq. (2.76). It uses a property of the Pauli paths called orthogonality (Lemma 2.11), which follows from a property of the gate set which we call gate-set orthogonality (Lemma 2.10). Gate-set orthogonality says that in the Pauli basis, if we consider averaging over two copies of a random gate in the gate set, then it effectively forces the input Pauli to be identical across the two copies. Lemma 2.10 shows that this holds as long as the gate set is closed under random Pauli.

However, in Google and USTC's experiments [7, 16, 17] this condition is violated. They considered random circuits with fixed 2-qubit gates and random single-qubit gates, where the 2-qubit gates are called fSim and are roughly parameterized as follows,

$$\mathrm{fSim}(\omega_1, \omega_2, \omega_3) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & e^{-i\omega_1} & 0 \\ 0 & e^{-i\omega_2} & 0 & 0 \\ 0 & 0 & 0 & e^{-i\omega_3} \end{bmatrix}. \tag{2.98}$$

These angles are site-dependent and are determined by benchmarking experiments. The single-qubit gates are chosen randomly[6] from $\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$, where $W = (X+Y)/\sqrt{2}$.

Here we consider a related gate set shown in LHS of Fig. 2.3 where the main difference is that we insert random Z rotations. The fSim gates have a special property that allows us to borrow randomness from $R_Z(\theta_3)$, $R_Z(\theta_4)$ and create additional random gates as $R_Z(\theta_5)$, $R_Z(\theta_6)$, leading to the equivalent gate set in RHS of Fig. 2.3. This is because of the following commutation property. By definition, we can check that for any angles $\theta_1, \theta_2, \omega = (\omega_1, \omega_2, \omega_3)$,

$$R_Z(\theta_1) \otimes R_Z(\theta_2) \cdot \mathrm{fSim}(\omega) = \mathrm{fSim}(\omega) \cdot R_Z(\theta_2) \otimes R_Z(\theta_1). \tag{2.99}$$

Therefore we can consider the effective single qubit gate set

$$R_Z(\theta_1) V R_Z(\theta_2),\ V \in \{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}.$$

By direct calculation, we can verify that this single-qubit gate set is invariant under random Pauli and thus satisfies gate-set orthogonality.

[6] Google's single qubit gates V are not independent across each layer; neighboring layers does not repeat. This is still covered by Lemma 2.20 as it holds even for any fixed $V \in \{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$.

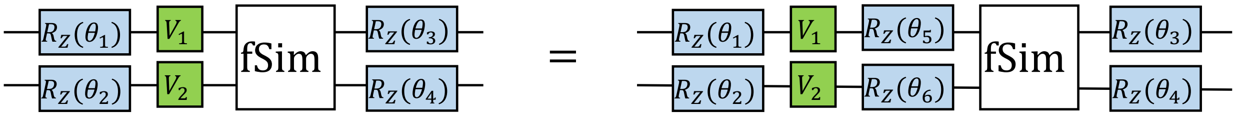


Figure 2.3: A gate set related to Google and USTC's experiments, for which our main result holds. LHS: the gate set consists of a fixed fSim gate surrounded by random gates from $\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$ as well as random Z rotations. RHS: this is equivalent to LHS due to a special property of the fSim gates.

Lemma 2.20. *Let $\mathcal{D}$ be a distribution over single-qubit unitary defined as $R_Z(\theta_1)VR_Z(\theta_2)$ where $\theta_1, \theta_2 \sim [-\pi, \pi]$ and $V \sim \{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$. Then for any $P, Q \in \{I, X, Y, Z\}$ such that $P \neq Q$, we have*

$$\mathop{\mathbb{E}}_{U\sim\mathcal{D}}\left[UPU^\dagger \otimes UQU^\dagger\right] = 0. \tag{2.100}$$

This implies the orthogonality condition in Lemma 2.11 which implies that our main result holds. An interesting open question is whether orthogonality is necessary for our main result, and whether our main result holds for the exact gate sets used in Google and USTC's experiments.

2.3.7 Refuting XQUATH for sublinear depth random circuits

Here we give a formal refutation of the XQUATH conjecture of [9] using the Pauli basis framework (a similar argument was first sketched in [18]; here we give a more formal treatment). In this section we only consider Haar random 2-qubit gates.

The XQUATH conjecture is about the hardness of estimating the output probability $p(C, 0^n)$ of an ideal random circuit C. It says that no efficient classical algorithm can achieve a slightly better variance compared with the trivial algorithm of outputting $\frac{1}{2^n}$.

Conjecture 2.3 (XQUATH [9])**.** *Let $\mathcal{D}$ be a distribution over quantum circuits. There is no polynomial-time classical algorithm that takes as input a quantum circuit $C \sim \mathcal{D}$ and produces a number $q(C, 0^n)$ such that*

$$\mathrm{XQ} := 2^{2n}\left(\mathop{\mathbb{E}}_{C\sim\mathcal{D}}\left[\left(p(C,0^n) - \frac{1}{2^n}\right)^2\right] - \mathop{\mathbb{E}}_{C\sim\mathcal{D}}\left[(p(C,0^n) - q(C,0^n))^2\right]\right) = \Omega\left(\frac{1}{2^n}\right). \tag{2.101}$$

The intuition behind this conjecture is the Feynman path integral in the computational basis (Eq. (2.52)). There are $2^{O(nd)}$ Feynman paths that are uniform, meaning that each path gives the same contribution on average. Therefore, a polynomial time classical algorithm cannot obtain a good estimate by calculating polynomial paths. In contrast, the Pauli path integral is highly non-uniform and the contribution of a path decays exponentially with the Hamming weight (Eq. (2.81)).

Remarkably, we show that using the Pauli path integral, a single path suffices to refute this conjecture below linear depth.

Theorem 2.7. *Let $\mathcal{D}$ be a distribution over quantum circuits with Haar random 2-qubit gates. Then there exists an algorithm that, on input C, outputs a number $q(C, 0^n)$ in time $O(nd)$ that achieves*

$$\mathrm{XQ} = \left(\frac{1}{15}\right)^d. \tag{2.102}$$

Therefore XQUATH is false for random circuits with depth $d = o(n)$.

Proof. On input $C = U_d \cdots U_1$, define the algorithm as computing

$$q(C, 0^n) := \frac{1}{2^n} + \langle\!\langle 0^n | s^* \rangle\!\rangle \langle\!\langle s^* | \mathcal{U}_d | s^* \rangle\!\rangle \cdots \langle\!\langle s^* | \mathcal{U}_1 | s^* \rangle\!\rangle \langle\!\langle s^* | 0^n \rangle\!\rangle = \frac{1}{2^n} + f(C, \vec{s^*}, 0^n) \tag{2.103}$$

where $s^* = \frac{1}{\sqrt{2^n}} Z_1 \otimes I^{\otimes n-1}$ (Z_1 acts on the first qubit). This takes time $O(nd)$. Then

$$\begin{aligned}
\mathrm{XQ} &= 2^{2n} \mathop{\mathbb{E}}_{C \sim \mathcal{D}} \left(\frac{1}{2^{2n}} - \frac{2}{2^n} p(C, 0^n) - q(C, 0^n)^2 + 2p(C, 0^n) q(C, 0^n) \right) \\
&= 2^{2n} \mathop{\mathbb{E}}_{C \sim \mathcal{D}} \left(-\frac{1}{2^{2n}} - q(C, 0^n)^2 + 2p(C, 0^n) q(C, 0^n) \right) \\
&= 2^{2n} \mathop{\mathbb{E}}_{C \sim \mathcal{D}} \left(-\frac{2}{2^{2n}} - f(C, \vec{s^*}, 0^n)^2 + 2p(C, 0^n) q(C, 0^n) \right) \\
&= 2^{2n} \mathop{\mathbb{E}}_{C \sim \mathcal{D}} \left(-f(C, \vec{s^*}, 0^n)^2 + 2p(C, 0^n) f(C, \vec{s^*}, 0^n) \right) \\
&= 2^{2n} \mathop{\mathbb{E}}_{C \sim \mathcal{D}} \left(-f(C, \vec{s^*}, 0^n)^2 + 2f(C, \vec{s^*}, 0^n)^2 \right) \\
&= 2^{2n} \mathop{\mathbb{E}}_{C \sim \mathcal{D}} f(C, \vec{s^*}, 0^n)^2 \\
&= \left(\frac{1}{15}\right)^d.
\end{aligned} \tag{2.104}$$

Here, the first line is by calculation; the second line is because $\mathbb{E}_{C \sim \mathcal{D}}[p(C, 0^n)] = \frac{1}{2^n}$; the third line is because $\mathbb{E}_{C \sim \mathcal{D}}[f(C, \vec{s^*}, 0^n)] = 0$ which follows from orthogonality with the all-identity path; the fourth line is again because $\mathbb{E}_{C \sim \mathcal{D}}[p(C, 0^n)] = \frac{1}{2^n}$; the fifth line follows from orthogonality (Lemma 2.11); the final step follows from the fact that there are exactly d 2-qubit gates that has a non-identity at the input and output, and each gate contributes a $\frac{1}{15}$ factor due to Lemma 2.9. □

As XQUATH is closely related to the XEB test, next we give a similar result by applying the above algorithm to XEB. Note that in actual experiments the XEB should be viewed as a statistical test, and our main result already implies that no such tests can distinguish between noisy RCS and the efficient classical algorithm in our main result. Thus the discussions below are for demonstration purposes, and for simplicity we only consider the expected value of XEB, and show that one Pauli path already suffices to achieve $2^{-O(d)}$ XEB.

For a random circuit C, let $p(C,x)$ be the output distribution of C, and let $q(C,x)$ be the output distribution of the noisy implementation of C or a simulation algorithm. The expected value of the linear cross entropy is defined as

$$\mathrm{XEB} := 2^n \mathop{\mathbb{E}}_{C} \sum_{x\in\{0,1\}^n} p(C,x)q(C,x) - 1. \tag{2.105}$$

First we consider noisy random circuits with the same model as in our main result. A useful property is that the XEB of noisy random circuits can be viewed as the Fourier weight polynomial.

$$\begin{aligned}\mathrm{XEB} &= 2^n \mathop{\mathbb{E}}_{C} \sum_{x\in\{0,1\}^n} p(C,x)q(C,x) - 1 \\ &= 2^n \mathop{\mathbb{E}}_{C} \sum_{x\in\{0,1\}^n} \sum_{s} (1-\gamma)^{|s|} f(C,s,x)^2 - 1 \\ &= 2^{2n} \mathop{\mathbb{E}}_{C} \sum_{s} (1-\gamma)^{|s|} f(C,s,0^n)^2 - 1 \\ &= \sum_{k>0} (1-\gamma)^k W_k.\end{aligned} \tag{2.106}$$

Here, the second line follows from the Pauli path integral and orthogonality (Lemma 2.11); the third line is by Eq. (2.62); the fourth line is by definition of Fourier weight and the fact that $W_0 = 1$.

Theorem 2.8. *Assuming anti-concentration, the linear cross entropy of a noisy random circuit with γ depolarizing noise satisfies*

$$(1-\gamma)^{d+1} \cdot n \cdot 2^d \cdot 3^{d-1} \cdot \left(\frac{1}{15}\right)^d \le \mathrm{XEB} \le O(1) \cdot e^{-\gamma d}. \tag{2.107}$$

Note that anti-concentration is only used for the upper bound; the lower bound does not require anti-concentration.

Proof. For the upper bound we use the upper bound on total Fourier weight (Lemma 2.12),

$$\mathrm{XEB} = \sum_{k\ge d+1} (1-\gamma)^k W_k \le O(1) \cdot (1-\gamma)^{d+1} \le O(1) \cdot e^{-\gamma d}. \tag{2.108}$$

The lower bound follows from $\mathrm{XEB} \ge (1-\gamma)^{d+1} W_{d+1}$. We have $W_{d+1} = n \cdot 2^d \cdot 3^{d-1} \cdot \left(\frac{1}{15}\right)^d$ because we have shown in Lemma 2.14 that the number of legal Pauli paths at degree $d+1$ equals $n \cdot 2^d \cdot 3^{d-1}$; each of them contributes $\left(\frac{1}{15}\right)^d$ to the Fourier weight. □

Next, consider a classical algorithm which samples from the same distribution as in the proof of Theorem 2.7, which is the following distribution

$$q(C,x) = \frac{1}{2^n} + \langle\!\langle x|s^*\rangle\!\rangle \langle\!\langle s^*|\mathcal{U}_d|s^*\rangle\!\rangle \cdots \langle\!\langle s^*|\mathcal{U}_1|s^*\rangle\!\rangle \langle\!\langle s^*|0^n\rangle\!\rangle = \frac{1}{2^n} + f(C,\vec{s^*},x) \tag{2.109}$$

where $s^* = \frac{1}{\sqrt{2^n}} Z_1 \otimes I^{\otimes n-1}$ (Z_1 acts on the first qubit). $\{q(C,x)\}_{x\in\{0,1\}^n}$ is a probability distribution because $\left|f(C,\vec{s^*},x)\right| \leq \frac{1}{2^n}$ and $\sum_{x\in\{0,1\}^n} f(C,\vec{s^*},x) = 0$.

The algorithm only samples the first qubit non-trivially, and uniformly on all other qubits.

Theorem 2.9. *There exists an efficient classical algorithm that, given a random circuit, outputs a sample in time $O(nd)$ that achieves*

$$\mathrm{XEB} = \left(\frac{1}{15}\right)^d. \tag{2.110}$$

Proof.

$$\begin{aligned}
\mathrm{XEB} &= 2^n \mathop{\mathbb{E}}_C \sum_{x\in\{0,1\}^n} p(C,x) f(C,s^*,x) \\
&= 2^n \mathop{\mathbb{E}}_C \sum_{x\in\{0,1\}^n} \sum_{s=(s_0,\ldots,s_d)\in\mathsf{P}_n} f(C,s,x) f(C,s^*,x) \\
&= 2^n \mathop{\mathbb{E}}_C \sum_{x\in\{0,1\}^n} f(C,s^*,x)^2 \\
&= 2^{2n} \mathop{\mathbb{E}}_C f(C,s^*,0^n)^2 = \left(\frac{1}{15}\right)^d.
\end{aligned} \tag{2.111}$$

Here, the first line is by definition of $q(C,x)$; the second line is by Pauli path integral; the third line follows from orthogonality (Lemma 2.11); the fourth line is by Eq. (2.62). □

2.3.8 Improved bounds on the convergence of noisy random circuits to the uniform distribution

In this section we develop improved bounds on the total variation distance between the output distribution of noisy random circuits and the uniform distribution, focusing on Haar random 2-qubit gates and depolarizing noise. The result can be directly applied to other depolarizing-like noise models as in [37, 38].

Let U denote the uniform distribution. In [38], the authors prove that the average total variation distance is lower bounded as

$$\mathop{\mathbb{E}}_C \left[\frac{1}{2}\left\|\tilde{p}(C) - U\right\|_1\right] \geq \frac{(1-\gamma)^{2d}}{4\cdot 30^d}. \tag{2.112}$$

Here we improve this bound by applying the Fourier weight and Pauli path integral technique. Note that in order to match with the notation in [38], we remove the layer of noise channels applied before the first layer of gates in Fig. 2.2 (b).

Theorem 2.10. *Let $\mathcal{D}$ be a distribution over quantum circuits on any parallel circuit architecture with Haar random 2-qubit gates. Let $\tilde{p}(C)$ be the output distribution of C subject to depolarizing noise with error rate γ on each qubit after each layer of gates, then we have*

$$\mathop{\mathbb{E}}_{C\sim\mathcal{D}}\left[\frac{1}{2}\left\|\tilde{p}(C)-U\right\|_1\right] \geq \frac{1}{12}\cdot(1-\gamma)^{2d}\cdot\left(\frac{2}{5}\right)^d. \tag{2.113}$$

Proof. Let $\delta := \frac{1}{2}\left\|\tilde{p}(C)-U\right\|_1$ be the total variation distance, and let

$$\tilde{p}_0 := \sum_{y\in\{0,1\}^{n-1}} \tilde{p}(C,0y) \tag{2.114}$$

be the marginal output probability of the first qubit being 0. As also noted by [38], the total variation distance can be lower bounded as $\delta \geq \left|\tilde{p}_0-\frac{1}{2}\right| \geq \left(\tilde{p}_0-\frac{1}{2}\right)^2$.

Following the proof of Lemma 2.17, the marginal output probability can be written as the Pauli path integral

$$\tilde{p}_0 = \sum_{s=(s_0,\ldots,s_d)\in\mathsf{P}_n^{d+1}} g(C,s) \tag{2.115}$$

where

$$\begin{aligned} g(C,s) &:= \langle\langle 0I^{\otimes n-1}|s_d\rangle\rangle\langle\langle s_d|\mathcal{E}^{\otimes n}\mathcal{U}_d|s_{d-1}\rangle\rangle\cdots\langle\langle s_1|\mathcal{E}^{\otimes n}\mathcal{U}_1|s_0\rangle\rangle\langle\langle s_0|0^n\rangle\rangle \\ &= (1-\gamma)^{|s_d|+\cdots+|s_1|}\langle\langle 0I^{\otimes n-1}|s_d\rangle\rangle\langle\langle s_d|\mathcal{U}_d|s_{d-1}\rangle\rangle\cdots\langle\langle s_1|\mathcal{U}_1|s_0\rangle\rangle\langle\langle s_0|0^n\rangle\rangle. \end{aligned} \tag{2.116}$$

The trivial all-identity path contributes $\frac{1}{2}$ to $\tilde{p}_0$. Therefore,

$$\begin{aligned} \mathop{\mathbb{E}}_{C\sim\mathcal{D}}\left[\frac{1}{2}\left\|\tilde{p}(C)-U\right\|_1\right] &\geq \mathop{\mathbb{E}}_{C\sim\mathcal{D}}\left(\tilde{p}_0-\frac{1}{2}\right)^2 \\ &= \mathop{\mathbb{E}}_{C\sim\mathcal{D}}\left(\sum_{s:|s|>0} g(C,s)\right)^2 \\ &= \mathop{\mathbb{E}}_{C\sim\mathcal{D}}\sum_{s:|s|>0} g(C,s)^2, \end{aligned} \tag{2.117}$$

where the third line is by orthogonality (Lemma 2.11). To lower bound the above sum, we consider all Pauli paths of weight $d+1$. Any such path s that gives a non-zero contribution will have Fourier weight

$$\mathop{\mathbb{E}}_{C\sim\mathcal{D}} g(C,s)^2 = (1-\gamma)^{2d}\cdot 2^{n-2}\cdot\left(\frac{1}{15}\right)^d\cdot\frac{1}{2^n} = \frac{(1-\gamma)^{2d}}{4\cdot 15^d}. \tag{2.118}$$

It remains to count the number of such paths. We have shown in Lemma 2.14 that the number of legal Pauli paths with weight $d+1$ equals $n\cdot 2^d\cdot 3^{d-1}$. However, here we lose a factor of n because the final layer has to be $s_d = \frac{1}{\sqrt{2^n}} Z_1 \otimes I^{\otimes n-1}$ (Z_1 acts on the first qubit) and therefore the path has a fixed ending point, giving $2^d \cdot 3^{d-1}$ paths in total. This gives

$$\begin{aligned}\mathop{\mathbb{E}}_{C\sim\mathcal{D}}\left[\frac{1}{2}\left\|\tilde{p}(C)-U\right\|_1\right] &\geq \mathop{\mathbb{E}}_{C\sim\mathcal{D}} \sum_{s:|s|>0} g(C,s)^2 \\ &\geq \mathop{\mathbb{E}}_{C\sim\mathcal{D}} \sum_{s:|s|=d+1} g(C,s)^2 \\ &= \frac{(1-\gamma)^{2d}}{4\cdot 15^d}\cdot 2^d \cdot 3^{d-1} = \frac{1}{12}\cdot(1-\gamma)^{2d}\cdot\left(\frac{2}{5}\right)^d .\end{aligned} \tag{2.119}$$

This bound can be further improved by considering more paths in Eq. (2.117). □

For completeness, we also summarize known upper bounds for the total variation distance. [36] showed that the KL divergence is upper bounded by $D_{\mathrm{KL}}(\tilde{p}(C)\|U) \leq n\cdot e^{-\gamma d}$. Thus by Pinsker's inequality

$$\frac{1}{2}\left\|\tilde{p}(C)-U\right\|_1 \leq \sqrt{\frac{n}{2}e^{-\gamma d}}. \tag{2.120}$$

Note that the above result holds for any circuit C, without averaging. In the anti-concentration regime, [37] showed an improved upper bound which gives

$$\mathop{\mathbb{E}}_{C\sim\mathcal{D}}\left[\frac{1}{2}\left\|\tilde{p}(C)-U\right\|_1\right] \leq O(1)\cdot e^{-\gamma d}. \tag{2.121}$$

2.4 Benchmarking near-term quantum computers via random circuit sampling

With the recent exciting progress in NISQ (Noisy Intermediate Scale Quantum) experiments, the characterization of noise in quantum devices has become a central challenge in the field [46, 47]. This goes beyond the benchmarking of individual gates; the characterization of many-body quantum noise can be expected to play an important role in building scalable quantum computers [48, 49]. This is because with scaling, understanding crosstalk between gates assumes increasing significance, in both improving fidelity of near-term experiments as well as making progress towards fault tolerance [50, 51, 52, 53].

Randomized benchmarking [54, 55, 56, 57, 58] has been the standard approach in traditional noise benchmarking, but in the context of characterizing quantum computers it does not scale beyond 2-3 qubits due to large circuit depth [59], and is therefore most useful for the benchmarking of individual gates. Cycle benchmarking and its variants [48, 60, 49, 61, 62, 63] provide a scalable approach to benchmark Clifford circuits. However, the characterization of noise in general circuits with non-Clifford gates has been an unreachable task due to the lack of group structure.

Here we address this challenge by drawing inspiration from Google's "quantum supremacy" experiment [7]. Google observed that their experimental estimate of the linear cross entropy benchmark – a proxy for the global fidelity of random circuits – was consistent with an uncorrelated noise model defined by multiplying individual gate fidelities, and they claimed that these experimental results could be considered a way of verifying that the noise channel acting on a layer of gates is uncorrelated across each gate. This observation is remarkable in the sense that the fidelity of highly complex random circuits could be predicted by such a simple noise model, but also intriguing as little theoretical evidence has been shown that supports this observation [64, 65]. A natural question is how convincing is this observation from the theoretical perspective, and more importantly, could this new observation be the germ of a new way to benchmark noise in general quantum circuits, even without assuming locality and independence in the noise model?

In this section we develop a noise benchmarking algorithm based on random circuit sampling (RCS). We show that the total amount of noise in a global and arbitrarily correlated noise model can be sample-efficiently extracted by measuring the linear cross entropy. While the noise parameter to be estimated is the same as cycle benchmarking [48], RCS benchmarking works for arbitrary non-Clifford gates and therefore can be used to benchmark general circuits. As an application, our results imply that if the noise in Google's device were highly non-local and correlated, this would cause the linear cross entropy and the uncorrelated noise model to deviate from each other in Google's experiment. This provides some formal evidence that supports Google's claim that the coincidence they observed between the two metrics indicated that the noise in their device was uncorrelated [7], which raises the potential for achieving fault tolerance.

RCS benchmarking. The noise benchmarking problem can be formulated as follows. Consider the Pauli noise channel induced by a layer of arbitrary non-Clifford gates $\mathcal{N}(\rho) = \sum_{\alpha\in\{0,1,2,3\}^n} p_\alpha \sigma_\alpha \rho \sigma_\alpha$ where σ_α are n-qubit Pauli operators and p_α are the corresponding Pauli error rates. The assumption that noise can be described by a Pauli channel is without loss of generality, as we show below that general noise channels will be effectively twirled into a Pauli channel by RCS. The main challenge then is designing an efficient experimental procedure to estimate the total error $\lambda = \sum_{\alpha\neq 0^n} p_\alpha$. In RCS benchmarking (Algorithm 2), the gates to be characterized are applied in an alternating architecture and interleaved by Haar random single-qubit gates (see Fig. 2.4), followed by a measurement in the computational basis. The algorithm works by estimating the average fidelity of these random circuits at several different depths, and then fit the average fidelity as an exponential decay function of depth, which gives the noise parameter λ. This algorithm can be implemented on today's hardware and is robust to state preparation and measurement (SPAM) errors.

Result 1: exponential decay of average fidelity. Let $\mathrm{RQC}(n, d)$ denote the ensemble of random quantum circuits with n qubits and depth d as shown in Fig. 2.4. An ideal implementation of a random circuit $C \sim \mathrm{RQC}(n, d)$ creates a pure state $|\psi\rangle = C\,|0^n\rangle$,

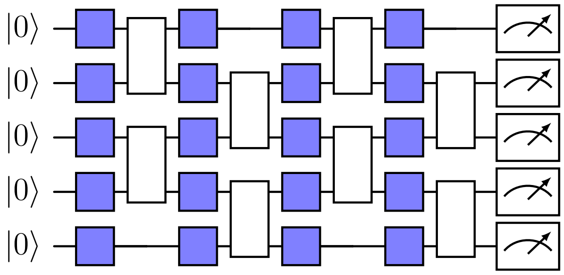


Figure 2.4: RCS benchmarking is an efficient algorithm to estimate the total amount of noise, including all crosstalks, on a layer of arbitrary two-qubit gates (white boxes) by implementing an alternating architecture interleaved with Haar random single qubit gates (blue boxes).

Algorithm 2 RCS benchmarking (simplified)

Input: number of qubits n, maximum circuit depth D, number of circuits L

Output: effective noise rate (ENR)

```
for $d = 1 \dots D$ do
    for $i = 1 \dots L$ do
        sample a random circuit $C_i \sim \mathrm{RQC}(n, d)$
        estimate the fidelity of $C_i$, denote as $\hat{F}_{d,i}$
                                        ▷ fidelity estimation
    end for
    $\hat{F}_d := \frac{1}{L} \sum_{i=1}^{L} \hat{F}_{d,i}$
end for
fit exponential decay $F = Ae^{-\lambda d}$ using data $\{\hat{F}_d\}_{d=1}^{D}$
Return $\lambda$
```

while due to noise the experimental implementation corresponds to a mixed state ρ, and the fidelity of C is $F = \langle\psi|\rho|\psi\rangle = \langle 0^n| C^\dagger \rho C |0^n\rangle$. The average fidelity is given by $\mathbb{E}_{C\sim\mathrm{RQC}(n,d)} F$ which we denote by $\mathbb{E}\, F$. Consider an arbitrary n-qubit noise channel acting on each layer of gates, which can be described as $\mathcal{N}(\rho) = \sum_{\alpha,\beta\in\{0,1,2,3\}^n} \chi_{\alpha\beta}\sigma_\alpha\rho\sigma_\beta$, where $(\chi_{\alpha\beta})$ is a positive semi-definite matrix known as the process matrix. We show that only the Pauli-diagonal component of the noise channel $\mathcal{N}^{\mathrm{diag}}(\rho) = \sum_{\alpha\in\{0,1,2,3\}^n} \chi_{\alpha\alpha}\sigma_\alpha\rho\sigma_\alpha$ is effective due to the twirling effect of random circuits (Section 2.4.2.1), in the sense that the average fidelity with noise channel $\mathcal{N}$ equals the average fidelity with $\mathcal{N}^{\mathrm{diag}}$, so without loss of generality we assume a Pauli noise channel $\mathcal{N}(\rho) = \sum_{\alpha\in\{0,1,2,3\}^n} p_\alpha\sigma_\alpha\rho\sigma_\alpha$, and the goal is to estimate the effective noise rate (ENR) $\lambda = \sum_{\alpha\neq 0^n} p_\alpha$. Our main result shows that $\mathbb{E}\, F \approx e^{-\lambda d}$, which implies that λ can be extracted via estimating $\mathbb{E}\, F$.

Theorem 2.11 (Main result)**.** *For random quantum circuits in 1D and 3-local noise channel with effective noise rate λ, the average fidelity is given by $e^{-\lambda d} \leq \mathbb{E}\, F \leq e^{-\lambda d}(1 + K\lambda)$ up to a first-order approximation in λ. Here K is a universal constant, and we assume $d \ll 2^n$.*

For simplicity here we focus on random circuits with Haar random 2-qubit gates. The same result is expected to hold for circuits with fixed 2-qubit gates interleaved by Haar random single-qubit gates as in Fig. 2.4 as they share similar properties such as convergence to unitary t-designs [45, 66, 67, 68, 69], which we also confirm via numerical simulations. The starting point of the proof (Section 2.4.2.2) is to decompose $\mathbb{E}\, F$ via the law of total expectation by conditioning on the number of errors that happen in the circuit, $\mathbb{E}\, F = \mathbb{E}\, F_0 + \mathbb{E}\, F_1 + \sum_{k\geq 2} \mathbb{E}\, F_k$, where $\mathbb{E}\, F_k := \Pr[k \text{ errors happen}] \cdot \mathbb{E}[F|k \text{ errors happen}]$. Note

Description	Lindblad	λ_F	λ_{uXEB}
T_1, T_ϕ	$\gamma D[\|0\rangle\langle 1\|] + 2\gamma D[\|1\rangle\langle 1\|]$	0.0511(2)	0.0511(2)
Pauli-X	$\gamma D[X]$	0.0508(2)	0.0509(2)
Corr-XX	$\gamma D[X_i X_{i+1}]$	0.0505(3)	0.0505(3)
$n-1$ Weight X	$\gamma D[\prod_{i\neq j} X_i]$	0.0506(3)	0.0506(3)

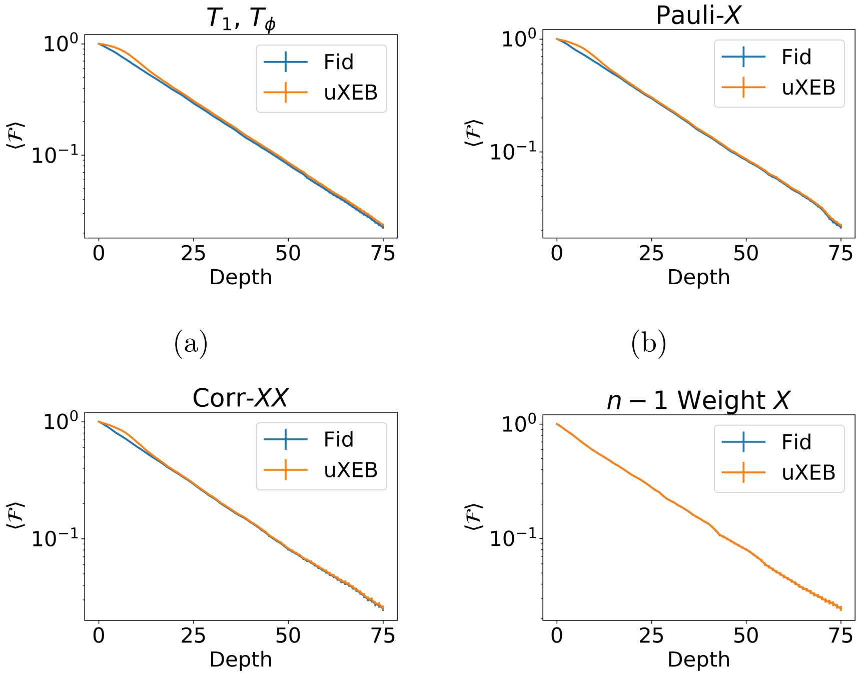
Table 2.2: Curve fitting results for the numerical simulation in Fig. 2.5. λ_F and λ_{uXEB} shows the simulated RCS benchmarking result, which corresponds to the decay rate of fidelity and unbiased linear cross entropy, respectively.

Description	Lindblad	λ_F	λ_{uXEB}
T_1, T_ϕ	$\gamma D[\sigma] + 2\gamma D[\sigma^\dagger \sigma]$	0.0531(2)	0.0536(2)
Corr-T_1	$\gamma D[\sigma_i \sigma_{i+1}]$	0.0495(3)	0.0502(3)
Pauli-X	$\gamma D[X]$	0.0492(3)	0.0500(3)
Corr-XX	$\gamma D[X_i X_{i+1}]$	0.0486(3)	0.0490(3)

Table 2.3: Curve fitting results for the numerical simulation in Fig. 2.6. λ_F and λ_{uXEB} shows the simulated RCS benchmarking result, which corresponds to the decay rate of fidelity and unbiased linear cross entropy, respectively.

that $\mathbb{E}\, F_0 = \Pr[\text{no error happens}] \approx e^{-\lambda d}$, so the goal is to prove that all $\mathbb{E}\, F_k$ with $k \geq 1$ are small compared with $\mathbb{E}\, F_0$. Intuitively $\mathbb{E}\, F_k$ has a λ^k factor and should decrease quickly with k when λ is small, so we make a first-order approximation by ignoring the $k \geq 2$ terms and focusing on the first-order contribution $\mathbb{E}\, F_1$. It is easy to prove that this approximation is valid when circuit depth $d \leq c/\lambda$ for some small constant c, and our numerical simulations suggest that it remains valid with experimentally relevant noise rates and circuit depths. Next, we show that $\mathbb{E}\, F_1 = O(\lambda \cdot e^{-\lambda d})$ by mapping this quantity to the partition function of a classical spin model and then use a domain wall counting argument to analytically bound the partition function. This technique has been used to analyze properties of random quantum circuits [70, 71, 72, 73, 39, 35] and here we generalize this to the context of noise benchmarking. While our rigorous arguments for showing $\mathbb{E}\, F_1 = O(\lambda \cdot e^{-\lambda d})$ apply to arbitrary 3-local errors which capture most error sources in quantum devices, numerical simulations suggest that it holds for arbitrary errors. Combining the above arguments gives the exponential decay $\mathbb{E}\, F \approx e^{-\lambda d}$ when λ is upper bounded by a small constant, which we also directly verify by simulating the average fidelity with different correlated noise models and gate sets. Our numerical simulation results are presented in Fig. 2.5 and Sections 2.4.2 and 2.4.6.

Result 2: fidelity estimation and variance. The above result outlines a procedure to extract λ via estimating $\mathbb{E}\, F$ (Algorithm 2). Here the depth-independent coefficient A



(a)

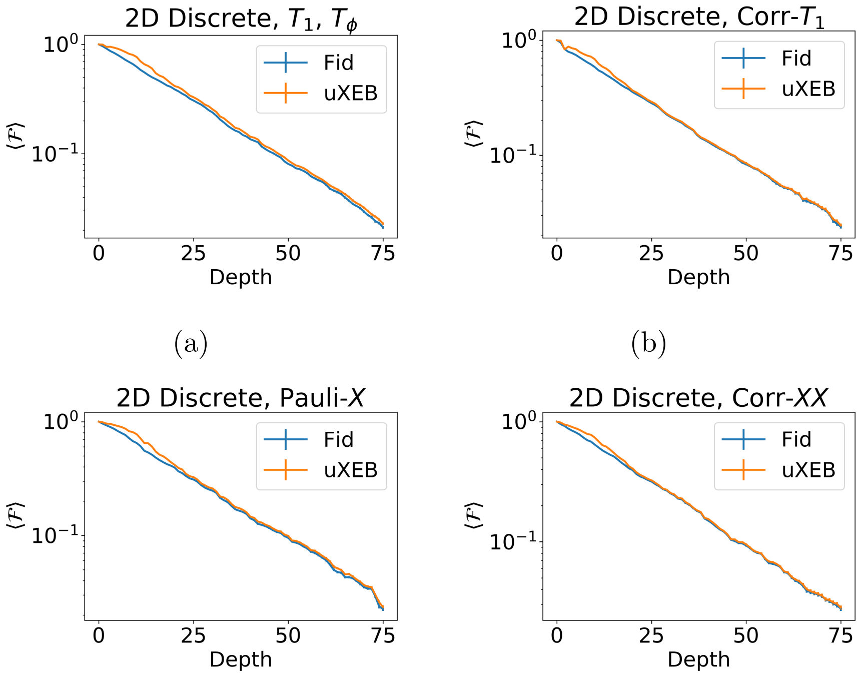
(b)

(c)

(d)

Figure 2.5: Numerical simulations using the Monte Carlo wave function (MCWF) method for various noise models. The system is modeled as perfect gates followed by evolution for one time unit under noisy channels [74] using the Lindblad master equation [75] $\frac{\mathrm{d}\rho}{\mathrm{d}t} = \sum_i \gamma_i D[J_i](\rho)$, where the sum is over different noise channels, $D[J_i](\rho) = J_i\rho J_i^\dagger - \frac{1}{2}(J_i^\dagger J_i\rho + \rho J_i^\dagger J_i)$ is a Lindblad superoperator for generic collapse operator J_i, and γ_i controls the noise strength. The unbiased linear cross entropy agrees with the fidelity for all depths above a small threshold and correctly predicts the ENR. The noise models include: (a) single qubit amplitude-decay and pure dephasing, (b) single qubit Pauli-X noise, (c) nearest-neighbor correlated XX noise, (d) correlated X noise with weight $n-1=19$, which is an artificial high-weight noise model. Here we simulate $n=20$ qubits on a 1D ring with noise strength $\gamma = 0.0025$. Each global noise channel has an ENR of $\lambda_{\text{true}} = n\gamma = 0.05$ by design. We average over 100 random circuits consisting of layers of two-qubit Haar-random unitaries and use 400 noise trajectories for each circuit at each depth. We fit the uXEB curves from depths 20 to 50.

corresponds to SPAM errors, and both A and λ can be extracted via curve fitting. To complete this we need a sample-efficient estimator of fidelity (line 4 and 5 of Algorithm 2). This is non-trivial as direct fidelity estimation (DFE) [76, 77] requires an exponential number of samples in the worst case. It has been suggested through heuristic arguments [20, 7, 23] that (unbiased) linear cross entropy appears to be a sample-efficient estimator for the



(a) (b)

(c) (d)

Figure 2.6: Numerical simulations using the Monte Carlo wave function (MCWF) method for various noise models. The system is modeled as perfect gates followed by evolution for one time unit under noisy channels [74] using the Lindblad master equation [75] $\frac{\mathrm{d}\rho}{\mathrm{d}t} = \sum_i \gamma_i D[J_i](\rho)$, where the sum is over different noise channels, $D[J_i](\rho) = J_i \rho J_i^\dagger - \frac{1}{2}(J_i^\dagger J_i \rho + \rho J_i^\dagger J_i)$ is a Lindblad superoperator for generic collapse operator J_i, and γ_i controls the noise strength. The unbiased linear cross entropy agrees with the fidelity for all depths above a small threshold and correctly predicts the ENR. The noise models include: (a) single qubit amplitude-decay and pure dephasing, (b) nearest-neighbor correlated amplitude-decay, (c) single qubit Pauli-X noise, (d) nearest-neighbor correlated XX noise. Here we simulate $n = 16$ qubits on a 2D lattice with noise strength $\gamma = 0.003125$. Each global noise channel has an ENR of $\lambda_{\text{true}} = n\gamma = 0.05$ by design. We average over 100 random circuits consisting of layers of random single qubit gates from the set $\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$, where $W = (X+Y)/\sqrt{2}$ and we use a fixed two-qubit SQiSWP gate over 400 noise trajectories for each circuit at each depth. We fit the uXEB curves from depths 20 to 50.

fidelity of noisy random circuits. For a random circuit C with output distribution $p_C(x) = |\langle x|C|0^n\rangle|^2$, the linear cross entropy estimator with M samples $S = \{x_i\}_{i=1}^M$ is given by $\hat{F}_{\text{XEB}}(S; C) = \frac{2^n}{M}\sum_{i=1}^M p_C(x_i) - 1$. In experiments, after collecting the output samples, we perform exact classical simulation of the ideal circuit C to compute the probabilities. We also consider the unbiased linear cross entropy estimator [78] defined as

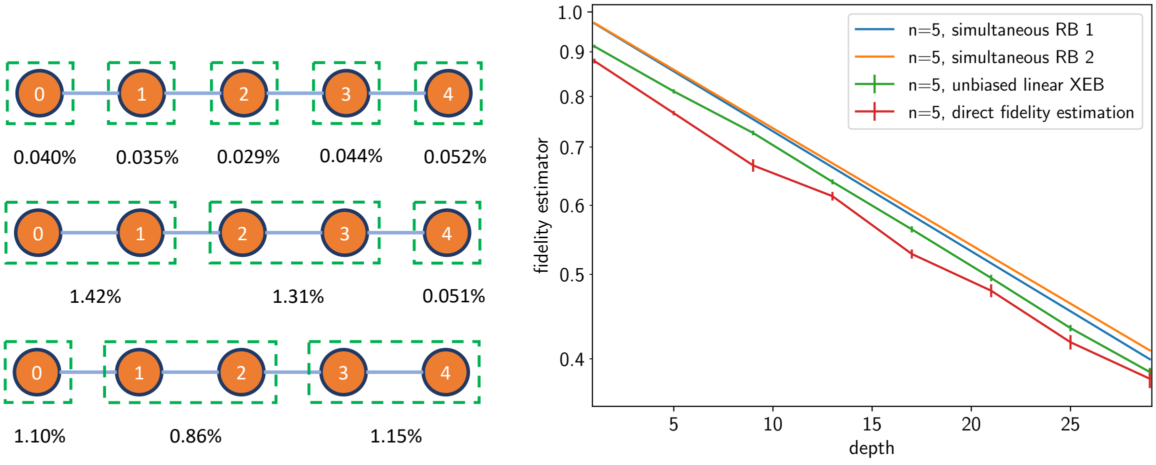
$$\hat{F}_{\mathrm{uXEB}}(S;C)=\frac{\frac{2^n}{M}\sum_{i=1}^{M}p_C(x_i)-1}{2^n\sum_{x\in\{0,1\}^n}p_C(x)^2-1}. \tag{2.122}$$

The term “unbiased” can be understood as follows: when the samples S come from the ideal distribution $p_C(x)$, we have $\mathbb{E}_S\,\hat{F}_{\mathrm{uXEB}}(S;C)=1$, while $\mathbb{E}_S\,\hat{F}_{\mathrm{XEB}}(S;C)$ can be exponentially large. Note that for random quantum circuits the denominator $2^n\sum_{x\in\{0,1\}^n}p_C(x)^2-1$ approaches 1 in log depth [39, 35], and therefore the two estimators give the same value as depth increases. In our experiments we use the unbiased linear cross entropy estimator by default, as it is more accurate at small constant depth. The main advantage of cross entropy estimators compared with DFE is that $O(1/\varepsilon^2)$ measurement samples suffice for estimating the fidelity of a random circuit within ε additive error.

To further justify the connection between unbiased linear cross entropy and fidelity, we perform extensive numerical simulations under correlated noise models. Fig. 2.5 shows the results of these simulations, which include the exponential decay curves of fidelity and the unbiased linear cross entropy, and error bars correspond to the standard error of the mean across different circuits which are too small to be seen on the plot. Note that the unbiased linear cross entropy estimates the true fidelity very well in all noise models except for very small depths. The curve fitting results are shown in Table 2.2. Here, both the decay rate of fidelity and unbiased linear cross entropy agree very well with the effective noise rate of the underlying noise model. These results verify our theoretical argument on the exponential decay of fidelity, as well as the correctness of unbiased linear cross entropy as a fidelity estimator, for both i.i.d. and highly correlated noise models. Figure 2.6 Shows a similar set of experiments using a 2D grid of qubits and we have additionally included a non-Pauli, nearest-neighbor correlated amplitude-decay noise model (Corr-T_1). Table 2.3 demonstrates that the 1D results extend to 2D lattices and to non-trivial non-Pauli noise channels. Additional simulation results with other system sizes, fidelity estimators, noise rates and gate sets are presented in Section 2.4.6.

Next, we show evidence that the variance of cross entropy estimators for a random circuit scale as $O\left(1/M+\lambda^2\left(\mathbb{E}\,F\right)^2\right)$, where M is the number of samples collected for each circuit (Section 2.4.2.5). In a large scale experiment the second term is much smaller than the first term due to the exponential decay of $\mathbb{E}\,F$, and therefore it suffices to collect a large number of samples for few circuits to estimate $\mathbb{E}\,F$ within small additive error. These results provide evidence that supports Google’s claim that only 10 random circuits with a large number of samples per circuit are sufficient to estimate the linear cross entropy in their experiment. On the other hand, the second term dominates for RCS benchmarking with a small number of qubits, and it is necessary to average over many (~ 100) different circuits to obtain good error bars, such as in our experiments below.

Result 3: experiments on IBM Quantum hardware. We implement RCS benchmarking via experiments on IBM Quantum hardware [79] with up to 20 superconducting



(a) Simultaneous RB

(b) RCS benchmarking

Figure 2.7: Experimental implementation of RCS benchmarking on `ibmq_athens` with 5 qubits. (a) Simultaneous RB, where parallel RB sequences are implemented for three gate patterns (green boxes) used in RCS benchmarking. The error rates underlying the single/two qubit green boxes are the RB results represented in Pauli error for Haar random single qubit gates and CNOT gates, respectively. (b) Exponential decay curves for RCS benchmarking with direct fidelity estimation and cross entropy. The decay rates, which represent the total amount of quantum noise per layer, are $\lambda_{\text{DFE}} = 3.05(5)\%$ and $\lambda_{\text{uXEB}} = 3.08(3)\%$. As a reference, the simultaneous RB estimator gives $\lambda_{\text{sRB}} = 3.13(4)\%$. We select 8 depth values ranging from 1 to 29. For DFE we implement 30 random circuits for each depth, and estimate the fidelity of each circuit by measuring 20 Fourier coefficients, which is 600 circuits for each depth. For cross entropy, we implement 100 random circuits for each depth. 8192 measurement samples are collected for each circuit. For both DFE and uXEB experiments, we use the standard error of the mean across different circuits as the error bar. The decay rate and its standard error are computed from the data points and error bars via standard least squares fitting. For the simultaneous RB estimator, we use half the difference between the two RB experiments as the standard error.

qubits in one dimension. On these devices, CNOT is the only 2-qubit gate available, and arbitrary single-qubit gates are easy to implement, which have error rates that are roughly 2 orders of magnitude smaller than CNOT. Therefore, in RCS benchmarking we are effectively measuring the total amount of quantum noise in a layer of CNOT gates. Fig. 2.7 shows experiment results on a 5-qubit device, see Section 2.4.3 for more details and larger experiments.

Here we perform three types of experiments: simultaneous RB, RCS with direct fidelity estimation, and RCS with cross entropy. In simultaneous RB [80], the main idea is to perform different RB sequences in parallel instead of performing RB on one pair of qubits while all other qubits are idle. A similar experiment was performed in Google's experiment [7] where linear cross entropy was used as the post processing method instead of standard RB. We implement simultaneous RB on each of the three patterns shown in Fig. 2.7a. The numbers underlying single qubit boxes represent the error rates of two X90 pulses ($\sqrt{X}$), which can represent the Pauli error rate of a Haar random single qubit gate. The numbers underlying two qubit boxes represent the Pauli error rate of CNOT gates. The results demonstrate some crosstalk behavior. For example, notice that in the third pattern in Fig. 2.7a, there is a large (1.10%) error rate on qubit 0, which is not present in the single qubit simultaneous RB. This suggests that a CNOT gate on qubit 1 and 2 can introduce a large error on qubit 0 due to crosstalk. Interestingly, in this experiment a CNOT gate on qubit 2 and 3 did not introduce additional errors on qubit 4.

Next we show results of RCS benchmarking with direct fidelity estimation and cross entropy. Note that DFE is not scalable due to the exponential sample complexity, and is implemented here mainly to verify our theoretical predictions. The results are shown in Fig. 2.7b, where all curves are exponential decays with roughly the same decay rate. The curves have different intercepts because DFE and uXEB experiments have different SPAM errors due to the additional overhead of DFE. From the curve fitting results, we can see that λ_{DFE} and λ_{uXEB} agrees with each other within the standard error. This confirms our theoretical results on the exponential decay of fidelity, and also verifies the validity of cross entropy as an efficient fidelity estimator. Also, note that uXEB is much more sample efficient than DFE, where the error bars for DFE are larger even when we collect 6 times the amount of samples in uXEB. As a reference, we implement simultaneous RB both before and after the RCS experiments, which can also be used to evaluate the error drift during the experiment period. In Fig. 2.7b we present heuristic fidelity estimators defined by multiplying individual gate fidelities measured by simultaneous RB experiments. Note that the simultaneous RB estimator gives a slightly larger prediction for the effective noise rate (also see below).

Application to diagnosing crosstalk. Diagnosing and reducing crosstalk errors is a central step towards achieving fault-tolerance. These errors can be characterized as two types [50]: the first is non-locality, where a correlated noise channel acts non-locally on multiple gates in the same layer; the second is dependence, where the noise channel on some gate depends on the other gates being implemented simultaneously. The second type can be demonstrated by comparing simultaneous RB results (such as in Fig. 2.7a). We show RCS benchmarking can provide information about the first type of crosstalk in a layer of arbitrary two-qubit gates.

In Google's experiment [7], the second type of crosstalk is clearly present, as the average two-qubit gate error increases from 0.36% as measured by individual RB to 0.62% as measured by simultaneous RB. Regarding the first type of crosstalk, Google observed in their paper that their experimental linear cross entropy was consistent with a simple uncorrelated noise model $\hat{F}_{\text{sRB}} = \prod_{i=1}^{m}(1 - e_i)$, where m is the number of gates and e_i is the error rate

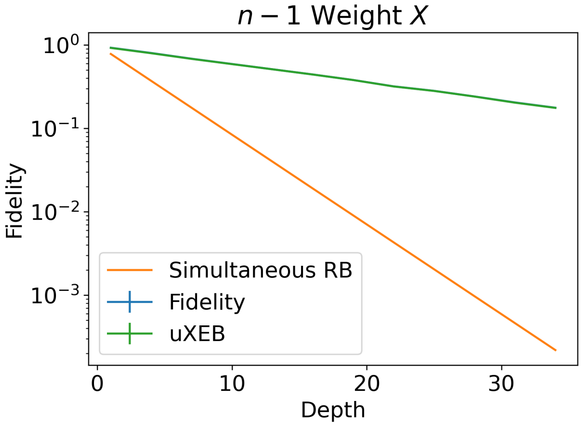


Figure 2.8: Simultaneous RB greatly overestimates noise when high-weight Pauli errors (the noise model in Fig. 2.5d) are used. Here fidelity and uXEB curves overlap and are indistinguishable on the scale of the plot, and error bars are too small to be seen.

of the ith gate measured by simultaneous RB. A similar behavior also happens in our experiment in Fig. 2.7b. Based on these experimental results, they claimed this coincidence indicated that the noise in their device was uncorrelated, i.e. the first type of crosstalk was not significant. Our results on RCS benchmarking provides formal evidence to support such a conclusion. This can be understood as follows. Imagine that a correlated (or high weight) Pauli error occurs in the device that acts on multiple gates, then it will be captured by multiple 2-qubit RB sequences. For example, if the error happens with 1% chance and is supported on gate i and gate j, then it will contribute 1% to both e_i and e_j. On the other hand, our result suggests that this error will only contribute 1% in the effective noise rate λ. Therefore correlated errors across multiple gates will be counted multiple times by $\hat{F}_{\text{sRB}}$ but only one time by the linear cross entropy and fidelity, so an experiment with significant correlated noise will demonstrate a deviation between $\hat{F}_{\text{sRB}}$ and linear cross entropy. In Fig. 2.8 we show a concrete simulated experiment with correlated noise to demonstrate such a deviation. This provides formal evidence that correlated noise was not significant in Google's experiment.

Discussion. We show that random circuit sampling is a powerful tool for benchmarking quantum noise, which can be viewed as a practical application of sampling-based quantum supremacy experiments. As larger scale quantum devices are being built, an important task is to develop efficient benchmarking protocols that can jointly test the performance of all of the qubits. Several holistic benchmarking protocols are proposed for this purpose, including quantum volume [81], accreditation protocols [82, 83] and others [84, 85, 86, 87, 88]. RCS benchmarking can also be understood in this context, where the effective noise rate characterizes the global noise strength when all two-qubit couplings are turned on, which can help predict the fidelity of running large scale circuits as well as inform the design of error correction codes.

While we have demonstrated the feasibility of RCS benchmarking for ~ 20 qubits, there are two main challenges when considering running RCS benchmarking in a larger scale. First, as shown in our main result, a necessary condition for the correctness of RCS benchmarking is the effective noise rate λ upper bouned by a small constant, that is, the effective noise rate per qubit scales as $O(1/n)$. This can be achieved if gate errors decrease as the number of qubits increases in hardware development. Second, the computation in RCS benchmarking becomes intractable when the number of qubits exceeds classical simulability, as computing the linear cross entropy (as well as other fidelity estimators) requires exact simulation of random quantum circuits, which makes the computation steps in RCS benchmarking inefficient.

Here we discuss two potential ways to overcome the computation barrier of RCS benchmarking in order to scale to 50+ qubits, assuming the effective noise rate is sufficiently small. First, for RCS benchmarking with generic gate sets, we can design variants of the fidelity estimation procedure which requires simulating groups of correlated amplitudes, where tractable tensor network simulation algorithms exist [12]. Second, when the native 2-qubit gate is a Clifford gate, such as the CNOT gate in IBM's hardware platform, the computation barrier can be avoided by using random single qubit Clifford gates in between CNOT layers, where the output probabilities are easy to compute. We expect the exponential decay of fidelity to still hold in this case, as the analysis in our main result only involves second moments while Clifford circuits can generate unitary 3-design.

2.4.1 Overview of RCS benchmarking

In this section, we give an overview of the RCS benchmarking protocol after introducing basic notations, and then briefly introduce the results, including theory of RCS fidelity decay, fidelity and variance estimation. We also discuss the relationship between RCS and other benchmarking protocols.

2.4.1.1 Setup and notations

Fig. 2.9 shows the ensemble of random quantum circuits used throughout. The system of qubits considered in our theoretical results and experiments are in one dimension, as shown in the figure, and we expect our results to be generalizable to higher dimensional lattices, where a similar alternating circuit architecture can be used such as in Google's experiment, as well as more general connectivity graphs. In the theory model of RCS (Fig. 2.9a), we consider circuits which consist of random 2-qubit gates drawn independently from the Haar measure on $\mathbb{U}(4)$. This model is used for the analysis of fidelity decay in section 2.4.2. In practice, as random 2-qubit gates are hard to implement, we consider RCS with layers of fixed 2-qubit gates with random single qubit gates in between (Fig. 2.9b). Our numerical simulation and experiments suggest that this model also creates an exponential decay in the same way as the theory model. The architecture in Fig. 2.9b is suitable for implementation on current quantum platforms, which usually optimize for a fixed 2-qubit gate. For example, our experiments on IBM Quantum hardware use CNOT gates with random single qubit gates drawn from the Haar measure on $\mathbb{U}(2)$.

Let $\mathrm{RQC}(n, d)$ denote the ensemble of random quantum circuits with n qubits and depth d as shown in Fig. 2.9. Here we define depth as the number of layers of 2-qubit gates, and both Fig. 2.9a and 2.9b correspond to $n = 5$ and $d = 4$. An ideal implementation of a random circuit $C \sim \mathrm{RQC}(n, d)$ creates a pure state $|\psi\rangle = C\,|0^n\rangle$, while due to noise the experimental implementation corresponds to a mixed state ρ, and the fidelity of the circuit C is defined as

$$F = \langle\psi|\rho|\psi\rangle = \langle 0^n|\, C^\dagger \rho C\,|0^n\rangle\,, \tag{2.123}$$

which is a random variable that depends on C. The average fidelity is then given by $\mathbb{E}_{C\sim\mathrm{RQC}(n,d)}\, F$, and we drop the subscript when unnecessary.

When assuming a gate-independent and Markovian noise channel between the layers, we show in section 2.4.2 that $\mathbb{E}\, F \approx e^{-\lambda d}$ for shallow depth circuits using the model in Fig. 2.9a, where λ is the total amount of Pauli noise for each layer, which we define as the effective noise rate (ENR). More specifically, consider an n-qubit noise channel $\mathcal{N}$ which can be uniquely specified as

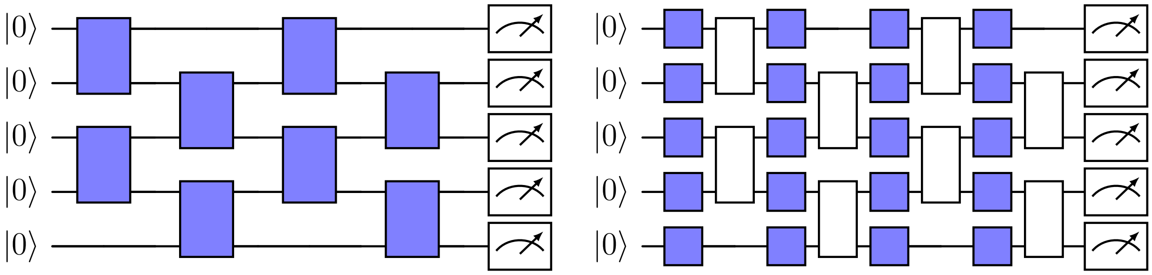
$$\mathcal{N}(\rho) = \sum_{\alpha,\beta\in\{0,1,2,3\}^n} \chi_{\alpha\beta}\sigma_\alpha\rho\sigma_\beta, \tag{2.124}$$

where $(\chi_{\alpha\beta})$ is a positive semi-definite matrix known as the process matrix, and $\sigma_\alpha \in \{I, X, Y, Z\}^{\otimes n}$ is a n-qubit Pauli operator. The effective noise rate is given by the sum of diagonal elements of the process matrix which corresponds to non-zero Pauli errors,

$$\mathrm{ENR}(\mathcal{N}) = \sum_{\alpha\in\{0,1,2,3\}^n\setminus\{0^n\}} \chi_{\alpha\alpha}. \tag{2.125}$$

We also consider the effective noise rate per qubit (ENRq) defined as $\mathrm{ENRq} = \mathrm{ENR}/n$, and also denote these quantities by λ and λ_q, respectively. When the noise channel is a Pauli channel (that is, χ is a diagonal matrix), the effective noise rate is simply the sum of probabilities of all non-zero Pauli operators.

In practice, however, noise is highly gate-dependent. For example, in today's quantum hardware the noise rate of 2-qubit gates are roughly two orders of magnitude higher than the noise rate of single qubit gates. In the practical implementation of RCS as in Fig. 2.9b, as two qubit gates are fixed and only single qubit gates (with much smaller error rates) are random, RCS benchmarking can be effectively viewed as benchmarking the effective noise rate of the noise channel introduced by the layer of 2-qubit gates, or more precisely, the average ENR of the two alternating layers of 2-qubit gates. This noise rate extracted from RCS benchmarking captures all local and cross talk errors among all qubits.



(a) Theory model (b) Experimental implementation

Figure 2.9: RCS benchmarking with circuits consist of local random gates in an alternating architecture. (a) Theory model of RCS, where each two-qubit gate (blue box) is independently drawn from the Haar measure on $\mathbb{U}(4)$. (b) Experimental implementation of RCS, where only random single qubit gates are used (blue box) with fixed two-qubit gates (white box).

Algorithm 3 RCS benchmarking (simplified)

Input: number of qubits n, maximum circuit depth D, number of circuits L

Output: effective noise rate (ENR)

```
for $d = 1 \ldots D$ do
    for $i = 1 \ldots L$ do
        sample a random circuit $C_i \sim \mathrm{RQC}(n, d)$
        estimate the fidelity of $C_i$, denote as $\hat{F}_{d,i}$        ▷ fidelity estimation
    end for
    $\hat{F}_d := \frac{1}{L}\sum_{i=1}^{L} \hat{F}_{d,i}$
end for
fit exponential decay $F = Ae^{-\lambda d}$ using data $\{\hat{F}_d\}_{d=1}^{D}$
Return $\lambda$
```

2.4.1.2 Fidelity estimation and variance

The core step of the RCS benchmarking protocol (Algorithm 3) is fidelity estimation. Next we describe the fidelity estimation methods that we consider in RCS benchmarking and analyze the variance. Here we focus on sample complexity, which is the main bottleneck of the entire procedure for the scale that we consider ($\sim 20 - 30$ qubits). See section 2.4.4 for discussions on computational complexity when considering RCS benchmarking in a scale that is beyond classical simulability.

For a small number of qubits, fidelity estimation can be done by running direct fidelity estimation (DFE) for each random circuit [76, 77]. Consider the Fourier expansion of the ideal state $|\psi\rangle\langle\psi| = \frac{1}{2^{n/2}} \sum_\alpha \gamma_\alpha \sigma_\alpha$ with $\gamma_\alpha = \frac{1}{2^{n/2}} \operatorname{Tr}[\sigma_\alpha |\psi\rangle\langle\psi|]$ and $\sum_\alpha \gamma_\alpha^2 = 1$. Let the noisy state from experiment be $\rho = \frac{1}{2^{n/2}} \sum_\alpha \gamma'_\alpha \sigma_\alpha$, then

$$F = \langle\psi|\rho|\psi\rangle = \sum_\alpha \gamma_\alpha \gamma'_\alpha = \mathop{\mathbb{E}}_{\alpha \sim \gamma_\alpha^2} \frac{\gamma'_\alpha}{\gamma_\alpha}. \tag{2.126}$$

Therefore F can be estimated by sampling a few Fourier coefficients from the distribution $\{\gamma_\alpha^2\}$, measuring them experimentally, and taking the empirical mean of $\gamma'_\alpha/\gamma_\alpha$. This procedure in general requires $O(2^n/\varepsilon^2)$ measurement samples in the worst case to obtain an estimate of F within ε additive error. In particular, if the circuit depth in RCS is sufficiently deep to generate a unitary 2-design, then the Fourier distribution of the output state is flat, which corresponds to the worst case in DFE. An interesting question is to study whether DFE can be improved for very low-depth random circuits.

Given the exponential sample complexity of DFE, we also consider sample efficient fidelity estimators based on cross entropy [20, 89, 7, 78]. In this work we justify the validity of cross entropy estimators by numerical simulation with different noise models and gate sets, and we leave the theoretical proof that cross entropy agrees with fidelity as important future work.

For a random circuit C with output distribution $p_C(x) = |\langle x|C|0^n\rangle|^2$, the linear cross entropy estimator with M samples $S = \{x_i\}_{i=1}^M$ is given by

$$\hat{F}_{\mathrm{XEB}}(S; C) = \frac{2^n}{M} \sum_{i=1}^M p_C(x_i) - 1. \tag{2.127}$$

In experiments, after collecting the output samples, we perform exact classical simulation of the ideal circuit C to compute the probabilities. We also consider the unbiased linear cross entropy estimator [78] defined as

$$\hat{F}_{\mathrm{uXEB}}(S; C) = \frac{\frac{2^n}{M} \sum_{i=1}^M p_C(x_i) - 1}{2^n \sum_{x \in \{0,1\}^n} p_C(x)^2 - 1}. \tag{2.128}$$

The term "unbiased" can be understood as follows: when the samples S come from the ideal distribution $p_C(x)$, we have $\mathbb{E}_S \hat{F}_{\mathrm{uXEB}}(S; C) = 1$, while $\mathbb{E}_S \hat{F}_{\mathrm{XEB}}(S; C)$ can be exponentially large. Note that for random quantum circuits the denominator $2^n \sum_{x \in \{0,1\}^n} p_C(x)^2 - 1$ approaches 1 in log depth [39, 35], and therefore the two estimators give the same value as depth increases. Different from the standard linear cross entropy, we need to classically simulate all 2^n output probabilities in order to compute the unbiased linear cross entropy estimator from experiment samples. In our experiments we use the unbiased linear cross entropy estimator by default, as it is more accurate at small constant depth. The main advantage of cross entropy estimators compared with DFE is that $O(1/\varepsilon^2)$ measurement samples suffice for estimating the fidelity of a random circuit within ε additive error, which follows from the property that the output probabilities obey the Porter-Thomas distribution.

In the RCS benchmarking protocol (Algorithm 3), an estimator for the average fidelity $\mathbb{E}\, F$ at depth d is obtained by taking the empirical mean of cross entropy estimators of different random circuits independently drawn from $\mathrm{RQC}(n, d)$, given by $\frac{1}{L} \sum_{i=1}^L \hat{F}_{\mathrm{uXEB}}(S; C_i)$.

The effective noise rate λ is extracted by fitting the curve $\mathbb{E}\, F = Ae^{-\lambda d}$ using the estimators of $\mathbb{E}\, F$ with increasing depth. Similar to RB protocols, this fitting procedure decouples the decay rate λ from SPAM errors, which is reflected in the depth-independent coefficient A.

In order to estimate the uncertainty of the benchmarking result λ, we need to estimate the variance of the estimator of $\mathbb{E}\, F$. The total variance is the sum of two parts: the variance of finite sampling for the fidelity estimation of each circuit, and the variance of fidelity across different circuits. In section 2.4.2.5 we give a theoretical model of the total variance,

$$\mathrm{Var}\left(\frac{1}{L}\sum_{i=1}^{L}\hat{F}_{\mathrm{uXEB}}(S;C_i)\right) = \frac{1}{L}O\left(\frac{1}{M} + \lambda^2\left(\mathbb{E}\, F\right)^2\right), \tag{2.129}$$

which suggests that the strategy for choosing parameters (including the number of samples for each circuit M and the number of circuits L) depends on the fidelity of the system. For a small number of qubits, it is necessary to choose a large L due to the second term, and for a large number of qubits such as in Google's experiment, it suffices to choose a small L and large M. In our numerical simulations and experiments on IBM Quantum hardware, we use the sample variance of $\hat{F}_{\mathrm{uXEB}}(S;C_i)$ across different circuits as the error bar, which is an unbiased estimator of the total variance.

The flexibility of RCS with respect to gate sets allows us to leverage the maximum amount of randomness that is efficiently implementable in the underlying hardware architecture in experimental implementations. In practice, the constants in Eq. (2.129) depends on the gate set, and we observe that the constant can be decreased by increasing the randomness in the gate set, which decreases the sample complexity of the experiment. For example, an arbitrary single qubit gate can be implemented with two X90 pulses combined with phase control of microwave drive on the IBM Quantum hardware platform, therefore we use Haar random single qubit gates between CNOT layers in order to achieve the smallest variance.

2.4.1.3 Relationship with other benchmarking protocols

In RCS benchmarking (Algorithm 3), the (unbiased) linear cross entropy plays the role of fidelity estimator for random quantum circuits. The idea that (unbiased) linear cross entropy is a sample efficient fidelity estimator for random circuits was originally proposed by Google [20, 7]. Ref. [7] also considered a form of unbiased linear cross entropy that is different from the one we use (Eq. (2.128)). In Ref. [23], a different form of unbiased linear cross entropy was used to estimate the fidelity of time-independent Hamiltonian evolution for Hamiltonians that lead to thermalization, where they develop an argument that connects the cross entropy estimator to the 2-design property of the projected state ensemble of a subsystem. Ref. [22] proved the exponential decay of linear cross entropy above log depth under i.i.d. noise models, which can be viewed as theoretical evidence for the agreement between linear cross entropy and fidelity. Ref. [18] also provides evidence that the effective noise rate λ upper bounded by a small constant is necessary and sufficient for the agreement between linear cross entropy and fidelity, which is consistent with our results.

In this work we further justify the observation that cross entropy is a good fidelity estimator for random quantum circuits by performing numerical simulation with practically motivated noise models which capture amplitude and phase decay in superconducting qubits as well as correlated noise. Meanwhile, instead of focusing on estimating fidelity itself, our main result is to prove the exponential decay of fidelity under correlated noise and use this to extract the effective noise rate.

In addition, in Google's experiment [7] it was shown that linear cross entropy can also be used as a post-processing method in benchmarking 2-qubit gates with a RB-like protocol. A detailed analysis of this method using global Haar random unitary gates was shown in [90]. In this work, we focus on the application of linear cross entropy in benchmarking a non-trivial number of qubits, where global randomness is hard to implement and only local randomness is accessible. As a simple comparison, note that RCS benchmarking only requires two-qubit gate errors to be smaller than order $1/n$, while for standard RB with global Cliffords the gate errors must be smaller than order $1/n^2$.

Several scalable RB variants were proposed in order to overcome the scalability bottleneck of standard RB, including simultaneous RB [80], direct RB [91], and cycle benchmarking [48], among others (see Ref. [90] for a comprehensive overview). A recent line of work uses a variant of simultaneous RB to obtain the full description of a gate-independent Pauli noise channel [60, 49, 61, 62]. Instead of using random global Cliffords, these works use random elements from small subgroups (subsets) of the global Clifford group, while it is unclear if they can be generalized to arbitrary gate sets without any group structure.

On the one hand, RCS benchmarking can also be put into context of the general framework of RB. Consider *two* layers in our random circuit architecture as a random element of a small subset of the global unitary group (note that this subset is not necessarily a group). This is similar to other "subset RB" protocols [92, 93, 91, 94, 48] where the distribution of the element is only supported on a small subset of the global unitary/Clifford group. General conditions for subset RB to create exponential decays were also formulated [90]. It is unclear if these previous analysis can be applied to RCS benchmarking which does not have a group structure.

On the other hand, RCS benchmarking is different from RB variants in two main aspects. First, unlike RCS, many scalable RB variants do not scramble across the entire system, and the state of the system remains in a tensor product of few qubit states. Therefore these RB variants create exponential decays that are different from RCS [7]. Second, RCS is very flexible with the gate set, while most RB variants assume Clifford gates with some cases extended to more general gates and finite groups [95, 96, 97, 98, 93, 94, 99, 90, 100, 101]. These differences mainly come from a special property of random quantum circuits: while the analysis of RB and variants rely heavily on the group structure, the fast convergence to unitary designs for random quantum circuits only requires generic gate sets [45, 66, 67, 68, 69]. This flexibility allows our RCS protocol to be implemented directly with the native

[7] Note that an exception is direct RB [91] which gives a heuristic argument of fidelity decay using random Clifford layers

gate set available on any hardware platform, including non-Clifford gates and gates with continuous parameters.

2.4.2 Theory of RCS benchmarking

Next we develop the theory of fidelity decay and variance estimation in RCS benchmarking. Throughout this section we work in the theory model shown in Fig. 2.9a, which assumes Haar random 2-qubit gates and a gate-independent noise channel acting on n qubits. Our numerical simulation and experiments suggest that RCS benchmarking also works well in practice with less randomness and gate-dependent noise.

2.4.2.1 Reducing to Pauli noise

We start by showing that general noise channels can be reduced to Pauli noise channels, then develop theoretical and numerical results under Pauli noise. Consider an arbitrary n-qubit noise channel $\mathcal{N}(\rho) = \sum_{\alpha,\beta\in\{0,1,2,3\}^n} \chi_{\alpha\beta}\sigma_\alpha\rho\sigma_\beta$ with process matrix $(\chi_{\alpha\beta})$, we define $\mathcal{N}^{\mathrm{diag}}$ as the noise channel that has process matrix $\mathrm{diag}(\chi_{\alpha\beta})$. By definition, $\mathcal{N}^{\mathrm{diag}}$ is a Pauli noise channel that stochastically applies a Pauli operator σ_α with probability $\chi_{\alpha\alpha}$. In addition, $\mathcal{N}$ and $\mathcal{N}^{\mathrm{diag}}$ have the same effective noise rate which is given by

$$\mathrm{ENR}(\mathcal{N}) = \mathrm{ENR}(\mathcal{N}^{\mathrm{diag}}) = \sum_{\alpha\neq 0^n} \chi_{\alpha\alpha}. \tag{2.130}$$

Consider two RCS benchmarking experiments with the same circuit architecture, where one has noise channel $\mathcal{N}$ and the other has $\mathcal{N}^{\mathrm{diag}}$. We show that these experiments have the same average fidelity.

Theorem 2.12. *The average fidelity of RCS benchmarking with noise channel $\mathcal{N}$ is equal to the average fidelity with noise channel $\mathcal{N}^{\mathrm{diag}}$. That is, without loss of generality we can assume that the underlying noise channel is Pauli noise.*

Therefore to study the fidelity decay of RCS for general noise channels, it suffices to only consider Pauli noise. Intuitively, Theorem 2.12 holds because of linearity, where we consider diagonal terms $\sigma_\alpha\rho\sigma_\alpha$ and off-diagonal terms $\sigma_\alpha\rho\sigma_\beta$ $(\alpha \neq \beta)$ separately, and show that the off-diagonal terms are "killed" by the Haar random 2-qubit gates after averaging. Details of the proof are presented in Section 2.4.5. Although RCS benchmarking is insensitive to off-diagonal terms, this is in general not a limitation, as there exist techniques such as randomized compiling [102] that can convert the underlying noise channel to Pauli noise for Clifford+T circuits.

2.4.2.2 Fidelity decay

Next we develop an argument showing that the average fidelity in RCS benchmarking decays as $\mathbb{E}\, F \approx e^{-\lambda d}$, where λ is the effective noise rate of a global Pauli noise channel acting on all qubits after each layer of two-qubit gates, provided that λ is a small constant.

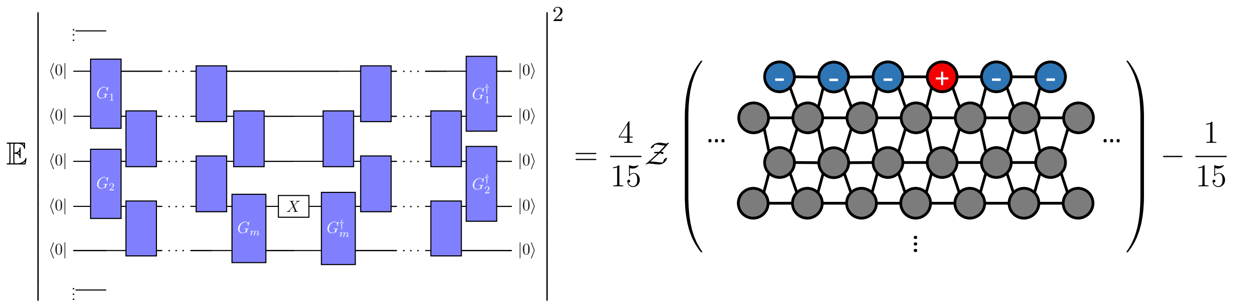


Figure 2.10: Mapping random quantum circuits to a statistical mechanical model. LHS: a tensor network diagram for $\mathbb{E}\,|\langle\psi_l|\psi\rangle|^2$, where an error happens at depth l after gate G_m. RHS: this expectation value equals to the partition function of a classical statistical mechanical spin model. Here, each spin corresponds to a 2-qubit gate. On the top boundary, the + spin (red) corresponds to the gate where the error happens (G_m), and the other gates at depth l correspond to $-$ spins (blue). All the other spins (grey) can be either + or $-$, and the partition function is the sum of weights of all possible configurations of the grey spins.

Throughout the theoretical analysis we consider a 1D system of qubits with periodic boundary condition (i.e. they are placed on a ring). For simplicity, below we first present a proof sketch where we consider single qubit Pauli-X noise channel acting on each qubit after each layer of gates, and then discuss generalizations afterwards. More detailed proofs as well as the extensions are presented in Section 2.4.5.

Consider single qubit Pauli-X noise channel with Pauli error rate ε defined as

$$\mathcal{E}(\rho) = (1-\varepsilon)\rho + \varepsilon X\rho X. \tag{2.131}$$

It is easy to see that the global noise channel $\mathcal{N} = \mathcal{E}^{\otimes n}$ has effective noise rate $\lambda = 1-(1-\varepsilon)^n \approx n\varepsilon$. We can view each Pauli-X noise channel as a stochastic process: with probability ε an X error is applied. We can then write the output density matrix of a n-qubit depth-d noisy random circuit as a weighted sum of all possible error patterns,

$$\begin{aligned}\rho = (1-\varepsilon)^{nd}\,|\psi\rangle\langle\psi| &+ \sum_{i=1}^{nd}\varepsilon(1-\varepsilon)^{nd-1}\,|\psi_i\rangle\langle\psi_i| \\ &+ \sum_{i=1}^{nd-1}\sum_{j=i+1}^{nd}\varepsilon^2(1-\varepsilon)^{nd-2}\,|\psi_{ij}\rangle\langle\psi_{ij}| + \cdots\end{aligned} \tag{2.132}$$

where $|\psi\rangle\langle\psi|$ denotes the ideal output state, $|\psi_i\rangle\langle\psi_i|$ denotes the ideal output state with an X error at location i, etc. Each term in the above sum denotes the density matrix with a fixed number of errors happened in all possible (nd) locations. For example, the first term denotes the state with no error and is weighted by the corresponding probability $(1-\varepsilon)^{nd}$. Similarly the other terms denote the state with $1, 2, \ldots$ errors. We can therefore write the average fidelity as

$$\begin{aligned}\mathbb{E}\, F &= \mathbb{E}\ \langle\psi|\rho|\psi\rangle \\ &= (1-\varepsilon)^{nd} + \sum_{i=1}^{nd} \varepsilon(1-\varepsilon)^{nd-1}\, \mathbb{E}\, |\langle\psi_i|\psi\rangle|^2 \\ &+ \sum_{i=1}^{nd-1} \sum_{j=i+1}^{nd} \varepsilon^2 (1-\varepsilon)^{nd-2}\, \mathbb{E}\, |\langle\psi_{ij}|\psi\rangle|^2 + \cdots \\ &:= F_0 + \mathbb{E}\, F_1 + \sum_{k\geq 2} \mathbb{E}\, F_k.\end{aligned} \tag{2.133}$$

Note that $F_0 \approx e^{-\varepsilon nd}$, our goal is therefore to prove that $\mathbb{E}\, F_1 + \sum_{k\geq 2} \mathbb{E}\, F_k$ is small compared with F_0.

Next we make a first order approximation by ignoring the term $\sum_{k\geq 2} \mathbb{E}\, F_k$ and focus on $\mathbb{E}\, F_1$. Intuitively, the contribution to fidelity should decrease with the number of errors, provided that noise rate is sufficiently small. We verify the validity of this first order approximation via extensive numerical simulations in the next subsection and Section 2.4.5 and 2.4.6.

Next we focus on proving that $\mathbb{E}\, F_1$ is small compared with F_0. Formally speaking, this is a necessary condition to our main goal $\mathbb{E}\, F \approx F_0$, as all the higher order terms are positive. First, to simplify $\mathbb{E}\, F_1$, notice that by assuming a periodic boundary condition, at a fixed depth the specific qubit where the error happens does not matter. We can simplify $\mathbb{E}\, F_1$ as

$$\mathbb{E}\, F_1 = n\varepsilon(1-\varepsilon)^{nd-1} \sum_{l=1}^{d} \mathbb{E}\, |\langle\psi_l|\psi\rangle|^2\,, \tag{2.134}$$

simplifying the sum and bringing an extra factor n, where $|\psi_l\rangle$ denotes the state with an X error at depth l at the first qubit. The problem is then reduced to bounding the sum $\sum_{l=1}^{d} \mathbb{E}\, |\langle\psi_l|\psi\rangle|^2$. See LHS of Fig. 2.10 for a demonstration of each term $\mathbb{E}\, |\langle\psi_l|\psi\rangle|^2$.

Second, note that this sum is at least a constant, $\sum_{l=1}^{d} \mathbb{E}\, |\langle\psi_l|\psi\rangle|^2 = \Omega(1)$. This is simply because all terms are positive, and in the first term where an error happens at depth 1, most gates cancel with the conjugate except one 2-qubit gate, and

$$\mathbb{E}\, |\langle\psi_1|\psi\rangle|^2 = \mathop{\mathbb{E}}_{U\sim\mathbb{U}(4)} |\, \langle 00|\, U^\dagger X U\, |00\rangle\, |^2 = \frac{1}{5}. \tag{2.135}$$

Our main result proves a tight upper bound $\sum_{l=1}^{d} \mathbb{E}\, |\langle\psi_l|\psi\rangle|^2 = O(1)$. This implies that

$$\mathbb{E}\, F_1/F_0 = O(n\varepsilon) = O(\lambda). \tag{2.136}$$

Third, note that all of the above arguments can be directly extended to general Pauli noise channels, where the only difference is that $|\psi_l\rangle$ has a general Pauli error at depth l instead of a single qubit Pauli. We extend our rigorous analysis of $\sum_{l=1}^{d} \mathbb{E}\,|\langle\psi_l|\psi\rangle|^2$ up to 3-local errors. While this captures most error sources in quantum device, we expect the same result to hold for general Pauli errors with arbitrary weight and locality, as supported by numerical evidence shown in Section 2.4.5 and 2.4.6.

Theorem 2.13. *For random quantum circuits in 1D with Haar random 2-qubit gates and 3-local noise channel with effective noise rate λ, the average fidelity is given by*

$$e^{-\lambda d} \le \mathbb{E}\,F \le e^{-\lambda d}(1 + K\lambda) \tag{2.137}$$

up to a first-order approximation in λ. Here K is a universal constant, and we assume $d \ll 2^n$.

Our results suggest that the fidelity decay $\mathbb{E}\,F \approx e^{-\lambda d}$ is a good approximation when λ is small, that is, when the effective noise rate per qubit scales like $1/n$. This requirement can be satisfied by the error rates in current quantum hardware. In addition, it is easy to see that Theorem 2.13 does not hold when circuit depth $d \to \infty$, as in this case $\mathbb{E}\,F \to \frac{1}{2^n}$ under depolarizing noise while the bounds goes to 0. Here our results mainly focus on low-depth random quantum circuits, which corresponds to the setting that is experimentally implementable.

Our results may seem surprising when comparing with other known results about random quantum circuits. In particular, many properties about random quantum circuits, such as convergence to unitary t-designs, are known to hold only above a certain depth [66], while our results hold already at shallow depth without any threshold requirement. This is partly because our fine grained analysis of $\mathbb{E}\,|\langle\psi_l|\psi\rangle|^2$ works for any $l = 1, \dots, d$, and in particular we show that

$$\mathbb{E}\,|\langle\psi_l|\psi\rangle|^2 \le e^{-\Delta l} + \frac{1}{2^n} \tag{2.138}$$

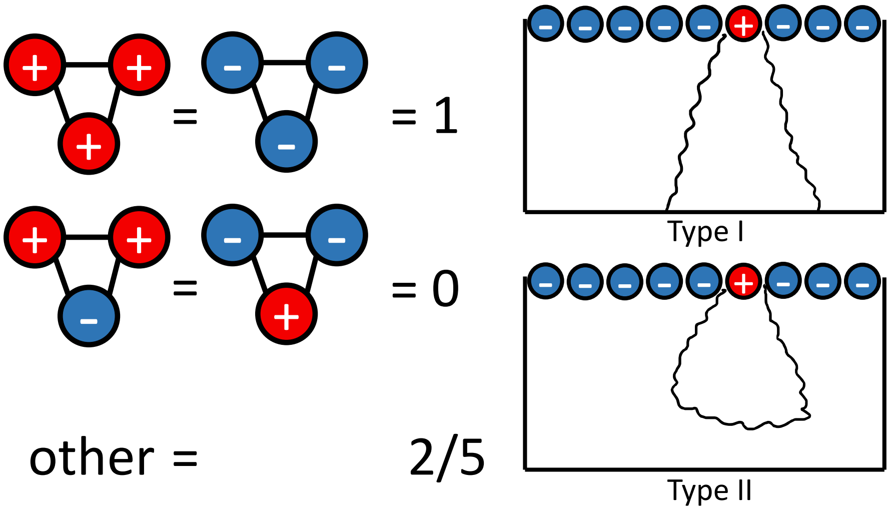
for some constant $\Delta > 0$. This suggests that for any l, random quantum circuits scramble the error at depth l exponentially fast.

It is easy to see that Eq. (2.138) implies Theorem 2.13. Next we give a simplified proof of Eq. (2.138) for single qubit errors; see Section 2.4.5 for details and extensions to higher weight errors. First, note that in $\mathbb{E}\,|\langle\psi_l|\psi\rangle|^2$ the gates that are applied after the error are canceled with the conjugate. Write the state as $|\psi_l\rangle = C_2 X C_1 |0^n\rangle$ and $|\psi\rangle = C_2 C_1 |0^n\rangle$ with unitary operators C_1 and C_2, then C_2 is canceled out and we have

$$\mathbb{E}\,|\langle\psi_l|\psi\rangle|^2 = \mathop{\mathbb{E}}_{C_1 \sim \mathrm{RQC}(n,l)} \left|\langle 0^n| C_1^\dagger X C_1 |0^n\rangle\right|^2, \tag{2.139}$$

therefore Eq. (2.138) is a property about depth-l random quantum circuits.

Second, we evaluate this expectation (LHS of Fig. 2.10) by taking the expectation of each 2-qubit gate, resulting in a classical statistical mechanical spin model (RHS of Fig. 2.10).



(a) Weights

(b) Domain wall

Figure 2.11: Details of the classical spin model. (a) Weights of the three-body interaction. (b) A non-zero weight configuration corresponds to the configuration of two domain walls, which either intersect and annihilate (Type II) or never intersect (Type I).

This technique of mapping to classical spin models has been widely used in recent study of random quantum circuits, see e.g. [70, 71, 72, 73, 39, 35]. Here, as the expectation $\mathbb{E}\left|\langle\psi_l|\psi\rangle\right|^2$ is a second moment property, that is, by linearity of expectation, each 2-qubit gate appears as a rank-2 projector $\mathbb{E}_{U\sim\mathbb{U}(4)}\left[U^{\otimes 2}\otimes U^{*\otimes 2}\right]$, we can represent each 2-qubit gate with a classical spin with 2 degrees of freedom. As a result, the expectation value is related to the partition function of a classical spin model (RHS of Fig. 2.10), which lives on a triangular lattice with three-body interactions. The weights of the interactions are given in Fig. 2.11a, and boundary conditions are presented in Fig. 2.10. See Section 2.4.5 for detailed derivations.

Finally, we prove Eq. (2.138) by showing an analytical bound for the partition function of the spin model. The proof follows from a domain wall (boundary between clusters of $+$ and $-$ spins) argument. Note that from the constraints given by the weights (Fig. 2.11a), if two spins at the top of a triangle have the same sign, then the spin at the bottom also has the same sign. Therefore the spin configuration that has a non-zero weight must correspond to two domain walls due to the top boundary condition. The domain wall configuration has two cases (Fig. 2.11b): they either intersect and annihilate (Type II) or never intersect (Type I). We can therefore evaluate the expectation as

$$\mathbb{E}\left|\langle\psi_l|\psi\rangle\right|^2 = \frac{4}{15}\left(\mathcal{Z}_1+\mathcal{Z}_2\right) - \frac{1}{15}, \tag{2.140}$$

where $\mathcal{Z}_i$ denotes the sum of weights of domain wall configuration of Type i. By a simple counting argument, $\mathcal{Z}_1 \le (4/5)^{2(l-1)}$, and we also show that $\frac{4}{15}\mathcal{Z}_2 - \frac{1}{15} \le \frac{1}{2^n}$, which concludes the proof.

2.4.2.3 Fidelity estimation

Our main results (Theorem 2.12 and 2.13) show rigorous evidence that $\mathbb{E}\, F \approx e^{-\lambda d}$ for noisy random circuits, where λ is the effective noise rate for the global noise channel. The main idea of RCS benchmarking is to efficiently estimate $\mathbb{E}\, F$ for varying depth, and then extract λ by fitting an exponential decay curve. Next we provide a detailed investigation of fidelity estimators and develop numerical simulation techniques to verify their correctness.

Let C be a random circuit. The goal is to estimate the fidelity of an unknown quantum state ρ prepared in experiment, with respect to the pure state $|\psi\rangle = C\,|0^n\rangle$. As discussed partly in section 2.4.1.2, estimating the fidelity of an unknown quantum state is in general a hard task, which either requires exponential copies of ρ via direct fidelity estimation [76, 77], or requires additional circuit depth overhead that is unrealistic with current hardware [103].

Motivated by the insight developed by Google [20, 7] that the special property of random quantum circuits allows fidelity to be efficiently estimated from output samples via cross entropy, we consider fidelity estimators of the following general form,

$$\langle\psi|\rho|\psi\rangle \approx \hat{F}(S;C) \tag{2.141}$$

where $S = \{x_1, \ldots, x_M\}$ are samples from the output distribution of the noisy circuit, i.e. computational basis measurement results of ρ. For example, the linear cross entropy $\hat{F}_{\text{XEB}}(S;C) = \frac{2^n}{M}\sum_{i=1}^{M} p_C(x_i) - 1$ has the above form which is related to the sum of probabilities of output samples. Several unbiased and non-linear variants of cross entropy were also proposed [7, 20, 78, 23], and we use the unbiased version defined in Eq. (2.128) which we find has the best performance. In general these fidelity estimators are sample efficient, requiring only $M = O(1/\varepsilon^2)$ measurement samples to achieve ε additive accuracy (see section 2.4.2.5), and their correctness is based on heuristic arguments and numerical simulation. An important future direction is to provide rigorous justifications as well as develop other efficient fidelity estimators.

Next we show numerical simulation results for noisy random circuits to verify our results on the exponential decay of fidelity, and also show that cross entropy estimators agree well with the fidelity. To accomplish this, we model the system as perfect gates followed by evolution for one time unit under noisy channels [74] using the Lindblad master equation [75],

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \sum_i \gamma_i D[J_i](\rho), \tag{2.142}$$

where the sum is over different noise channels, $D[J_i](\rho) = J_i\rho J_i^\dagger - \frac{1}{2}(J_i^\dagger J_i\rho + \rho J_i^\dagger J_i)$ is a Lindblad superoperator for generic collapse operator J_i, and γ_i is a parameter that controls the noise strength.

We implement the evolution of this open quantum system in the open-source simulator, QuaC [104]. Importantly, for the numerical simulation of 20 noisy qubits, the naive density matrix simulation is inefficient. We instead use the Monte Carlo wave function (MCWF) method [105], which simulates random "quantum jumps", rather than the full dynamics of the density matrix, to reduce the total computational cost of the simulations. We simulate 1D rings of n qubits with periodic boundary conditions. Full details of the computational method can be found in Section 2.4.6.

Using the MCWF technique, we simulate a variety of noise models, as summarized in Table 2.2. These noise models are:

1. T_1 and T_ϕ, which includes single qubit amplitude decay and pure dephasing, and represent the primary noise sources in superconducting qubits [106];

2. i.i.d. single qubit Pauli-X noise, which models single qubit bit-flip;

3. nearest-neighbor correlated XX noise, where a Pauli-XX noise channel is applied to all neighboring qubit pairs, which models two-body incoherent coupling;

4. $n-1$ - weight Pauli-X noise, where a Pauli-$X^{\otimes n-1}$ noise channel is applied to all subsets of size $n-1$, which is an artificial noise model used to test RCS benchmarking with extremely high-weight noise.

Note that the identity of the Pauli operators in the noise models (whether it's X, Y or Z) does not matter due to averaging over random circuits. The effective noise rate of each of these noise channels is related to the noise strength γ in the Lindblad superoperator by a constant factor, while the constant differs for each noise model (see Section 2.4.6 for details). We manually adjust the coefficients γ_i such that the effective noise rate for all noise models in Table 2.2 are equal to $\lambda = n\gamma$, where γ is a parameter we control. We simulate 1D rings of $n = 20$ qubits, averaging over 100 random circuits consisting of layers of two-qubit Haar-random unitaries and use 400 noise trajectories for each circuit at each depth. We fit the uXEB curves from depths 20 to 50. Here the fidelity and cross entropy for each circuit is calculated by averaging over the stochastic noise trajectories, which is more efficient than simulating individual measurement samples (see Section 2.4.6 for more details).

Fig. 2.5 and Table 2.2 show the results of these simulations, which include the exponential decay curves of fidelity and the unbiased linear cross entropy, and error bars correspond to the standard error of the mean across different circuits which are too small to be seen on the plot. Note that the unbiased linear cross entropy estimates the true fidelity very well in all noise models. In Fig. 2.17 in Section 2.4.6 we also plot several other fidelity estimators using the same data, where they are far from true fidelity until they converge at around depth 15. As discussed in section 2.4.1.2, this is because the unbiased version corrects an additional normalization factor which converges to 1 very quickly. Additional simulation results with other system sizes, fidelity estimators, noise rates and gate sets can be found in Section 2.4.6.

For comparison, we also consider simultaneous RB [80], an alternative benchmarking method that can also be used to heuristically estimate the fidelity. Here, two-qubit RB sequences are simultaneously executed for all neighboring qubit pairs, and a Pauli error rate e_i can be extracted for all two qubit couplings from the RB decay curve. See section 2.4.3.1 for a concrete example. Then, the fidelity of a random circuit with m 2-qubit gates can be estimated by

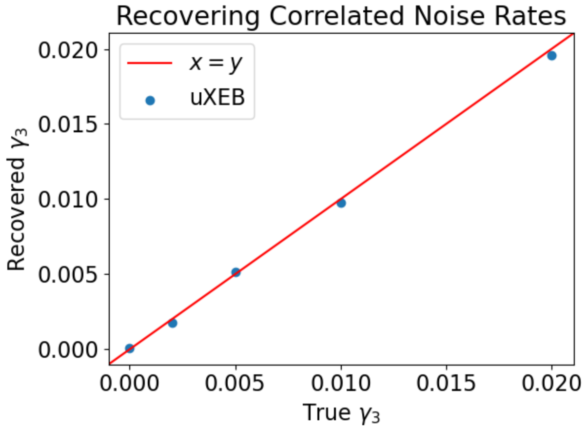


Figure 2.12: The amount of correlated noise, γ_3, can be readily extracted by combining amplitude decay, Ramsey, and RCS experiments.

$$\hat{F}_{\mathrm{sRB}} = \prod_{i=1}^{m} (1 - e_i). \tag{2.143}$$

First, note that $\hat{F}_{\mathrm{sRB}}$ is an accurate fidelity estimator for random quantum circuits when the noise sources are limited to single qubits, as the error rates can be estimated from the RB decay, and the estimator $\hat{F}_{\mathrm{sRB}}$ agrees with $\mathbb{E}\, F$ due to our Theorem 2.13. Second, when there is crosstalk, intuitively simultaneous RB can capture some correlated noise. For example, suppose when qubit pair 1 is turned on, it generates an error on a neighboring qubit pair 2. Since the RB sequences are executed simultaneously, this error can be captured by the RB sequence on qubit pair 2. However, this intuition can fail when we have high weight correlated errors, as the error could be over counted by multiple RB sequences (see "Application to diagnosing crosstalk").

2.4.2.4 Virtual experiment for extracting correlated noise

Next, we demonstrate a virtual experiment, where we show that RCS benchmarking can be used to extract the amount of correlated noise in conjunction with other benchmarking methods. Here, we assume that the underlying noise model is known to be a combination of T_1, T_ϕ, and a correlated ZZ noise channel, but the values of the individual noise rates is unknown. For simplicity, we assume that all noise channels of the same type (e.g., all T_1 processes) have the same noise rate across different qubits, but the noise rates can differ between types (T_1, T_ϕ, and the correlated ZZ noise channels can have different strengths).

This noise model is described by the Lindblad

$$\mathcal{L}(\rho) = \sum_{i=1}^{n} \Big(\gamma_1 D(\sigma_i)[\rho] + \gamma_2 D(\sigma_i^\dagger \sigma_i)[\rho] + \gamma_3 D(Z_i Z_{i+1})[\rho] \Big) . \tag{2.144}$$

Because we know the form of the noise channel, we can write down an explicit equation for the total effective noise rate in terms of the amplitude decay rate $\gamma_1 = \frac{1}{T_1}$, the pure dephasing rate $\gamma_2 = \frac{1}{T_\phi}$ times, as well as the correlated noise rate, γ_3,

$$\lambda = n\Big(\frac{\gamma_1}{2} + \frac{\gamma_2}{4} + \gamma_3\Big), \tag{2.145}$$

where the prefactors for γ_1 and γ_2 are calculated in Section 2.4.6. The T_1 time can be measured via a simple amplitude decay experiment, giving an equation

$$\Gamma_1 = \gamma_1, \tag{2.146}$$

The T_ϕ time can similarly be calculated by a Ramsey technique [107]; however, for this noise model, the T_ϕ time will not be recovered accurately due to the additional correlated dephasing caused by the incoherent ZZ coupling. The observed decay in the Ramsey experiment is a combination of the T_1, T_ϕ, and ZZ terms,

$$\Gamma_2 = \gamma_1 + \gamma_2 + 8\gamma_3, \tag{2.147}$$

where the prefactor 8 comes from arguments about the relative strengths of noise channels; see Section 2.4.6. These two standard single qubit benchmarking techniques can be combined with the ENR over the whole system provided by RCS. This gives three equations (from the relaxation decay experiment (2.146), the Ramsey experiment (2.147), and the RCS experiment (2.145)) and three unknowns (γ_1, γ_2 and γ_3), and can be readily solved to extract the correlated noise rate,

$$\gamma_3 = \frac{\Gamma_1}{4} + \frac{\Gamma_2}{4} - \frac{\lambda}{n}, \tag{2.148}$$

which directly combines the experimentally measured Γ_1, Γ_2, and λ. We simulate the noise model of Eq. (2.144) using $n = 10$ qubits with a ring geometry with $\gamma_1 = 0.01$, $\gamma_2 = 0.02$, and $\gamma_3 = 0.02\alpha$, with $\alpha \in [0.0, 0.1, 0.25, 0.5, 1.0]$. We extract the total ENR by using two-qubit Haar random circuits, fitting the resulting uXEB curve from depths 12 to 40, except in the case of $\alpha = 1.0$, where we fit from 12 to 22, due to the large total ENR of 0.4. The result of performing this extraction is shown in Figure 2.12.

2.4.2.5 Variance analysis

We provide an analysis of the variance of fidelity estimators in RCS benchmarking. In particular, this can help to decide how many random circuits to implement and how many measurement samples to collect in a RCS benchmarking experiment. Our analysis is a generalization of the statistical analysis of Google's quantum supremacy experiment [20, 7, 78] and provides a more accurate model for small scale RCS experiments.

In the following we focus on providing a theoretical model for the variance of cross entropy estimators. In experiments, once we have an estimate for the mean and variance of cross entropy estimators of different circuit depth, we can use standard least squares fitting to extract the effective noise rate λ as well as its standard error.

Recall the definition of the unbiased linear cross entropy

$$\hat{F}_{\mathrm{uXEB}}(S;C) = \frac{\frac{2^n}{M}\sum_{i=1}^{M} p_C(x_i) - 1}{2^n \sum_{x\in\{0,1\}^n} p_C(x)^2 - 1} \tag{2.149}$$

which has two sources of randomness: the random circuit C and the measurement samples $S = \{x_1, \ldots, x_M\}$. By the law of total variance, we have

$$\mathrm{Var}_{C,S}(\hat{F}_{\mathrm{uXEB}}) = \mathop{\mathbb{E}}_{C}\left[\mathrm{Var}_S\left(\hat{F}_{\mathrm{uXEB}}|C\right)\right] + \mathrm{Var}_C\left(\mathop{\mathbb{E}}_{S}\left[\hat{F}_{\mathrm{uXEB}}|C\right]\right). \tag{2.150}$$

This can be understood as the following: the total variance of $\hat{F}_{\mathrm{uXEB}}$ is the sum of the variance across different measurement samples and the variance across different random circuits.

In the previous analysis [20, 7, 78], the second term is ignored and only the first term is considered. In particular, they showed that the first term scales like $\left(1 + 2\,\mathbb{E}\,F - (\mathbb{E}\,F)^2\right)/M$ where M is the number of samples collected for each circuit. This can be derived by assuming that the output distribution of the noisy circuit is a linear combination of an ideal Porter-Thomas distribution with the uniform distribution. When $\mathbb{E}\,F \ll 1$ this can be well approximated by $1/M$.

For the second term, we assume that $\hat{F}_{\mathrm{uXEB}}$ is an unbiased estimator of the fidelity of the random circuit, and $\mathbb{E}_S\left[\hat{F}_{\mathrm{uXEB}}|C\right] = F_C$ equals to the fidelity of C. This implies that

$$\mathrm{Var}_C\left(\mathop{\mathbb{E}}_{S}\left[\hat{F}_{\mathrm{uXEB}}|C\right]\right) = \mathrm{Var}_C\left(F_C\right), \tag{2.151}$$

which is the variance of fidelity across different random circuits. This term was ignored based on a concentration of measure argument [7], where they argued that for a large number of qubits the fidelity is the same across different random circuits. However, it is unclear how this term scales compared with the first term $1/M$, especially for a small number of qubits. In the following we drop the subscript C as in the previous subsection, and provide an analysis for $\mathrm{Var}(F)$.

For simplicity we work with the same model as in section 2.4.2.2, where we consider single qubit Pauli-X noise and assume a first-order approximation. Then the fidelity can be written as

$$F \approx (1-\varepsilon)^{nd} + n\varepsilon(1-\varepsilon)^{nd-1}\sum_{l=1}^{d} A_l \tag{2.152}$$

where $A_l := |\langle\psi_l|\psi\rangle|^2$. Then the variance can be approximated by

$$\begin{aligned}\mathrm{Var}(F) &\approx (n\varepsilon)^2 (1-\varepsilon)^{2nd-2}\,\mathrm{Var}\left(\sum_{l=1}^{d} A_l\right)\\ &\approx \lambda^2 (\mathbb{E}\,F)^2 \sum_{k,l=1}^{d} \left(\mathbb{E}\left[A_k A_l\right] - \mathbb{E}[A_k]\,\mathbb{E}[A_l]\right).\end{aligned} \tag{2.153}$$

Here we have used the fact that $\lambda \approx n\varepsilon$ and $\mathbb{E}\,F \approx e^{-\lambda d}$. Recall we have proven in section 2.4.2.2 that $\mathbb{E}\,A_l \leq e^{-\Delta l} + 1/2^n$ decays exponentially with l, and we therefore expect that each term in the above sum also decays exponentially, and the sum can be bounded by a constant,

$$\mathrm{Var}(F) = O\left(\lambda^2 (\mathbb{E}\,F)^2\right). \tag{2.154}$$

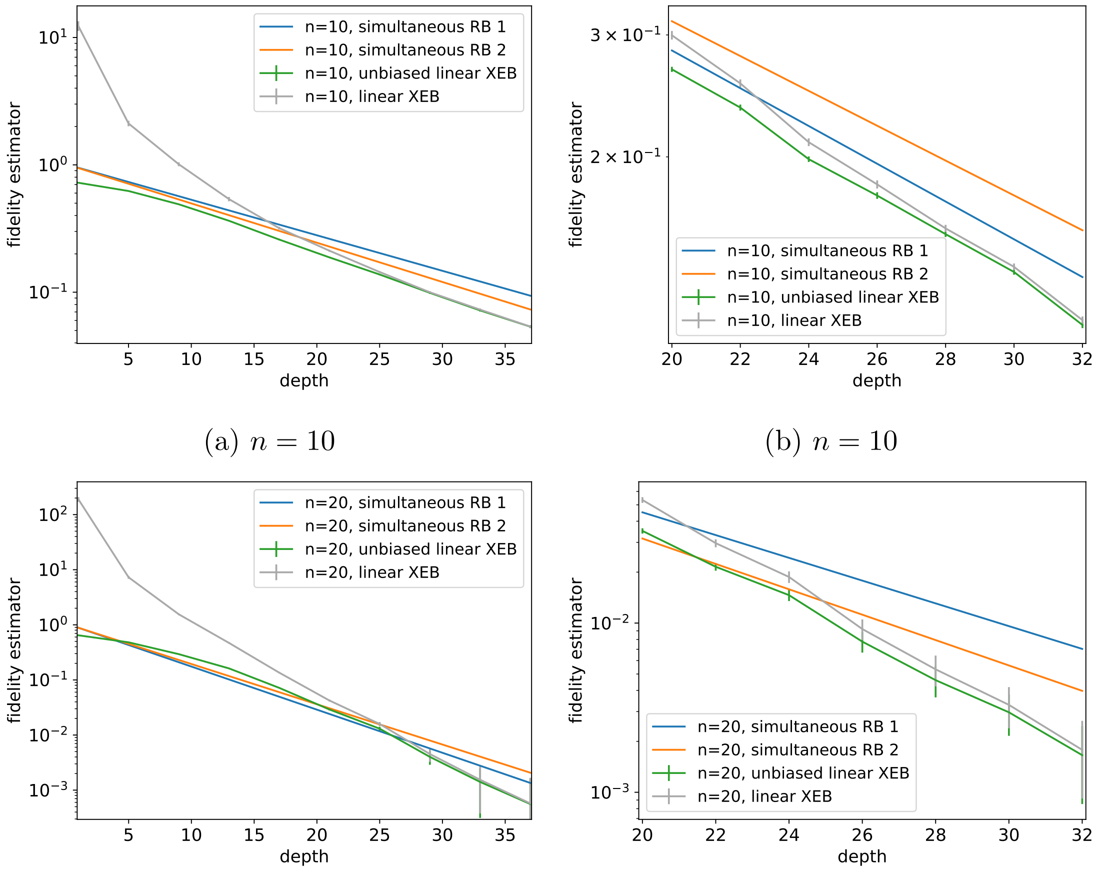
We leave the rigorous proof of this statement to future work. Note that the same proof technique for the analysis of $\mathbb{E}\,A_l$ can be applied here; however, the variance contains a fourth moment property $\mathbb{E}\left[A_k A_l\right]$, which corresponds to a more complicated spin model with negative weights, making it difficult to analytically bound the partition function.

Combining both terms, we obtain a theoretical model for the total variance

$$\mathrm{Var}_{C,S}(\hat{F}_{\mathrm{uXEB}}) = O\left(\frac{1}{M} + \lambda^2 (\mathbb{E}\,F)^2\right). \tag{2.155}$$

In the next section, we show that this model is well supported by experiment data on IBM Quantum hardware. Additional numerical simulation results verifying our conjecture that $\sum_{k,l=1}^{d} \left(\mathbb{E}\left[A_k A_l\right] - \mathbb{E}[A_k]\,\mathbb{E}[A_l]\right)$ can be upper bounded by a constant are presented in Section 2.4.6.

Finally, recall that in RCS benchmarking, L random circuits are implemented for a given depth, and M measurement samples are collected for each circuit. The empirical mean of the L cross entropy estimators is used to estimate $\mathbb{E}\,F$ at the given depth. The variance of this estimator is then given by $1/L \cdot O\left(1/M + \lambda^2 (\mathbb{E}\,F)^2\right)$. Assuming M is given, then the number of circuits required to estimate $\mathbb{E}\,F$ within ε additive error is $L = O\left(1/M\varepsilon^2 + \lambda^2 (\mathbb{E}\,F)^2/\varepsilon^2\right)$. This suggests that the strategy for choosing parameters for a RCS benchmarking experiment depends on the system size. Suppose that the effective noise rate per qubit λ/n is fixed, then $\mathbb{E}\,F \approx e^{-\lambda d}$ decays exponentially with the number of qubits. Therefore, for a large number of qubits, the variance across different circuits $\mathrm{Var}(F) = O\left(\lambda^2 (\mathbb{E}\,F)^2\right)$ is exponentially small, and a small L with a large number of samples M is sufficient for obtaining a small total variance. For example, in Google's experiment with 53 qubits only 10 random circuits are implemented. However, for a small number of qubits, the variance can be large even if we collect infinite number of samples for few random circuits, due to the second term. Therefore a large number of random circuits are needed, while the number of samples for each circuit can be smaller. For example, in our 5-qubit RCS benchmarking experiment shown below, we implement 100 random circuits and collect 8192 samples for each circuit for a given depth.



(a) $n = 10$

(b) $n = 10$

(c) $n = 20$

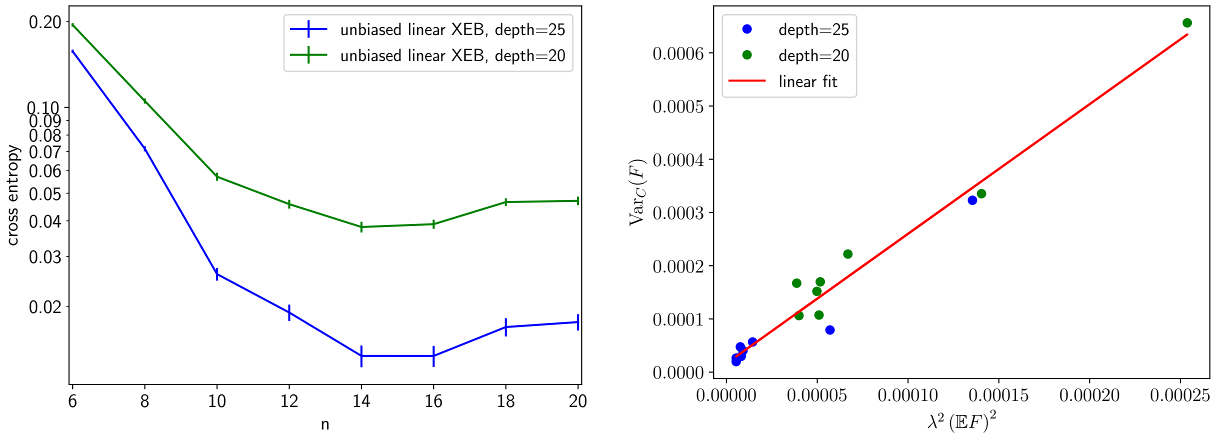
(d) $n = 20$

Figure 2.13: Experimental implementation of RCS benchmarking on `ibmq_mumbai` with 10 and 20 qubits. (a)(c) RCS benchmarking with long depth range (1-37) on 10 and 20 qubits, respectively. The linear cross entropy starts from a large value and converges to the unbiased one at depth 20-25. (b) RCS benchmarking with short depth range (20-32) on 10 qubits. Curve fitting results are $\lambda_{\text{uXEB}} = 6.9(1)\%$ and $\lambda_{\text{sRB}} = 6.0(2)\%$. (d) RCS benchmarking with short depth range (20-32) on 20 qubits. Curve fitting results are $\lambda_{\text{uXEB}} = 24.4(1)\%$ and $\lambda_{\text{sRB}} = 16.4(9)\%$.

These observations suggest that our generalized variance model is necessary, especially for characterizing RCS benchmarking experiments for a small number of qubits.

2.4.3 Experiments on IBM Quantum hardware

In the following we show experiment results of RCS benchmarking on IBM's superconducting qubits. We implement RCS benchmarking protocols on Qiskit platform [108] and access the



(a) Cross entropy

(b) Linear fit

Figure 2.14: Measuring cross entropy on `ibmq_montreal` with up to 20 qubits and varying depth. (a) Cross entropy as a function of the number of qubits, which does not obey a simple exponential decay. (b) Using these data points, we can verify our variance model $\mathrm{Var}_C(F) = O\left(\lambda^2\,(\mathbb{E}\,F)^2\right)$. A linear dependence can be observed even with noisy estimates for both sides. The results of the linear fit are slope $= 2.4(1)$, $r = 0.982$.

devices through cloud [79]. All of our experiments are performed on a 1D system of qubits, where we use a 5-qubit system `ibmq_athens` as well as 1D subsets of 27-qubit systems `ibmq_montreal` and `ibmq_mumbai` with up to 20 qubits. On these devices, CNOT is the only 2-qubit gate available, and arbitrary single-qubit gates are easy to implement, which have error rates that are roughly 2 orders of magnitude smaller than CNOT. Therefore, in RCS benchmarking we are effectively measuring the total amount of quantum noise of a layer of CNOT gates, where due to crosstalk this is larger than the sum of individual gate errors measured from individual RB. Details of the gate set and architecture can be found in Section 2.4.7.

Similar to our numerical simulation results presented in section 2.4.2, we perform three types of experiments: simultaneous RB, RCS with direct fidelity estimation, and RCS with cross entropy. These experimental results support our theoretical arguments for both fidelity decay and variance estimation.

By default, all error rates reported here are represented in Pauli error rates, which is a constant factor larger than the error rates based on average fidelity that is used in RB.

2.4.3.1 Simultaneous RB

In simultaneous RB [80], the main idea is to perform RB simultaneously with the gate pattern that is used in applications. That is, instead of performing RB on one pair of qubits while all other qubits are idle, different RB sequences are executed in parallel, which can capture correlated errors between different qubit pairs. A similar experiment was performed in Google's experiment [7] where linear cross entropy was used as the post processing method instead of standard RB.

As shown in Fig. 2.9b, in RCS benchmarking we consider three patterns for applying gates: one layer of single qubit gates, one layer of CNOT starting with qubit 0, and one layer of CNOT starting with qubit 1. We implement simultaneous RB on each of these three patterns, and the results are shown in Fig. 2.7. Here each row shows one of the three patterns, and a standard Clifford RB sequence is applied to each of the green boxes. After fitting the RB curve to an exponential decay, we obtain gate error rates shown below the green boxes. The numbers underlying single qubit boxes represent the error rates of two X90 pulses ($\sqrt{X}$), which can represent the Pauli error rate of a Haar random single qubit gate. The numbers underlying two qubit boxes represent the Pauli error rate of CNOT gates.

2.4.3.2 RCS benchmarking

Next we show results of RCS benchmarking with direct fidelity estimation and cross entropy. Note that DFE is not scalable due to the exponential sample complexity, and is implemented here mainly to verify our theoretical predictions. As a reference, we implement simultaneous RB both before and after the RCS experiments, which can also be used to evaluate the error drift during the experiment period.

We implement RCS benchmarking on a 5-qubit device `ibmq_athens` where direct fidelity estimation is tractable. Following our variance model, a large number of random circuits are implemented in order to achieve small total variance. We select 8 depth values ranging from 1 to 29. By default, we take the maximum amount of measurement samples allowed (8192) for all circuits submitted to the hardware platform. For direct fidelity estimation, we implement 30 random circuits for each depth, and estimate the fidelity of each circuit by measuring 20 Fourier coefficients according to the sampling procedure described in section 2.4.1.2. That is, 600 circuits are implemented for each depth. For cross entropy, we implement 100 random circuits for each depth.

The results of RCS benchmarking are shown in Fig. 2.7, where all curves are exponential decays with roughly the same decay rate. Note that for both DFE and uXEB experiments, the sample variance across different circuits is an unbiased estimator of the total variance. Therefore we use the standard error of the mean across different circuits as the error bar. The decay rate and its standard error can be computed from the data points and error bars in Fig. 2.7 via standard least squares fitting. For the simultaneous RB estimator, we use half the difference between the two RB experiments as the standard error.

From the curve fitting results, we can see that λ_{DFE} and λ_{uXEB} agrees with each other within the standard error. This confirms our theoretical results on the exponential decay of fidelity, and also verifies the validity of cross entropy as an efficient fidelity estimator. Also, note that uXEB is much more sample efficient than DFE, where the error bars for DFE are larger even when we collect 6 times the amount of samples in uXEB. In addition, note that the simultaneous RB estimator gives a slightly larger prediction for the effective noise rate. As we have discussed before, this could come from overestimating high weight correlated errors.

Similar RCS benchmarking results for 10 and 20 qubits are presented in Fig. 2.13, where we do not implement DFE as it requires too many samples. Detailed parameter settings for these experiments are given in Section 2.4.7. We perform two types of experiments: long range with depth 1-37 (Fig. 2.13a and c) for demonstrating the overall behavior, and short range with depth 20-32 (Fig. 2.13b and d) for fitting the curve. From the long range results, we can observe that the linear cross entropy starts from a large value at low depth, and converges to the unbiased version at depth around 20-25. Also, the unbiased linear cross entropy has a small "bump" at low depth, which is more evident for larger system size. We can observe similar behavior in our numerical simulations with 20 qubits (Fig. 2.5). Therefore, to obtain accurate predictions for the effective noise rate, we measure cross entropy for a short depth range starting from 20, and fit the unbiased linear cross entropy using these data points. Interestingly, in both 10 and 20 qubit results, the effective noise rate given by RCS benchmarking with the unbiased linear cross entropy is larger than the simultaneous RB predictions. Assuming our depth fitting range is deep enough so that uXEB correctly estimates fidelity (which is the case in our simulations), this suggests that some error sources were captured by RCS but not by simultaneous RB. These errors should not have come from high weight correlated noise, as in this case simultaneous RB will overestimate instead of underestimate. We leave the identification of these error sources as future work. One possible explanation for these additional errors captured by RCS benchmarking is that in RCS all CNOT gates are running at the same time in each layer, which is not the case in sRB which has independent Clifford sequences. The simultaneous CNOT gates in RCS could introduce additional couplings, leading to a larger effective noise rate. In addition, note that the effective noise rate for 20 qubits is much larger than twice the effective noise rate for 10 qubits on the same device, which suggests that the errors are highly non-uniformly distributed across the device.

Finally, taking the result $\lambda_{\text{uXEB}} = 3.08(3)\%$ from Fig. 2.7 as an example, we can compute the effective noise rate per qubit $\lambda_q = \lambda/n = 0.62(1)\%$. This number represents the average quality of the quantum system, which we can understand as the error rate of "half" of a 2-qubit gate plus a single qubit gate when the gates are implemented in parallel. As a comparison, we can compute from the curve fitting data of Google's experiment [7, 109] that $\lambda_q = 0.43(4)\%$. Recall that our Theorem 2.13 suggests that a smaller λ_q means that a larger number of qubits can be benchmarked together via RCS benchmarking, although the specific constants are unclear. Google's experiment can be interpreted as evidence that $\lambda_q = 0.43\%$ is small enough to implement RCS benchmarking on 53 qubits, with their gate set and architecture. We expect that a smaller error rate is needed to benchmark the same number of qubits with 1D connectivity compared with 2D. This is because in some cases random circuits with 2D connectivity are known to scramble faster than 1D [67], and therefore we expect 2D circuits to have a smaller constant K in our Theorem 2.13.

2.4.3.3 Cross entropy with increasing number of qubits

Next we present results on cross entropy with increasing number of qubits. This experiment shows how cross entropy changes with more qubits added to the subset, and also gives data points that allow us to verify our variance model. We start with a 1D subset of 6 qubits on a 27-qubit device `ibmq_montreal`, and add more qubits to the subset until it forms a 1D chain of 20 qubits.

Fig. 2.14a shows the cross entropy results for depth 20 and 25. Each data point in this figure is collected by implementing 100 random circuits. The two curves correspond to different depths have consistent shapes, and the error bars for the blue curve is larger than the error bars in the green curve, which qualitatively agrees with our prediction. Interestingly, different from the results shown in Google's experiment [7], here the cross entropy does not decay as a simple exponential function with the number of qubits. This could potentially suggest that adding more qubits to the subset can dramatically change the error pattern, indicating complicated error correlations among the qubits. In particular, as more qubits are added to the subset, the pattern of edge effects and dangling qubits also changes, which can affect the noise of the subset of qubits being benchmarked. For example, in Fig. 2.14a the two system sizes with lowest cross entropy ($n = 14, 16$) correspond to having additional dangling qubits (qubit 20 and 13 as shown in Fig. 2.22c) at the edge.

The data points collected in Fig. 2.14a can be used to verify our variance model. Recall that our variance model $\mathrm{Var}_{C,S}(\hat{F}_{\mathrm{uXEB}}) = \mathbb{E}_C\left[\mathrm{Var}_S\left(\hat{F}_{\mathrm{uXEB}}|C\right)\right] + \mathrm{Var}_C\left(\mathbb{E}_S\left[\hat{F}_{\mathrm{uXEB}}|C\right]\right) = O\left(1/M + \lambda^2\left(\mathbb{E}\,F\right)^2\right)$ (Eq. (2.155)) states that the total variance equals the variance across samples and the variance across circuits, and we would like to verify our model for the second term $\mathrm{Var}_C(F) = O\left(\lambda^2\left(\mathbb{E}\,F\right)^2\right)$. For this purpose, we use a rough estimate for both sides from experiment data, and test the linear dependence. For the right hand side, we use $\lambda^2\left(\mathbb{E}\,F\right)^2 \approx (\mathbb{E}\,F\log\mathbb{E}\,F)^2/d^2$ and substitute $\mathbb{E}\,F$ with the empirical mean of $\hat{F}_{\mathrm{uXEB}}$. For the left hand side, we use the sample variance of $\hat{F}_{\mathrm{uXEB}}$ across different circuits minus the average sample variance across the measurement outcome probabilities for each circuit. This value is an unbiased estimator for $\mathrm{Var}_C(F)$. Even though the estimates for both sides are very noisy, we can still observe a linear dependence shown in Fig. 2.14b. This verifies our model $\mathrm{Var}_C(F) = O\left(\lambda^2\left(\mathbb{E}\,F\right)^2\right)$, which suggests that $\mathrm{Var}_C(F)$ does not directly depend on system size or depth, while the constant depends on circuit architecture and gate set.

2.4.4 Additional discussions

We develop RCS benchmarking, which allows sample efficient benchmarking of the total amount of quantum noise of a many-body system using only shallow circuits and local randomness. Unlike RB and variants where the fidelity decay comes from the group structure, RCS benchmarking uses the scrambling properties of random quantum circuits, which can be directly implemented with the native gate set on any hardware platform without additional compilation. While we consider qubits with 1D connectivity, we expect our results to be naturally generalized to more general connectivity graphs and higher local Hilbert space dimensions.

An interesting future direction is to design variants of RCS benchmarking such that more information about the noise channel can be extracted. For example, recent work [110] showed that improved post-processing techniques can be used to classify incoherent vs coherent noise. In addition, we can consider RCS benchmarking with a growing number of qubits (such as in Fig. 2.14) and extract noise correlations from the data. Also, similar to simultaneous RB, we can also consider simultaneous RCS where different subsets of the qubits are benchmarked at the same time. This allows more flexible subset choices than simultaneous RB, and also provides a possible way to avoid the computation barrier. Finally, it is interesting to consider if the learning algorithms for Pauli noise channels [60, 49, 61, 62] can be extended to general gate sets, using ideas from RCS benchmarking.

2.4.5 Detailed proofs

2.4.5.1 Mapping random quantum circuits to a classical spin model

We first describe the basic idea in our analysis of RCS benchmarking which uses the technique that maps random quantum circuits to a classical spin model. This technique has been used to study several other properties of random quantum circuits, see e.g. [70, 71, 72, 73, 39, 35].

In general our analysis of RCS benchmarking concerns the following object

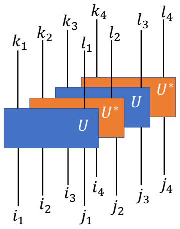
$$\mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,d)}\left[\text{A second moment property of } C\right]. \tag{2.156}$$

Here the second moment property can be understood as the following: for each individual gate U_i in C, this property is a linear function in the second moment of U_i, which is $U_i^{\otimes 2}\otimes U_i^{*\otimes 2}$, where U_i^* is the complex conjugate of U_i. By linearity of expectation, we have

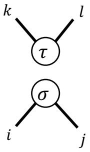
$$\begin{aligned}
&\mathop{\mathbb{E}}_{C=U_m\cdots U_1\sim\mathrm{RQC}(n,d)}\left[\text{A second moment property of } C\right]\\
&=\mathop{\mathbb{E}}_{U_m\sim\mathbb{U}(4)}\cdots\mathop{\mathbb{E}}_{U_1\sim\mathbb{U}(4)}\left[\text{A second moment property of } C\right]\\
&=\text{A linear function of}\mathop{\mathbb{E}}_{U_i\sim\mathbb{U}(4)}\left[U_i^{\otimes 2}\otimes U_i^{*\otimes 2}\right].
\end{aligned}\tag{2.157}$$

After taking the expectation, the above equation can be viewed as a tensor network, where the basic element in the above object is the following local tensor (re-ordered for convenience):

$$\mathop{\mathbb{E}}_{U\sim\mathbb{U}(4)}\left[U\otimes U^*\otimes U\otimes U^*\right]=\mathop{\mathbb{E}}_{U\sim\mathbb{U}(4)}\Bigg[$$



$$\Bigg]=\sum_{\tau,\sigma\in\{-1,1\}}w(\tau,\sigma)\cdot$$

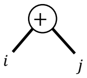


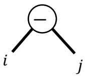
(2.158)

where τ, σ are classical variables taking values in $\{-1, 1\}$, and the white tensors depend on their values. $w(\tau, \sigma)$ are additional weights given by

$$w(\tau, \sigma) = \frac{1}{12}\delta_{\tau\sigma} - \frac{1}{60} = \begin{cases} \frac{1}{15}, & \tau = \sigma, \\ -\frac{1}{60}, & \tau \neq \sigma. \end{cases} \tag{2.159}$$

The white σ/τ tensors have the following form:

- $\sigma/\tau = +1$: 

$= \delta_{i_1 i_2} \cdot \delta_{i_3 i_4} \cdot \delta_{j_1 j_2} \cdot \delta_{j_3 j_4}$,
- $\sigma/\tau = -1$: 

$= \delta_{i_1 i_4} \cdot \delta_{i_2 i_3} \cdot \delta_{j_1 j_4} \cdot \delta_{j_2 j_3}$.

After replacing all two-qubit gates with the above tensor, we have mapped any second moment property of random quantum circuits to a classical spin model with the τ, σ variables. The original quantity in Eq. (2.156) corresponds to the partition function of this spin model. For each specific assignment of τ, σ, Eq. (2.158) is a separable tensor, and the value corresponding to the assignment is easy to calculate. However, this spin model has negative weights, making it hard to directly analyze in this form.

2.4.5.2 Proof of Theorem 2.12

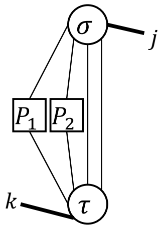
Let $C \sim \mathrm{RQC}(n, d)$ be a random circuit, $|\psi\rangle = C|0^n\rangle$ be the ideal output state, and ρ be the output state of the noisy circuit. It is easy to see that the fidelity

$$F = \langle\psi|\rho|\psi\rangle = \mathrm{Tr}\left[\rho \cdot |\psi\rangle\langle\psi|\right] \tag{2.160}$$

is a second moment property, because trace is a linear operation and each two-qubit gate U_i appears in the fidelity as the form $U_i^{\otimes 2} \otimes U_i^{*\otimes 2}$.

Let $\mathcal{N}(\rho) = \sum_{\alpha,\beta \in \{0,1,2,3\}^n} \chi_{\alpha\beta}\sigma_\alpha \rho \sigma_\beta$ be an arbitrary noise channel which acts on all qubits after each layer of gates. Due to the linearity of quantum channels, we only need to prove that the off-diagonal terms $\sigma_\alpha \rho \sigma_\beta$ ($\alpha \neq \beta$) have 0 contribution in the calculation of $\mathbb{E}\, F$, which implies that the average fidelity with $\mathcal{N}$ is equal to the average fidelity of $\mathcal{N}^{\mathrm{diag}}$, which proves Theorem 1.

Let $\sigma_\alpha \neq \sigma_\beta \in \{I, X, Y, Z\}^{\otimes n}$ be two distinct Pauli operators. They have to differ by at least one location, and we denote the single qubit Pauli operator as $P_1 \neq P_2 \in \{I, X, Y, Z\}$. Without loss of generality, we assume P_1 and P_2 act on the l_1, l_2 index of a two-qubit gate and the i_1, i_2 index of another two-qubit gate. We perform the mapping described in the previous subsection which maps the average fidelity to the partition function of a classical spin model. The relevant local tensor for P_1 and P_2 is given by



(2.161)

and it is easy to see that this tensor always equals to 0 no matter what value σ and τ takes, which makes the average fidelity equal to 0 under the off-diagonal term.

2.4.5.3 Analysis of average fidelity

Next we analyze $\mathbb{E}\,F$ under Pauli noise channels. As discussed above, $\mathbb{E}\,F$ can be directly mapped to the partition function of a classical spin model with $2m$ spins, where m is the number of two-qubit gates. However, as this spin model has negative weights due to the noise channel, it appears difficult to directly analyze its partition function.

Our first idea is to separate the effect of noise from random quantum circuits via an expansion in the noise rate. As described in section 2.4.2.2, we first focus on i.i.d. single qubit Pauli-X noise and later address more general cases. Following section 2.4.2.2 we write the average fidelity as

$$\begin{aligned}\mathbb{E}\,F &= (1-\varepsilon)^{nd} + \sum_{i=1}^{nd} \varepsilon(1-\varepsilon)^{nd-1}\,\mathbb{E}\,|\langle\psi_i|\psi\rangle|^2 + \sum_{i=1}^{nd-1}\sum_{j=i+1}^{nd} \varepsilon^2(1-\varepsilon)^{nd-2}\,\mathbb{E}\,|\langle\psi_{ij}|\psi\rangle|^2 + \cdots \\ &:= F_0 + \mathbb{E}\,F_1 + \sum_{k\geq 2}\mathbb{E}\,F_k.\end{aligned} \tag{2.162}$$

Here $|\psi_i\rangle$ denotes the ideal state with an X error at location i. We can understand $\mathbb{E}\,F_k$ as the contribution to average fidelity with k errors in the circuit, which has an ε^k factor. Therefore we expect the higher order contributions to be small compared with $\mathbb{E}\,F_1$, when ε is sufficiently small.

To verify this intuition, we present numerical simulation results shown in Fig. 2.15. Here the simulation is performed by simulating the partition function of the corresponding spin model and therefore has no error bar. For $n = 20$ qubits with $\varepsilon \leq 0.01$ Pauli-X noise per qubit, we plot the ratio $\mathbb{E}\,F_1/F_0$ as well as higher order terms $\mathbb{E}\,F_{\geq 2}/F_0$, where $\mathbb{E}\,F_{\geq 2} = \sum_{k\geq 2}\mathbb{E}\,F_k$. From the simulation results we can observe the following:

- First, as we will rigorously prove next, $\mathbb{E}\,F_1/F_0$ converges to a constant that is proportional to ε as depth increases.
- Second, note that $\mathbb{E}\,F_{\geq 2}/F_0$ is much smaller compared with $\mathbb{E}\,F_1/F_0$ for shallow depth circuits. This verifies our intuition to throw away higher order terms in our analysis

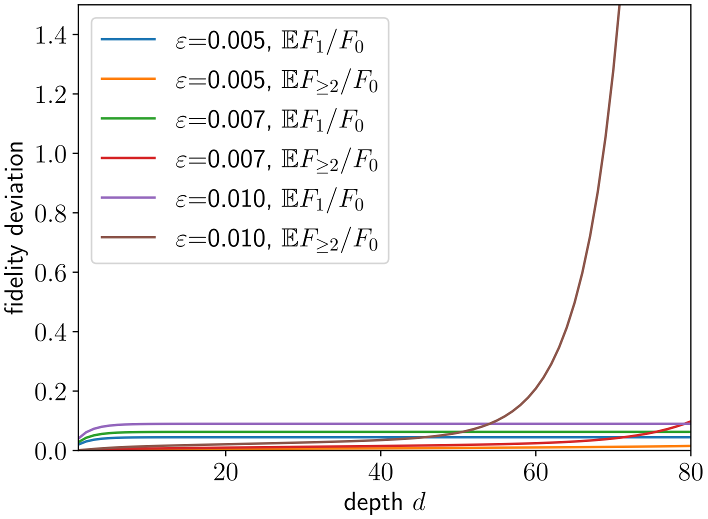


Figure 2.15: Simulation of the first and higher order terms in average fidelity. Here we consider $n = 20$ qubits on a 1D ring and i.i.d. Pauli-X noise with effective noise rate per qubit ε.

of RCS benchmarking experiments, as in experiments we are limited to implementing shallow depth circuits.

- Third, note that $\mathbb{E}\, F_{\geq 2}/F_0 \to \infty$ as $d \to \infty$.

Next we provide a simple argument to explain the third point. As $d \to \infty$, we have $F_0 = (1-\varepsilon)^{nd} \to 0$, and $\mathbb{E}\, F_1 \leq O(F_0) \to 0$. However, we know that $\mathbb{E}\, F \to \frac{1}{2^n}$ as the state converges to the maximally mixed state on average. Taking the limit $d \to \infty$ on both sides of Eq. (2.162), we have $\mathbb{E}\, F_{\geq 2} \to \frac{1}{2^n}$, and therefore $\mathbb{E}\, F_{\geq 2}/F_0 \to \infty$. This observation suggests a potential improvement to RCS benchmarking. In particular, we can consider other fitting schemes, such as fitting an exponential decay with $\mathbb{E}\, F - \frac{1}{2^n}$ instead of $\mathbb{E}\, F$, which might extend the depth range for which RCS benchmarking gives accurate results.

Next we proceed by analyzing the first order term $\mathbb{E}\, F_1$ and prove Theorem 2.13. Following the proof sketch in section 2.4.2.2, we focus on proving

$$\mathbb{E}\, |\langle \psi_l | \psi \rangle|^2 = \mathop{\mathbb{E}}_{C \sim \mathrm{RQC}(n,l)} \left| \langle 0^n | \, C^\dagger \sigma_p C \, | 0^n \rangle \right|^2 \leq e^{-\Delta l} + \frac{1}{2^n}, \tag{2.163}$$

where $\sigma_p \in \{I, X, Y, Z\}^{\otimes n}$ is a Pauli operator. We will formally prove Eq. (2.163) for arbitrary 3-local Pauli errors, and show numerical evidence that shows this holds in general.

It is easy to see that Eq. (2.163) implies that $\mathbb{E}\, F_1/F_0 = O(\lambda)$ where λ is the effective noise rate for the global noise channel. For example, for i.i.d. single qubit Pauli-X noise channel where $\lambda \approx n\varepsilon$, Eq. (2.163) with $\sigma_p = X \otimes I^{\otimes n-1}$ implies that

$$\begin{aligned}\mathbb{E}\,F_1/F_0 &= \frac{n\varepsilon}{1-\varepsilon}\sum_{l=1}^{d}\mathbb{E}\,|\langle\psi_l|\psi\rangle|^2 \\ &\leq \frac{n\varepsilon}{1-\varepsilon}\sum_{l=1}^{d}\left(e^{-\Delta l}+\frac{1}{2^n}\right) \\ &= O(n\varepsilon)\end{aligned} \tag{2.164}$$

when $d \leq 2^n$. We can similarly prove $\mathbb{E}\,F_1/F_0 = O(\lambda)$ for general noise channels, as long as Eq. (2.163) still holds, where we replace the single qubit Pauli-X with the Pauli operator in the noise channel.

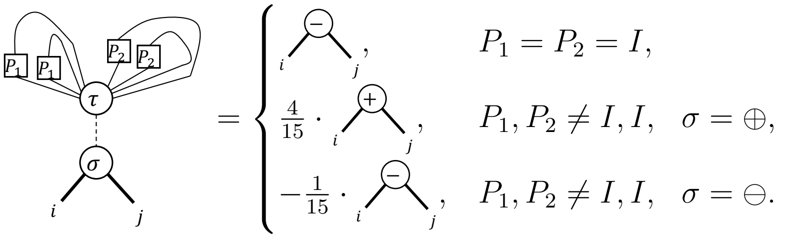
To complete our analysis on the first order term $\mathbb{E}\,F_1$, it remains to prove the upper bound on $\mathbb{E}_{C\sim\mathrm{RQC}(n,l)}\left|\langle 0^n|\,C^\dagger\sigma_p C\,|0^n\rangle\right|^2$. First note that this is a second moment property of ideal depth-l random quantum circuits, as

$$\mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,l)}\left|\langle 0^n|\,C^\dagger\sigma_p C\,|0^n\rangle\right|^2 = \mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,l)}\mathrm{Tr}\left[\sigma_p C\,|0^n\rangle\langle 0^n|\,C^\dagger\sigma_p\cdot C\,|0^n\rangle\langle 0^n|\,C^\dagger\right]. \tag{2.165}$$

We therefore map this quantity to the partition function of a classical spin model, where the bulk of the spin model is a hexagonal lattice consists of local tensors given in Eq. (2.158).

This was first developed by [70], where it was noted that this spin model can be further simplified by first summing over the τ variables, and as a result eliminates the negative weights in $w(\tau,\sigma)$. After performing this summation over τ variables, the spin model corresponds to a triangular lattice of the σ variables shown in Fig. 2.10, and the bulk of the weights corresponding to the three-body interaction are given in Fig. 2.11. We refer to Refs. [70, 71, 72, 73, 39, 35] for the detailed calculations.

Next we derive the boundary conditions in order to fully specify the spin model for Eq. (2.165). The bottom boundary corresponds to open boundary conditions for all σ variables, as the input is fixed to be 0. At the top boundary, for each two-qubit gate there is a Pauli operator $P_1 \otimes P_2 \in \{I,X,Y,Z\}^{\otimes 2}$ acting on it. Summing over the τ variables for each two-qubit gate at the top, we derive the following boundary condition:



(2.166)

Here, the dashed line denotes the $w(\tau,\sigma)$ interaction. This can be summarized as follows:

1. When there is no error, the two-qubit gate corresponds to a fixed $\ominus$ spin.

2. When there is one or two Pauli errors, the σ spin has an additional weight $\frac{4}{15}$ when it equals to $\oplus$, and $-\frac{1}{15}$ when it equals to $\ominus$.

Now we are ready to calculate the expectation value in Eq. (2.165). Define $\mathcal{Z}(n,l;b)$ as the partition function of the spin model which corresponds to n-qubit, depth-l random quantum circuits, where the top boundary condition is given by $b \in \{\oplus,\ominus\}^{n/2}$. We start with single qubit Pauli noise, where $\sigma_p = X \otimes I^{\otimes n-1}$. In this case, there is one spin at the top which has the second boundary condition, while all other spins are fixed to be $\ominus$. Therefore

$$\begin{aligned}\mathop{\mathbb{E}}_{C\sim \mathrm{RQC}(n,l)} \left|\langle 0^n| C^\dagger X C |0^n\rangle\right|^2 &= \frac{4}{15}\mathcal{Z}(n,l;\cdots\ominus\oplus\ominus\cdots) - \frac{1}{15}\mathcal{Z}(n,l;\cdots\ominus\ominus\ominus\cdots)\\ &= \frac{4}{15}\mathcal{Z}(n,l;\cdots\ominus\oplus\ominus\cdots) - \frac{1}{15},\end{aligned} \tag{2.167}$$

where in the second line we use the observation that the partition function equals to 1 when all spins at the top boundary are equal to $\ominus$. This proves the equation in Fig. 2.10.

Following the argument in section 2.4.2.2 and demonstrated in Fig. 2.11, we can divide the partition function into two parts,

$$\mathcal{Z}(n,l;\cdots\ominus\oplus\ominus\cdots) = \mathcal{Z}_1(n,l;\cdots\ominus\oplus\ominus\cdots) + \mathcal{Z}_2(n,l;\cdots\ominus\oplus\ominus\cdots) \tag{2.168}$$

where $\mathcal{Z}_i$ denotes the sum of weights of domain wall configuration of type i shown in Fig. 2.11b. First, note that the analysis of $\mathcal{Z}_1$ is simple. Each of the two domain walls has length $l-1$ and weight $\left(\frac{2}{5}\right)^{l-1}$. For each domain wall, the number of possible configurations is at most 2^{l-1}. Thus, the contribution of two domain walls is at most

$$\mathcal{Z}_1(n,l;\cdots\ominus\oplus\ominus\cdots) \le \left(\frac{4}{5}\right)^{2(l-1)}. \tag{2.169}$$

In particular, this implies that $\mathcal{Z}_1(n,\infty;\cdots\ominus\oplus\ominus\cdots) = 0$. Next, we use the following useful fact: when $l \to \infty$, depth-l random quantum circuits converge to an exact unitary 2-design [66], which implies that

$$\lim_{l\to\infty} \mathop{\mathbb{E}}_{C\sim \mathrm{RQC}(n,l)} \left|\langle 0^n| C^\dagger \sigma_p C |0^n\rangle\right|^2 = \mathop{\mathbb{E}}_{C\sim \mathbb{U}(2^n)} \left|\langle 0^n| C^\dagger \sigma_p C |0^n\rangle\right|^2 = \frac{1}{2^n+1}, \tag{2.170}$$

for any non-zero $\sigma_p \in \{I,X,Y,Z\}^{\otimes n}$. See e.g. [45] for a proof of the second equality. Therefore,

$$\begin{aligned}\frac{1}{2^n+1} &= \lim_{l\to\infty} \mathop{\mathbb{E}}_{C\sim \mathrm{RQC}(n,l)} \left|\langle 0^n| C^\dagger X C |0^n\rangle\right|^2\\ &= \frac{4}{15}\left(\mathcal{Z}_1(n,\infty;\cdots\ominus\oplus\ominus\cdots) + \mathcal{Z}_2(n,\infty;\cdots\ominus\oplus\ominus\cdots)\right) - \frac{1}{15}\\ &= \frac{4}{15}\mathcal{Z}_2(n,\infty;\cdots\ominus\oplus\ominus\cdots) - \frac{1}{15}.\end{aligned} \tag{2.171}$$

Combining the above facts, we have

$$\begin{aligned}\mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,l)}\left|\langle 0^n|\, C^\dagger X C\,|0^n\rangle\right|^2 &= \frac{4}{15}\left(\mathcal{Z}_1(n,l;\cdots\ominus\oplus\ominus\cdots)+\mathcal{Z}_2(n,l;\cdots\ominus\oplus\ominus\cdots)\right)-\frac{1}{15}\\ &\le \frac{4}{15}\left(\frac{4}{5}\right)^{2(l-1)}+\frac{4}{15}\mathcal{Z}_2(n,l;\cdots\ominus\oplus\ominus\cdots)-\frac{1}{15}\\ &\le \frac{4}{15}\left(\frac{4}{5}\right)^{2(l-1)}+\frac{4}{15}\mathcal{Z}_2(n,\infty;\cdots\ominus\oplus\ominus\cdots)-\frac{1}{15}\\ &= \frac{4}{15}\left(\frac{4}{5}\right)^{2(l-1)}+\frac{1}{2^n+1}.\end{aligned} \tag{2.172}$$

Here, the third line follows from the simple observation that $\mathcal{Z}_2(n,l;\cdots\ominus\oplus\ominus\cdots)$ is monotonically non-decreasing with respect to l. This proves Eq. (2.163) for arbitrary 1-local errors.

Next we extend the above argument to arbitrary 3-local errors, of the form $X\otimes X\otimes X\otimes I_{\text{else}}$, $X\otimes I\otimes Z\otimes I_{\text{else}}$, etc. The 2-local case either reduces to 1-local (when the two errors act on the same two-qubit gate) or 3-local (when the two errors act on different two-qubit gates). For arbitrary 3-local errors, the top boundary has two neighboring spins with the second boundary condition, and all other spins are fixed to be $\ominus$. After calculating the boundary conditions, we have

$$\mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,l)}\left|\langle 0^n|\, C^\dagger \sigma_{\text{3-local}} C\,|0^n\rangle\right|^2 = \frac{16}{225}\mathcal{Z}(n,l;\cdots\ominus\oplus\oplus\ominus\cdots)-\frac{8}{225}\mathcal{Z}(n,l;\cdots\ominus\oplus\ominus\cdots)+\frac{1}{225}. \tag{2.173}$$

To relate the first two terms, we derive the following recursive formula for the partition function:

$$\mathcal{Z}(n,l;\cdots\ominus\oplus\ominus\cdots)=\frac{4}{25}\mathcal{Z}(n,l-1;\cdots\ominus\oplus\oplus\ominus\cdots)+\frac{8}{25}\mathcal{Z}(n,l-1;\cdots\ominus\oplus\ominus\cdots)+\frac{4}{25}. \tag{2.174}$$

Combining both equations above, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,l)}\left|\langle 0^n|\, C^\dagger \sigma_{\text{3-local}} C\,|0^n\rangle\right|^2\\ &=\frac{4}{9}\mathcal{Z}(n,l+1;\cdots\ominus\oplus\ominus\cdots)-\frac{8}{45}\mathcal{Z}(n,l;\cdots\ominus\oplus\ominus\cdots)-\frac{1}{15}\\ &\le O\left(e^{-\Delta l}\right)+\frac{4}{9}\mathcal{Z}_2(n,l+1;\cdots\ominus\oplus\ominus\cdots)-\frac{8}{45}\mathcal{Z}_2(n,l;\cdots\ominus\oplus\ominus\cdots)-\frac{1}{15}.\end{aligned} \tag{2.175}$$

Here in the inequality we have combined the exponentially decaying $\mathcal{Z}_1$ terms. Now, notice that the difference between $\mathcal{Z}_2(n,l+1;\cdots\ominus\oplus\ominus\cdots)$ and $\mathcal{Z}_2(n,l;\cdots\ominus\oplus\ominus\cdots)$ can be bounded: it corresponds to domain walls with depth l, and therefore has weight that is

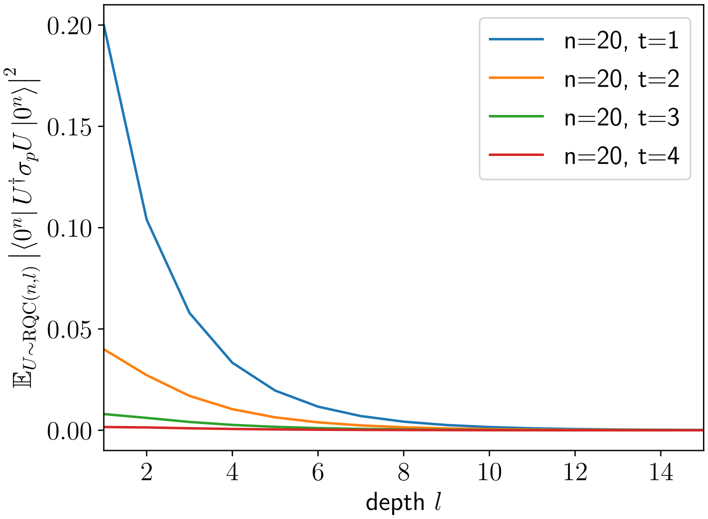


Figure 2.16: Simulation of the second moment of Pauli observable for depth-l random quantum circuits on $n = 20$ qubits, which corresponds to a classical spin model with depth l and width 10. Here t denotes the number of consecutive spins that has the second boundary condition.

exponentially small in l, which gives

$$\begin{aligned}
&\mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,l)} \left|\langle 0^n| C^\dagger \sigma_{\text{3-local}} C |0^n\rangle\right|^2 \\
&\leq O\left(e^{-\Delta l}\right) + \frac{4}{9}\mathcal{Z}_2(n, l+1; \cdots \ominus \oplus \ominus \cdots) - \frac{8}{45}\mathcal{Z}_2(n, l; \cdots \ominus \oplus \ominus \cdots) - \frac{1}{15} \\
&\leq O\left(e^{-\Delta l}\right) + \frac{4}{9}\left(\mathcal{Z}_2(n, l; \cdots \ominus \oplus \ominus \cdots) + O\left(e^{-\Delta l}\right)\right) - \frac{8}{45}\mathcal{Z}_2(n, l; \cdots \ominus \oplus \ominus \cdots) - \frac{1}{15} \\
&= O(e^{-\Delta l}) + \frac{4}{15}\mathcal{Z}_2(n, l; \cdots \ominus \oplus \ominus \cdots) - \frac{1}{15} \\
&\leq O(e^{-\Delta l}) + \frac{4}{15}\mathcal{Z}_2(n, \infty; \cdots \ominus \oplus \ominus \cdots) - \frac{1}{15} \\
&= O(e^{-\Delta l}) + \frac{1}{2^n + 1}.
\end{aligned} \tag{2.176}$$

This proves Eq. (2.163) for arbitrary 3-local errors, and concludes the proof of Theorem 2.13.

Finally we present numerical simulation results which verify our proof and also show the correctness of Eq. (2.163) for higher weight errors. In Fig. 2.16 we show simulation results for depth-l random quantum circuits on 20 qubits. Here we simulate the spin model of width 10, with t consecutive spins at the top which have the second boundary condition. t corresponds to the locality of the Pauli error, as for a k-local Pauli operator we have $\frac{k}{2} \leq t \leq \frac{k}{2} + 1$.

From Fig. 2.16 we can observe exponential decays as l increases, which eventually converge to $\frac{1}{2^n+1}$. Interestingly, the curve also decreases as t increases. We have already proven the exponential decay for $t = 1$ (blue curve). The simulation results suggest that Eq. (2.163) can actually be strengthened for higher weight errors. Here we conjecture that

$$\mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,l)} \left|\langle 0^n|\, C^\dagger \sigma_{k\text{-local}} C\, |0^n\rangle\right|^2 \le e^{-O(k)} e^{-\Delta l} + \frac{1}{2^n}, \tag{2.177}$$

which follows intuitively because of the additional factor of $\sim (4/15)^t$ due to the second boundary condition.

2.4.5.4 Alternative proof idea

Here we also present an alternative proof idea for analyzing $\mathbb{E}_{C\sim\mathrm{RQC}(n,l)} \left|\langle 0^n|\, C^\dagger \sigma_p C\, |0^n\rangle\right|^2$ using the result of [66]. This proof idea works for arbitrary Pauli observables but has an additional 2^n factor due to converting between different norms, and therefore does not satisfy our purpose for proving Eq. (2.163). It is interesting to see whether this idea can be improved to directly prove Eq. (2.163) with arbitrary locality.

First, recall the following result from [66],

$$\left\| \mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,l)} C^{\otimes t} \otimes (C^*)^{\otimes t} - \mathop{\mathbb{E}}_{C\sim\mathbb{U}(2^n)} C^{\otimes t} \otimes (C^*)^{\otimes t} \right\|_{2\to 2} \le e^{-\Delta l}, \tag{2.178}$$

where $\mathrm{RQC}(n,l)$ denotes depth-l random quantum circuits on n qubits, $\|\cdot\|_{2\to 2}$ denotes the super-operator 2 norm, and Δ is a constant that depends on t. When $t = 2$, $\mathbb{E}[C^{\otimes 2} \otimes (C^*)^{\otimes 2}]$ can be understood as a quantum channel that acts on $2n$ qubits. As we have a fixed input $|0^n\rangle\!\langle 0^n|^{\otimes 2}$, Eq. (2.178) implies that the output states are close in 2 norm, which combined with Lemma 2.21 gives

$$\mathop{\mathbb{E}}_{C\sim\mathrm{RQC}(n,l)} \left|\langle 0^n|\, C^\dagger \sigma_p C\, |0^n\rangle\right|^2 \le 2^n e^{-\Delta l} + \frac{1}{2^n+1}, \tag{2.179}$$

which has an additional dimension factor 2^n due to the conversion between 1 and 2 norm. However, our numerical simulation results in Fig. 2.16 suggest that this additional 2^n factor can be replaced by $O(1)$, for single qubit as well as high weight Pauli errors.

Lemma 2.21. *For a Hermitian observable W and n-qubit quantum states ρ, σ, we have*

$$\mathrm{Tr}[W(\rho-\sigma)] \le 2^{n/2} \|W\|_\infty \cdot \|\rho-\sigma\|_2, \tag{2.180}$$

where $\|\cdot\|_p$ denotes the Schatten p-norm.

Proof. Let $\lambda_1 \ge \cdots \ge \lambda_{2^n}$ denote the absolute eigenvalues of $\rho - \sigma$. Then

$$\begin{aligned} \mathrm{Tr}[W(\rho-\sigma)] &\le \|W\|_\infty (\lambda_1 + \cdots + \lambda_{2^n}) \\ &\le 2^{\frac{n}{2}} \|W\|_\infty (\lambda_1^2 + \cdots + \lambda_{2^n}^2)^{\frac{1}{2}} \\ &= 2^{\frac{n}{2}} \|W\|_\infty \cdot \|\rho-\sigma\|_2. \end{aligned} \tag{2.181}$$

□

	Name	Lindblad Superoperator	Description	ENR
Single qubit noise	T_1	$\gamma D[\sigma]$	Amplitude Decay	$\gamma/2$
	T_ϕ	$\gamma D[\sigma^\dagger\sigma]$	Dephasing	$\gamma/4$
	Pauli-X	$\gamma D[X]$	Pauli-X	γ
Correlated noise	Corr-T_1	$\gamma D[\sigma_i\sigma_j]$	Amplitude Decay	$\gamma/4$
	Corr-T_ϕ	$\gamma D[\sigma_i^\dagger\sigma_i\sigma_j^\dagger\sigma_j]$	Dephasing	$3\gamma/16$
	Corr-Pauli	$\gamma D[\sigma_p]$	Arbitrary Pauli noise	γ

Table 2.4: Effective noise rate for various noise models, including those studied here. Here $\sigma = |0\rangle\langle 1|$, $\sigma_p \in \{I, X, Y, Z\}^{\otimes n}$. γ is the noise strength used in simulating the Lindblad evolution, and ENR stands for effective noise rate.

2.4.6 Additional numerical simulation results

2.4.6.1 Computing the effective noise rate

In this section we show how to compute the effective noise rate given the description of a noise model, where the results are given in Table 2.4. These results are used to control the total effective noise rate in our numerical simulations. For example, in the first noise model of Table 2.6, the total effective noise rate is $\gamma/2 + 2 \times \gamma/4 = \gamma$.

Next we show how to calculate the numbers in Table 2.4 with an example. Consider the T_ϕ noise model with Lindblad operator $\gamma D[|1\rangle\langle 1|]$. The evolution of the density matrix is given by

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \gamma \left(|1\rangle\langle 1| \rho |1\rangle\langle 1| - \frac{1}{2} |1\rangle\langle 1| \rho - \frac{1}{2}\rho |1\rangle\langle 1| \right). \tag{2.182}$$

In our simulation the evolution continues for 1 time unit. Using a first order approximation, we can write the noise channel as

$$\mathcal{N}(\rho) = \rho + \gamma \left(|1\rangle\langle 1| \rho |1\rangle\langle 1| - \frac{1}{2} |1\rangle\langle 1| \rho - \frac{1}{2}\rho |1\rangle\langle 1| \right). \tag{2.183}$$

To represent the noise channel in Pauli basis, we use $|1\rangle\langle 1| = \frac{I-Z}{2}$ and get

$$\mathcal{N}(\rho) = \rho + \frac{\gamma}{4}(Z\rho Z - \rho) = \left(1 - \frac{\gamma}{4}\right)\rho + \frac{\gamma}{4} Z\rho Z. \tag{2.184}$$

Therefore the effective noise rate of T_ϕ is $\gamma/4$, and the other numbers can be calculated similarly.

2.4.6.2 Simulation algorithm

For a generic noise source, the incoherent dynamics of the density matrix between gates can be described by the master equation,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \sum_l \gamma_l D[J_l](\rho), \tag{2.185}$$

where ρ is the density matrix, $D[J_i](\rho) = J_i\rho J_i^\dagger - \frac{1}{2}(J_i^\dagger J_i\rho + \rho J_i^\dagger J_i)$ is a Lindblad superoperator for generic collapse operator J_i, and we use units where $\hbar = 1$. A Lindblad superoperator can represent various Markovian noise sources, such as amplitude decay, dephasing, correlated noise, etc. Lindblad operators for the various noise sources studied in this work are given in Table 2.4. We take the system Hamiltonian, H, to be zero and assume that gates are applied perfectly and instantaneously one time unit apart.

For 10 qubit simulations, we evolve the density matrix according to the Lindblad master equation. For a large number of qubits (such as the 20 qubit runs), storing the full density matrix ρ becomes unfeasible. Instead, we simulate the action of the noise sources using the Monte Carlo wave function (MCWF) method [105]. We evolve under an effective, non-Hermitian Hamiltonian

$$H_{\text{eff}} = \frac{-i}{2}\sum_l \gamma_l J_l^\dagger J_l. \tag{2.186}$$

As the state evolves, it will gradually lose norm. A random number p is drawn, and the state is evolved under H_{eff} until the norm falls below p. At this point a "quantum jump" occurs. At this point, one of the noise channels is randomly chosen. The probability of the jump happening due to noise channel l is given by

$$P_l = \frac{\gamma_l\langle\psi|J_l^\dagger J_l|\psi\rangle}{\sum_l \gamma_l\langle\psi|J_l^\dagger J_l|\psi\rangle}. \tag{2.187}$$

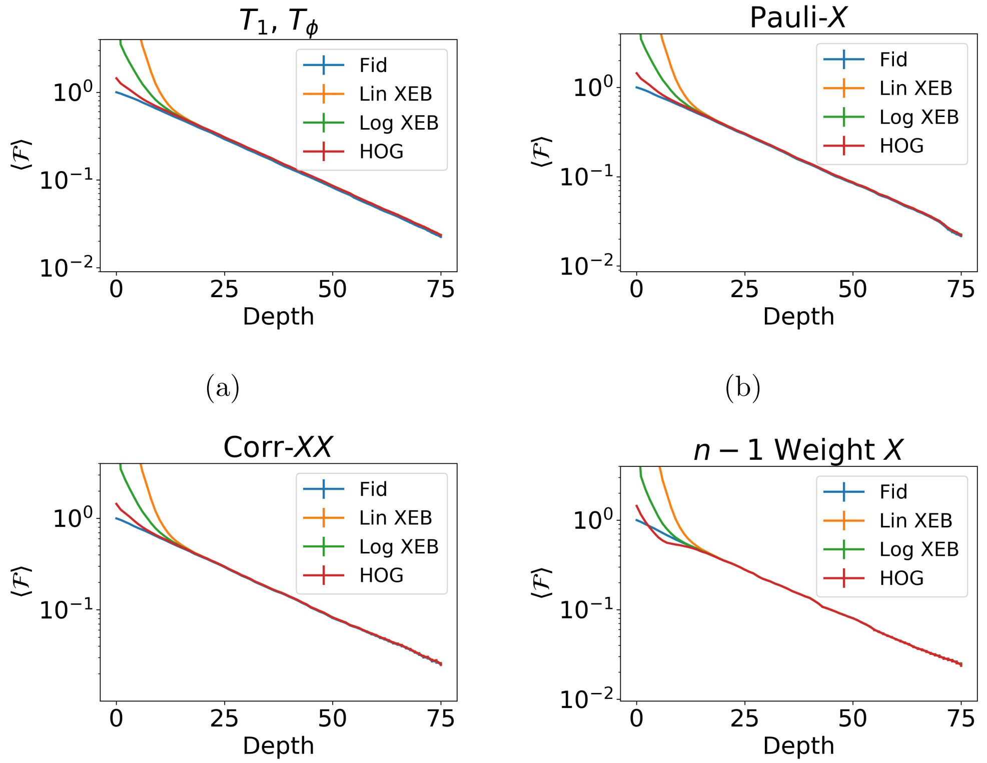
Once a specific noise channel is chosen, the jump is applied according to

$$|\psi\rangle = \frac{J_l|\psi\rangle}{\langle\psi|J_l^\dagger J_l|\psi\rangle}. \tag{2.188}$$

A new random number, p, is drawn and the evolution then proceeds under the effective non-Hermitian Hamiltonian H_{eff} until the next jump occurs. This process is applied between gates until all gates have been applied. This process is repeated many times, where each time a pure state trajectory of the noisy circuit is generated. Due to linearity, the fidelity and fidelity estimators of a noisy circuit can be computed by averaging over the pure state trajectories.

2.4.6.3 Additional simulation results

In addition to the unbiased linear cross entropy estimator, we also run three other fidelity estimators: estimators based on the standard linear cross entropy [7], cross entropy [20], and



(a)

(b)

(c)

(d)

Figure 2.17: Numerical simulations of XEB, log XEB, and HOG fidelity using the Monte Carlo wave function (MCWF) method for various noise models. As opposed to unbiased XEB, these estimators only converge with sufficient depth. (a) Single qubit amplitude-decay and pure dephasing. (b) Single qubit Pauli-X noise. (c) Nearest-neighbor correlated XX noise. (d) Correlated X noise with weight $n_q - 1 = 19$.

heavy-output-generation (HOG) score [19, 7]. All estimators are summarized in Table 2.5, and they take a slightly different form than the equations used earlier due to the simulations having access to the full wavefunction for a single noise trajectory. We use these estimators to study the $n = 20$ linear ring and plot the results in Fig. 2.17 and summarize the results in Table 2.6. These simulations, as well as the $n = 20$ results earlier, were sampled over 100 circuits and 400 noise trajectories for each circuit at each given depth. All estimators are fit from depths 20 to 50.

We also compare the results from both true fidelity and the unbiased linear cross entropy for $n = 10$ and $n = 20$ qubits with both two qubit Haar random unitaries and CNOT+single qubit Haar random unitaries at three different effective noise rates for our four different noise models. The results are summarized in Table 2.7. We find that the results are consistent

Name	Formula
uXEB	$\hat{F}_{\text{uXEB}} = \mathbb{E}\,\frac{\sum_x Dp(x)q(x)-1}{D\sum_x p(x)^2-1}$
XEB	$\hat{F}_{\text{XEB}} = \mathbb{E}\sum_x Dp(x)q(x) - 1$
Log XEB	$\hat{F}_{\text{log}} = \log D + \gamma + \mathbb{E}\sum_x q(x)\log(p(x))$
HOG Fidelity	$\hat{F}_{\text{HOG}} = \mathbb{E}(2\sum_x q(x)1[p(x) \geq \frac{\log 2}{D}] - 1)/\log 2$

Table 2.5: Fidelity estimators studied here. Since our numerical simulations have access to the full wavefunction, these formulas are slightly different than those used when using experimental samples. $x \in \{0,1\}^n$ represents all possible bitstrings, $D = 2^n$ is the dimension of the Hilbert space, $p(x)$ is the ideal output probability, $q(x)$ is the output probability of the noisy circuit, γ is Euler's constant, and $1[\cdot]$ is the indicator function. The expectation value is taken over both random circuits C and, in the case of the MCWF solver, noise trajectories.

Description	Lindblad	λ_F	λ_{uXEB}	λ_{XEB}	λ_{LOG}	λ_{HOG}
T_1, T_ϕ	$\gamma D[\sigma] + 2\gamma D[\sigma^\dagger\sigma]$	0.0511(2)	0.0511(2)	0.0511(2)	0.0511(2)	0.0510(2)
Pauli-X	$\gamma D[X]$	0.0508(2)	0.0509(2)	0.0509(2)	0.0509(2)	0.0508(2)
Corr-XX	$\gamma D[X_i X_{i+1}]$	0.0505(3)	0.0505(3)	0.0505(3)	0.0505(3)	0.0505(3)
$n-1$ Weight X	$\gamma D[\prod_{i\neq j} X_i]$	0.0506(3)	0.0506(3)	0.0506(3)	0.0506(3)	0.0506(3)

Table 2.6: Numerical simulation of RCS benchmarking using the Monte Carlo wave function (MCWF) technique. Here we simulate $n = 20$ qubits with noise strength $\gamma = 0.0025$, and $\sigma = |0\rangle\langle 1|$. Each global noise model has a total ENR of $\lambda_{\text{true}} = n\gamma = 0.05$ by design. λ_F and λ_{uXEB} shows the simulated RCS benchmarking result, which corresponds to the decay rate of fidelity and unbiased linear cross entropy, respectively. λ_{XEB}, λ_{LOG}, and λ_{HOG} correspond to the alternative fidelity estimators of standard linear cross entropy, cross entropy, and a HOG-score based estimator.

over all variations. We fit the $n = 20$, two qubit Haar random results from depths 20 to 50. We fit the $n = 10$, two qubit Haar random results from depths 10 to 25. We fit the $n = 20$ and $n = 10$ single qubit Haar random results from depths 20 to 34.

We provide additional results on $n = 16$, 2D 4×4 lattices using two different gate sets: SQiSWP + $\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$ gate set, where $W = (X+Y)/\sqrt{2}$ and single qubit Haar random gates followed by fixed two-qubit SQiSWP gates. Figure 2.18 shows the collection of various fidelity estimators not included earlier for the 2D discrete gate set (SQiSWP + $\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$). Similar to the 1D results, the other estimators require sufficient depth to converge. The estimates of all of the estimators are found in Table 2.8. Estimates for uXEB at different noise strengths are summarized in Tabke 2.9. For the single qubit Haar random gates followed by fixed two-qubit SQiSWP gates, uXEB results are plotted in Fig. 2.19 and the other estimators are plotted in Fig. 2.20. The results for this continuous gate set are

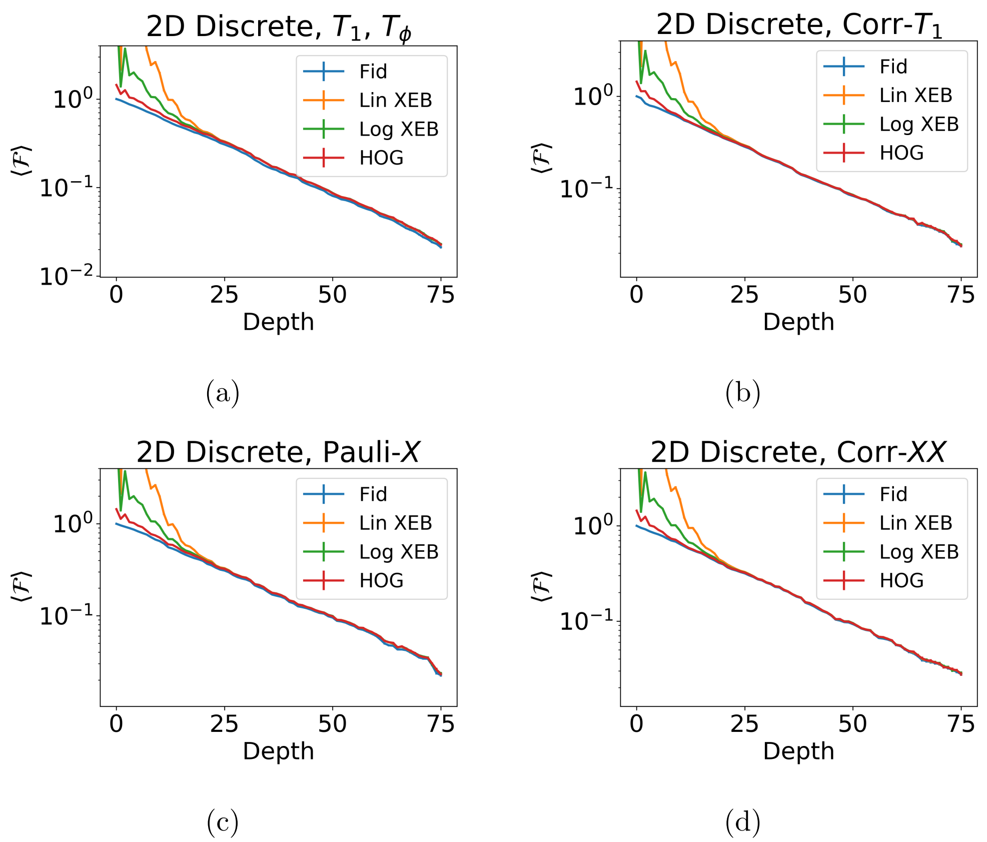
Noise rate	Description	$n = 10$, single qubit		$n = 10$, two qubit		$n = 20$, single qubit		$n = 20$, two qubit	
		λ_F	λ_{uXEB}	λ_F	λ_{uXEB}	λ_F	λ_{uXEB}	λ_F	λ_{uXEB}
$\lambda = 0.05$	T_1, T_ϕ	0.0491(2)	0.0511(7)	0.04978(4)	0.04990(2)	0.0495(4)	0.0523(1)	0.0511(2)	0.0511(2)
	Pauli-X	0.0488(6)	0.0515(9)	0.04970(4)	0.04983(2)	0.0482(4)	0.0527(1)	0.0508(2)	0.0509(2)
	Corr-XX	0.0494(2)	0.0517(9)	0.04974(2)	0.04987(1)	0.0494(4)	0.0527(2)	0.0505(3)	0.0505(3)
	$n-1$ Weight X	0.0500(2)	0.0509(1)	0.049761(1)	0.049879(6)	0.0499(3)	0.0505(8)	0.0506(3)	0.0506(3)
$\lambda = 0.1$	T_1, T_ϕ	0.0968(9)	0.1029(2)	0.09902(9)	0.09964(1)	0.0971(8)	0.1127(3)	0.0998(6)	0.1000(6)
	Pauli-X	0.0968(6)	0.1049(4)	0.09871(8)	0.09929(4)	0.0964(6)	0.1112(2)	0.1020(8)	0.1020(7)
	Corr-XX	0.0973(5)	0.1042(2)	0.09888(5)	0.09946(2)	0.0980(7)	0.1163(3)	0.0996(8)	0.1000(8)
	$n-1$ Weight X	0.100(1)	0.099(1)	0.098954(2)	0.099512(1)	0.0998(7)	0.1013(3)	0.0948(8)	0.0948(8)
$\lambda = 0.15$	T_1, T_ϕ	0.1436(4)	0.1557(5)	0.1471(1)	0.1498(5)	0.1435(8)	0.1811(5)	0.150(1)	0.150(1)
	Pauli-X	0.146(3)	0.150(9)	0.1464(1)	0.1484(6)	0.143(1)	0.182(6)	0.144(2)	0.145(2)
	Corr-XX	0.145(1)	0.152(1)	0.14677(7)	0.14878(4)	0.142(1)	0.175(6)	0.164(2)	0.164(2)
	$n-1$ Weight X	0.150(1)	0.144(4)	0.146874(4)	0.148911(2)	0.146(1)	0.159(9)	0.155(2)	0.155(2)

Table 2.7: Additional numerical simulation results of RCS benchmarking. We simulate three effective noise rates with different noise models, two system sizes (10 and 20 qubits), and two gate sets (two-qubit Haar random gates, CNOT+single qubit Haar random gates).

Description	Lindblad	λ_F	λ_{uXEB}	λ_{XEB}	λ_{LOG}	λ_{HOG}
T_1, T_ϕ	$\gamma D[\sigma] + 2\gamma D[\sigma^\dagger \sigma]$	0.0531(2)	0.0536(2)	0.0536(2)	0.0534(2)	0.0529(2)
Corr-T_1	$\gamma D[\sigma_i \sigma_{i+1}]$	0.0495(3)	0.0502(3)	0.0502(3)	0.0499(3)	0.0494(3)
Pauli-X	$\gamma D[X]$	0.0492(3)	0.0500(3)	0.0500(3)	0.0498(3)	0.0493(3)
Corr-XX	$\gamma D[X_i X_{i+1}]$	0.0486(3)	0.0490(3)	0.0490(3)	0.0488(3)	0.0484(3)

Table 2.8: Numerical simulation of RCS benchmarking using the Monte Carlo wave function (MCWF) technique. Here we simulate $n = 16$ qubits in a 4×4 lattice using the SQiSWP+$\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$ gate set, where $W = (X + Y)/\sqrt{2}$, with noise strength $\gamma = 0.003125$, and $\sigma = |0\rangle\langle 1|$. Each global noise model has a total ENR of $\lambda_{\text{true}} = n\gamma = 0.05$ by design. λ_F and λ_{uXEB} shows the simulated RCS benchmarking result, which corresponds to the decay rate of fidelity and unbiased linear cross entropy, respectively. λ_{XEB}, λ_{LOG}, and λ_{HOG} correspond to the alternative fidelity estimators of standard linear cross entropy, cross entropy, and a HOG-score based estimator.

similar to the discrete 2D gate set and all of the 1D results. The estimates of all of the estimators are found in Table 2.10. Estimates for uXEB at different noise strengths are summarized in Table 2.11.



(a)

(b)

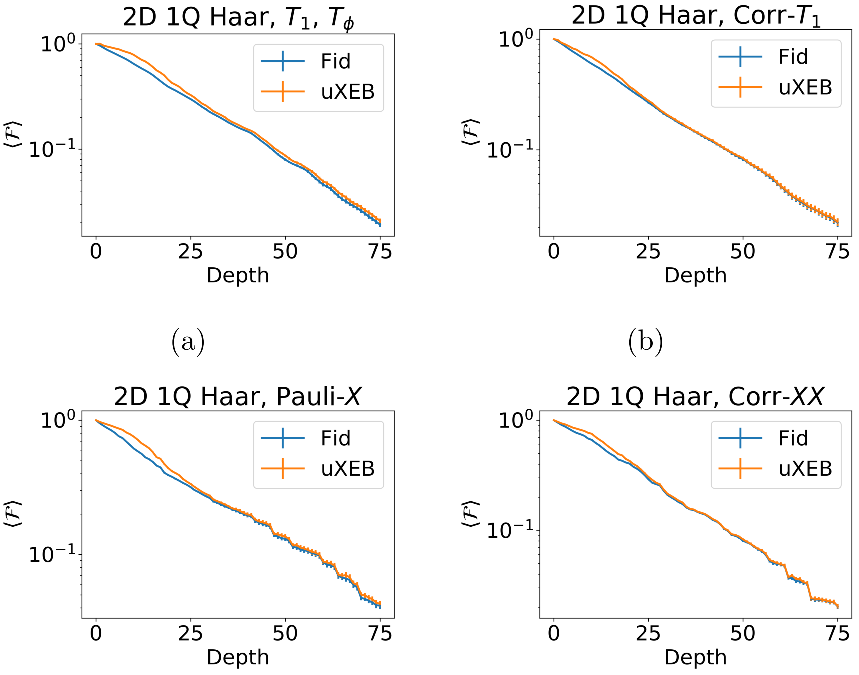
(c)

(d)

Figure 2.18: Numerical simulations of XEB, log XEB, and HOG fidelity using the Monte Carlo wave function (MCWF) method for various noise models on a 4 × 4 lattice of qubits using the SQiSWP+$\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$ gate set, where $W = (X+Y)/\sqrt{2}$. As opposed to unbiased XEB, these estimators only converge with sufficient depth. (a) Single qubit amplitude-decay and pure dephasing. (b) Nearest-neighbor correlated amplitude decay. (c) Single qubit Pauli-X noise. (d) Nearest-neighbor correlated XX noise.

Description	$\lambda = 0.05$		$\lambda = 0.10$		$\lambda = 0.15$	
	λ_F	λ_{uXEB}	λ_F	λ_{uXEB}	λ_F	λ_{uXEB}
T_1, T_ϕ	0.0531(2)	0.0536(2)	0.1003(5)	0.1030(4)	0.1511(8)	0.1591(8)
Corr-T_1	0.0495(3)	0.0502(3)	0.1016(9)	0.1042(8)	0.135(2)	0.143(2)
Pauli-X	0.0492(3)	0.0500(3)	0.0926(9)	0.0963(8)	0.177(2)	0.177(2)
Corr-XX	0.0486(3)	0.0490(3)	0.0982(8)	0.1006(7)	0.138(2)	0.145(2)

Table 2.9: Additional numerical simulation results of RCS benchmarking. We simulate three effective noise rates with different noise models using the using the SQiSWP+$\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$ gate set, where $W = (X+Y)/\sqrt{2}$. on a 4 × 4 square lattice.



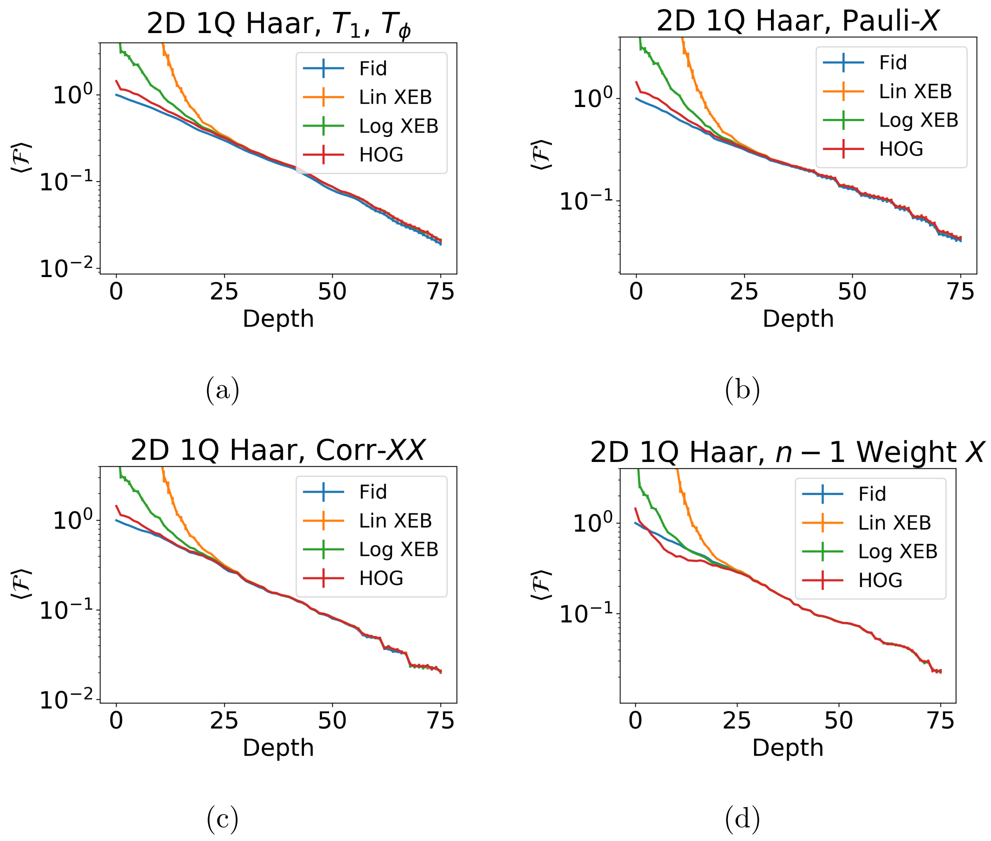
(a)

(b)

(c)

(d)

Figure 2.19: Numerical simulations using the Monte Carlo wave function (MCWF) method for various noise models. The system is modeled as perfect gates followed by evolution for one time unit under noisy channels [74] using the Lindblad master equation [75] $\frac{\mathrm{d}\rho}{\mathrm{d}t} = \sum_i \gamma_i D[J_i](\rho)$, where the sum is over different noise channels, $D[J_i](\rho) = J_i \rho J_i^\dagger - \frac{1}{2}(J_i^\dagger J_i \rho + \rho J_i^\dagger J_i)$ is a Lindblad superoperator for generic collapse operator J_i, and γ_i controls the noise strength. The unbiased linear cross entropy agrees with the fidelity for all depths above a small threshold and correctly predicts the ENR. The noise models include: (a) single qubit amplitude-decay and pure dephasing, (b) nearest-neighbor correlated amplitude-decay, (c) single qubit Pauli-X noise, (d) nearest-neighbor correlated XX noise. Here we simulate $n = 16$ qubits on a 2D lattice with noise strength $\gamma = 0.003125$. Each global noise channel has an ENR of $\lambda_{\text{true}} = n\gamma = 0.05$ by design. We average over 100 random circuits consisting of layers of single qubit Haar random gates followed by fixed two-qubit SQiSWP gates over 400 noise trajectories for each circuit at each depth. We fit the uXEB curves from depths 20 to 50.



(a)

(b)

(c)

(d)

Figure 2.20: Numerical simulations of XEB, log XEB, and HOG fidelity using the Monte Carlo wave function (MCWF) method for various noise models on a 4×4 lattice of qubits using the SQiSWP+ single qubit Haar random gate set. As opposed to unbiased XEB, these estimators only converge with sufficient depth. (a) Single qubit amplitude-decay and pure dephasing. (b) Nearest-neighbor correlated amplitude decay. (c) Single qubit Pauli-X noise. (d) Nearest-neighbor correlated XX noise.

2.4.6.4 Variance analysis

Next we present numerical simulation results which support our analysis of the variance of fidelity in section 2.4.2.5. Recall the setting in section 2.4.2.5 where we consider i.i.d. single qubit Pauli-X noise and use a first-order approximation of the fidelity. For a random circuit $C \sim \mathrm{RQC}(n, d)$, we can write the fidelity as

$$F \approx (1-\varepsilon)^{nd} + \sum_{i=1}^{n} \sum_{l=1}^{d} \varepsilon (1-\varepsilon)^{nd-1} \left| \langle \psi_{i,l} | \psi \rangle \right|^2 , \tag{2.189}$$

Description	Lindblad	λ_F	λ_{uXEB}	λ_{XEB}	λ_{LOG}	λ_{HOG}
T_1, T_ϕ	$\gamma D[\sigma] + 2\gamma D[\sigma^\dagger \sigma]$	0.0492(3)	0.0513(3)	0.0513(3)	0.0509(3)	0.0497(3)
Corr-T_1	$\gamma D[\sigma_i \sigma_{i+1}]$	0.0491(5)	0.0509(5)	0.0509(5)	0.0503(5)	0.0488(5)
Pauli-X	$\gamma D[X]$	0.0351(5)	0.0379(5)	0.0379(5)	0.0375(5)	0.0360(5)
Corr-XX	$\gamma D[X_i X_{i+1}]$	0.0528(4)	0.0546(4)	0.0546(4)	0.0542(4)	0.0530(4)

Table 2.10: Numerical simulation of RCS benchmarking using the Monte Carlo wave function (MCWF) technique. Here we simulate $n = 16$ qubits in a 2D lattice using the SQiSWP+ single qubit Haar random gate set with noise strength $\gamma = 0.003125$, and $\sigma = |0\rangle\langle 1|$. Each global noise model has a total ENR of $\lambda_{\text{true}} = n\gamma = 0.05$ by design. λ_F and λ_{uXEB} shows the simulated RCS benchmarking result, which corresponds to the decay rate of fidelity and unbiased linear cross entropy, respectively. λ_{XEB}, λ_{LOG}, and λ_{HOG} correspond to the alternative fidelity estimators of standard linear cross entropy, cross entropy, and a HOG-score based estimator.

Description	$\lambda = 0.05$		$\lambda = 0.10$		$\lambda = 0.15$	
	λ_F	λ_{uXEB}	λ_F	λ_{uXEB}	λ_F	λ_{uXEB}
T_1, T_ϕ	0.0492(3)	0.0513(3)	0.0972(6)	0.1039(7)	0.145(1)	0.160(1)
Corr-T_1	0.0491(5)	0.0509(5)	0.103(1)	0.108(1)	0.135(2)	0.149(3)
Pauli-X	0.0351(5)	0.0379(5)	0.0984(8)	0.1053(8)	0.178(2)	0.195(2)
Corr-XX	0.0528(4)	0.0546(4)	0.0960(7)	0.1024(7)	0.143(2)	0.160(2)

Table 2.11: Additional numerical simulation results of RCS benchmarking. We simulate three effective noise rates with different noise models using the using the using the SQiSWP+ single qubit Haar random gate set. on a 4×4 square lattice.

where $|\psi_{i,l}\rangle$ denotes the ideal state with an X error on qubit i at depth l. Let

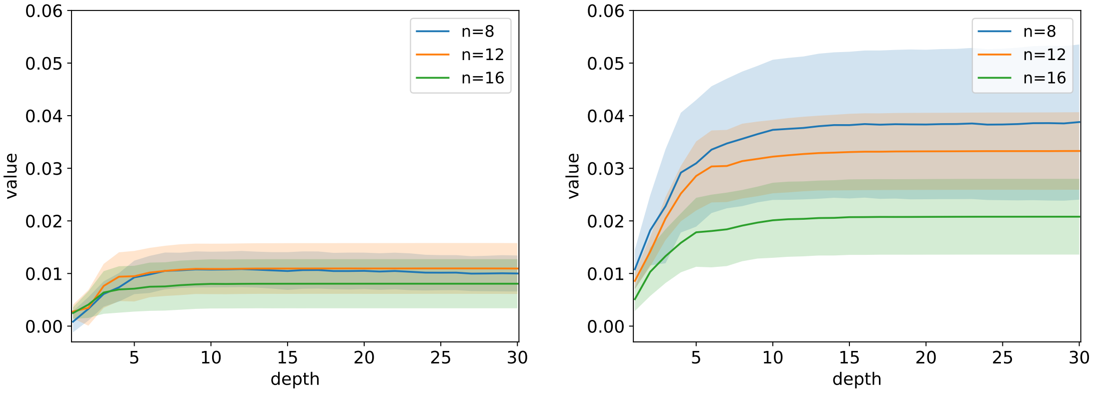
$$A_l := \frac{1}{n} \sum_{i=1}^{n} |\langle \psi_{i,l} | \psi \rangle|^2 , \tag{2.190}$$

then

$$F \approx (1 - \varepsilon)^{nd} + n\varepsilon(1 - \varepsilon)^{nd-1} \sum_{l=1}^{d} A_l. \tag{2.191}$$

Therefore,

$$\mathrm{Var}(F) \approx (n\varepsilon)^2 (1 - \varepsilon)^{2nd-2} \, \mathrm{Var}\left(\sum_{l=1}^{d} A_l \right) \approx \lambda^2 (\mathbb{E}\, F)^2 \, \mathrm{Var}\left(\sum_{l=1}^{d} A_l \right) , \tag{2.192}$$



(a) two-qubit Haar (b) CNOT+single qubit Haar

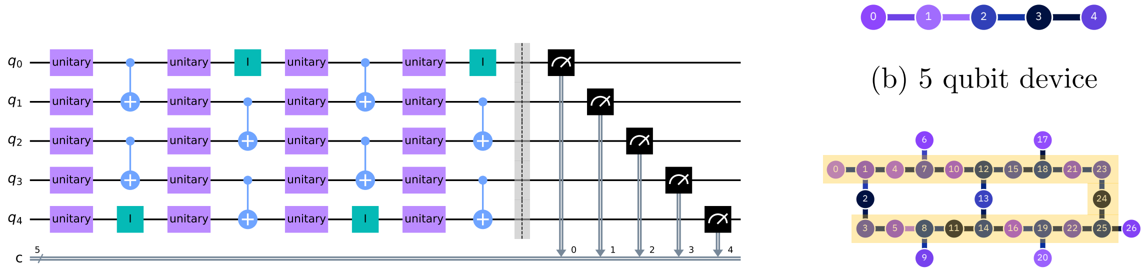
Figure 2.21: Numerical simulation of $\mathrm{Var}\left(\sum_{l=1}^{d} A_l\right) = \sum_{k,l=1}^{d} \left(\mathbb{E}[A_k A_l] - \mathbb{E}[A_k]\,\mathbb{E}[A_l]\right)$ which verifies our variance model $\mathrm{Var}(F) = O\left(\lambda^2 \left(\mathbb{E}\,F\right)^2\right)$. (a) Two-qubit Haar random gates. (b) CNOT+single qubit Haar random gates. $\mathrm{Var}\left(\sum_{l=1}^{d} A_l\right)$ converges to a constant with both gate sets, while the constant for single qubit Haar random gates is larger.

where we have used the fact that $\lambda \approx n\varepsilon$ and $\mathbb{E}\,F \approx (1-\varepsilon)^{nd}$. Next, we show with numerical simulation that

$$\mathrm{Var}\left(\sum_{l=1}^{d} A_l\right) = \sum_{k,l=1}^{d} \left(\mathbb{E}[A_k A_l] - \mathbb{E}[A_k]\,\mathbb{E}[A_l]\right) = O(1). \tag{2.193}$$

In the following we present simulation results to verify Eq. (2.193) with two gate sets: 2-qubit Haar random gates (as in Fig. 2.9a) and CNOT+single qubit Haar random gates (as in Fig. 2.9b). For each $k, l = 1, \ldots, 30$, we simulate $\mathbb{E}[A_k A_l]$, $\mathbb{E}[A_k]$ and $\mathbb{E}[A_l]$ by averaging over 100 random circuits. For each random circuit, we randomly sample 100 error locations. The solid lines and color regions in Fig. 2.21 denote mean and standard error across 5 independent experiments.

In Fig. 2.21 we present simulation results for $n = 8, 12, 16$ qubits. We observe that $\mathrm{Var}\left(\sum_{l=1}^{d} A_l\right)$ converges to a constant as depth increases for both gate sets. Interestingly, although the simulation results are very noisy with large error bars, we can observe a clear difference between the two gate sets. That is, $\mathrm{Var}\left(\sum_{l=1}^{d} A_l\right)$ converges to a larger constant with single qubit Haar random gates. This verifies our intuition that a gate set with more randomness has smaller variance in RCS benchmarking.



(a)

(b) 5 qubit device

(c) 27 qubit device

Figure 2.22: Architecture of the IBM Quantum superconducting qubit devices used in our experiments. (a) An example circuit on 5 qubits with depth 4. Each purple box represents a Haar random single qubit unitary gate. Figure generated by Qiskit [108]. (b)(c) Architecture for the 5 and 27 qubit devices. Figures retrieved from https://quantum-computing.ibm.com/services?services=systems. The subset of qubits used in the 27 qubit device is highlighted. The experiment in Fig. 2.14 is performed by adding qubits in the order given by [0,1,4,7,10,12,15,18,21,23,24,25,22,19,16,14,11,8,5,3].

2.4.7 Additional experiment details

Fig. 2.22 shows the architecture of the devices used in our experiments. For the 27 qubit device, we consider a 20 qubit subset indexed by

[0,1,4,7,10,12,15,18,21,23,24,25,22,19,16,14,11,8,5,3]

for the 20 qubit experiments, and use the first 10 qubits in this list for the 10 qubit experiments.

We give an example of the circuits implemented in RCS benchmarking in Fig. 2.22a. We implement Haar random single qubit gates between CNOT layers. Note that an arbitrary single qubit gate can be decomposed by

$$U(\theta,\phi,\lambda) = R_z(\phi-\pi/2)R_x(\pi/2)R_z(\pi-\theta)R_x(\pi/2)R_z(\lambda-\pi/2). \tag{2.194}$$

(See https://qiskit.org/documentation/stubs/qiskit.circuit.library.UGate.html.) Here $R_x(\pi/2)$ is the $\sqrt{X}$ gate, and also called a X90 pulse. The R_z gates, which are rotations around z-axis, are implemented *virtually* in hardware via framechanges and do not have any error. Therefore the error rate of a Haar random single qubit gate equals twice the error rate of a X90 pulse. $R_z(\theta)$ and $\sqrt{X}$ are the native single qubit gates supported by the devices. Therefore, when submitting circuits to the devices through cloud, each single qubit gate is first decomposed to the form in Eq. (2.194).

By default, for any circuit submitted to the cloud, we take the maximum amount of measurement samples allowed (8192). We can submit the same circuit multiple times if more samples are needed. Below we give the parameters used in the experiments presented in Fig. 2.13. The parameters for the other experiments are already given earlier. Here "repeat" refers to how many times each circuit is submitted. The total circuit count includes the repeated ones, so the total amount of samples collected equals this number times 8192.

- Fig. 2.13a: `depth=[1, 5, 9, 13, 17, 21, 25, 29, 33, 37]`, 90 circuits for each depth, `repeat=[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]`, 900 circuits in total.
- Fig. 2.13b: `depth=[20, 22, 24, 26, 28, 30, 32]`, 100 circuits for each depth, `repeat=[1, 1, 1, 1, 2, 2, 2]`, 1000 circuits in total.
- Fig. 2.13c: `depth=[1, 5, 9, 13, 17, 21, 25, 29, 33, 37]`, 100 circuits for each depth, `repeat=[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]`, 1000 circuits in total.
- Fig. 2.13d: `depth=[20, 22, 24, 26, 28, 30, 32]`, 100 circuits for each depth, `repeat=[1, 1, 1, 1, 2, 2, 2]`, 1000 circuits in total.

Chapter 3

Learning algorithms

This chapter studies the following two fundamental problems in quantum complexity theory and quantum learning theory:

1. Given access to an unknown constant depth quantum circuit U on a finite dimensional lattice, learn a constant depth circuit that is close to U (in diamond distance).

   In Section 3.2 we give a polynomial time algorithm for this problem, based on joint work with Hsin-Yuan Huang, Michael Broughton, Isaac Kim, Anurag Anshu, Zeph Landau, and Jarrod R. McClean [111].

2. Given copies of an unknown quantum state $|\psi\rangle = U|0^n\rangle$ that is prepared by an unknown constant depth circuit U on a finite dimensional lattice, learn a constant depth circuit that prepares $|\psi\rangle$ (within small trace distance).

   In Section 3.3 we give a polynomial time algorithm for this problem, based on joint work with Zeph Landau [112].

Both algorithms extend to the case when the depth of U is polylog(n), with quasi-polynomial running time. In addition, we also give a polynomial time algorithm for the first problem, when U has arbitrary unknown architecture.

The above problems are also motivated by the development of quantum algorithms in NISQ. NISQ computation can be modeled as shallow quantum circuits. Depite their simplicity, shallow quantum circuits can already generate probability distributions that are classically hard to simulate. This has motivated the development of NISQ algorithms toward achieving useful quantum advantage, where the key idea can be summarized as trying to discover a shallow circuit as a solution to an interesting problem (assuming the circuit exists). This approach can be formulated as a learning problem, and the main challenge is to develop efficient and provable learning algorithms. In this context, this chapter presents two new learning algorithms that provably work in simple settings, which could be used as primitives for new NISQ algorithms.

3.1 Introduction

The question of how to efficiently learn expressive classes of quantum states and circuits features prominently in quantum complexity theory, quantum algorithm design, and the experimental characterization of quantum devices. As a first step, one might consider the efficiency of learning shallow (constant depth) quantum circuits, where, to date, there has been no resolution despite considerable interest from a number of angles. From a complexity perspective, shallow quantum circuits are known to be more powerful than their classical counterparts [113, 114, 115, 116], and under widely accepted complexity assumptions, sampling from the output distribution of shallow quantum circuits is classically hard to simulate [117, 118, 119, 120, 8]. This computational power provides the basis for quantum computational advantage with NISQ (noisy intermediate-scale quantum) devices and supports the quest for developing quantum algorithms based on learning parameterized shallow quantum circuits [121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135]. Within an experimental setting focused on coherent errors or gate calibration, characterizing a NISQ device can be modeled as learning what shallow quantum circuit the device is performing. Despite substantial interest in the question of learning shallow quantum circuits from these directions, to date, no polynomial time algorithm for learning shallow quantum circuits has been found. In this work, we introduce several efficient algorithms for two related tasks.

Theorem (Summary of main results)**.** *There are polynomial time algorithms for (1) learning the description of an unknown n-qubit shallow quantum circuit U (with arbitrary unknown architecture) within a small diamond distance, given access to U; (2) learning the description of an unknown n-qubit state $|\psi\rangle = U|0^n\rangle$ prepared by a shallow quantum circuit U (on a finite dimensional lattice) within a small trace distance, given copies of $|\psi\rangle$.*

The main challenges in learning shallow quantum circuits are twofold. While foundational results in computational learning theory have established the efficient learnability of shallow classical circuits [136, 137, 138], these techniques may not apply to shallow quantum circuits, as these circuits can generate distributions with nontrivial correlations over the entire system that are classically hard to simulate [119, 120, 8]. Furthermore, even when the structure of a shallow quantum circuit is known up to parameterization, the optimization landscape for learning shallow quantum circuits is swamped with exponentially many suboptimal local minima [134]. The bad optimization landscape causes standard optimization methods, such as gradient descent algorithms and Newton methods, to fail in learning shallow quantum circuits.

To address these challenges, we consider a quantum circuit representation based on *local inversions*, which yields an optimization landscape that can be efficiently navigated. The local inversions disentangle qubits in each local region in a way that does not perturb the remaining system. We then show how these local inversions may be combined to build up the entire circuit without having to solve a computationally hard problem. Together, this new technique enables us to learn a natural class of quantum circuits that are classically hard to simulate.

3.1.1 Background

Learning shallow classical circuits. Although the shallow quantum case has many conceptual challenges resulting from non-locality, the learnability of shallow classical circuits is a fundamental question in computational learning theory that has been well-studied and resolved in many cases. Learning constant-depth classical circuits with bounded fan-in gates (NC^0) is equivalent to learning juntas and can be performed in polynomial time from uniform samples [137]. In addition, quasi-polynomial time algorithms are known for learning constant-depth classical circuits with unbounded fan-in AND/OR gates (AC^0) [136], as well as mod p gates ($\mathsf{AC}^0[p]$) [138] in the PAC model. The problem of learning shallow quantum circuits (QNC^0) and their output states are natural quantum analogs of learning Boolean circuits. As QNC^0 can be exponentially more powerful than AC^0 for some computational problems [116], it is natural to ask if shallow quantum circuits can be learned efficiently from random data samples.

Quantum machine learning. When one parameterizes the gates in a quantum circuit, the parameterized quantum circuit forms an ML model, known as a *quantum neural network*, that can learn from data and make predictions on new inputs [121, 122, 123, 124, 125, 126, 127]. Since deep parameterized quantum circuits suffer from having *barren plateaus* in the optimization landscape [139, 140] and are challenging to implement on noisy quantum devices [141, 142], shallow quantum circuits have been subject to extensive study in recent years [128, 129, 130, 131, 132, 133, 134, 135]. Various applications of learning shallow quantum circuits have been explored, ranging from compressing quantum circuits for implementing a unitary [143, 144, 145, 127, 146], speeding up quantum dynamics [147, 148, 149, 150, 151], to learning generative models for sampling from predicted distributions [152, 153, 154, 155, 156, 157]. While the optimization landscape for learning shallow quantum circuits is free from barren plateau [128], the landscape is swamped with exponentially many suboptimal local minima [134]. The presence of a large number of suboptimal local minima causes standard local optimization methods, such as gradient descent or Newton's method, to fail in learning parameterized shallow quantum circuits.

Efficient quantum tomography. While quantum state and process tomography generally require exponential resources, performing tomography over some restricted families of states or processes can be made computationally efficient. Examples of such families include matrix product states [158, 159, 160], high-temperature Gibbs states [161, 162, 163], stabilizer states [164, 165, 166, 167], quantum phase states [168], noninteracting Fermionic states [169], Clifford circuits with a small number of T gates [165, 170, 167], Pauli channels under structural assumptions [60, 62, 171, 172], and interacting Hamiltonian dynamics [173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183] (see [184] for a recent survey). Most of these examples correspond to quantum circuit families that are classically easy to simulate [185, 186, 187, 188, 189]. In contrast, sampling from the output distribution of constant-depth quantum circuits is classically hard even when restricted to a 2D lattice [118, 119]. The experimental effort to characterize NISQ devices motivates the question of how to perform tomography for states and processes generated by shallow quantum circuits. While these states can be learned sample-efficiently using shadow tomography [190, 191, 103], no computationally efficient algorithms are known.

3.1.2 Our Results

We first focus on cases where one is given black-box access to the unknown unitary in (1) learning general shallow quantum circuits and (2) learning geometrically-local shallow quantum circuits. We then consider the more restricted model where one is only provided access to copies of an unknown state and focus on (3) learning quantum states prepared by geometrically-local shallow quantum circuits on finite-dimensional lattices.

3.1.2.1 Learning general shallow quantum circuits

Let U be an unknown n-qubit unitary generated by a shallow quantum circuit. The learning algorithm uses a randomized measurement dataset consisting of N samples about U [192, 193, 194, 195, 127, 150, 151]. This dataset has been proposed as the classical shadow of U [192, 193, 194]. Each classical data sample specifies a random n-qubit product input state $|\psi_\ell\rangle = \bigotimes_{i=1}^n |\psi_{\ell,i}\rangle$ and a randomized Pauli measurement outcome $|\phi_\ell\rangle = \bigotimes_{i=1}^n |\phi_{\ell,i}\rangle$ on the output states $U|\psi_\ell\rangle$, where $|\psi_{\ell,i}\rangle, |\phi_{\ell,i}\rangle \in \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |y+\rangle, |y-\rangle\}$ are single-qubit stabilizer states. Each data sample can be generated by a single query to U. Our goal is to learn U within a small diamond distance. The following results have the form of learning a circuit V acting on $2n$ qubits, such that $\|V - U \otimes U^\dagger\|_\diamond \leq \varepsilon$. Hence, V can be used to implement U by tracing out the n-qubit ancilla system.

Our first main result shows that one can learn U with a polynomial sample and computational complexity, with only the assumption that U is constant-depth (i.e., U has arbitrary unknown connectivity). Furthermore, the result applies even when the circuit generating U can have any number m of ancilla qubits used as working space and can have arbitrary two-qubit gates in SU(4) between any pair of the $n + m$ qubits so long as the resulting operation on the n system qubits is unitary. The learning algorithm is fully classical given the randomized measurement dataset.

Theorem 3.1 (Learning shallow quantum circuits; see Theorem 3.6)**.** *Given an unknown n-qubit unitary U generated by a constant-depth circuit over any two-qubit gates between any pair of qubits. One can learn a constant-depth circuit approximating U to diamond distance ε with high probability from $N = \mathcal{O}(n^2 \log(n)/\varepsilon^2)$ samples about U and* $\mathrm{poly}(n)/\varepsilon^2$ *classical running time.*

When the circuit is over a finite gate set, U can be learned to zero error with high probability from $N = \mathcal{O}(\log n)$ samples and $\mathrm{poly}(n)$ *time.*

3.1.2.2 Learning geometrically-local shallow quantum circuits

The algorithm for learning general shallow quantum circuits runs in polynomial time but with a large exponent. Furthermore, the depth of the learned circuit V, while constant, could be substantially greater than the depth of U. Motivated by the fact that most realistic quantum systems are geometrically local on a finite-dimensional lattice, it is natural to wonder if these aspects can be improved when learning geometrically-local quantum circuits on lattices. Next, we show that this is indeed the case.

See Theorem 3.7 for a related result on learning shallow circuits over any geometry represented by a bounded-degree graph.

Theorem 3.2 (Learning geometrically-local shallow circuits; see Theorem 3.8)**.** *Given an unknown n-qubit geometrically-local depth-d quantum circuit U over a k-dimensional lattice with $d, k = \mathcal{O}(1)$. One can learn a geometrically-local shallow circuit that approximates U to diamond distance ε with high probability from $N = \mathcal{O}(n^2 \log(n)/\varepsilon^2)$ classical data samples and either*

- $\mathcal{O}(n^3 \log(n)/\varepsilon^2)$ *classical running time with a learned circuit depth of* $(k+1)4^{4(8kd)^k}+1$.
- $(n/\varepsilon)^{\mathcal{O}((8kd)^{k+1})}$ *classical running time with a learned circuit depth of* $(k+1)(2d+1)+1$.

When the circuit is over a finite gate set, U can be learned to zero error with high probability from $N = \mathcal{O}(\log n)$ samples and $\mathcal{O}(n \log(n))$ time with a learned circuit depth of $(k+1)(2d+1)+1$.

This shows that in the geometrically local setting, the learned circuit depth can achieve a linear blow-up. Furthermore, the learning algorithm works for $d = \mathrm{polylog}(n)$ depth circuits at the cost of quasipolynomial running time.

We remark that the more formal statement of the above theorem, which is labeled in this work as Theorem 3.8, can be straightforwardly generalized to a larger class of unitaries called *quantum cellular automata* (QCA), which play an important role in understanding quantum phases of matter [196, 197, 198, 199]. These are unitaries that map any geometrically local operator to a geometrically local operator in the Heisenberg picture. For any such unitary, our proof technique applies without any modification, yielding an efficient algorithm for learning any QCAs. Interestingly, while shallow quantum circuits are QCAs by definition, the converse statement is not necessarily true. For instance, shifting a set of qubits on a one-dimensional lattice trivially maps local operators to local operators. However, it is impossible to decompose this unitary into a geometrically local shallow quantum circuit [197]; see Ref. [198, 199] for other nontrivial examples of QCA. Therefore, our algorithm is applicable beyond shallow quantum circuits.

So far, we have been focusing on learning a shallow quantum circuit from a classical randomized measurement dataset. A natural question asks if further improvement is possible when we allow more general quantum query access to U. In the following, we show that by using quantum queries to U, an exponential improvement in query complexity is possible and this result is *asymptotically-optimal* in both time and query complexity for learning geometrically-local shallow circuits over finite gate sets. Surprisingly, quantum access also allows these circuits to be *with certainty*, dropping the familiar qualifier of high probability. The matching lower bounds stem from the need to query at least $\Omega(1)$ times to obtain any information about U and to write down the learned n-qubit circuit, which requires $\Omega(n)$ time.

Theorem 3.3 (Learning shallow circuits with quantum queries; see Theorem 3.9)**.** *An unknown n-qubit geometrically-local shallow quantum circuit U over a finite gate set can be learned to zero error with zero failure probability using $\Theta(1)$ queries to U and $\Theta(n)$ quantum computational time.*

3.1.2.3 Learning output states of geometrically-local shallow quantum circuits

Besides learning the n-qubit unitary U using input-output queries, it is natural to study the problem of learning a pure quantum state $|\psi\rangle$ prepared by a shallow quantum circuit U, i.e., $|\psi\rangle = U|0^n\rangle$. Here, instead of given access to U, we are only given copies of the pure state $|\psi\rangle$ as in quantum state tomography [200, 158]. As discussed in Section 3.1.1, most families of efficient learnable quantum states, such as matrix product states [158, 159, 160] and stabilizer states [164, 165, 166, 167], correspond to quantum circuit families that are classically easy to simulate [186, 187]. In contrast, constant-depth quantum circuits are classically hard to simulate even when restricted to a 2D lattice [118, 119].

Learning $|\psi\rangle = U|0^n\rangle$ from copies of $|\psi\rangle$ has an incomparable difficulty to the earlier results because it has a less stringent requirement (learning an output state of U) but a more restricted access model (accessing copies of $|\psi\rangle$ instead of U). While $|\psi\rangle = U|0^n\rangle$ can be learned from polynomially many copies [162, 201], the restricted access model makes the problem computationally more challenging, and the question of whether there exists a polynomial time algorithm remains open.

Here we give two efficient algorithms. The first works for general finite dimensional lattices, and the learned circuit uses linear number of ancilla qubits.

Theorem 3.4 (Learning quantum states prepared by shallow circuits; see Theorem 3.13)**.** *There is an algorithm that, given copies of an unknown state $|\psi\rangle$, with the promise that $|\psi\rangle = U|0^n\rangle$ where U is an unknown depth-d circuit acting on a k-dimensional lattice (using arbitrary 2-qubit gates), outputs a depth-$(2k+1)d$ circuit W that prepares $|\psi\rangle$ up to 0.01 trace distance, with success probability 0.99. The algorithm uses M copies of $|\psi\rangle$ and runs in time T, where*

$$M = \tilde{O}(n^4) \cdot 2^{O(c)}, \quad T = \tilde{O}(n^4) \cdot 2^{O(c)} + (nkd \cdot c)^{O(d \cdot c)}. \tag{3.1}$$

Here, $c = O((3k)^{k+2}d)^k$, and W uses $r \cdot n$ ancilla qubits where $r > 0$ can be chosen to be an arbitrarily small constant.

Note that the running time is polynomial when $d = O(1)$, and quasi-polynomial when $d = \text{polylog}(n)$. In addition, the dominating term in the running time (second term in T) can be significantly improved when assuming a discrete gate set.

The second algorithm is specialized to 2D lattice. This achieves an improvement relative to the above result in the sense that the learned circuit does not use any ancilla qubits.

Theorem 3.5 (Learning quantum states prepared by shallow circuits in 2D; see Theorem 3.15)**.** *Given copies of an unknown state $|\psi\rangle$, with the promise that $|\psi\rangle = U|0^n\rangle$ for an unknown n-qubit circuit U with circuit depth $d = \mathcal{O}(1)$ acting on a 2-dimensional lattice, and assume that each two-qubit gate in U is chosen from a finite gateset of constant size. Then there is an algorithm that learns a circuit V with depth $2^{c \cdot d^2}$ (for some universal constant c) acting on n qubits (without using any ancilla), such that $\left|\langle 0^n| V^\dagger |\psi\rangle\right|^2 \geq 1 - \varepsilon$ with probability $1 - \delta$, using $\mathcal{O}(\log(n/\delta))$ copies of $|\psi\rangle$ and time $(n/\varepsilon)^{\mathcal{O}(1)}$.*

3.1.3 Discussion

Higher circuit depth. In the general setting without geometric locality, we show that log-depth circuits require exponentially many quantum queries to learn within a small diamond distance (see Prop. 3.3), which is proven by showing that log-depth circuits can implement Grover's oracle over 2^n elements and applying the Grover lower bound [202]. Therefore, our result for efficiently learning general constant-depth quantum circuits cannot be extended to much higher depth.

In the geometrically-local setting, Theorem 3.8 implies polynomial-time learnability for quantum circuits on a k-dimensional lattice up to $\log(n)^{1/k}$ depth, and quasi-polynomial time for up to polylog(n) depth. What structural assumptions allow us to efficiently learn quantum circuits beyond polylog-depth remains an important open question.

Worst-case vs average-case distance. Motivated by the above discussion, it is natural to consider learning quantum circuits under weaker notions of distance, analogous to the classical notion of PAC learning. The standard notion of average-case distance in the literature [203, 204] is defined as the distance between output states when averaging over input states generated by Haar random unitaries. While learning polynomial-size quantum circuits to small average-case distance can be achieved with polynomial sample complexity [127, 150], the computational complexity of achieving a small average-case distance remains an open question.

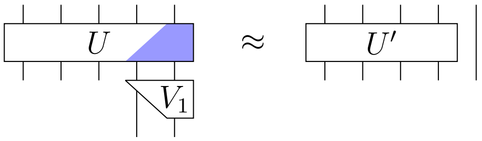
In addition, Ref. [193] considered a weaker notion of an average-case error where the goal is to learn observables of the output state for random input states and showed that under this notion, any quantum circuit (even those with exponential depth) could be learned in quasi-polynomial time.

3.2 Learning shallow quantum circuits

3.2.1 Technical overview

Let U be an unknown n-qubit circuit of depth $d = \mathcal{O}(1)$. We consider the following two tasks in this chapter: (1) Learn a constant-depth circuit $\hat{U}$ from random data samples from U or query access to U, such that U and $\hat{U}$ are close in diamond distance. (2) Learn a constant-depth circuit $\hat{U}$ from measuring copies of the n-qubit state $|\psi\rangle = U|0^n\rangle$, such that $\hat{U}|0^n\rangle$ and $|\psi\rangle$ are close in trace distance.

A basic idea to learn U is to produce a guess $\hat{U}$ and check if $\hat{U}$ is close to U (i.e., $\hat{U}^\dagger \cdot U$ is close to identity). While the search space over $\hat{U}$ is exponentially large, the *locality* of shallow circuits allows us to search more efficiently. For example, in the following figure, we can find a small *local inversion* circuit V_1, that disentangles qubit 1 (the rightmost qubit), i.e., $UV_1 \approx U' \otimes I_1$. Here, the input wires are at the bottom, and the output wires are at the top; V_1 is applied before applying U.



(3.2)

This follows from a two-step argument. First, the existence of such a local inversion circuit is guaranteed by the locality of U, as undoing the gates in the backward lightcone (shaded blue region) of qubit 1 forms such a local inversion. Second, given a guess V_1, we develop an efficient procedure to check *approximate local identity*, i.e. $UV_1 \approx U' \otimes I_1$ for some $n-1$ qubit unitary U'. This allows us to find local inversions via brute force enumerate-and-test since the search space is small (as V_1 has depth d and is supported within a constant size region). Note that after this exhaustive process, we may find a list of valid local inversions. The "ground truth" local inversion compatible with the unique global inverse of the unitary is among them, but we do not know which one. Similarly, given copies of a state $|\psi\rangle = U|0^n\rangle$ we can find small local inversion circuits V_1 to disentangle qubit 1, $V_1|\psi\rangle \approx |\psi'\rangle \otimes |0\rangle_1$ for some $n-1$ qubit state $|\psi'\rangle$.

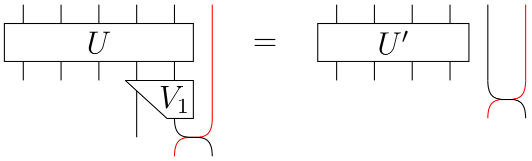
The above argument shows a procedure to efficiently learn local inversions for each qubit for both of our learning problems. The central question is whether this suffices to reconstruct the circuit and, if so, whether the reconstruction can be done efficiently. The main obstacle is that local inversions for each qubit are not unique, and two local inversions on neighboring qubits may not be consistent in the overlapping regions. Finding a consistent set of local inversions may require solving a constraint satisfaction problem that is computationally hard. Next, we show how to overcome this obstacle.

3.2.1.1 Sewing local inversions

Suppose we have learned a set of local inversions $\mathcal{C}_i$ for an unknown shallow quantum circuit U for each qubit i. Here, we show how to reconstruct the circuit using the learned local information. Surprisingly, the algorithm only requires an *arbitrary* element $V_i \in \mathcal{C}_i$ for each qubit i, without the need to search for the element compatible with the global inverse, which could require solving a complicated constraint satisfaction problem. The formal statements on this algorithmic technique are given in Section 3.2.4.3.

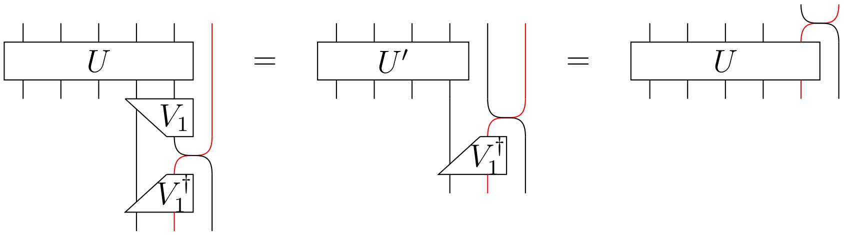
For simplicity, here we first assume all the local inversions are found exactly without any approximation. Take any $V_1 \in \mathcal{C}_1$, applying it to the unknown circuit U gives $UV_1 = U' \otimes I_1$, see Eq. (3.2), where we imagine qubit 1 to be the rightmost qubit and use a simple 1D geometry for illustration. This represents some progress: applying V_1 reduces the unknown n-qubit unitary U to an unknown $(n-1)$-qubit unitary U' (note that U' may not be a shallow circuit). A natural thought is whether we can keep making this progress by applying local inversion on other qubits. The main issue here is that now the unitary has changed. For example, consider qubit 2 which is right next to qubit 1. Due to the fact that they have overlapping lightcones, some local inversion $V_2 \in \mathcal{C}_2$ may no longer work for the new circuit UV_1. Separately, we can attempt to find local inversion for qubit 2 with respect to this new circuit UV_1; however, doing so might disturb the progress we have made on qubit 1 and therefore requires coordinated effort across different qubits. This is exactly the type of constraint satisfaction problem that we want to avoid.

Here we introduce a general approach to keep making progress: the idea is to introduce a fresh ancilla qubit, swap it with qubit 1, and then *undo* the local inversion V_1. We show this in two steps: first, introduce a fresh ancilla qubit (red) and swap it with qubit 1,



(3.3)

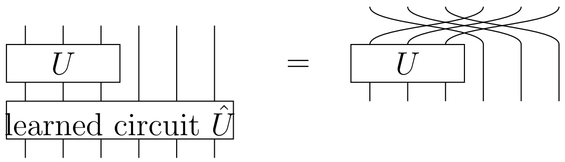
and then apply $V_1^\dagger$,



(3.4)

To explain the second equality of Eq. (3.4), note that without the swap operation, the above procedure is not doing anything (since we just perform some operation and undo it). In the second picture of Eq. (3.4), after experiencing $V_1^\dagger$, the red wire corresponds to the first output wire of U, but then it gets swapped out to the ancilla. Therefore, the overall effect is equivalent to performing a swap at the end after applying U.

The key reason that the above procedure is useful is because it *repairs* the circuit. This allows us to continue doing the same operation on qubit 2 because even though a lot of operations were applied before U (see the first picture in Eq. (3.4)), it is equivalent to as if nothing were applied before U (see the last picture in Eq. (3.4)); therefore we can similarly apply $V_2^\dagger$, swap with a new fresh qubit, and V_2 before U, achieving the effect of swapping qubit 2 at the end. Repeating the above procedure for all qubits, we have learned a circuit $\hat{U}$ acting on $2n$ qubits that satisfies



(3.5)

which implies that $\hat{U} = S \cdot (U \otimes U^\dagger)$, where S denotes the global swap operation between the system and ancilla qubits. To implement U using the learned circuit, on input ρ we initialize an ancilla register with some arbitrary state (say $|0^n\rangle$), apply $S \cdot \hat{U}$ and trace out the ancilla register, and the output state equals $U\rho U^\dagger$. We can use a similar procedure to implement $U^\dagger$. Thus, the above procedure simultaneously learns to implement U and $U^\dagger$, using access only to U.

Finally, we remark that the learned circuit $S \cdot \hat{U}$ is shallow. To see this, note that $S = \mathrm{SWAP}^{\otimes n}$ is depth-1. $\hat{U}$ consists of unitaries of the form $W_i := V_i \cdot \mathrm{SWAP} \cdot V_i^\dagger$ that are *local*: each of them supports on the lightcone of qubit i, as well as an extra ancilla qubit. Therefore we can implement non-overlapping W_is simultaneously, and all of the W_is can be stacked into a constant number of layers since, at most, a constant number of qubits share overlapping lightcones.

To achieve the optimal query and time complexity of $\Theta(1), \Theta(n)$ for learning geometrically-local shallow quantum circuits over finite gate sets in Theorem 3.3, we present a quantum learning algorithm that finds the exact local inversions for all n qubits with zero failure probability by querying U for only $\mathcal{O}(1)$ times. This surprising scaling is achieved by combining a few ideas: (a) coloring the geometry described by a bounded-degree graph, (b) decoupling the n-qubit unitary U into $\mathcal{O}(n)$ few-qubit channels based on the coloring, and (c) designing a tournament to perfectly distinguish between two classes of few-qubit quantum channels: those that form an exact local identity versus those that do not. The tournament uses the perfect distinguishability of certain pairs of CPTP maps shown in [205], where we design the few-qubit channels to ensure perfect distinguishability. Then, the learning algorithm finds a good order to sew the local inversions to produce a constant-depth circuit implementation for the unknown constant-depth n-qubit circuit U.

3.2.1.2 Sewing Heisenberg-evolved Pauli operators

Next, we describe a simpler technique based on directly sewing the Heisenberg-evolved Pauli operators $U^\dagger P_i U$ (P_i is a single-qubit Pauli acting on qubit i) and discuss how it is closely related to local inversion. Section 3.2.4.4 provides a detailed discussion of this technique.

We first describe how to learn the Heisenberg-evolved Pauli operators. Because U is a shallow quantum circuit, each operator $U^\dagger P_i U$ acts on a constant number of qubits. The few-qubit observable $U^\dagger P_i U$ can be reconstructed from the randomized measurement dataset. Let the random input product state be $|\psi\rangle = |\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle$, where $|\psi_i\rangle$ is a random one-qubit stabilizer state. Because each qubit in the output state is measured in a random X, Y, Z basis with equal probability, we will measure P_i on the output state $U |\psi\rangle\!\langle\psi| U^\dagger$ with probability 1/3. This allows us to estimate $\langle\psi|U^\dagger P_i U|\psi\rangle$. Then, we show that we can efficiently reconstruct $U^\dagger P_i U$ from a small number of different random input states.

After learning the $3n$ Heisenberg-evolved Pauli operators $U^\dagger P_i U$, we present a direct approach for sewing them into a circuit. This approach uses the identity

$$\mathrm{SWAP} = \frac{1}{2} \sum_{P \in \{I,X,Y,Z\}} P \otimes P.$$

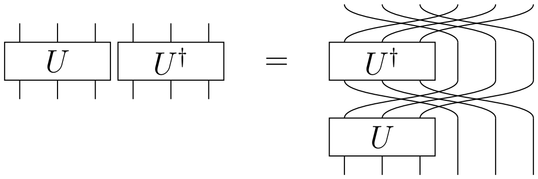
Let S_i be the SWAP gate acting on the i-th system qubit and the i-th ancilla qubit, let $S = \otimes_{i=1}^n S_i$ be the global swap between system and ancilla, and let $W_i := U^\dagger S_i U = \frac{1}{2} \sum_{P \in \{I,X,Y,Z\}} U^\dagger P_i U \otimes P, \forall i = 1, \ldots, n$. From the previous technique for sewing local inversion, we have proven the identity

$$U \otimes U^\dagger = S \cdot \prod_{i=1}^{n} \left(V_i \cdot S_i \cdot V_i^\dagger \right), \tag{3.6}$$

where V_i satisfies $UV_i = U'^{(i)} \otimes I_i$ is an arbitrary exact local inversion on qubit i. We can see that

$$V_i \cdot S_i \cdot V_i^\dagger = U^\dagger U V_i \cdot S_i \cdot V_i^\dagger U^\dagger U = U^\dagger S_i U = W_i \Longrightarrow U \otimes U^\dagger = S \cdot \prod_{i=1}^{n} W_i = S \cdot \prod_{i=1}^{n} \left(U^\dagger S_i U \right). \tag{3.7}$$

The new equation can also be seen by itself: simply cancel U with $U^\dagger$ in the product so that the right-hand side becomes $SU^\dagger SU$, and observe that



(3.8)

As we can see, the Heisenberg-evolved Pauli operators can be directly sewn into $U \otimes U^\dagger$.

This outlines the following procedure to learn U: first learn the Heisenberg-evolved Pauli operators $\{U^\dagger P_i U\}_{i=1}^n$, combine them to form $\{W_i\}_{i=1}^n$ according to

$$W_i = \frac{1}{2} \sum_{P \in \{I,X,Y,Z\}} U^\dagger P_i U \otimes P_i,$$

and reconstruct the circuit using $\{W_i\}_{i=1}^n$. Note that each W_i acts on a constant number k of qubits and can be directly compiled into a circuit of depth $2^{O(k)}$. To further optimize the depth of the learned circuit, notice that each W_i has the form $W_i = U^\dagger S_i U = V_i S_i V_i^\dagger$, i.e., it can be represented by a depth-$(2d+1)$ circuit. We can find such a representation for W_i by brute-force enumerating all depth-$(2d+1)$ circuits acting on k qubits, and the learned circuit has the same form as in Section 3.2.1.1. This thus provides a simpler framework for learning an unknown shallow quantum circuit U using a classical dataset containing random samples about U.

To prove Theorem 3.1 and 3.2 on learning general and geometrically-local shallow quantum circuits, we combine this framework with some additional ideas on (a) coloring the k-dimensional lattices to ensure all qubits with the same color has nonoverlapping lightcone, (b) truncating small Fourier coefficients to ensure the learned observables acts only on qubits in the support of the true observables, (c) compiling the Heisenberg-evolved Pauli operator when over a finite gate set, and (d) finding a good order to sew the Heisenberg-evolved Pauli operators into a short-depth circuit.

3.2.2 Preliminaries

Let $\mathrm{stab}_1 = \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |y+\rangle, |y-\rangle\}$ be the set of single-qubit stabilizer states. Given an n-qubit unitary U, we use the Catholic letter $\mathcal{U}$ to denote the corresponding CPTP map $\mathcal{U}(X) = UXU^\dagger$. We denote $\mathcal{I}$ as the identity CPTP map. Given a Pauli operator $P \in \{X, Y, Z\}$, we consider P_i to be a multi-qubit operator that is equal to the tensor product of P on the i-th qubit and identity on the rest of the qubits. We also consider the following definitions.

Definition 3.1 (Reduced channel). *Given $n > 0$, $i \in \{1, \ldots, n\}$, and an n-qubit CPTP map $\mathcal{C}$. The reduced channel $\mathcal{E}^{\mathcal{C}}_{\neq i}$ of the CPTP map $\mathcal{C}$ with the i-th qubit removed is*

$$\mathcal{E}^{\mathcal{C}}_{\neq i}(\rho_{\neq i}) = \mathrm{Tr}_i\left(\mathcal{C}\left(\frac{I^{(i)}}{2} \otimes \rho_{\neq i}\right)\right), \tag{3.9}$$

where $\rho_{\neq i}$ is a density matrix on all except the i-th qubit, $I^{(i)}$ is the identity on the i-th qubit, and Tr_i is the partial trace over the i-th qubit. For $k \in \{0, 1, \ldots, n\}$, we define

$$\mathcal{E}^{\mathcal{C}}_{>k}(\rho_{>k}) = \mathrm{Tr}_{\leq k}\left(\mathcal{C}\left(\frac{I^{(1,\ldots,k)}}{2^k} \otimes \rho_{>k}\right)\right), \tag{3.10}$$

where $\rho_{>k}$ is a density matrix on all except the first k qubits, $I^{(1,\ldots,k)}$ is the identity on the first k qubits, and $\mathrm{Tr}_{\leq k}$ is the partial trace over the first k qubits. Given a subset of qubits $S \subseteq \{1,\ldots,n\}$, we define

$$\mathcal{E}^{\mathcal{C}}_S(\cdot) = \mathrm{Tr}_{\notin S}\left(\mathcal{C}\left(\frac{I^{(\notin S)}}{2^{n-|S|}} \otimes (\cdot)\right)\right), \tag{3.11}$$

where $I^{(\notin S)}$ is the identity on qubits not in S and $\mathrm{Tr}_{\notin S}$ is the partial trace over qubits not in S.

Definition 3.2 (Fidelity)**.** *Given two quantum states ρ, σ. The fidelity $\mathcal{F}(\rho,\sigma) \in [0,1]$ between the two states is defined as $\mathrm{Tr}\big(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\big)^2$. If $\sigma = |\psi\rangle\!\langle\psi|$, then $\mathcal{F}(\rho,\sigma) = \langle\psi|\,\rho\,|\psi\rangle$.*

Fact 3.1 (Properties of fidelity [206])**.** *The function $1 - F(\rho,\sigma)$ satisfies*

$$1 - F(\rho,\sigma) = 1 - F(\sigma,\rho) \qquad \textit{(symmetric)}; \tag{3.12}$$

$$1 - F(\rho,\sigma) \geq 0 \qquad \textit{(nonnegative)}; \tag{3.13}$$

$$1 - F(\rho,\sigma) = 0 \iff \rho = \sigma \qquad \textit{(identity of indiscernible)}. \tag{3.14}$$

But $1-F$ does not satisfy triangle inequality. In contrast, $\Theta(\rho,\sigma) := \arcsin\Big(\sqrt{1 - F(\rho,\sigma)}\Big) \in [0,\pi/2]$ is symmetric, nonnegative, and satisfies identity of indiscernible and triangle inequality,

$$\Theta(\rho,\sigma) \leq \Theta(\rho,\tau) + \Theta(\tau,\sigma). \tag{3.15}$$

Hence, $\theta(\rho,\sigma)$ is a metric (known as the Fubini-Study metric), but $1 - F(\rho,\sigma)$ is not. In addition to the metric properties, we also have

$$1 - F(\psi,\rho) \leq \frac{1}{2}\|\psi - \rho\|_{\mathrm{tr}}, \tag{3.16}$$

for any state ρ and any pure state ψ, where $\|\cdot\|_{\mathrm{tr}}$ is the trace norm. Also, the fidelity is monotonic increasing under CPTP maps,

$$F(\mathcal{E}(\rho),\mathcal{E}(\sigma)) \geq F(\rho,\sigma), \tag{3.17}$$

for any CPTP map $\mathcal{E}$ and any state ρ,σ.

Definition 3.3 (Average-case distance)**.** *Given two n-qubit CPTP maps $\mathcal{E}_1, \mathcal{E}_2$. The average-case distance $\mathcal{D}_{\mathrm{ave}}(\mathcal{E}_1,\mathcal{E}_2)$ between the two CPTP maps is defined as*

$$\mathop{\mathbb{E}}_{|\psi\rangle:\mathrm{Unif}}\big[1 - \mathcal{F}(\mathcal{E}_1(|\psi\rangle\!\langle\psi|), \mathcal{E}_2(|\psi\rangle\!\langle\psi|))\big], \tag{3.18}$$

where $\mathbb{E}_{|\psi\rangle:\mathrm{Unif}}$ considers averaging under the uniform measure over pure states.

Fact 3.2 (Haar average for average-case distance [203]). *Given an n-qubit CPTP map $\mathcal{E}$ and an n-qubit unitary U. We have the following identity,*

$$\mathcal{D}_{\text{ave}}(\mathcal{E},\mathcal{U}) = \frac{2^n}{2^n+1}\left(1 - \frac{1}{4^n}\sum_{i,j} \langle i|\, \mathcal{E}\left(U^\dagger \,|i\rangle\!\langle j|\, U\right)|j\rangle\right), \tag{3.19}$$

after averaging over the uniform measure over pure states.

Proposition 3.1 (Normalized Frobenius norm). *Given two n-qubit unitaries U_1, U_2. We have*

$$\frac{1}{3}\min_{\phi\in\mathbb{R}} \frac{\left\|e^{i\phi}U_1 - U_2\right\|_F^2}{2^n} \leq \mathcal{D}_{\text{ave}}(\mathcal{U}_1,\mathcal{U}_2) \leq \min_{\phi\in\mathbb{R}} \frac{\left\|e^{i\phi}U_1 - U_2\right\|_F^2}{2^n}, \tag{3.20}$$

where $\|X\|_F = \sqrt{\mathrm{Tr}(X^\dagger X)}$ is the Frobenius norm of X.

Proof. From [203], the average-case distance (also known as the average gate fidelity) satisfies

$$\mathcal{D}_{\text{ave}}(\mathcal{U}_1,\mathcal{U}_2) = \frac{2^n}{2^n+1}\left(1 - \frac{1}{4^n}\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|^2\right). \tag{3.21}$$

Expanding the definition of Frobenius norm, we have

$$\min_{\phi\in\mathbb{R}} \frac{\left\|e^{i\phi}U_1 - U_2\right\|_F^2}{2^n} = 2\left(1 - \frac{\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|}{2^n}\right). \tag{3.22}$$

Recall that

$$0 \leq \frac{\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|}{2^n} \leq 1. \tag{3.23}$$

Hence, we have

$$\left(1 - \frac{\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|}{2^n}\right) \leq \left(1 + \frac{\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|}{2^n}\right)\left(1 - \frac{\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|}{2^n}\right) \leq 2\left(1 - \frac{\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|}{2^n}\right). \tag{3.24}$$

This immediately implies that

$$\frac{2}{3}\left(1 - \frac{\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|}{2^n}\right) \leq \frac{2^n}{2^n+1}\left(1 - \frac{\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|^2}{4^n}\right) \leq 2\left(1 - \frac{\left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|}{2^n}\right) \tag{3.25}$$

which is equivalent to

$$\frac{1}{3}\min_{\phi\in\mathbb{R}} \frac{\left\|e^{i\phi}U_1 - U_2\right\|_F^2}{2^n} \leq \mathcal{D}_{\text{ave}}(\mathcal{U}_1,\mathcal{U}_2) \leq \min_{\phi\in\mathbb{R}} \frac{\left\|e^{i\phi}U_1 - U_2\right\|_F^2}{2^n}. \tag{3.26}$$

This concludes the proof. □

Definition 3.4 (Worse-case distance / diamond distance)**.** *Given two n-qubit CPTP maps $\mathcal{E}_1, \mathcal{E}_2$. The worst-case distance $\mathcal{D}_\diamond(\mathcal{E}_1, \mathcal{E}_2)$ between the two CPTP maps is defined as*

$$\frac{1}{2}\max_\rho \|(\mathcal{E}_1 \otimes \mathcal{I})(\rho) - (\mathcal{E}_2 \otimes \mathcal{I})(\rho)\|_1 \triangleq \frac{1}{2}\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond, \tag{3.27}$$

where ρ is maximized over $2n$-qubit states and $\mathcal{I}^{(>n)}$ is an identity map acting on the n qubits. $\mathcal{D}_\diamond(\mathcal{E}_1, \mathcal{E}_2)$ is also known as diamond distance and $\|\cdot\|_\diamond$ is the diamond norm.

Fact 3.3 (Diamond distance for unitaries; Prop. 1.6 of [207])**.** *For any two unitaries U_1, U_2, we have*

$$\min_{\phi\in\mathbb{R}} \left\|e^{i\phi}U_1 - U_2\right\|_\infty \leq \|\mathcal{U}_1 - \mathcal{U}_2\|_\diamond \leq 2\min_{\phi\in\mathbb{R}} \left\|e^{i\phi}U_1 - U_2\right\|_\infty. \tag{3.28}$$

Fact 3.4 (Exact unitary synthesis; see e.g. [208, 209])**.** *Given any unitary U acting on k qubits, there is an algorithm that outputs a circuit (acting on k qubits) consisting of at most 4^k two-qubit gates, which exactly implements the unitary U, in time $2^{O(k)}$.*

Corollary 3.1 (Exact unitary synthesis in geometrically-local circuit)**.** *Given any unitary U acting on k qubits and a connected graph G over k qubits, there is an algorithm that outputs a geometrically-local circuit (acting on k qubits and consists only of gates between connected qubits) consisting of at most $2k4^k$ two-qubit gates, which exactly implements the unitary U, in time $2^{O(k)}$.*

Proof. For each two-qubit gate in the original synthesis protocol, which may not be geometrically-local under the connectivity graph G, we consider at most $k-1$ swap gates to move one of the qubits from the original location to a location next to the other qubit, apply the two-qubit gate, then perform at most $k-1$ swap gates to move the qubit back to the original location. □

3.2.3 Approximate local identity

A central concept that we will use to define local inversion for representing n-qubit unitaries is the *ε-approximate local identity*. In this section, we provide the properties for understanding the concept of approximate local identity. In particular, we will consider a strong and a weak form of local identity in Section 3.2.3.1 and 3.2.3.2. In each section, we state the definition, show how to characterize if a unitary map forms a strong/weak ε-approximate local identity, and prove how local identity relates to global identity.

3.2.3.1 Strong ε-approximate local identity

We begin by looking at a strong form of approximate local identity. The idea is that the action of the n-qubit unitary U on the i-th qubit is close to the identity map, while the action on the other qubits is close to the reduced channel of U with the i-th qubit removed (feed in a maximally mixed state on qubit i and trace out qubit i at the end). Recall Definition 3.1 of reduced channel,

$$\mathcal{E}^{\mathcal{U}}_{\neq i}(\rho_{\neq i}) = \mathrm{Tr}_i \left(\mathcal{U} \left(\frac{I^{(i)}}{2} \otimes \rho_{\neq i} \right) \right), \tag{3.29}$$

where $\rho_{\neq i}$ is a density matrix on all except the i-th qubit, $I^{(i)}$ is the identity on the i-th qubit, and Tr_i is the partial trace over the i-th qubit.

Definition 3.5 (Strong ε-approximate local identity)**.** *Given $n > 0, \varepsilon \geq 0,$ and $i \in \{1, \ldots, n\}$. An n-qubit unitary U is a strong ε-approximate local identity on the i-th qubit if*

$$\mathcal{D}_\diamond \left(\mathcal{U}, \mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i} \right) \leq \varepsilon, \tag{3.30}$$

where $\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i}$ is an n-qubit CPTP map that acts as identity on the i-th qubit.

While diamond distances are typically hard to characterize, the strong ε-approximate local identity can be characterized up to a constant factor by studying the Heisenberg evolution of single-qubit Pauli observables under the n-qubit unitary U. Hence, in order to check if an n-qubit unitary U strong approximate local identity on the i-th qubit, all we need to check is whether the three Pauli observables X_i, Y_i, Z_i remains approximately unchanged after Heisenberg evolution under U.

Lemma 3.1 (Characterization of strong ε-approximate local identity)**.** *Given $n > 0$, $\varepsilon \geq 0$, and an n-qubit unitary $\mathcal{U}$. If $\mathcal{U}$ is a strong ε-approximate local identity on the i-th qubit, then*

$$\frac{1}{2} \left\| U^\dagger P_i U - P_i \right\|_\infty \leq \varepsilon, \forall P \in \{X, Y, Z\}, \tag{3.31}$$

where P_i is the Pauli operator P acting only on qubit i, and $U^\dagger P_i U$ is the Heisenberg evolution of P_i under U. Furthermore, if the following holds,

$$\frac{1}{2} \sum_{P \in \{X,Y,Z\}} \left\| U^\dagger P_i U - P_i \right\|_\infty \leq \varepsilon, \tag{3.32}$$

then $\mathcal{U}$ is a strong ε-approximate local identity on the i-th qubit.

Proof. We start by showing the first claim. Consider any n-qubit pure state $|\psi\rangle$. We have

$$\left\| U^\dagger P_i U - P_i \right\|_\infty = \max_{|\psi\rangle} \left| \langle\psi| \left(U^\dagger P_i U - P_i \right) |\psi\rangle \right|. \tag{3.33}$$

By the definition of CPTP maps, we have

$$\langle\psi| U^\dagger P_i U |\psi\rangle = \mathrm{Tr} \left(P_i \mathcal{U} \left(|\psi\rangle\!\langle\psi| \right) \right). \tag{3.34}$$

From the definition of diamond distance and of strong ε-approximate local identity on the i-th qubit, we have the following inequality,

$$\frac{1}{2} \left| \mathrm{Tr} \left(P_i \mathcal{U} \left(|\psi\rangle\!\langle\psi| \right) \right) - \mathrm{Tr} \left(P_i \left(\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i} \right) \left(|\psi\rangle\!\langle\psi| \right) \right) \right| \leq \varepsilon. \tag{3.35}$$

By the definition of a CPTP map, we have

$$\mathrm{Tr}_{\neq i}\left(\mathcal{E}^{\mathcal{U}}_{\neq i}(\rho)\right) = \rho \tag{3.36}$$

for any quantum state ρ, where $\mathrm{Tr}_{\neq i}$ traces out all qubits except for qubit i. Hence, we have $\mathrm{Tr}\left(P_i\left(\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i}\right)(|\psi\rangle\langle\psi|)\right) = \mathrm{Tr}(P_i\,|\psi\rangle\langle\psi|)$. Together, we obtain the first claim.

The second claim uses the following equality defined over an $n+1$-qubit system,

$$\frac{1}{2}\left(I_{n+1} + \sum_{P\in\{X,Y,Z\}} P_i \otimes P\right) = S_{i,n+1}, \tag{3.37}$$

where I_{n+1} is an $n+1$-qubit identity, P_i is an n-qubit unitary that acts as the Pauli operator P on the i-th qubit, and $S_{i,n+1}$ is the swap operator between qubit i in the first n qubits and the last qubit (qubit $n+1$). We interpret the error in the Heisenberg-evolved single-qubit Pauli observables as an error in commuting the Pauli observable P_i and the n-qubit unitary U,

$$\left\|U^\dagger P_i U - P_i\right\|_\infty = \|P_i U - U P_i\|_\infty. \tag{3.38}$$

From this interpretation, we have the following inequalities,

$$\|S_{i,n+1}(U\otimes I) - (U\otimes I)S_{i,n+1}\|_\infty \leq \frac{1}{2}\sum_{P\in\{X,Y,Z\}} \|(P_i\otimes P)(U\otimes I) - (U\otimes I)(P_i\otimes P)\|_\infty \tag{3.39}$$

$$\leq \frac{1}{2}\sum_{P\in\{X,Y,Z\}} \|(P_i U - U P_i)\otimes P\|_\infty \tag{3.40}$$

$$= \frac{1}{2}\sum_{P\in\{X,Y,Z\}} \|P_i U - U P_i\|_\infty \leq \varepsilon. \tag{3.41}$$

The above inequality can be easily generalized to any of the following,

$$\|S_{i,j}(U\otimes I_m) - (U\otimes I_m)S_{i,j}\|_\infty \leq \varepsilon, \tag{3.42}$$

where $m \geq 1$, $n+1 \leq j \leq n+m$, and I_m is the identity operator on m qubits. Recall the formal definition diamond distance from Definition 3.4,

$$\mathcal{D}_\diamond(\mathcal{E}_1, \mathcal{E}_2) = \frac{1}{2}\max_\rho \|(\mathcal{E}_1\otimes\mathcal{I}_n)(\rho) - (\mathcal{E}_2\otimes\mathcal{I}_n)(\rho)\|_1, \tag{3.43}$$

where ρ is a density matrix over $2n$ qubits, and $\mathcal{I}_n$ is the identity map over n qubits. From Fact 3.3, for any two unitaries U_1, U_2, we have $\|\mathcal{U}_1 - \mathcal{U}_2\|_\diamond \leq 2\|U_1 - U_2\|_\infty$. We obtain the following from Eq. (3.42),

$$\left\|\mathcal{S}_{i,j}(\mathcal{U}\otimes I_m) - (\mathcal{U}\otimes I_m)\mathcal{S}_{i,j}\Big)\right\|_\diamond \leq 2\|S_{i,j}(U\otimes I_m) - (U\otimes I_m)S_{i,j}\|_\infty \leq 2\varepsilon. \tag{3.44}$$

The strong ε-approximate local identity considers

$$\mathcal{D}_\diamond\left(\mathcal{U},\, \mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i}\right) = \frac{1}{2} \max_\rho \left\|(\mathcal{U} \otimes \mathcal{I}_n)(\rho) - (\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i} \otimes \mathcal{I}_n)(\rho)\right\|_1. \tag{3.45}$$

We add one more qubit to form $2n+1$ qubits. The additional qubit begins in a maximally mixed state $I/2$, stays in $I/2$, and is traced out at the end. Let us now consider the following series of analysis,

$$\begin{aligned}
&\left\|(\mathcal{U} \otimes \mathcal{I}_n)(\rho) - (\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i} \otimes \mathcal{I}_n)(\rho)\right\|_1 && (3.46)\\
&= \left\|\mathrm{Tr}_{2n+1}\left[(\mathcal{U} \otimes \mathcal{I}_{n+1})(\rho \otimes (I/2))\right] - (\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i} \otimes \mathcal{I}_n)(\rho)\right\|_1 && (3.47)\\
&= \left\|\mathrm{Tr}_i\left[\left(\mathcal{S}_{i,2n+1} \circ (\mathcal{U} \otimes \mathcal{I}_{n+1})\right)(\rho \otimes (I/2))\right] - (\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i} \otimes \mathcal{I}_{n+1})(\rho \otimes (I/2))\right\|_1 && (3.48)\\
&\leq \left\|\mathrm{Tr}_i\left[\left((\mathcal{U} \otimes \mathcal{I}_{n+1}) \circ \mathcal{S}_{i,2n+1}\right)(\rho \otimes (I/2))\right] - (\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i} \otimes \mathcal{I}_{n+1})(\rho \otimes (I/2))\right\|_1 + 2\varepsilon && (3.49)\\
&= \left\|(\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i} \otimes \mathcal{I}_{n+1})(\rho \otimes (I/2)) - (\mathcal{I}^{(i)} \otimes \mathcal{E}^{\mathcal{U}}_{\neq i} \otimes \mathcal{I}_{n+1})(\rho \otimes (I/2))\right\|_1 + 2\varepsilon = 2\varepsilon. && (3.50)
\end{aligned}$$

The only inequality above uses Eq. (3.44). We have proved the claim. □

The following two lemmas give the relationships between global and local identity checks. The basic idea is to check whether a map is close to identity by checking whether the map forms approximate local identities on all the n qubits. If the map is far from identity, then the map is not an approximate local identity for some qubits. If the map is an approximate local identity for all qubits, then the map is close to the identity.

Lemma 3.2 (Global non-identity check from local non-identity checks)**.** *Given an integer $n > 0$ and an n-qubit unitary U. If there exists $\varepsilon > 0$ and $i \in \{1, \ldots, n\}$, such that $\mathcal{U}$ is not a strong ε-approximate local identity on the i-th qubit, then $\|\mathcal{U} - \mathcal{I}\|_\diamond \geq \varepsilon/2$.*

Lemma 3.3 (Global identity check from local identity checks)**.** *Given an integer $n > 0$ and an n-qubit unitary U. If there exists $\varepsilon_1, \ldots, \varepsilon_n > 0$, such that $\mathcal{U}$ is a strong ε_i-approximate local identity on the i-th qubit for all $i \in \{1, \ldots, n\}$, then $\|\mathcal{U} - \mathcal{I}\|_\diamond \leq 3 \sum_{i=1}^n \varepsilon_i$.*

We give proofs of these two lemmas at the end of this subsection. Lemma 3.2 is proven by contradiction. To prove Lemma 3.3, we consider a stabilizer decomposition for a single qubit.

Proposition 3.2 (Single-qubit stabilizer decomposition)**.** *Given an integer $n > 0$ and an n-qubit density matrix ρ. For any $S \subseteq \{1, \ldots, n\}$, ρ can be written as a linear combination of $R = 10^{|S|}$ n-qubit density matrices $\rho_1, \ldots, \rho_R$, $\rho = \sum_{r=1}^R \alpha_r \rho_r$, where $\alpha_r \in \mathbb{R}$ and ρ_r is a density matrix that satisfies*

$$\rho_r = \bigotimes_{j \in S} |s_j\rangle\langle s_j| \otimes \mathrm{Tr}_S(\rho_r), \tag{3.51}$$

for some $|s_j\rangle \in \mathrm{stab}_1$. We also have $\sum_{r=1}^R \alpha_r = 1$ and $\sum_{r=1}^R |\alpha_r| = 3^{|S|}$.

Proof. Given an integer $i \in \{1, \ldots, n\}$, consider the following linear map $\mathcal{M}_i$ which equals to the identity channel on i-th qubit,

$$\begin{aligned}\mathcal{M}_i(\rho) := &\left|0\middle\rangle\!\middle\langle 0\right|_i \otimes \langle 0|\,\rho\,|0\rangle_i + \left|1\middle\rangle\!\middle\langle 1\right|_i \otimes \langle 1|\,\rho\,|1\rangle_i \\ &+ \frac{1}{2}\left|+\middle\rangle\!\middle\langle +\right|_i \otimes \langle +|\,\rho\,|+\rangle_i - \frac{1}{2}\left|+\middle\rangle\!\middle\langle +\right|_i \otimes \langle -|\,\rho\,|-\rangle_i \\ &- \frac{1}{2}\left|-\middle\rangle\!\middle\langle -\right|_i \otimes \langle +|\,\rho\,|+\rangle_i + \frac{1}{2}\left|-\middle\rangle\!\middle\langle -\right|_i \otimes \langle -|\,\rho\,|-\rangle_i \\ &+ \frac{1}{2}\left|y+\middle\rangle\!\middle\langle y+\right|_i \otimes \langle y+|\,\rho\,|y+\rangle_i - \frac{1}{2}\left|y+\middle\rangle\!\middle\langle y+\right|_i \otimes \langle y-|\,\rho\,|y-\rangle_i \\ &- \frac{1}{2}\left|y-\middle\rangle\!\middle\langle y-\right|_i \otimes \langle y+|\,\rho\,|y+\rangle_i + \frac{1}{2}\left|y-\middle\rangle\!\middle\langle y-\right|_i \otimes \langle y-|\,\rho\,|y-\rangle_i\,, \end{aligned} \tag{3.52}$$

$$= \sum_{r=1}^{10} b_r \left|s_r\middle\rangle\!\middle\langle s_r\right|_i \otimes \langle s'_r|\,\rho\,|s'_r\rangle_i\,. \tag{3.53}$$

where $|s\rangle\!\langle s|_i$ is a single-qubit stabilizer state on the i-th qubit, $\langle s|\,\rho\,|s\rangle_i$ is a partial inner product on the i-th qubit, s_r, s'_r, b_r takes on the corresponding values in stab_1, stab_1, $\{1, 1/2, -1/2\}$, respectively. The fact that $\mathcal{M}_i$ equals to the identity CPTP map $\mathcal{I}$ is because of the following identity

$$\rho = \sum_{P \in \{I,X,Y,Z\}} \mathrm{Tr}_i(P_i \rho) \otimes \frac{P_i}{2}, \tag{3.54}$$

where P_i acts on the i-th qubit, and Eq. (3.52) follows by further decomposing the Pauli operators into their eigenstates.

Without loss of generality, we consider $k = |S|$ and $S = \{1, \ldots, k\}$. The identity $\rho = (\circ_{i \in S} \mathcal{M}_i)(\rho)$ gives rise to the equality

$$\rho = \sum_{r_1=1}^{10} \cdots \sum_{r_k=1}^{10} \left(\prod_{i=1}^{k} b_{r_i} \right) \left|s_{r_1}, \ldots, s_{r_k}\middle\rangle\!\middle\langle s_{r_1}, \ldots, s_{r_k}\right| \otimes \left\langle s'_{r_1}, \ldots, s'_{r_k}\right|\rho\left|s'_{r_1}, \ldots, s'_{r_k}\right\rangle. \tag{3.55}$$

We define $r = \sum_{i=1}^{k} 10^{i-1} r_i$, $R = 10^k$, $Z_r = \mathrm{Tr}\big(\left\langle s'_{r_1}, \ldots, s'_{r_k}\right|\rho\left|s'_{r_1}, \ldots, s'_{r_k}\right\rangle\big) \geq 0$, and

$$\rho_r = \begin{cases} \left|s_{r_1}, \ldots, s_{r_k}\middle\rangle\!\middle\langle s_{r_1}, \ldots, s_{r_k}\right| \otimes \frac{\left\langle s'_{r_1}, \ldots, s'_{r_k}\right|\rho\left|s'_{r_1}, \ldots, s'_{r_k}\right\rangle}{Z_r} & \text{if } Z_r > 0, \\ \left|s_{r_1}, \ldots, s_{r_k}\middle\rangle\!\middle\langle s_{r_1}, \ldots, s_{r_k}\right| \otimes \frac{I}{2^{n-k}} & \text{if } Z_r = 0, \end{cases} \tag{3.56}$$

and $\alpha_r = Z_r \prod_{i=1}^{k} b_{r_i}$. It is not hard to check that $\sum_r |\alpha_r| = 3^k$. Together, we have the single-qubit stabilizer decomposition $\rho = \sum_{r=1}^{R} \alpha_r \rho_r$. □

Proof of Lemma 3.2. We consider proof by contradiction. Assume $\|\mathcal{U} - \mathcal{I}\|_\diamond < \varepsilon/2$. For any integer $m \geq 0$, for any state $|s\rangle_i \in \mathrm{stab}_1$ on the i-th qubit, and for any $(n-1+m)$-qubit density matrix ρ,

$$
\begin{aligned}
&\left\|(\mathcal{U}\otimes\mathcal{I}^{(>n)})\left(|s\rangle\langle s|_i\otimes\rho\right) - |s\rangle\langle s|_i\otimes(\mathcal{E}^U_{\neq i}\otimes\mathcal{I}^{(>n)})(\rho)\right\|_1 && (3.57)\\
&\leq \left\|(\mathcal{U}\otimes\mathcal{I}^{(>n)})\left(|s\rangle\langle s|_i\otimes\rho\right) - |s\rangle\langle s|_i\otimes\rho\right\|_1 + \left\||s\rangle\langle s|_i\otimes\rho - |s\rangle\langle s|_i\otimes(\mathcal{E}^U_{\neq i}\otimes\mathcal{I}^{(>n)})(\rho)\right\|_1 && (3.58)\\
&\leq \|\mathcal{U}-\mathcal{I}\|_\diamond + \|\mathcal{U}-\mathcal{I}\|_\diamond < \varepsilon. && (3.59)
\end{aligned}
$$

The first inequality follows from putting in $|s\rangle\langle s|_i\otimes\rho$ and using triangle inequality. The second inequality follows from the definition of diamond distance, the identity

$$
\begin{aligned}
&\left\||s\rangle\langle s|_i\otimes\rho - |s\rangle\langle s|_i\otimes(\mathcal{E}^U_{\neq i}\otimes\mathcal{I}^{(>n)})(\rho)\right\|_1 && (3.60)\\
&= \left\||s\rangle\langle s|_i\otimes\mathrm{Tr}_i\left(\frac{I^{(i)}}{2}\otimes\rho\right) - |s\rangle\langle s|_i\otimes\mathrm{Tr}_i\left((\mathcal{U}\otimes\mathcal{I}^{(>n)})\left(\frac{I^{(i)}}{2}\otimes\rho\right)\right)\right\|_1, && (3.61)
\end{aligned}
$$

and the two facts: $\|\rho_A\otimes\rho_B - \rho_A\otimes\rho_C\|_1 = \|\rho_B-\rho_C\|_1, \|\mathrm{Tr}_i(\rho_A)\|_1 \leq \|\mathrm{Tr}(\rho_A)\|_1$ for any density matrix ρ_A,ρ_B,ρ_C. The above derivation shows that U is an ε-approximate local identity on the i-th qubit, which is a contradiction. Therefore, $\|\mathcal{U}-\mathcal{I}\|_\diamond \geq \varepsilon/2$. □

Proof of Lemma 3.3. From Theorem 3.55 in [210], we have

$$
\|\mathcal{U}-\mathcal{I}\|_\diamond = \left\|U|\psi\rangle\langle\psi|U^\dagger - |\psi\rangle\langle\psi|\right\|_1 \tag{3.62}
$$

for some n-qubit state $|\psi\rangle$. Let $\mathcal{I}^{(\leq k)}$ be the identity CPTP map acting on the first k qubit. We use a telescoping sum of the form,

$$
U|\psi\rangle\langle\psi|U^\dagger - |\psi\rangle\langle\psi| = \sum_{k=0}^{n-1}\left[\left(\mathcal{I}^{(\leq k)}\otimes\mathcal{E}^U_{>k}\right)(|\psi\rangle\langle\psi|) - \left(\mathcal{I}^{(\leq k+1)}\otimes\mathcal{E}^U_{>k+1}\right)(|\psi\rangle\langle\psi|)\right]. \tag{3.63}
$$

By triangle inequality, we obtain

$$
\|\mathcal{U}-\mathcal{I}\|_\diamond \leq \sum_{k=0}^{n-1}\left\|\left(\mathcal{I}^{(\leq k)}\otimes\mathcal{E}^U_{>k}\right)(|\psi\rangle\langle\psi|) - \left(\mathcal{I}^{(\leq k+1)}\otimes\mathcal{E}^U_{>k+1}\right)(|\psi\rangle\langle\psi|)\right\|_1. \tag{3.64}
$$

In the next step, we will bound each term in the above telescoping sum.

To bound the term corresponding to $k\in\{0,\ldots,n-1\}$ in Eq. (3.64), we consider an $(k+(n-k)+k)$-qubit density matrix $\rho^{(k)}$. The first k qubits of $\rho^{(k)}$ is the maximally mixed state $\frac{I^{(1,\ldots,k)}}{2^k}$. The next $(n-k)$ qubits of $\rho^{(k)}$ corresponds to all except the first k qubits in $|\psi\rangle\langle\psi|$. The last k qubits of $\rho^{(k)}$ corresponds to the first k qubits in $|\psi\rangle\langle\psi|$. Under this definition of $\rho^{(k)}$, we have

$$
\begin{aligned}
&\left\|\left(\mathcal{I}^{(\leq k)}\otimes\mathcal{E}^U_{>k}\right)(|\psi\rangle\langle\psi|) - \left(\mathcal{I}^{(\leq k+1)}\otimes\mathcal{E}^U_{>k+1}\right)(|\psi\rangle\langle\psi|)\right\|_1 && (3.65)\\
&= \left\|\left(\mathcal{U}\otimes\mathcal{I}^{(>n)}\right)(\rho^{(k)}) - \left(\mathcal{I}^{(k+1)}\otimes\mathcal{E}^U_{\neq k+1}\otimes\mathcal{I}^{(>n)}\right)(\rho^{(k)})\right\|_1, && (3.66)
\end{aligned}
$$

where $\left(\mathcal{I}^{(k+1)} \otimes \mathcal{E}^U_{\neq k+1} \otimes \mathcal{I}^{(>n)}\right)\left(\rho^{(k)}\right)$ is the output state after applying the $(n-1)$-qubit CPTP map $\mathcal{E}^U_{\neq k+1}$ to the first n qubits except the $(k+1)$-th qubit of $\rho^{(k)}$. We now use the single-qubit stabilizer decomposition with $S = \{k+1\}$ given in Prop. 3.2 to obtain $\rho^{(k)} = \sum_{r=1}^{10} \alpha_r \rho_r^{(k)}$ with $\sum_r |\alpha_r| = 3$ and the reduced density matrix of $\rho_r^{(k)}$ on the $(k+1)$-th qubit is a single-qubit stabilizer state. We can now bound each term by

$$\left\|\left(\mathcal{U} \otimes \mathcal{I}^{(>n)}\right)\left(\rho^{(k)}\right) - \left(\mathcal{I}^{(k+1)} \otimes \mathcal{E}^U_{\neq k+1} \otimes \mathcal{I}^{(>n)}\right)\left(\rho^{(k)}\right)\right\|_1 \tag{3.67}$$

$$\leq \sum_{r=1}^{10} |\alpha_r| \left\|\left(\mathcal{U} \otimes \mathcal{I}^{(>n)}\right)\left(\rho_r^{(k)}\right) - \left(\mathcal{I}^{(k+1)} \otimes \mathcal{E}^U_{\neq k+1} \otimes \mathcal{I}^{(>n)}\right)\left(\rho_r^{(k)}\right)\right\|_1 \tag{3.68}$$

$$\leq \sum_{r=1}^{10} |\alpha_r| \varepsilon_{k+1} = 3\varepsilon_{k+1}. \tag{3.69}$$

The first line is the triangle inequality. The second line uses the assumption that U is an ε_{k+1}-approximate local identity on the $(k+1)$-th qubit. Combining Eq. (3.64), Eq. (3.66), Eq. (3.69),

$$\left\|\mathcal{U} - \mathcal{I}\right\|_\diamond \leq 3 \sum_{k=0}^{n-1} \varepsilon_{k+1}, \tag{3.70}$$

which establishes the stated result. □

3.2.3.2 Weak ε-approximate local identity

We next look at another definition of approximate local identity: the reduced channel of U on the i-th qubit is close to the identity map. This definition is very easy to check but only guarantees that the unitary U is close to the identity in the average-case distance (instead of the worst-case distance, i.e., the diamond distance). Hence, we will refer to this as the weak ε-approximate local identity. Recall Definition 3.1 of reduced channel,

$$\mathcal{E}_i^{\mathcal{U}}(\rho_i) = \mathrm{Tr}_{\neq i}\left(U\left(\frac{I^{(\neq i)}}{2^{n-1}} \otimes \rho_i\right) U^\dagger\right), \tag{3.71}$$

where ρ_i is a density matrix on the i-th qubit, $I^{(\neq i)}$ is the identity on all except the i-th qubit, and $\mathrm{Tr}_{\neq i}$ is the partial trace over all except the i-th qubit.

Definition 3.6 (Weak ε-approximate local identity; unitary version). *Given $n > 0, \varepsilon \geq 0$, and $i \in \{1, \ldots, n\}$. An n-qubit unitary U is a weak ε-approximate local identity on the i-th qubit if*

$$\mathcal{D}_{\text{ave}}\left(\mathcal{E}_i^{\mathcal{U}}, \mathcal{I}\right) \leq \varepsilon, \tag{3.72}$$

where $\mathcal{I}$ is a 1-qubit CPTP map that acts as an identity.

In the literature of quantum junta learning [211], one defines the influence of a qubit i in an n-qubit unitary $U = \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \alpha_P P$, where $\alpha_P \in \mathbb{C}$ to be

$$\sum_{\substack{P \in \{I,X,Y,Z\}^{\otimes n} \\ P_i \neq I}} \left|\alpha_P\right|^2 . \tag{3.73}$$

The following lemma shows that weak approximate local identity is equivalent to low influence.

Lemma 3.4 (Characterization of weak ε-approximate local identity)**.** *Given $n > 0$, $\varepsilon \geq 0$, and an n-qubit unitary U. Consider the Pauli representation of $U = \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \alpha_P P$, where $\alpha_P \in \mathbb{C}$. $\mathcal{U}$ is a weak ε-approximate local identity on the i-th qubit if and only if*

$$\sum_{\substack{P \in \{I,X,Y,Z\}^{\otimes n} \\ P_i \neq I}} \left|\alpha_P\right|^2 \leq \frac{3}{2}\varepsilon. \tag{3.74}$$

From the definition of influence in quantum junta learning [211], we have qubit i has influence bounded above by 1.5ε in the unitary U.

Proof. From the definition of the reduced channel, we have

$$\mathcal{E}_i^{\mathcal{U}}(\rho_i) = \sum_{s_1, s_2 \in \{I,X,Y,Z\}} \left(\sum_{\substack{P,Q \in \{I,X,Y,Z\}^{\otimes n} \\ P_i = s_1, Q_i = s_2, P_{\neq i} = Q_{\neq i}}} \alpha_P^* \alpha_Q \right) s_1 \rho_i s_2, \tag{3.75}$$

where $P_{\neq i}, Q_{\neq i}$ is an $(n-1)$-qubit Pauli observable equal to P, Q with qubit i removed. From Fact 3.2 characterizing the average-case distance $\mathcal{D}_{\text{ave}}$, we have

$$\mathcal{D}_{\text{ave}}\left(\mathcal{E}_i^{\mathcal{U}}, \mathcal{I}\right) = \frac{2}{3}\left(1 - \sum_{\substack{P,Q \in \{I,X,Y,Z\}^{\otimes n} \\ P_i = I, Q_i = I, P_{\neq i} = Q_{\neq i}}} \alpha_P^* \alpha_Q \right) = \frac{2}{3}\left(1 - \sum_{\substack{P \in \{I,X,Y,Z\}^{\otimes n} \\ P_i = I}} \left|\alpha_P\right|^2 \right). \tag{3.76}$$

Furthermore, we note that $\mathrm{Tr}\left(U^\dagger U\right) = 2^n = 2^n \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} |\alpha_P|^2$. Hence, we have

$$1 - \sum_{\substack{P \in \{I,X,Y,Z\}^{\otimes n} \\ P_i = I}} \left|\alpha_P\right|^2 = \sum_{\substack{P \in \{I,X,Y,Z\}^{\otimes n} \\ P_i \neq I}} \left|\alpha_P\right|^2 . \tag{3.77}$$

The lemma follows from the two identities given above. □

Weak ε-approximate local identity naturally generalizes to any quantum process (channel) by using the definition of reduced channels for channels. The formal definition is given below.

Definition 3.7 (Weak ε-approximate local identity; channel version)**.** *Given $n > 0, \varepsilon \geq 0$, and $i \in \{1, \ldots, n\}$. An n-qubit CPTP map $\mathcal{C}$ is a weak ε-approximate local identity on the i-th qubit if*

$$\mathcal{D}_{\text{ave}}\left(\mathcal{E}_i^{\mathcal{C}}, \mathcal{I}\right) \leq \varepsilon, \tag{3.78}$$

where $\mathcal{I}$ is a 1-qubit CPTP map that acts as an identity.

The following two lemmas give the relationships between global and local identity checks. The basic idea is to check whether a map is close to identity by checking whether the map forms approximate local identities on all the n qubits.

Lemma 3.5 (Global non-identity check from local non-identity checks)**.** *Given an integer $n > 0$ and an n-qubit CPTP map $\mathcal{C}$. If there exists $\varepsilon > 0$ and $i \in \{1, \ldots, n\}$, such that $\mathcal{C}$ is not a weak ε-approximate local identity on the i-th qubit, then $\mathcal{D}_{\text{ave}}(\mathcal{C}, \mathcal{I}) \geq \varepsilon$.*

Lemma 3.6 (Global identity check from local identity checks)**.** *Given an integer $n > 0$ and an n-qubit CPTP map $\mathcal{C}$. If there exists $\varepsilon_1, \ldots, \varepsilon_n > 0$, such that $\mathcal{C}$ is a weak ε_i-approximate local identity on the i-th qubit for all $i \in \{1, \ldots, n\}$, then $\mathcal{D}_{\text{ave}}(\mathcal{C}, \mathcal{I}) \leq \frac{3}{2} \sum_{i=1}^n \varepsilon_i$.*

Proof of Lemma 3.5 and 3.6. Let us define $|\Omega_1\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, and $|\Omega_n\rangle = |\Omega_1\rangle^{\otimes n}$. From Fact 3.2 characterizing the average-case distance $\mathcal{D}_{\text{ave}}$, we have

$$\mathcal{D}_{\text{ave}}(\mathcal{C}, \mathcal{I}) = \frac{2^n}{2^n + 1} \left(1 - \langle\Omega_n| \left(\mathcal{C} \otimes \mathcal{I}\right) \left(|\Omega_n\rangle\!\langle\Omega_n|\right) |\Omega_n\rangle\right). \tag{3.79}$$

We can think of the term $\langle\Omega_n| \left(\mathcal{C} \otimes \mathcal{I}\right) \left(|\Omega_n\rangle\!\langle\Omega_n|\right) |\Omega_n\rangle$ as the probability of getting $|\Omega_1\rangle$ on all n parallel two-qubit Bell-basis measurements on the $2n$-qubit state $\left(\mathcal{C} \otimes \mathcal{I}\right) \left(|\Omega_n\rangle\!\langle\Omega_n|\right)$. From standard probability theory, we have the following inequality,

$$1 - \langle\Omega_n| \left(\mathcal{E} \otimes \mathcal{I}\right) \left(|\Omega_n\rangle\!\langle\Omega_n|\right) |\Omega_n\rangle \geq 1 - \mathrm{Tr}\left(\left(|\Omega_1\rangle\!\langle\Omega_1| \otimes I_{\neq i}^{\otimes 2}\right) \left(\mathcal{C} \otimes \mathcal{I}\right) \left(|\Omega_n\rangle\!\langle\Omega_n|\right)\right), \tag{3.80}$$

where $|\Omega_1\rangle\!\langle\Omega_1| \otimes I_{\neq i}^{\otimes 2}$ is a projection onto $|\Omega_1\rangle\!\langle\Omega_1|$ on the i-th and $(n+i)$-th qubit for any i. Also, from union bound, we have

$$1 - \langle\Omega_n| \left(\mathcal{E} \otimes \mathcal{I}\right) \left(|\Omega_n\rangle\!\langle\Omega_n|\right) |\Omega_n\rangle \leq 1 - \sum_{i=1}^n \left(1 - \mathrm{Tr}\left(\left(|\Omega_1\rangle\!\langle\Omega_1| \otimes I_{\neq i}^{\otimes 2}\right) \left(\mathcal{C} \otimes \mathcal{I}\right) \left(|\Omega_n\rangle\!\langle\Omega_n|\right)\right)\right). \tag{3.81}$$

By reorganizing using the reduced channel of $\mathcal{C}$ on the i-th qubit, we have

$$\mathrm{Tr}\left(\left(|\Omega_1\rangle\!\langle\Omega_1| \otimes I_{\neq i}^{\otimes 2}\right) \left(\mathcal{C} \otimes \mathcal{I}\right) \left(|\Omega_n\rangle\!\langle\Omega_n|\right)\right) = \langle\Omega_1| \left(\mathcal{E}_i^{\mathcal{C}} \otimes \mathcal{I}\right) \left(|\Omega_1\rangle\!\langle\Omega_1|\right) |\Omega_1\rangle. \tag{3.82}$$

Therefore, we have

$$\frac{3}{2} \times \frac{2^n}{2^n + 1} \sum_{i=1}^n \mathcal{D}_{\text{ave}}(\mathcal{E}_i^{\mathcal{C}}, \mathcal{I}) \geq \mathcal{D}_{\text{ave}}(\mathcal{C}, \mathcal{I}) \geq \frac{3}{2} \times \frac{2^n}{2^n + 1} \mathcal{D}_{\text{ave}}(\mathcal{E}_i^{\mathcal{C}}, \mathcal{I}). \tag{3.83}$$

By noting that $\frac{3}{2} \geq \frac{3}{2} \times \frac{2^n}{2^n+1}$ and $\frac{3}{2} \times \frac{2^n}{2^n+1} \geq 1$, we obtain Lemma 3.5 and 3.6. □

3.2.4 Learning shallow quantum circuits from a classical dataset

In this section, we present algorithms for learning shallow quantum circuits that achieve a small diamond distance. All algorithms in this section use a classical dataset obtained from performing randomized measurements on the unknown shallow quantum circuit (defined below) to classically reconstruct the unknown circuit. The learning algorithms only require classical computation.

Definition 3.8 (Randomized measurement dataset for an unknown unitary)**.** *The learning algorithm accesses an unknown n-qubit unitary U via a randomized measurement dataset of the following form,*

$$\mathcal{T}_U(N) = \left\{ |\psi_\ell\rangle = \bigotimes_{i=1}^{n} |\psi_{\ell,i}\rangle \,, |\phi_\ell\rangle = \bigotimes_{i=1}^{n} |\phi_{\ell,i}\rangle \right\}_{\ell=1}^{N} . \tag{3.84}$$

A randomized measurement dataset of size N is constructed by obtaining N samples from the unknown unitary U. One sample is obtained from one experiment given as follows.

1. *Sample an input state $|\psi_\ell\rangle = \bigotimes_{i=1}^{n} |\psi_{\ell,i}\rangle$, which is a product state consisting of uniformly random single-qubit stabilizer states in* stab_1*.*

2. *Apply the unknown unitary U to $|\psi_\ell\rangle$.*

3. *Measure every qubit of $U|\psi_\ell\rangle$ under a random Pauli basis. The measurement collapses the state $U|\psi_\ell\rangle$ to a state $|\phi_\ell\rangle = \bigotimes_{i=1}^{n} |\phi_{\ell,i}\rangle$, where $|\phi_{\ell,i}\rangle$ is a single-qubit stabilizer state* stab_1*.*

Together, N queries to U construct a dataset $\mathcal{T}_U(N)$ with N samples. The dataset can be represented efficiently on a classical computer with $\mathcal{O}(Nn)$ bits.

An interesting question is whether quantum learning algorithms that have access to the unknown quantum circuit U could be much more efficient. In Section 3.2.5, we present a quantum learning algorithm that achieves the optimal scaling in query complexity and computational time for learning geometrically-local shallow quantum circuits over finite gate sets.

3.2.4.1 Results

We present the results for learning general and geometrically-local shallow quantum circuits consisting of two-qubit gates over SU(4) and over a finite gate set using a classical dataset.

Learning general shallow quantum circuits. We consider the problem of learning an n-qubit unitary U created by a general shallow quantum circuit C with arbitrary circuit connectivity, i.e., every qubit can be connected to any other qubit by a quantum gate, and an arbitrary number m of ancilla qubits initialized in $|0^m\rangle$ and ended up in $|0^m\rangle$ after C. Formally, we have the following identity for U,

$$U \otimes |0^m\rangle = C(I_n \otimes |0^m\rangle), \tag{3.85}$$

where I_n is an identity on n qubits.

We have the following theorems for learning the unknown unitary U. We can see that the sample/query complexity is very similar to learning geometric-local circuits. However, the computational complexity becomes higher, and we can only guarantee a polynomial scaling with system size n. The learning algorithm and proof are given in Section 3.2.4.5.

Theorem 3.6 (Learning general shallow quantum circuits)**.** *Given a failure probability δ, an approximation error ε, and an unknown n-qubit unitary U generated by a constant-depth circuit over any two-qubit gates in $\mathrm{SU}(4)$ with an arbitrary number of ancilla qubits. With a randomized measurement dataset $\mathcal{T}_U(N)$ of size*

$$N = \mathcal{O}\left(\frac{n^2 \log(n/\delta)}{\varepsilon^2}\right), \tag{3.86}$$

we can learn an n-qubit quantum channel $\hat{\mathcal{E}}$ that can be implemented by a constant-depth quantum circuit over $2n$ qubits, such that

$$\left\|\hat{\mathcal{E}} - \mathcal{U}\right\|_\diamond \leq \varepsilon, \tag{3.87}$$

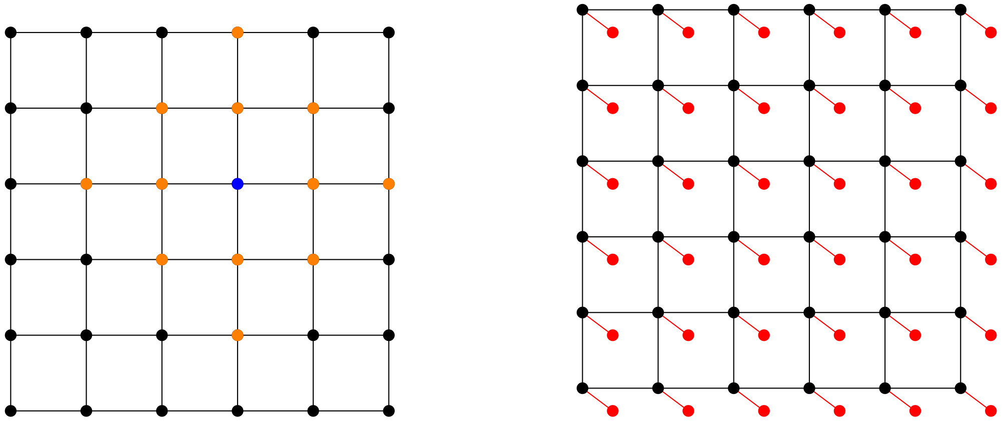
with probability at least $1-\delta$. The classical computational time to learn $\hat{\mathcal{E}}$ is $\mathrm{poly}(n)\log(1/\delta)/\varepsilon^2$.

In addition, if each two-qubit gate in the unknown circuit is chosen from a finite gate set of a constant size, then the algorithm learns an exact description $\hat{\mathcal{E}} = \mathcal{U}$ with probability $1-\delta$, using $N = \mathcal{O}(\log(n/\delta))$ samples and $\mathcal{O}(\mathrm{poly}(n)\log(1/\delta))$ time.

Remark 3.1 (Implementation of learned n-qubit channel)**.** *The n-qubit channel $\hat{\mathcal{E}}$ is the reduced channel $\mathcal{E}^{\hat{V}}_{\leq n}$ of the constant-depth $2n$-qubit circuit $\hat{V}$ on the first n qubits.*

Learning geometrically-local shallow quantum circuits. We consider the problem of learning geometrically-local shallow quantum circuits. Here, we consider a generalized definition of geometric locality, which includes quantum circuits over 1D, 2D, and 3D geometry. The generalization enables more exotic geometry over the qubits and is formally represented by a fixed constant-degree graph. See Fig. 3.1(a) for an illustration of the definitions.

Definition 3.9 (Geometric locality)**.** *A geometry over n qubits is defined by a graph $G = (V, E)$ with $n = |V|$ vertices, and each vertex has a degree of at most $\kappa = O(1)$. A*



(a) Original geometry

(b) Learned geometry

Figure 3.1: Learning geometrically-local shallow quantum circuits. (a) In this example, the geometry is a 2D lattice where each vertex has a degree at most 4. The lightcone of the blue qubit (for depth $d = 2$) is the union of the blue and orange qubits. (b) The learned circuit acts on an extended geometry with $2n$ qubits, where each system qubit (black) is attached to an ancilla qubit (red). Note that each ancilla qubit is connected only with its corresponding system qubit (red edges).

geometrically-local two-qubit gate can only act on an edge of G. A geometrically-local quantum circuit is a circuit with only geometrically-local two-qubit quantum gates. A depth-d geometrically-local quantum circuit has d layers, where each layer consists of non-overlapping geometrically-local two-qubit gates.

Definition 3.10 (Lightcone in a geometry)**.** *Given a geometry over n qubits represented by a graph $G = (V, E)$ with degree κ and an integer d. The lightcone $L_d(i)$ of a qubit i with depth d is the set of qubits with distance at most d from qubit i in the graph G. We have $|L_d(i)| \leq (\kappa + 1)^d$.*

Definition 3.11 (Geometrically-local set)**.** *Given a geometry over n qubits represented by a graph $G = (V, E)$. A set S of qubits is geometrically local if all qubits in S are of $\mathcal{O}(1)$ distance in G.*

Under this more general definition of geometry, our proposed algorithm can still learn very efficiently. The following theorem quantifies the efficiency in terms of both the query complexity and the computational complexity. The learning algorithm and proof are given in Section 3.2.5.2.

Theorem 3.7 (Learning geometrically-local shallow quantum circuits)**.** *Given an unknown geometrically local constant-depth n-qubit circuit U over any two-qubit gates in* SU(4)*. With* *a randomized measurement dataset $\mathcal{T}_U(N)$ of size*

$$N = \mathcal{O}\left(\frac{n^2 \log(n/\delta)}{\varepsilon^2}\right), \tag{3.88}$$

we can learn an n-qubit quantum channel $\hat{\mathcal{E}}$ that can be implemented by a geometrically local constant-depth quantum circuit over $2n$ qubits, such that

$$\left\| \hat{\mathcal{E}} - \mathcal{U} \right\|_\diamond \leq \varepsilon, \tag{3.89}$$

with probability at least $1 - \delta$. The computational time to learn $\hat{\mathcal{E}}$ is $\mathcal{O}(n^3 \log(n/\delta)/\varepsilon^2)$.

In addition, if each two-qubit gate in the unknown circuit is chosen from a finite gate set of a constant size, then the algorithm learns an exact description $\hat{\mathcal{E}} = \mathcal{U}$ with probability $1 - \delta$, using $N = \mathcal{O}(\log(n/\delta))$ samples and $\mathcal{O}(n \log(n/\delta))$ time.

Remark 3.2 (Implementation of learned n-qubit channel)**.** *The n-qubit channel $\hat{\mathcal{E}}$ is equal to the reduced channel $\mathcal{E}^{\hat{V}}_{\leq n}$ of the geometrically-local constant-depth $2n$-qubit circuit $\hat{V}$ on the first n qubits.*

Next, we look at a result, where we optimize the circuit depth in the learned circuit for implementing $\hat{\mathcal{E}}$. While the depth in the learned circuit can be controlled, the computational complexity becomes substantially worse. The learning algorithm and proof are given in Section 3.2.4.7.

Theorem 3.8 (Learning geometrically-local shallow circuits on k-dimensional lattice with optimized circuit depth)**.** *Given an unknown n-qubit circuit U over any two-qubit gates in* $\mathrm{SU}(4)$ *with circuit depth $d = \mathcal{O}(1)$ acting on a k-dimensional lattice with $k = \mathcal{O}(1)$. With a randomized measurement dataset $\mathcal{T}_U(N)$ of size*

$$N = 2^{\mathcal{O}((8kd)^k)} \frac{n^2 \log(n/\delta)}{\varepsilon^2}, \tag{3.90}$$

we can learn an n-qubit quantum channel $\hat{\mathcal{E}}$ that can be implemented by a quantum circuit over $2n$ qubits on an extended k-dimensional lattice (see Fig. 3.1(b)), such that

$$\left\| \hat{\mathcal{E}} - \mathcal{U} \right\|_\diamond \leq \varepsilon, \tag{3.91}$$

with probability at least $1 - \delta$.

- *With computational time $\mathcal{O}(n) \cdot N$, the learned circuit has depth at most*

$$(k+1)4^{4(8kd)^k} + 1. \tag{3.92}$$

- *With computational time $\mathcal{O}(n) \cdot N + (n/\varepsilon)^{\mathcal{O}((8kd)^{k+1})}$, the learned circuit has depth at most*
$$(k+1)(2d+1)+1. \tag{3.93}$$

In addition, if each two-qubit gate in the unknown circuit is chosen from a finite gate set of a constant size, then the algorithm learns an exact description $\hat{\mathcal{E}} = \mathcal{U}$ with probability $1-\delta$, using $N = \mathcal{O}(\log(n/\delta))$ samples, $\mathcal{O}(n\log(n/\delta))$ time, and a learned circuit of depth $(k+1)(2d+1)+1$.

Remark 3.3 (The geometry in the doubled system)**.** *In the two theorems given above, we mentioned geometrically-local circuits over $2n$ qubits, while the geometry is defined over n qubits. Given the geometry represented as a graph $G = (V, E)$ over n qubits with $V = \{1, \ldots, n\}$. We extend the graph to $2n$ qubits $G_{\text{ext}} = (V_{\text{ext}}, E_{\text{ext}})$ as follows.*

$$V_{\text{ext}} = \{1, \ldots, n, n+1, \ldots, 2n\}, \quad E_{\text{ext}} = E \cup \{(i, n+i) | 1 \leq i \leq n\}. \tag{3.94}$$

Each qubit $n+i$ in the added system is connected only to qubit i in the original system; See Fig. 3.1(b).

3.2.4.2 Techniques

We present two sets of closely related techniques for learning an n-qubit unitary U. The first set in Section 3.2.4.3 uses an idea called local inversion unitary, which follows from the concept of strong approximate local identity given in Section 3.2.3. As we have shown earlier, strong local identity checks can be performed by using Heisenberg-evolved single-qubit Pauli observables $U^\dagger P_i U$. The second set in Section 3.2.4.4 directly uses the Heisenberg-evolved Pauli observables $U^\dagger P_i U$.

3.2.4.3 Learning using local inversion

We begin by defining the concept of an approximate local inversion unitary.

Definition 3.12 (Strong ε-approximate local inversion)**.** *Given $n \in \mathbb{N}, \varepsilon \in (0,1)$, $i \in \{1, \ldots, n\}$, and n-qubit unitaries U and V_i. We say V_i is a strong ε-approximate local inversion of U on the i-th qubit if $\mathcal{U}\mathcal{V}_i$ is a strong ε-approximate local identity on the i-th qubit.*

Corollary 3.2 (Local inversion from Heisenberg-evolved Pauli observables)**.** *Given $n \in \mathbb{N}, \varepsilon \in (0,1)$, $i \in \{1, \ldots, n\}$, and n-qubit unitaries U and V_i. If V_i satisfies*

$$\sum_{P \in \{X,Y,Z\}} \left\| V_i^\dagger U^\dagger P_i U V_i - P_i \right\|_\infty \leq \varepsilon, \tag{3.95}$$

where P_i acts as $P \in \{X, Y, Z\}$ on the i-th qubit and as identity on the rest of the qubits, then V_i is a strong ε-approximate local inversion of U on the i-th qubit.

Proof. This corollary follows from Lemma 3.1, which characterizes the strong ε-approximate local identity with Heisenberg evolution of single-qubit Pauli observables. □

Instead of learning the unitary U alone, we consider learning the n local inversion unitaries $V_1, \ldots, V_n$. From the corollary given above, a straightforward way to learn V_i is to first learn the Heisenberg-evolved single-qubit Pauli observable $U^\dagger P_i U$ for all $P = X, Y, Z$, then try to find a unitary V_i that evolves $U^\dagger P_i U$ approximately back to P_i. This could be a much simpler task than learning the entire n-qubit unitary altogether.

While local inversion could potentially make the learning easier, it is *a priori* unclear if learning these local inversions is sufficient to learn U. In the following, we define a formalism for sewing these local inversion unitaries into a $2n$-qubit unitary (instead of n qubits).

Definition 3.13 (Sewing the local inversions)**.** *Given $n \in \mathbb{N}$ and n-qubit unitaries $V_1, \ldots, V_n$. We define the sewed $2n$-qubit unitary consisting of two sets of n qubits to be the following,*

$$U_{\text{sew}}(V_1, \ldots, V_n) := S \left[\prod_{i=1}^{n} \left(V_i^{(1)} \right) S_i \left(V_i^{(1)} \right)^\dagger \right], \tag{3.96}$$

where $V_i^{(1)}$ corresponds to applying the n-qubit unitary V_i on the first n qubits, S_i is the swap operator for the i-th qubit between the two sets of n qubits, S is the swap operator for all n qubits.

Remark 3.4 (Sewing order)**.** *The order for $\left(V_i^{(1)} \right) \mathrm{S}_i \left(V_i^{(1)} \right)^\dagger$ in sewing the local inversions does not matter. We can choose the order to optimize the resulting circuit, e.g., to minimize the circuit depth.*

Lemma 3.7 (Form of the sewed local inversions)**.** *Given $n \in \mathbb{N}$ and n-qubit unitaries $U, V_1, \ldots, V_n$. Assume V_i is a strong ε_i-approximate local inversion of U on the i-th qubit. Let $U_{\text{sew}} = U_{\text{sew}}(V_1, \ldots, V_n)$.*

$$\mathcal{D}_\diamond(\mathcal{U}_{\text{sew}}, \mathcal{U} \otimes \mathcal{U}^\dagger) = \frac{1}{2} \left\| \mathcal{U}_{\text{sew}} - \mathcal{U} \otimes \mathcal{U}^\dagger \right\|_\diamond \leq \sum_{i=1}^{n} \varepsilon_i, \tag{3.97}$$

where the first/second set of n qubits is on the left/right of the tensor product.

Proof. From Theorem 3.55 in [210], we have

$$\left\| \mathcal{U}_{\text{sew}} - \mathcal{U} \otimes \mathcal{U}^\dagger \right\|_\diamond = \left\| (U^\dagger \otimes U) U_{\text{sew}} \left|\psi\right\rangle\!\left\langle\psi\right| U_{\text{sew}}^\dagger (U \otimes U^\dagger) - \left|\psi\right\rangle\!\left\langle\psi\right| \right\|_1 \tag{3.98}$$

for some $2n$-qubit state $|\psi\rangle$. We define the following mathematical object,

$$\left|\psi_i\right\rangle\!\left\langle\psi_i\right| := \left[(\mathcal{U}^\dagger \otimes \mathcal{I}) \left(\mathcal{S}_1 \ldots \mathcal{S}_i \right) (\mathcal{I} \otimes \mathcal{U}) \mathcal{S} \left(\left(\mathcal{V}_{i+1}^{(1)} \right) \mathcal{S}_{i+1} \left(\mathcal{V}_{i+1}^{(1)} \right)^\dagger \ldots \left(\mathcal{V}_n^{(1)} \right) \mathcal{S}_n \left(\mathcal{V}_n^{(1)} \right)^\dagger \right) \right] \left(\left|\psi\right\rangle\!\left\langle\psi\right| \right) \tag{3.99}$$

for each $i = 0, \ldots, n$. Note that we have the following identities,

$$|\psi_0\rangle\!\langle\psi_0| = (U^\dagger \otimes U)U_{\text{sew}} \, |\psi\rangle\!\langle\psi| \, U_{\text{sew}}^\dagger (U \otimes U^\dagger), \tag{3.100}$$

$$|\psi_n\rangle\!\langle\psi_n| = \left[(\mathcal{U}^\dagger \otimes \mathcal{I})\mathcal{S}(\mathcal{I} \otimes \mathcal{U})\mathcal{S}\right](|\psi\rangle\!\langle\psi|) = |\psi\rangle\!\langle\psi| \, . \tag{3.101}$$

By the triangle inequality, we can obtain the following telescoping sum,

$$\left\|\mathcal{U}_{\text{sew}} - \mathcal{U} \otimes \mathcal{U}^\dagger\right\|_\diamond = \left\| |\psi_0\rangle\!\langle\psi_0| - |\psi_n\rangle\!\langle\psi_n| \right\|_1 \leq \sum_{i=1}^{n} \left\| |\psi_i\rangle\!\langle\psi_i| - |\psi_{i-1}\rangle\!\langle\psi_{i-1}| \right\|_1 . \tag{3.102}$$

Each summand can be bounded as follows,

$$\begin{aligned}
\left\| |\psi_i\rangle\!\langle\psi_i| - |\psi_{i-1}\rangle\!\langle\psi_{i-1}| \right\|_1 &\leq \left\| \mathcal{S}_i(\mathcal{I} \otimes \mathcal{U})\mathcal{S} - (\mathcal{I} \otimes \mathcal{U})\mathcal{S}\,(\mathcal{V}_i \otimes \mathcal{I})\,\mathcal{S}_i\,(\mathcal{V}_i \otimes \mathcal{I})^\dagger \right\|_\diamond & (3.103)\\
&= \left\| \mathcal{S}\mathcal{S}_i(\mathcal{U} \otimes \mathcal{I}) - \mathcal{S}(\mathcal{U} \otimes \mathcal{I})\,(\mathcal{V}_i \otimes \mathcal{I})\,\mathcal{S}_i\,(\mathcal{V}_i \otimes \mathcal{I})^\dagger \right\|_\diamond & (3.104)\\
&\leq \left\| \mathcal{S}_i(\mathcal{U} \otimes \mathcal{I}) - \left(\left(\mathcal{I}_i \otimes \mathcal{E}^{\mathcal{U}\mathcal{V}_i}_{\neq i}\right) \otimes \mathcal{I}\right)\mathcal{S}_i\,(\mathcal{V}_i \otimes \mathcal{I})^\dagger \right\|_\diamond + \varepsilon_i & (3.105)\\
&= \left\| \mathcal{S}_i(\mathcal{U} \otimes \mathcal{I}) - \mathcal{S}_i\left(\left(\mathcal{I}_i \otimes \mathcal{E}^{\mathcal{U}\mathcal{V}_i}_{\neq i}\right) \otimes \mathcal{I}\right)(\mathcal{V}_i \otimes \mathcal{I})^\dagger \right\|_\diamond + \varepsilon_i & (3.106)\\
&= \left\| (\mathcal{U}\mathcal{V}_i \otimes \mathcal{I}) - \left(\left(\mathcal{I}_i \otimes \mathcal{E}^{\mathcal{U}\mathcal{V}_i}_{\neq i}\right) \otimes \mathcal{I}\right) \right\|_\diamond + \varepsilon_i \leq 2\varepsilon_i. & (3.107)
\end{aligned}$$

Together, we obtain the desired statement. □

Remark 3.5 (A basic identity for $U \otimes U^\dagger$)**.** *A trivial example of an exact local inversion of U on the i-th qubit is $V_i = U^\dagger$. In this case, Lemma 3.7 yields the following basic identity,*

$$U \otimes U^\dagger = S\left[\prod_{i=1}^{n} \left(U^\dagger \otimes I\right) S_i \left(U \otimes I\right)\right], \tag{3.108}$$

which can also be shown by canceling all the intermediate $(U \otimes I)\left(U^\dagger \otimes I\right)$.

3.2.4.4 Learning using Heisenberg-evolved Pauli observables

We have seen earlier that one direct approach to learning local inversion is to first learn the Heisenberg-evolved single-qubit Pauli observables $U^\dagger P_i U$. In the following, we define an alternative formalism that directly sews the Heisenberg-evolved Pauli observables into a $2n$-qubit unitary (instead of n qubits) that approximates $U \otimes U^\dagger$. One can flexibly choose either approach. Typically, learning the Heisenberg-evolved Pauli observables is computationally simpler, but yields higher depth in the learned circuit.

Definition 3.14 (Approximate Heisenberg-evolved Paui observables)**.** *Given $n \in \mathbb{N}, \varepsilon \in (0, 1)$, $i \in \{1, \ldots, n\}$, $P \in \{X, Y, Z\}$, an n-qubit unitary U, and an n-qubit observable $O_{i,P}$. We say $O_{i,P}$ is an ε-approximate Heisenberg-evolved Pauli observable P on qubit i under U if $\left\|O_{i,P} - U^\dagger P_i U\right\|_\infty \leq \varepsilon$.*

Given a set of $3n$ Heisenberg-evolved Pauli observables, we use the following definition to sew them into a $2n$-qubit unitary.

Definition 3.15 (Sewing the Heisenberg-evolved observables)**.** *Given $n \in \mathbb{N}$ and $3 \times n$ n-qubit observables $O_{i,P}, \forall i = 1, \ldots, n, P \in \{X, Y, Z\}$. Let $\mathrm{Proj}_{\mathrm{U}}(A)$ be the projection of a matrix A to a unitary matrix minimizing the operator norm $\|\cdot\|_\infty$, i.e.,*

$$\mathrm{Proj}_{\mathrm{U}}(A) := \operatorname*{arg\,min}_{B:\textit{unitary}} \|A - B\|_\infty. \tag{3.109}$$

We define the sewed $2n$-qubit unitary consisting of two sets of n qubits to be the following,

$$U_{\mathrm{sew}}(\{O_{i,P}\}_{i,P}) := S \prod_{i=1}^{n} \left[\mathrm{Proj}_{\mathrm{U}} \left(\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} O_{i,P} \otimes P_i \right) \right], \tag{3.110}$$

where $V_i^{(1)}$ corresponds to applying the n-qubit unitary V_i on the first n qubits, S_i is the swap operator for the i-th qubit between the two sets of n qubits, S is the swap operator for all n qubits.

Remark 3.6 (Sewing order)**.** *The order for sewing $\mathrm{Proj}_{\mathrm{U}}\left(\frac{1}{2} I \otimes I + \frac{1}{2} \sum_P O_{i,P} \otimes P_i\right)$ is arbitrary.*

In the above, we have utilized the projection function $\mathrm{Proj}_{\mathrm{U}}$. In the following lemma, we show that this function can be computed efficiently on a classical computer.

Lemma 3.8 (Projection onto unitary matrices)**.** *Consider the singular value decomposition $A = U \Sigma V^\dagger$, where Σ is diagonal, nonnegative, and U, V is unitary. The projection can be defined as*

$$\mathrm{Proj}_{\mathrm{U}}(A) = U V^\dagger. \tag{3.111}$$

The computational time is polynomial in the dimension of A.

Proof. Consider any unitary B. We have $\|A - B\|_\infty = \left\|\Sigma - U^\dagger B V\right\|_\infty$. Let W be the unitary $U^\dagger B V$. We can use the definition of $\|M\|_\infty = \sup_v \|Mv\|_2 / \|v\|_2$ to see that

$$\begin{aligned} \|\Sigma - W\|_\infty &\geq \max_i \|\Sigma_{ii} \hat{e}_i - W \hat{e}_i\|_2 \\ &\geq \max_i \sqrt{1 + \Sigma_{ii}^2 - 2\Sigma_{ii} \mathrm{Re}[\hat{e}_i^T W \hat{e}_i]} \geq \max_i |1 - \Sigma_{ii}| = \|\Sigma - I\|_\infty, \end{aligned} \tag{3.112}$$

where $\hat{e}_i$ is the unit vector with a nonzero entry on the i-th coordinate. Because $\|\Sigma - I\|_\infty = \left\|A - UV^\dagger\right\|_\infty$, we have obtained $\|A - B\|_\infty \geq \left\|A - UV^\dagger\right\|_\infty$. □

Similar to sewing local inversions, the sewed unitary accurately approximates $U \otimes U^\dagger$.

Lemma 3.9 (Form of the sewed Heisenberg-evolved observables)**.** *Given $n \in \mathbb{N}$, an n-qubit unitary U, and $3 \times n$ n-qubit observables $O_{i,P}, \forall i = 1, \ldots, n, P \in \{X, Y, Z\}$. Assume $O_{i,P}$ is an $\varepsilon_{i,P}$-approximate Heisenberg-evolved Pauli observable P on qubit i under U. Let $U_{\mathrm{sew}} = U_{\mathrm{sew}}(\{O_{i,P}\}_{i,P})$. Then*

$$\mathcal{D}_\diamond(\mathcal{U}_{\mathrm{sew}}, \mathcal{U} \otimes \mathcal{U}^\dagger) = \frac{1}{2}\left\|\mathcal{U}_{\mathrm{sew}} - \mathcal{U} \otimes \mathcal{U}^\dagger\right\|_\diamond \leq \sum_{i=1}^{n} \sum_{P \in \{X,Y,Z\}} \varepsilon_{i,P}, \tag{3.113}$$

where the first/second set of n qubits is on the left/right of the tensor product.

Proof. From Eq. (3.108), we have the following identity,

$$U \otimes U^\dagger = S\left[\prod_{i=1}^{n} (U^\dagger \otimes I) S_i (U \otimes I)\right]. \tag{3.114}$$

Using the fact that $S_i = \frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} P_i \otimes P_i$, we can rewrite the above identity as

$$U \otimes U^\dagger = S \prod_{i=1}^{n} \left[\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} (U^\dagger P_i U) \otimes P_i\right]. \tag{3.115}$$

Let us denote the following unitaries,

$$V_i := \frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} (U^\dagger P_i U) \otimes P_i, \tag{3.116}$$

$$\widetilde{W}_i := \frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} O_{i,P} \otimes P_i \tag{3.117}$$

$$W_i := \mathrm{Proj}_{\mathrm{U}}\left(\widetilde{W}_i\right). \tag{3.118}$$

We can upper bound the diamond distance as follows,

$$\left\|\mathcal{U}_{\mathrm{sew}} - \mathcal{U} \otimes \mathcal{U}^\dagger\right\|_\diamond = \|\mathcal{V}_n \ldots \mathcal{V}_1 - \mathcal{W}_n \ldots \mathcal{W}_1\|_\diamond \tag{3.119}$$

$$\leq \sum_{i=1}^{n} \|\mathcal{V}_n \ldots \mathcal{V}_{i+1} \mathcal{W}_i \ldots \mathcal{W}_1 - \mathcal{V}_n \ldots \mathcal{V}_i \mathcal{W}_{i-1} \ldots \mathcal{W}_1\|_\diamond \tag{3.120}$$

$$\leq \sum_{i=1}^{n} \|\mathcal{V}_n \ldots \mathcal{V}_{i+1} \mathcal{W}_i \ldots \mathcal{W}_1 - \mathcal{V}_n \ldots \mathcal{V}_i \mathcal{W}_{i-1} \ldots \mathcal{W}_1\|_\diamond \tag{3.121}$$

$$= \sum_{i=1}^{n} \|\mathcal{W}_i - \mathcal{V}_i\|_\diamond \leq 2 \sum_{i=1}^{n} \|W_i - V_i\|_\infty. \tag{3.122}$$

The last inequality uses the fact that $\mathcal{W}_i$ and $\mathcal{V}_i$ are unitary channels. From triangle inequality and the definition of $\mathrm{Proj}_{\mathrm{U}}(\cdot)$, we have the following inequality,

$$\begin{aligned} \|W_i - V_i\|_\infty &\leq \left\|W_i - \widetilde{W}_i\right\|_\infty + \left\|\widetilde{W}_i - V_i\right\|_\infty \\ &= \min_{V:\text{unitary}} \left\|\widetilde{W}_i - V\right\|_\infty + \left\|\widetilde{W}_i - V_i\right\|_\infty \\ &\leq 2\left\|\widetilde{W}_i - V_i\right\|_\infty . \end{aligned} \tag{3.123}$$

We now use the specific form of $\widetilde{W}_i, V_i$ to upper bound the summand,

$$\|W_i - V_i\|_\infty \leq \sum_{P \in \{X,Y,Z\}} \left\|O_{i,P} - U^\dagger P_i U\right\|_\infty \leq \sum_P \varepsilon_{i,P}. \tag{3.124}$$

Together with Eq. (3.122), we can obtain the desired statement. □

Given an n-qubit observable O, we define $\mathrm{supp}(O)$ to be the set of qubits that the observable O acts on. We also define $|O|$ to be the size of $\mathrm{supp}(O)$. We have the following lemma for learning a few-body observable. The learned observable $\hat{O}$ has the property that it only acts on qubits that O acts on, hence $\mathrm{supp}(\hat{O}) \subseteq \mathrm{supp}(O)$.

Lemma 3.10 (Learning a few-body observable with an unknown support)**.** *Given an error ε, failure probability δ, an unknown n-qubit observable O with $\|O\|_\infty \leq 1$ that acts on an unknown set of k qubits, and a dataset $\mathcal{T}_O(N) = \{|\psi_\ell\rangle = \bigotimes_{i=1}^n |\psi_{\ell,i}\rangle, v_\ell\}_{\ell=1}^N$, where $|\psi_{\ell,i}\rangle$ is sampled uniformly from* stab_1 *and v_ℓ is a random variable with $\mathbb{E}[v_\ell] = \langle\psi_\ell| O |\psi_\ell\rangle$, $|v_\ell| = \mathcal{O}(1)$. Given a dataset size of*

$$N = \frac{2^{\mathcal{O}(k)} \log(n/\delta)}{\varepsilon^2}, \tag{3.125}$$

with probability at least $1-\delta$, we can learn an observable $\hat{O}$ such that $\left\|\hat{O} - O\right\|_\infty \leq \varepsilon$ and $\mathrm{supp}(\hat{O}) \subseteq \mathrm{supp}(O)$*. The computational complexity is $\mathcal{O}(n^k \log(n/\delta)/\varepsilon^2)$.*

Proof. Consider the observable O under the Pauli basis, $O = \sum_P \alpha_P P$. The α_P coefficients satisfy

$$\alpha_P = 3^{|P|} \mathop{\mathbb{E}}_{|\psi\rangle \sim \mathrm{stab}_1^{\otimes n}} \langle\psi|O|\psi\rangle \, \langle\psi|P|\psi\rangle \,, \tag{3.126}$$

which can be learned by replacing the expectation with averaging over the dataset.

We begin by defining the learned observable $\hat{O}$.

$$\hat{\alpha}_P := \frac{3^{|P|}}{N} \sum_{\ell=1}^{N} v_\ell \langle \psi_\ell | P | \psi_\ell \rangle, \qquad \forall P \in \{I, X, Y, Z\}^{\otimes n} : |P| \leq k, \tag{3.127}$$

$$\hat{\beta}_P := \begin{cases} \hat{\alpha}_P, & |\hat{\alpha}_P| \geq 0.5\varepsilon/(2\sqrt{2})^k, \\ 0, & |\hat{\alpha}_P| < 0.5\varepsilon/(2\sqrt{2})^k, \end{cases} \tag{3.128}$$

$$\hat{O} := \sum_{P \in \{I,X,Y,Z\}^{\otimes n} : |P| \leq k} \hat{\beta}_P P. \tag{3.129}$$

Because O acts on at most k qubits, $\alpha_P = 0$ for $|P| > k$. From Bernstein's inequality, given a dataset size of

$$N = \frac{2^{\mathcal{O}(k)} \log(n/\delta)}{\varepsilon^2}, \tag{3.130}$$

with probability at least $1 - \delta$, we have

$$|\alpha_P - \hat{\alpha}_P| < 0.5\varepsilon/(2\sqrt{2})^k, \qquad \forall P \in \{I, X, Y, Z\}^{\otimes n} : |P| \leq k. \tag{3.131}$$

In the following, we assume the above event holds, which happens with probability at least $1 - \delta$. We separately prove the following two statements.

$\mathrm{supp}(\hat{O}) \subseteq \mathrm{supp}(O)$: For a Pauli observable P with $\alpha_P = 0$, we have $|\hat{\alpha}_P| < 0.5\varepsilon/(2\sqrt{2})^k$ from Eq. (3.131). Hence, $\hat{\beta}_P = 0$. As a result, the set of qubits acted by $\hat{O}$ is a subset of $\mathrm{supp}(O)$.

$\left\| \hat{O} - O \right\|_\infty \leq \varepsilon$: From the fact that $\alpha_P = 0$ implies $\hat{\beta}_P = 0$, we have

$$\begin{aligned} \hat{O} - O &= \sum_{P \in \{I,X,YZ\}^{\otimes n} : \mathrm{supp}(P) \subseteq \mathrm{supp}(O)} \left(\hat{\beta}_P - \alpha_P \right) P \qquad (3.132) \\ &= \sum_{Q \in \{I,X,YZ\}^{\otimes k}} \left(\hat{\beta}_{P(Q)} - \alpha_{P(Q)} \right) P(Q), \qquad (3.133) \end{aligned}$$

where $P(Q) := Q \otimes I_{\{1,\ldots,n\} \setminus \mathrm{supp}(O)}$ and $k = |\mathrm{supp}(O)|$. Therefore, we can upper bound the spectral norm by

$$\left\| \hat{O} - O \right\|_\infty \leq \left\| \sum_{Q \in \{I,X,YZ\}^{\otimes k}} \left(\hat{\beta}_{P(Q)} - \alpha_{P(Q)} \right) P(Q) \right\|_\infty = \left\| \sum_{Q \in \{I,X,YZ\}^{\otimes k}} \left(\hat{\beta}_{P(Q)} - \alpha_{P(Q)} \right) Q \right\|_\infty . \tag{3.134}$$

Recall that $\|A\|_\infty \leq \sqrt{\mathrm{Tr}(A^2)}$ for any Hermitian matrix A, we have

$$\left\|\hat{O} - O\right\|_\infty \leq \sqrt{\sum_{Q \in \{I,X,YZ\}^{\otimes k}} \left(\hat{\beta}_{P(Q)} - \alpha_{P(Q)}\right)^2 \mathrm{Tr}(Q^2)} \leq (2\sqrt{2})^k \max_{|P|\leq k} \left|\hat{\beta}_P - \alpha_P\right|. \tag{3.135}$$

By the triangle inequality and Eq. (3.131), we have

$$\left|\hat{\beta}_P - \alpha_P\right| \leq \left|\hat{\beta}_P - \hat{\alpha}_P\right| + |\hat{\alpha}_P - \alpha_P| < \varepsilon/(2\sqrt{2})^k, \qquad \forall |P| \leq k. \tag{3.136}$$

Therefore, we have obtained the desired inequality $\left\|\hat{O} - O\right\|_\infty \leq \varepsilon$. □

Lemma 3.11 (Learning a few-body observable with a known support)**.** *Given an error ε, failure probability δ, an unknown n-qubit observable O with $\|O\|_\infty \leq 1$ that acts on an known set S of k qubits, and a dataset $\mathcal{T}_O(N) = \{|\psi_\ell\rangle = \bigotimes_{i=1}^n |\psi_{\ell,i}\rangle, v_\ell\}_{\ell=1}^N$, where $|\psi_{\ell,i}\rangle$ is sampled uniformly from* stab$_1$ *and v_ℓ is a random variable with $\mathbb{E}[v_\ell] = \langle\psi_\ell| O |\psi_\ell\rangle$, $|v_\ell| = \mathcal{O}(1)$. Given a dataset size of*

$$N = \frac{2^{\mathcal{O}(k)} \log(1/\delta)}{\varepsilon^2}, \tag{3.137}$$

with probability at least $1-\delta$, we can learn an observable $\hat{O}$ such that $\left\|\hat{O} - O\right\|_\infty \leq \varepsilon$ and $\mathrm{supp}(\hat{O}) \subseteq S$. *The computational complexity is $\mathcal{O}(2^{\mathcal{O}(k)} \log(1/\delta)/\varepsilon^2)$.*

Proof. We begin by defining the learned observable $\hat{O}$.

$$\hat{\alpha}_P := \frac{3^{|P|}}{N} \sum_{\ell=1}^N v_\ell \langle\psi_\ell| P |\psi_\ell\rangle, \qquad \forall P \in \{I,X,Y,Z\}^{\otimes n} : \mathrm{supp}(P) \subseteq S, \tag{3.138}$$

$$\hat{O} := \sum_{P \in \{I,X,Y,Z\}^{\otimes n} : \, \mathrm{supp}(P) \subseteq S} \hat{\alpha}_P P. \tag{3.139}$$

By definition, we can see that $\mathrm{supp}(\hat{O}) \subseteq S$. Consider the observable O under the Pauli basis, $O = \sum_P \alpha_P P$. Because O acts on the qubits in the set S, $\alpha_P = 0$ for $\mathrm{supp}(P) \not\subseteq S$. From Bernstein's inequality, given a dataset of size

$$N = \frac{2^{\mathcal{O}(k)} \log(1/\delta)}{\varepsilon^2}, \tag{3.140}$$

with probability at least $1-\delta$, we have

$$|\alpha_P - \hat{\alpha}_P| < \varepsilon/(2\sqrt{2})^k, \qquad \forall P \in \{I,X,Y,Z\}^{\otimes n} : \mathrm{supp}(P) \subseteq S. \tag{3.141}$$

In the following, we assume the above event holds, which happens with probability at least $1-\delta$. Using the same derivation as in Eq. (3.132) to Eq. (3.135) for the proof of Lemma 3.10, we have

$$\left\|\hat{O} - O\right\|_\infty \leq (2\sqrt{2})^k \max_{P : \mathrm{supp}(P) \subseteq S} |\hat{\alpha}_P - \alpha_P| < \varepsilon, \tag{3.142}$$

hence we have arrived at the desired statement. □

Remark 3.7 (Relation to learning quantum juntas)**.** *The two lemmas given above are related to quantum junta learning [211] but consider a much weaker access model. [211] requires that the unknown observable O be a unitary, and the learning algorithm can access the unitary coherently. In particular, [211] requires inputting half of the maximally entangled state to the unitary. Here, we consider access to O through a simple classical dataset consisting of random product input states and the outcome when measuring the input states with observable O. When the lemmas are used as a subroutine in learning algorithms given in Section 3.2.4, we do not have access to O as a unitary, so [211] cannot be used.*

3.2.4.5 Learning general shallow circuits (Proof of Theorem 3.6)

We present the algorithm for learning an unknown n-qubit unitary U generated by an arbitrary constant-depth quantum circuit C with arbitrarily many ancilla qubits. We separate the proof into two-qubit gates over SU(4) and over a finite gate set.

Arbitrary SU(4) **gates.** The algorithm utilizes a randomized measurement dataset $\mathcal{T}_U(N)$. The key ideas are using Lemma 3.10 to learn approximate Heisenberg-evolved Pauli observables, using Lemma 3.13 to sew the Heisenberg-evolved Pauli observables into a constant-depth quantum circuit, and using Lemma 3.9 to obtain the rigorous performance guarantee.

The following lemma shows how to reuse the randomized measurement dataset $\mathcal{T}_U(N)$ to create the datasets needed to learn approximate Heisenberg-evolved Pauli observables using Lemma 3.10.

Lemma 3.12 (Reusing the randomized measurement dataset)**.** *Given an unknown n-qubit unitary U, and a randomized measurement dataset $\mathcal{T}_U(N)$ given in Eq. (3.84). We can create $3n$ datasets $\mathcal{T}_{U^\dagger P_i U}(N)$, for each Pauli observable $P \in \{X, Y, Z\}$ and each qubit i,*

$$\mathcal{T}_{U^\dagger P_i U}(N) := \left\{ |\psi_\ell\rangle = \bigotimes_{j=1}^{n} |\psi_{\ell,j}\rangle, v_\ell^{U^\dagger P_i U} \right\}_{\ell=1}^{N}, \tag{3.143}$$

where $|\psi_{\ell,i}\rangle$ is sampled uniformly and independently from stab_1 *and $v_\ell^{U^\dagger P_i U}$ is a random variable with $\mathbb{E}[v_\ell^{U^\dagger P_i U}] = \langle\psi_\ell| U^\dagger P_i U |\psi_\ell\rangle$ and $|v_\ell^{U^\dagger P_i U}| = \mathcal{O}(1)$.*

Proof. Recall that from Eq. (3.84), we have

$$\mathcal{T}_U(N) = \left\{ |\psi_\ell\rangle = \bigotimes_{i=1}^{n} |\psi_{\ell,i}\rangle, |\phi_\ell\rangle = \bigotimes_{i=1}^{n} |\phi_{\ell,i}\rangle \right\}_{\ell=1}^{N}. \tag{3.144}$$

The input states are reused over the $3n$ datasets. For each Pauli observable $P \in \{X, Y, Z\}$ and each qubit i, we define the output value to be

$$v_\ell^{U^\dagger P_i U} := 3 \langle\phi_{\ell,i}| P |\phi_{\ell,i}\rangle. \tag{3.145}$$

We have $|v_\ell^{U^\dagger P_i U}| = |3 \langle \phi_{\ell,i}| P |\phi_{\ell,i}\rangle| \leq 3 = \mathcal{O}(1)$. Now, recall how $|\phi_{\ell,i}\rangle$ is defined. $|\phi_{\ell,i}\rangle$ is the measurement outcome when we measure the i-th qubit of the n-qubit state $U|\psi_\ell\rangle$ in a random Pauli basis: X basis gives $|X,0\rangle := |+\rangle, |X,1\rangle := |-\rangle$; Y basis gives $|Y,0\rangle := |y+\rangle, |Y,1\rangle := |y-\rangle$; Z basis gives $|Z,0\rangle := |0\rangle, |Z,1\rangle := |1\rangle$. Using the fact that

$$0 = \langle Q,b| P |Q,b\rangle, \qquad \forall P \neq Q \in \{X,Y,Z\}, b \in \{0,1\}, \tag{3.146}$$

$$P = \sum_{b\in\{0,1\}} (-1)^b |P,b\rangle\langle P,b|, \qquad \forall P \in \{X,Y,Z\}. \tag{3.147}$$

and that the randomized measurement measures X, Y, Z bases equally likely, we have

$$\mathbb{E}\left[3 \langle \phi_{\ell,i}| P |\phi_{\ell,i}\rangle\right] = \langle \psi_\ell| U^\dagger P_i U |\psi_\ell\rangle. \tag{3.148}$$

This concludes the proof. □

From Lemma 3.14 and the fact that $\mathrm{supp}\left(U^\dagger P_i U\right) \subseteq A(i) = \bigcup_{P\in\{X,Y,Z\}} \mathrm{supp}\left(U^\dagger P_i U\right)$, we have

$$\left|\mathrm{supp}\left(U^\dagger P_i U\right)\right| \leq |A(i)| = \mathcal{O}(1). \tag{3.149}$$

This enables us to combine Lemma 3.12 for constructing $\mathcal{T}_{U^\dagger P_i U}(N), \forall i, P$ from $\mathcal{T}_U(N)$ and Lemma 3.10 for learning few-body observables with unknown supports (since $A(i)$ is unknown) to show the following. For any constant value $\tilde{\varepsilon} = \mathcal{O}(1)$, given a dataset size of

$$N = \mathcal{O}\left(\frac{n^2 \log(n/\delta)}{\varepsilon^2}\right), \tag{3.150}$$

we can learn $\hat{O}_{i,P}, \forall i, P$, such that with probability at least $1-\delta$, for all $i \in \{1,\ldots,n\}$ and Pauli observable $P \in \{X,Y,Z\}$, we have

$$\left\|\hat{O}_{i,P} - U^\dagger P_i U\right\|_\infty \leq \frac{\varepsilon}{6n}, \quad \text{and} \quad \mathrm{supp}(\hat{O}_{i,P}) \subseteq \mathrm{supp}(U^\dagger P_i U) \subseteq A(i). \tag{3.151}$$

The computational time for learning all $\hat{O}_{i,P}$ is $\mathcal{O}(n^{\mathcal{O}(1)} \log(n/\delta)/\varepsilon^2) = \mathrm{poly}(n)\log(1/\delta/\varepsilon^2)$. From Lemma 3.14, we can characterize $\mathrm{supp}(\hat{O}_{i,P}) \subseteq \mathrm{supp}(U^\dagger P_i U)$ to apply Lemma 3.13.

Lemma 3.13 (Sewing into a constant-depth quantum circuit)**.** *Given $3n$ n-qubit observables $\hat{O}_{i,P}, \forall i \in \{1,\ldots,n\}, P \in \{X,Y,Z\}$, such that for any qubit i, $\left|\bigcup_P \mathrm{supp}\left(\hat{O}_{i,P}\right)\right| = \mathcal{O}(1)$ and there is only a constant number of qubit j with*

$$\left(\bigcup_P \mathrm{supp}\left(\hat{O}_{i,P}\right)\right) \cap \left(\bigcup_P \mathrm{supp}\left(\hat{O}_{j,P}\right)\right) \neq \varnothing. \tag{3.152}$$

There exists a sewing ordering for $U_{\mathrm{sew}}(\{\hat{O}_{i,P}\}_{i,P})$ given in Definition 3.15, such that $U_{\mathrm{sew}}(\{\hat{O}_{i,P}\}_{i,P})$ can be implemented by a constant-depth quantum circuit. The constant-depth quantum circuit is geometrically-local (see Definition 3.9) if $\bigcup_P \mathrm{supp}\left(\hat{O}_{i,P}\right), \forall i$ are geometrically-local sets (see Definition 3.11). The computational time for finding the circuit implementation is $\mathcal{O}(n)$.

Proof. For simplicity of notations, we define $A(i) := \bigcup_P \operatorname{supp}\left(\hat{O}_{i,P}\right)$. We can see that

$$\operatorname{supp}\left(\operatorname{Proj}_{\mathrm{U}}\left(\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \hat{O}_{i,P} \otimes P_i\right)\right) \subseteq A(i) \cup \{n+i\}, \tag{3.153}$$

Because $|A(i)| = \left|\bigcup_P \operatorname{supp}\left(\hat{O}_{i,P}\right)\right| = \mathcal{O}(1)$ and $\operatorname{Proj}_{\mathrm{U}}$ can be implemented in time polynomial in $2^{|A(i) \cup \{n+i\}|} = \mathcal{O}(1)$ as shown in Lemma 3.8, the following unitary

$$\operatorname{Proj}_{\mathrm{U}}\left(\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \hat{O}_{i,P} \otimes P_i\right) \tag{3.154}$$

can be implemented by a constant-depth circuit acting only on qubits in $A(i) \cup \{n+i\}$; see Fact 3.4 for exact unitary synthesis. Furthermore, if $A(i) = \bigcup_P \operatorname{supp}\left(\hat{O}_{i,P}\right)$ is a geometrically-local set, the constant-depth circuit is geometrically-local; see Corollary 3.1 for exact unitary synthesis given a connectivity graph. The geometric locality for the $2n$-qubit system is defined in Remark 3.3.

Consider an n-node graph (equivalently, an n-qubit graph), where each pair (i,j) of nodes (qubits) is connected by an edge if

$$A(i) \cap A(j) \neq \varnothing. \tag{3.155}$$

The graph only has $\mathcal{O}(n)$ edges and can be constructed as an adjacency list in time $\mathcal{O}(n)$. Because the graph has a constant degree, we can use a $\mathcal{O}(n)$-time greedy graph coloring algorithm to color the n-qubit graph using only a constant number $\chi = \mathcal{O}(1)$ of colors. For each node/qubit i, we consider $c(i)$ to be the color labeled from 1 to χ. The sewing order for the $3n$ observables $\hat{O}_{i,P}$ in Definition 3.15 are given by the greedy graph coloring, where we order from the smallest color to the largest color. By the definition of graph coloring, for any pair i, j of qubits with the same color, we have

$$A(i) \cap A(j) = \varnothing. \tag{3.156}$$

Therefore, for any color c', we can find an implementation of the $2n$-qubit unitary

$$\prod_{i : c(i) = c'} \left[\operatorname{Proj}_{\mathrm{U}}\left(\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} O_{i,P} \otimes P_i\right)\right] \tag{3.157}$$

with a constant-depth (and geometrically-local if $A(i), \forall i$ are geometrically-local) quantum circuit in time $\mathcal{O}(n)$. Since there is only a constant number of colors, the $2n$-qubit unitary $U_{\text{sew}}(\{\hat{O}_{i,P}\}_{i,P})$ in Eq. (3.110) with the color-based ordering can be implemented with a constant-depth (and geometrically-local if $A(i), \forall i$ are geometrically-local) quantum circuit in time $\mathcal{O}(n)$. □

Lemma 3.13 shows that there exists an ordering for sewing the approximate Heisenberg-evolved Pauli observables $\hat{O}_{i,P}$ to create $U_{\text{sew}}(\{\hat{O}_{i,P}\}_{i,P})$ given in Definition 3.15, such that $U_{\text{sew}}(\{\hat{O}_{i,P}\}_{i,P})$ can be implemented by a constant-depth quantum circuit. Given Eq. (3.151), we can use Lemma 3.9 on the form of the sewed Heisenberg-evolved Pauli observables to yield

$$\left\|\mathcal{U}_{\text{sew}}(\{\hat{O}_{i,P}\}_{i,P}) - \mathcal{U} \otimes \mathcal{U}^{\dagger}\right\|_{\diamond} \leq \varepsilon. \tag{3.158}$$

Finally, define an n-qubit channel $\hat{\mathcal{E}}$ as follows,

$$\hat{\mathcal{E}}(\rho) := \mathrm{Tr}_{>n}\left(\mathcal{U}_{\text{sew}}(\{\hat{O}_{i,P}\}_{i,P})(\rho \otimes |0^n\rangle\langle 0^n|)\right), \tag{3.159}$$

which can be implemented as a constant-depth quantum circuit over $2n$ qubits. Because Eq. (3.151) holds with probability at least $1-\delta$, we have

$$\left\|\hat{\mathcal{E}} - \mathcal{U}\right\|_{\diamond} \leq \varepsilon \tag{3.160}$$

with probability at least $1-\delta$. This concludes the proof of the first part of Theorem 3.6.

Finite gate sets. Let the circuit depth be $d = \mathcal{O}(1)$, the finite gate set be $\mathcal{G}$ with $|\mathcal{G}| = \mathcal{O}(1)$, and the number of ancilla qubits be m. The ancilla qubits are initialized as $|0\rangle$ and end up at $|0\rangle$ after applying C, i.e.,

$$U \otimes |0^m\rangle = C(I_n \otimes |0^m\rangle). \tag{3.161}$$

The Schrodinger evolution of an n-qubit state ρ under U is

$$U\rho U^{\dagger} = \mathrm{Tr}_{>n}(C(\rho \otimes |0^m\rangle\langle 0^m|)C^{\dagger}), \tag{3.162}$$

where C is a shallow quantum circuit over $n+m$ qubits and $\mathrm{Tr}_{>n}$ traces out the ancilla qubits. The Heisenberg evolution of an n-qubit observable O under U is

$$U^{\dagger}OU = (I_n \otimes \langle 0^m|)C^{\dagger}(O \otimes I_m)C(I_n \otimes |0^m\rangle), \tag{3.163}$$

where I_n is an identity on n qubits and I_m is an identity on m qubits.

The algorithm utilizes a randomized measurement dataset $\mathcal{T}_U(N)$. The key ideas are using Lemma 3.10 and a brute-force search algorithm over a constant number of choices to find the exact Heisenberg-evolved Pauli observables, using Lemma 3.13 to sew the Heisenberg-evolved Pauli observables into a constant-depth quantum circuit, and using Lemma 3.9 to obtain the rigorous guarantee.

Lemma 3.14 (Characterizing the support)**.** *Given an n-qubit unitary U generated by a constant-depth quantum circuit C with m ancilla qubits. For each qubit $i \in \{1, \ldots, n\}$, let us define a set of qubits*

$$A(i) := \bigcup_{P \in \{X,Y,Z\}} \mathrm{supp}\left(U^{\dagger}P_iU\right). \tag{3.164}$$

We have $|A(i)| = \mathcal{O}(1)$ and the number of qubits j such that $A(i) \cap A(j) \neq \varnothing$ is at most a constant.

Proof. From the definition of U, $U \otimes |0^m\rangle = C(I_n \otimes |0^m\rangle)$, we have

$$A(i) \subseteq \bigcup_{P \in \{X,Y,Z\}} \mathrm{supp}\left(C^\dagger P_i C\right). \tag{3.165}$$

Let $d = \mathcal{O}(1)$ be the depth of the circuit C. We say qubit i is connected to qubit j in the circuit C if there is a sequence of gates in C with strictly decreasing layers, such that each pair of consecutive gates share a qubit and the first gate acts on qubit i and the last gate acts on qubit j. Let $B(i)$ be the set of qubits connected to i. Because each pair of consecutive two-qubit gates share a qubit, the number of possible gate sequences for a fixed i grows at most twice as large at every step. Hence, $|B(i)| \leq 2^d$. Furthermore, for any Pauli operator P, $\mathrm{supp}\left(C^\dagger P_i C\right)$ only contains qubits connected to i, so $A(i) \subseteq B(i)$. Together, $|A(i)| \leq |B(i)| \leq 2^d = \mathcal{O}(1)$. This establishes the first claim.

Now, we show that for any i, the number of j such that $B(i) \cap B(j) \neq \varnothing$ is at most a constant. If $B(i) \cap B(j) \neq \varnothing$, we know that there is a sequence of gates in C with strictly decreasing layers and then strictly increasing layers, such that each pair of consecutive gates share a qubit and the first gate acts on qubit i and the last gate acts on qubit j. Similar to before, The number of possible gate sequences for a fixed i grows at most twice as large at every step. Hence the number of j with $B(i) \cap B(j) \neq \varnothing$ is at most $2^{2d} = \mathcal{O}(1)$. Because $A(i) \subseteq B(i)$, any j with $A(i) \cap A(j) \neq \varnothing$ satisfies $B(i) \cap B(j) \neq \varnothing$. Therefore, the number of qubits j such that $A(i) \cap A(j) \neq \varnothing$ is at most a constant. This establishes the second claim of the lemma. □

From the above lemma and the fact that $\mathrm{supp}\left(U^\dagger P_i U\right) \subseteq A(i)$, we have

$$\left|\mathrm{supp}\left(U^\dagger P_i U\right)\right| \leq |A(i)| = \mathcal{O}(1). \tag{3.166}$$

This enables us to combine Lemma 3.12 for constructing $\mathcal{T}_{U^\dagger P_i U}(N), \forall i, P$ from $\mathcal{T}_U(N)$ and Lemma 3.10 for learning few-body observables with unknown supports (since $A(i)$ is unknown) to show the following. For any constant value $\tilde{\varepsilon} = \mathcal{O}(1)$, given a dataset size of

$$N = \mathcal{O}\left(\log(n/\delta)\right), \tag{3.167}$$

we can learn $\hat{O}_{i,P}, \forall i, P$, such that with probability at least $1-\delta$, for all $i \in \{1, \ldots, n\}$ and Pauli observable $P \in \{X, Y, Z\}$, we have

$$\left\|\hat{O}_{i,P} - U^\dagger P_i U\right\|_\infty \leq \tilde{\varepsilon}, \quad \text{and} \quad \mathrm{supp}(\hat{O}_{i,P}) \subseteq \mathrm{supp}(U^\dagger P_i U). \tag{3.168}$$

The computational time for learning all $\hat{O}_{i,P}$ is $\mathcal{O}(n^{\mathcal{O}(1)} \log(n/\delta)) = \mathrm{poly}(n) \log(1/\delta)$.

Our goal now is to find $U^\dagger P_i U$ exactly using the approximate observable $\hat{O}_{i,P}$ satisfying Eq. (3.168) by choosing a sufficiently small $\tilde{\varepsilon}$ that is constant in system size n. To do so, we need to consider the backward lightcone of qubit i in circuit C defined below.

Definition 3.16 (Backward lightcone in a circuit)**.** *We say a gate g in circuit U is in the backward lightcone of qubit i in C if there is a sequence of gates in C with strictly decreasing layers, such that each pair of consecutive gates share a qubit, the first gate acts on qubit i, and the last gate is g.*

The circuit C_i corresponding to the backward lightcone of qubit i in circuit C is the circuit with all gates in the backward lightcone of qubit i in circuit C.

The set S_i of qubits corresponding to the backward lightcone of qubit i in circuit C is the set of all qubits acted by at least one of the gates in the backward lightcone of qubit i in circuit C.

From the definition of C_i, S_i corresponding to the backward lightcones given above, we have

$$\mathrm{supp}(U^\dagger P_i U) \subseteq \mathrm{supp}(C^\dagger P_i C) \subseteq S_i \quad \text{and} \quad U^\dagger P_i U = (I_n \otimes \langle 0^m|) C_i^\dagger P_i C_i (I_n \otimes |0^m\rangle). \tag{3.169}$$

Note one cannot guarantee $S_i = \mathrm{supp}(U^\dagger P_i U)$. By a counting argument similar to the proof of Lemma 3.14, we have the following fact.

Fact 3.5 (Size of backward lightcone)**.** *Given a depth-d circuit C. The circuit C_i corresponding to the backward lightcone of qubit i in C consists of at most 2^{d-1} gates. The set S_i of qubits corresponding to the backward lightcone of qubit i in C contains at most 2^d qubits.*

Recall that the depth of C is $d = \mathcal{O}(1)$, and the gate set is $\mathcal{G}$ with $|\mathcal{G}| = \mathcal{O}(1)$. Because $d = \mathcal{O}(1)$, $|S_i| \leq 2^d = \mathcal{O}(1)$. For any $(n+m)$-qubit constant-depth circuit $\tilde{C}$ over a finite gate set, given a fixed set $\tilde{S}_i$ of qubits corresponding to the backward lightcone of qubit i in $\tilde{C}$, the number of possible circuit $\tilde{C}_i$ corresponding to the backward lightcone of qubit i in circuit $\tilde{C}$ is a constant independent of n, m and $1/\delta$. Hence, there is a constant number of $\tilde{C}_i^\dagger P_i \tilde{C}_i = \tilde{C}^\dagger P_i \tilde{C}$. We denote the possible choices of the n-qubit observable given the set $\tilde{S}_i$ and qubit $i \in \{1, \ldots, n\}$ to be $\mathcal{S}_{\mathrm{obs}}(i, \tilde{S}_i)$,

$$\mathcal{S}_{\mathrm{obs}}(i, \tilde{S}_i) := \Big\{ (I_n \otimes \langle 0^m|) \tilde{C}^\dagger P_i \tilde{C} (I_n \otimes |0^m\rangle) \;\Big|\; \tilde{C} \text{ is a depth-}d \text{ circuit over gate set } \mathcal{G}, \tag{3.170}$$

$$\text{such that } \tilde{S}_i \text{ is the set of qubits corresponding to the backward lightcone of qubit } i \text{ in } \tilde{C} \Big\} \tag{3.171}$$

We have $|\mathcal{S}_{\mathrm{obs}}(i, \tilde{S}_i)| = \mathcal{O}(1)$. Furthermore, we can always consider a permutation $\Pi_{i,\tilde{S}_i}$ over the qubits that implements the following permutation mapping,

$$1 \to_{\Pi_{i,\tilde{S}_i}} i, \quad \{1, \ldots, |\tilde{S}_i|\} \to_{\Pi_{i,\tilde{S}_i}} \tilde{S}_i, \tag{3.172}$$

and $\Pi_{i,\tilde{S}_i}$ acts as identity on the m ancilla qubits. Given a permutation $\Pi_{i,\tilde{S}_i}$ over the qubits (which is itself a unitary), we have

$$\mathcal{S}_{\mathrm{obs}}(i, \tilde{S}_i) = \Big\{ \Pi_{i,\tilde{S}_i} O \Pi_{i,\tilde{S}_i} \;\Big|\; O \in \mathcal{S}_{\mathrm{obs}}(1, \{1, \ldots, |\tilde{S}_i|\}) \Big\}. \tag{3.173}$$

We note that O acts on n qubits, while $\Pi_{i,\tilde{S}_i}$ acts on $n+m$ qubits; hence, we implicitly extend O to $n+m$ qubits by acting as identity on the m ancilla qubits. The set $\mathcal{S}_{\mathrm{obs}}(1, \{1, \ldots, |\tilde{S}_i|\})$ contains all the possible observables (up to permutation of the qubits) with $|\tilde{S}_i|$ qubits in the backward lightcone of qubit $i \in \{1, \ldots, n\}$ in a depth-d circuit.

Recall from Fact 3.5 that the set $\tilde{S}_i$ of qubits corresponding to the backward lightcone of qubit i in a depth-d circuit satisfies $1 \leq |\tilde{S}_i| \leq 2^d$. We take the union over all possible values of $|\tilde{S}_i|$ to define

$$\mathcal{S}^*_{\mathrm{obs}} := \bigcup_{k=1}^{2^d} \mathcal{S}_{\mathrm{obs}}(1, \{1, \ldots, k\}). \tag{3.174}$$

Because $2^d = \mathcal{O}(1)$ and for all $k = \mathcal{O}(1)$, $|\mathcal{S}_{\mathrm{obs}}(1, \{1, \ldots, k\}| = \mathcal{O}(1)$, we have $|\mathcal{S}^*_{\mathrm{obs}}| = \mathcal{O}(1)$. We define the minimum distance between every pair of distinct observables in $\mathcal{S}^*_{\mathrm{obs}}$ as follows,

$$\varepsilon^{\mathrm{dist}} := \min_{O_1 \neq O_2 \in \mathcal{S}^*_{\mathrm{obs}}} \|O_1 - O_2\|_\infty. \tag{3.175}$$

The minimum distance $\varepsilon^{\mathrm{dist}}$ depends on the depth $d = \mathcal{O}(1)$ and the finite gate set $\mathcal{G}$ with $|\mathcal{G}| = \mathcal{O}(1)$, so ε^* is a constant independent of the system size n and failure probability δ. We also define the minimum distance to an observable with a strictly smaller support.

$$\varepsilon^{\mathrm{supp}} := \min_{O_1 \in \mathcal{S}^*_{\mathrm{obs}}} \min_{\substack{O_2,\text{ such that} \\ \mathrm{supp}(O_2) \subseteq \mathrm{supp}(O_1) \\ \mathrm{supp}(O_2) \neq \mathrm{supp}(O_1)}} \|O_1 - O_2\|_\infty. \tag{3.176}$$

Because the support of O_2 is strictly contained in the support of O_1, we have $\|O_1 - O_2\|_\infty > 0$. And since $|\mathcal{S}^*_{\mathrm{obs}}| = \mathcal{O}(1)$, we have $\varepsilon^{\mathrm{supp}}$ is a constant independent of n and δ.

Let $\tilde{\varepsilon} = \min(\varepsilon^{\mathrm{dist}}, \varepsilon^{\mathrm{supp}})/3$ in Eq. (3.168), and define $\hat{S}_i := \{i\} \cup \mathrm{supp}(\hat{O}_{i,P})$. Consider any permutation $\Pi_{i,\hat{S}_i}$ over n qubits that implements the following permutation mapping,

$$1 \to_{\Pi_{i,\hat{S}_i}} i, \quad \{1, \ldots, |\hat{S}_i|\} \to_{\Pi_{i,\hat{S}_i}} \hat{S}_i. \tag{3.177}$$

We consider the following observable

$$O^*_{i,P} := \Pi_{i,\hat{S}_i} \left(\mathop{\arg\min}_{O \in \mathcal{S}^*_{\mathrm{obs}}} \left\| \Pi^{-1}_{i,\hat{S}_i} \hat{O}_{i,P} \Pi^{-1}_{i,\hat{S}_i} - O \right\|_\infty \right) \Pi_{i,\hat{S}_i}. \tag{3.178}$$

Because $|\mathcal{S}^*_{\mathrm{obs}}| = \mathcal{O}(1)$ and the dimension of $O \in \mathcal{S}^*_{\mathrm{obs}}$ is a constant, the brute-force minimum over $\mathcal{S}^*_{\mathrm{obs}}$ takes $\mathcal{O}(1)$ time. Because there are $3n$ observables $O^*_{i,P}$, the computational time to find all $3n$ observables $O^*_{i,P}$ is $\mathcal{O}(n)$. The following lemma shows that $O^*_{i,P}$ is exactly equal to the desired Heisenberg-evolved Pauli observable $U^\dagger P_i U$.

Lemma 3.15 (Exact reconstruction)**.** *Given the definitions above, with probability at least $1 - \delta$, we have $O^*_{i,P} = U^\dagger P_i U$ for all qubits i and Pauli observable P.*

Proof. We condition on the event that Eq. (3.168) is true, which happens with probability at least $1-\delta$. Recall that $\mathrm{supp}(\hat{O}_{i,P}) \subseteq \mathrm{supp}(U^\dagger P_i U)$ and $\left\|\hat{O}_{i,P} - U^\dagger P_i U\right\|_\infty \leq \tilde{\varepsilon} \leq \varepsilon^{\mathrm{supp}}/3$. From the definition of $\varepsilon^{\mathrm{supp}}$, we have $\mathrm{supp}(\hat{O}_{i,P}) = \mathrm{supp}(U^\dagger P_i U)$. Hence,

$$\hat{S}_i = \left(\{i\} \cup \mathrm{supp}(U^\dagger P_i U)\right) \subseteq S_i, \tag{3.179}$$

where S_i is the set of qubits corresponding to the backward lightcone of qubit i in circuit C. Consider any permutation $\Pi_{i,\hat{S}_i,S_i}$ over n qubits that is equal to $\Pi_{i,\hat{S}_i}$ for inputs $1,\ldots,|\hat{S}_i|$ and implements the following permutation mapping,

$$\left\{|\hat{S}_i|+1,\ldots,|S_i|\right\} \to_{\Pi_{i,\hat{S}_i,S_i}} S_i \setminus \hat{S}_i, \tag{3.180}$$

and $\Pi_{i,\tilde{S}_i,S_i}$ acts as identity on the m ancilla qubits. Because $\mathrm{supp}(U^\dagger P_i U) \subseteq \hat{S}_i$, we have

$$\begin{aligned}\Pi^{-1}_{i,\hat{S}_i} U^\dagger P_i U \Pi^{-1}_{i,\hat{S}_i} &= \Pi^{-1}_{i,\hat{S}_i}(I_n \otimes \langle 0^m|)C^\dagger(P_i \otimes I_m)C(I_n \otimes |0^m\rangle)\Pi^{-1}_{i,\hat{S}_i} \\ &= (I_n \otimes \langle 0^m|)\left(\Pi^{-1}_{i,\hat{S}_i,S_i} C^\dagger \Pi^{-1}_{i,\hat{S}_i,S_i}\right) P_1 \left(\Pi^{-1}_{i,\hat{S}_i,S_i} C \Pi^{-1}_{i,\hat{S}_i,S_i}\right)(I_n \otimes |0^m\rangle). \end{aligned} \tag{3.181}$$

By the definition of the permutation $\Pi^{-1}_{i,\hat{S}_i,S_i}$, $\{1,\ldots,|S_i|\}$ is the set of qubits corresponding to the backward lightcone of qubit 1 in the circuit $\Pi^{-1}_{i,\hat{S}_i,S_i} C \Pi^{-1}_{i,\hat{S}_i,S_i}$. As a result, we have

$$O^* := (I_n \otimes \langle 0^m|)\left(\Pi^{-1}_{i,\hat{S}_i,S_i} C^\dagger \Pi^{-1}_{i,\hat{S}_i,S_i}\right) P_1 \left(\Pi^{-1}_{i,\hat{S}_i,S_i} C \Pi^{-1}_{i,\hat{S}_i,S_i}\right)(I_n \otimes |0^m\rangle) \tag{3.182}$$

$$\in \mathcal{S}_{\mathrm{obs}}(1, \{1,\ldots,|S_i|\}) \subseteq \mathcal{S}^*_{\mathrm{obs}}. \tag{3.183}$$

The last $\subseteq$ follows from the fact that $|S_i| \leq 2^d$ in Fact 3.5. We can use Eq. (3.181) and

$$\left\|\hat{O}_{i,P} - U^\dagger P_i U\right\|_\infty \leq \tilde{\varepsilon} \leq \varepsilon^{\mathrm{dist}}/3 \tag{3.184}$$

to see that

$$\left\|\Pi^{-1}_{i,\hat{S}_i} \hat{O}_{i,P} \Pi^{-1}_{i,\hat{S}_i} - O^*\right\|_\infty \leq \varepsilon^{\mathrm{dist}}/3. \tag{3.185}$$

For any $O \in \mathcal{S}^*_{\mathrm{obs}}$ with $O \neq O^*$, we have $\|O - O^*\|_\infty \geq \varepsilon^{\mathrm{dist}}$. By the triangle inequality, we have

$$\left\|\Pi^{-1}_{i,\hat{S}_i} \hat{O}_{i,P} \Pi^{-1}_{i,\hat{S}_i} - O\right\|_\infty \geq \|O - O^*\|_\infty - \left\|\Pi^{-1}_{i,\hat{S}_i} \hat{O}_{i,P} \Pi^{-1}_{i,\hat{S}_i} - O^*\right\|_\infty \geq 2\varepsilon^{\mathrm{dist}}/3. \tag{3.186}$$

Together, we can show that O^* is the unique global minimum,

$$O^* = \operatorname*{arg\,min}_{O \in \mathcal{S}^*_{\mathrm{obs}}} \left\|\Pi^{-1}_{i,\hat{S}_i} \hat{O}_{i,P} \Pi^{-1}_{i,\hat{S}_i} - O\right\|_\infty. \tag{3.187}$$

Using Eq. (3.181) again shows that

$$O^*_{i,P} = \Pi_{i,\hat{S}_i}\left(\operatorname*{arg\,min}_{O \in \mathcal{S}^*_{\mathrm{obs}}} \left\|\Pi^{-1}_{i,\hat{S}_i} \hat{O}_{i,P} \Pi^{-1}_{i,\hat{S}_i} - O\right\|_\infty\right)\Pi_{i,\hat{S}_i} = U^\dagger P_i U. \tag{3.188}$$

This concludes the proof. □

From Lemma 3.14, we can characterize the support of $O^*_{i,P} = U^\dagger P_i U$ to apply Lemma 3.13. Lemma 3.13 shows that there exists an ordering for sewing the Heisenberg-evolved Pauli observables $O^*_{i,P} = U^\dagger P_i U$ to create $U_{\mathrm{sew}}(\{O^*_{i,P}\}_{i,P})$ given in Definition 3.15, such that $U_{\mathrm{sew}}(\{O^*_{i,P}\}_{i,P})$ can be implemented by a constant-depth quantum circuit. Under the event that $O^*_{i,P} = U^\dagger P_i U$ (think of $O^*_{i,P}$ as 0-approximate Heisenberg-evolved Pauli observable P on qubit i under U) for all Pauli observable P and qubit i, Lemma 3.9 shows that

$$U_{\mathrm{sew}}(\{O^*_{i,P}\}_{i,P}) = U \otimes U^\dagger. \tag{3.189}$$

Finally, define an n-qubit channel $\hat{\mathcal{E}}$ as follows,

$$\hat{\mathcal{E}}(\rho) := \mathrm{Tr}_{>n}\left(\mathcal{U}_{\mathrm{sew}}(\{O^*_{i,P}\}_{i,P})(\rho \otimes |0^n\rangle\!\langle 0^n|)\right), \tag{3.190}$$

which can be implemented as a constant-depth $2n$ qubits circuit. Using Lemma 3.15, we have

$$\hat{\mathcal{E}} = \mathcal{U} \tag{3.191}$$

with probability at least $1-\delta$. This concludes the proof of Theorem 3.6.

3.2.4.6 Learning geometrically-local shallow circuits (Proof of Theorem 3.7)

We present the algorithm for learning an unknown geometrically-local shallow quantum circuit U. We separate the proof into two-qubit gates over SU(4) and over a finite gate set.

Arbitrary SU(4) **gates.** We present the algorithm for learning an unknown geometrically-local shallow quantum circuit U over any two-qubit gate in SU(4). The algorithm uses the randomized measurement dataset $\mathcal{T}_U(N)$. The key ideas are constructing a superset of the support of the Heisenberg-evolved Pauli observables using Lemma 3.16, finding the Heisenberg-evolved Pauli observables for every qubit using Lemma 3.11, and sewing the Heisenberg-evolved Pauli observables together using Definition 3.15 and Lemma 3.9.

Consider the lightcones $L_d(i)$ for each qubit i with depth d as given in Definition 3.10. We have the following lemma for characterizing the properties of $L_d(i)$.

Lemma 3.16 (Properties of lightcones)**.** *Given a geometry over n qubits represented by a graph $G = (V, E)$ with a degree $\kappa = \mathcal{O}(1)$, a depth-d geometrically-local circuit U as given in Definition 3.9 with $d = \mathcal{O}(1)$, and the lightcones $L_d(i)$ for each qubit i with depth d as given in Definition 3.10. For each qubit i, we have*

$$\mathrm{supp}\left(U^\dagger P_i U\right) \subseteq L_d(i), \tag{3.192}$$

for any Pauli operator $P \in \{X, Y, Z\}$. Furthermore, $L_d(i)$ is geometrically local (see Definition 3.11), $|L_d(i)| = \mathcal{O}(1)$, $L_d(i)$ is known, and the number of qubits j such that $L_d(i) \cap L_d(j) \neq \varnothing$ is at most a constant.

Proof. Because U is of depth d and P_i acts only on qubit i, $U^\dagger P_i U$ only acts only on qubits that are distance d away from qubit i according to the graph G. By the definition of $L_d(i)$, we have $\operatorname{supp}\left(U^\dagger P_i U\right) \subseteq L_d(i)$. Recall that $|L_d(i)| \leq (\kappa+1)^d = \mathcal{O}(1)$. Furthermore, since G is known, $L_d(i)$ is known. Now, consider a qubit j such that $L_d(i) \cap L_d(j) \neq \varnothing$. This condition shows that qubit j must be of distance at most $2d$ from qubit i in the graph G. Hence, the number of such j is bounded above by $(\kappa+1)^{2d} = \mathcal{O}(1)$. This concludes the proof of the lemma. □

Lemma 3.16 shows that $L_d(i)$ is a geometrically-local set, $|L_d(i)| = \mathcal{O}(1)$, $L_d(i)$ is known, and the number of qubits j such that $L_d(i) \cap L_d(j) \neq \varnothing$ is at most a constant.

Recall that we can use Lemma 3.12 to constructing $\mathcal{T}_{U^\dagger P_i U}(N), \forall i, P$ from the classical dataset $\mathcal{T}_U(N)$ given in Definition 3.8. Because $|L_d(i)| = \mathcal{O}(1)$ and $L_d(i)$ is known, from Lemma 3.11, with a dataset size of

$$N = \mathcal{O}\left(\frac{n^2 \log(3n/\delta)}{\varepsilon^2}\right), \tag{3.193}$$

we can use $\mathcal{T}_{U^\dagger P_i U}(N), \forall i, P$ constructed from $\mathcal{T}_U(N)$ to learn $\hat{O}_{i,P}, \forall i, P$ such that, with probability at least $1-\delta$, for all $i \in \{1, \ldots, n\}$ and Pauli observable $P \in \{X, Y, Z\}$, we have

$$\left\|\hat{O}_{i,P} - U^\dagger P_i U\right\|_\infty \leq \frac{\varepsilon}{6n} \quad \text{and} \quad \operatorname{supp}(\hat{O}_{i,P}) \subseteq \operatorname{supp}\left(U^\dagger P_i U\right) \subseteq L_d(i). \tag{3.194}$$

The computational time for learning all $\hat{O}_{i,P}$ is $\mathcal{O}(n^3 \log(n/\delta)/\varepsilon^2)$.

We now utilize Lemm 3.13 to sew the learned observables into a geometrically-local constant-depth quantum circuit. To use the lemma, we note the following relations from Eq. (3.194),

$$A(i) := \bigcup_P \operatorname{supp}(\hat{O}_{i,P}) \subseteq \bigcup_P \operatorname{supp}(U^\dagger P_i U) \subseteq L_d(i). \tag{3.195}$$

Because $L_d(i)$ is a geometrically-local set, $|L_d(i)| = \mathcal{O}(1)$ and the number of qubits j such that $L_d(i) \cap L_d(j) \neq \varnothing$ is at most a constant, we have $A(i)$ is a geometrically-local set, $|A(i)| = \mathcal{O}(1)$ and the number of qubits j such that $A(i) \cap A(j) \neq \varnothing$ is at most a constant. Hence Lemma 3.13 given above shows that we can find an implementation of $U_{\text{sew}}(\{\hat{O}_{i,P}\}_{i,P})$ as a geometrically-local constant-depth $2n$-qubit circuit in time $\mathcal{O}(n)$. Given Eq. (3.194), we can use Lemma 3.9 on the form of the sewed Heisenberg-evolved Pauli observables to yield

$$\left\|\mathcal{U}_{\text{sew}}(\{\hat{O}_{i,P}\}_{i,P}) - \mathcal{U} \otimes \mathcal{U}^\dagger\right\|_\diamond \leq \varepsilon. \tag{3.196}$$

Finally, define an n-qubit channel $\hat{\mathcal{E}}$ as follows,

$$\hat{\mathcal{E}}(\rho) := \operatorname{Tr}_{>n}\left(\mathcal{U}_{\text{sew}}(\{\hat{O}_{i,P}\}_{i,P})(\rho \otimes |0^n\rangle\!\langle 0^n|)\right), \tag{3.197}$$

which can be implemented as a geometrically-local constant-depth quantum circuit over $2n$ qubits. Because Eq. (3.194) holds with probability at least $1-\delta$, we have

$$\left\| \hat{\mathcal{E}} - \mathcal{U} \right\|_\diamond \leq \varepsilon \tag{3.198}$$

with probability at least $1-\delta$. This concludes the proof of the first part of Theorem 3.7.

Finite gate sets. We present the algorithm for learning an unknown geometrically-local shallow quantum circuit U over a finite gate set. Let the depth of the unknown shallow quantum circuit be $d = \mathcal{O}(1)$ and the finite gate set be $\mathcal{G}$ with $|\mathcal{G}| = \mathcal{O}(1)$. The algorithm uses the randomized measurement dataset $\mathcal{T}_U(N)$. The algorithm constructs a superset of the support of the Heisenberg-evolved Pauli observables using Lemma 3.16, finds the Heisenberg-evolved Pauli observables for every qubit exactly using Lemma 3.11 and the information about the finite gate set $\mathcal{G}$, and sew the Heisenberg-evolved Pauli observables together using Definition 3.15 and Lemma 3.9.

Consider the lightcones $L_d(i)$ for each qubit i with depth d as given in Definition 3.10. Lemma 3.16 shows that $L_d(i)$ is a geometrically-local set, $|L_d(i)| = \mathcal{O}(1)$, $L_d(i)$ is known, and the number of qubits j such that $L_d(i) \cap L_d(j) \neq \varnothing$ is at most a constant. The algorithm and the proof proceed similarly to the case of having arbitrary two-qubit gates in SU(4). The main difference is in defining the following set $\mathcal{S}_{\mathrm{obs}}(P_i)$ for all $i \in \{1, \ldots, n\}$ and Pauli observable $P \in \{X, Y, Z\}$,

$$\mathcal{S}_{\mathrm{obs}}(P_i) := \left\{ U^\dagger P_i U \mid U \text{ is a geometrically-local depth-}d \text{ circuit over the gate set } \mathcal{G} \right\}. \tag{3.199}$$

Because $|\mathcal{G}| = \mathcal{O}(1)$ and $d = \mathcal{O}(1)$, the set $\mathcal{S}_{\mathrm{obs}}(P_i)$ contains a constant number of observables that only act on qubits in $L_d(i)$. We can define the minimum distance to be

$$\varepsilon_0(P_i) := \min \left\{ \|O_1 - O_2\|_\infty \mid O_1 \neq O_2 \in \mathcal{S}_{\mathrm{obs}}(P_i) \right\} = \Omega(1). \tag{3.200}$$

We also define $\varepsilon_0 = \min_{i,P} \varepsilon_0(P_i) = \Omega(1)$, which is a constant.

Recall that we can use Lemma 3.12 to constructing $\mathcal{T}_{U^\dagger P_i U}(N), \forall i, P$ from the classical dataset $\mathcal{T}_U(N)$ given in Definition 3.8. Because $|L_d(i)| = \mathcal{O}(1)$ and $L_d(i)$ is known, from Lemma 3.11, with a dataset size of

$$N = \mathcal{O}\left(\frac{\log(3n/\delta)}{\varepsilon_0^2} \right) = \mathcal{O}(\log(n/\delta)), \tag{3.201}$$

we can use $\mathcal{T}_{U^\dagger P_i U}(N), \forall i, P$ constructed from $\mathcal{T}_U(N)$ to learn $\hat{O}_{i,P}, \forall i, P$ such that, with probability at least $1-\delta$, for all $i \in \{1, \ldots, n\}$ and Pauli observable $P \in \{X, Y, Z\}$, we have

$$\left\| \hat{O}_{i,P} - U^\dagger P_i U \right\|_\infty \leq \frac{\varepsilon_0}{3} \quad \text{and} \quad \mathrm{supp}(\hat{O}_{i,P}) \subseteq \mathrm{supp}\left(U^\dagger P_i U \right) \subseteq L_d(i). \tag{3.202}$$

The computational time for learning all $\hat{O}_{i,P}$ is $\mathcal{O}(n\log(n/\delta)/\varepsilon_0^2) = \mathcal{O}(n\log(n/\delta))$. Because $U^\dagger P_i U \in \mathcal{S}_{\mathrm{obs}}(P_i)$ only has a constant number of possibilities, we can find

$$O^*_{i,P} := \operatorname*{arg\,min}_{O \in \mathcal{S}_{\mathrm{obs}}(P_i)} \left\| O - \hat{O}_{i,P} \right\|_\infty \tag{3.203}$$

in time $\mathcal{O}(n)$. Because the pairwise distance in $\mathcal{S}_{\mathrm{obs}}(P_i)$ is at least ε_0 and $U^\dagger P_i U \in \mathcal{S}_{\mathrm{obs}}(P_i)$,

$$O^*_{i,P} = U^\dagger P_i U, \quad \forall i \in \{1,\ldots,n\}, P \in \{X,Y,Z\} \tag{3.204}$$

with probability at least $1-\delta$.

We now utilize Lemm 3.13 to sew the learned observables into a geometrically-local constant-depth quantum circuit. To use the lemma, we note the following relations from Eq. (3.194),

$$A(i) := \bigcup_P \mathrm{supp}(O^*_{i,P}) \subseteq \bigcup_P \mathrm{supp}(U^\dagger P_i U) \subseteq L_d(i). \tag{3.205}$$

Because $L_d(i)$ is a geometrically-local set, $|L_d(i)| = \mathcal{O}(1)$ and the number of qubits j such that $L_d(i) \cap L_d(j) \neq \varnothing$ is at most a constant, we have $A(i)$ is a geometrically-local set, $|A(i)| = \mathcal{O}(1)$ and the number of qubits j such that $A(i) \cap A(j) \neq \varnothing$ is at most a constant. Hence Lemma 3.13 given above shows that we can find an implementation of $U_{\mathrm{sew}}(\{O^*_{i,P}\}_{i,P})$ as a geometrically-local constant-depth $2n$-qubit circuit in time $\mathcal{O}(n)$. Given Eq. (3.204), we can use Lemma 3.9 on the form of the sewed Heisenberg-evolved Pauli observables to yield

$$\mathcal{U}_{\mathrm{sew}}(\{O^*_{i,P}\}_{i,P}) = \mathcal{U} \otimes \mathcal{U}^\dagger. \tag{3.206}$$

Finally, define an n-qubit channel $\hat{\mathcal{E}}$ as follows,

$$\hat{\mathcal{E}}(\rho) := \mathrm{Tr}_{>n}\left(\mathcal{U}_{\mathrm{sew}}(\{O^*_{i,P}\}_{i,P})(\rho \otimes |0^n\rangle\langle 0^n|)\right), \tag{3.207}$$

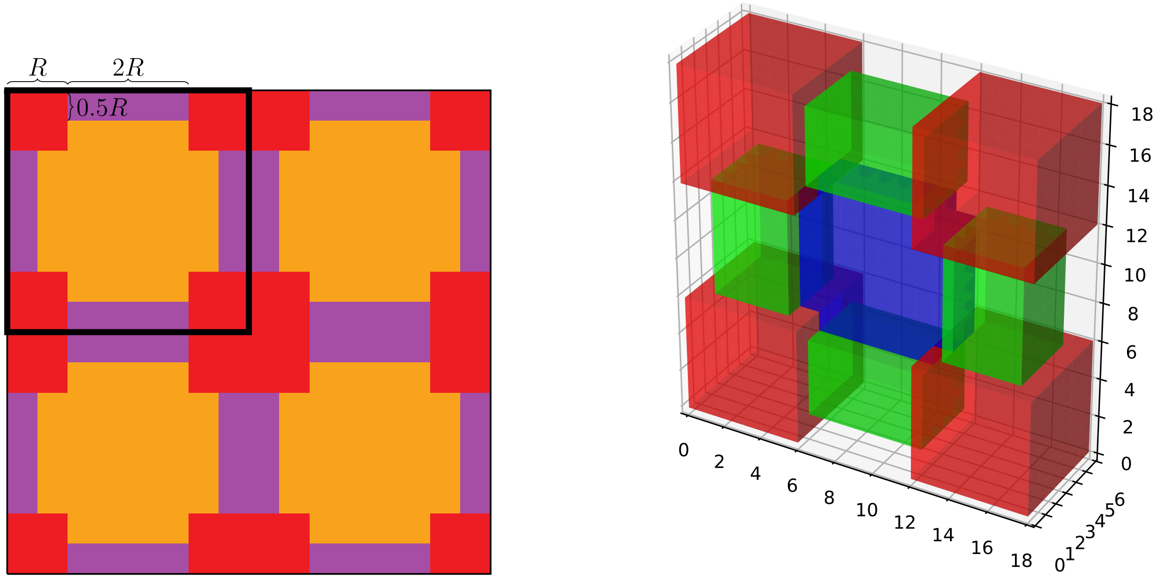
which can be implemented as a geometrically-local constant-depth quantum circuit over $2n$ qubits. Because Eq. (3.194) holds with probability at least $1-\delta$, we have

$$\hat{\mathcal{E}} = \mathcal{U} \tag{3.208}$$

with probability at least $1-\delta$. This concludes the proof of Theorem 3.7.

3.2.4.7 Learning shallow circuits on k-dimensional lattice with optimized circuit depth (Proof of Theorem 3.8)

Here we develop an approach to optimize the depth of the learned circuit. The main idea is to design a coloring scheme for the k-dimensional lattice with the fewest colors possible, such that gates supported on the same color can be implemented simultaneously.



(a) 2D

(b) 3D

Figure 3.2: A coloring of k-dimensional lattice with $k+1$ colors, where different regions of the same color are separated by distance at least R. (a) A coloring of 2-dimensional lattice. (b) A coloring of 3-dimensional lattice (the fourth color is not shown).

Definition 3.17 ($k+1$-coloring of k-dimensional lattice with distance R)**.** *Consider a graph representing a k-dimensional lattice (Fig. 3.1(a) shows $k = 2$). Each vertex is assigned a color, and the entire lattice is divided into many small regions with different colors. A $k+1$-coloring of k-dimensional lattice with distance R satisfies the following properties:*

1. *There are $k+1$ colors in total;*

2. *Each small region has constant size;*

3. *The distance between two regions with the same color is at least R.*

Here we give a construction of the above coloring (see Fig. 3.2). Similar approaches have been used in e.g. [212], although explicit constructions in 3D or above are not provided. The construction is based on "fattening" different t-cells in the lattice, from small to large t.[1] Consider a k-dimensional cube of length $2kR$ (the volume of the cube is $(2kR)^k$). Then we do the following:

- Fatten each 0-cell (vertices) to length kR, assign color 1.

- Fatten each 1-cell (edges) to length $(k-1)R$, assign color 2.

[1] We thank Jeongwan Haah for teaching this argument at PCMI 2023 Graduate Summer School.

- Fatten each 2-cell (faces) to length $(k-2)R$, assign color 3.
- ...
- Fill in the remaining k-cell with color $k+1$.

This is repeated in a translation-invariant way across the entire lattice.

This construction is illustrated in Fig. 3.2 for $k = 2, 3$. First, consider $k = 2$. A 2-dimensional square of size $4R \times 4R$ is shown in the top left corner (thick black box) of Fig. 3.2(a). In the first step, we fatten each of the 4 vertices into red squares of size $2R \times 2R$. Only a quarter of each red square remains within the original square. Next, we fatten each of the 4 edges into purple rectangles of size $R \times 2R$. This can be viewed as "growing" the edge until it has thickness R, but the regions that were colored red remain unchanged. Note the fact that the purple edges have a thickness of R, while the red vertices have a thickness of $2R$. This is crucial as it ensures that different purple regions are separated by a distance of at least R. Finally, the remaining regions are colored orange. Note that different orange regions are also separated by a distance of at least R due to the thickness of the purple edges.

The coloring of 3-dimensional lattices is shown in Fig. 3.2(b). Here we assign colors to a 3-dimensional cube of size $6R \times 6R \times 6R$, and Fig. 3.2(b) illustrates one of the six faces of that cube, which is the result of fattening the red vertices, green edges, and the blue face (the final coloring of the 3-cell is not shown in the figure). The thickness of the red vertices is larger than the thickness of the green edges, which guarantees that different green edges are separated by distance R. Similarly, the decrease in the thickness of the blue faces relative to the green edges guarantees the separation of different blue faces.

Choose $R = 3d$ in the above coloring scheme, and suppose the system is divided into L small regions $A_1, \ldots, A_L$ ($\sum_i |A_i| = n$). Two regions A_i, A_j that have the same color are separated by distance at least $3d$. Let A'_i be the ancilla system associated with A_i (see Fig. 3.1), and let S_{A_i} be the SWAP operator across A_i and A'_i. Let $S = \prod_{i=1}^{L} S_{A_i}$ be the global swap between system and ancilla. We are now ready to describe the learning algorithm. We separate the proof into two-qubit gates over SU(4) and over a finite gate set.

Arbitrary SU(4) **gates.** The learning algorithm proceeds in the same way as in Theorem 3.7; the only difference is that we need to learn Heisenberg-evolved Pauli operator $U^\dagger P U$ for P supported on each small regions in the coloring scheme instead of on each of the single qubits.

Our goal is to learn to implement the unitary

$$U \otimes U^\dagger = S \left[\prod_{i=1}^{L} (U^\dagger \otimes I) S_{A_i} (U \otimes I) \right]. \tag{3.209}$$

The algorithm learns each of the operators $W_{A_i} := (U^\dagger \otimes I) S_{A_i} (U \otimes I)$ and then multiply them together, followed by the global swap. The key idea to optimize the circuit depth of the learned circuit is to utilize the coloring scheme in the following sense:

Lemma 3.17 (Disjointness of supports)**.** *Let A_i, A_j be two regions with the same color. Then W_{A_i} and W_{A_j} have disjoint support.*

Proof. Recall that the operator W_{A_i} is supported on $L(A_i) \cup A_i'$, where $L(A_i)$ is the lightcone of A_i according to Definition 3.10. Therefore, W_{A_i} does not overlap with W_{A_j} when the lightcones $L(A_i)$ and $L(A_j)$ do not overlap. The coloring scheme has the property that A_i, A_j are separated by distance at least $3d$. Note that the lightcone of a region spreads the region by distance d. This implies that $L(A_i)$ and $L(A_j)$ are still separated by distance at least d and therefore do not overlap. □

Using the above lemma, we can construct the learned circuit by applying the learned operators $\{W_{A_i}\}$ with the same color simultaneously.

Lemma 3.18. *There is an implementation of $U \otimes U^\dagger$ via applying the operators $\{W_{A_i}\}$ in an appropriate order, such that the total circuit depth is $(k+1)(2d+1)+1$.*

Proof. We would like to implement

$$U \otimes U^\dagger = S \prod_{i=1}^{L} W_{A_i}. \tag{3.210}$$

Note that the operators $\{W_{A_i}\}$ pairwise commute, and we apply them in the following order: for each color $j \in \{1, 2, \ldots, k+1\}$, apply all operators W_{A_i} that has color j simultaneously. Finally, apply the global swap S.

Note that by definition, $W_{A_i} = (U^\dagger \otimes I) S_{A_i} (U \otimes I)$ can be viewed as a depth-$(2d+1)$ circuit acting on $L(A_i) \cup A_i'$. The total circuit depth is therefore $(k+1)(2d+1)+1$ (the final $+1$ comes from the global swap). □

The learning algorithm has two steps: learning and compiling.

1. (Learning) Learn an approximate classical description $\hat{W}_{A_i}$ for each W_{A_i}, such that $\|\hat{W}_{A_i} - W_{A_i}\|_\infty \leq \varepsilon_1$ for all i with high probability.
2. (Compiling) Compile the learned unitaries $\hat{W}_{A_i}$ from step one into depth-$(2d+1)$ circuits $\hat{W}'_{A_i}$, such that $\|\hat{W}_{A_i} - \hat{W}'_{A_i}\|_\infty \leq 2\varepsilon_1$ for all i.

The diamond distance between the learned circuit and the true circuit is at most $3L\varepsilon_1 \leq 3n\varepsilon_1$.

Step 1: Learning. The goal is to learn an approximation $\hat{O}_{i,P_{A_i}}$ of each operator $U^\dagger P_{A_i} U$, such that the following,

$$\left\| \hat{O}_{i,P_{A_i}} - U^\dagger P_{A_i} U \right\|_\infty \leq \frac{\varepsilon_1}{2^{|A_i|+1}}, \quad \forall i \in \{1, 2, \ldots, L\}, \quad P_{A_i} \in \{I, X, Y, Z\}^{|A_i|}, \tag{3.211}$$

holds with probability at least $1 - \delta$.

Using the fact that $S_{A_i} = \frac{1}{2^{|A_i|}} \sum_{P \in \{I,X,Y,Z\}^{|A_i|}} P \otimes P$, we have

$$W_{A_i} = \frac{1}{2^{|A_i|}} \sum_{P \in \{I,X,Y,Z\}^{|A_i|}} U^\dagger P U \otimes P. \tag{3.212}$$

Meanwhile, let

$$\hat{W}_{A_i} := \mathrm{Proj}_{\mathrm{U}} \left(\frac{1}{2^{|A_i|}} \sum_{P \in \{I,X,Y,Z\}^{|A_i|}} \hat{O}_{i,P_{A_i}} \otimes P_{A_i} \right). \tag{3.213}$$

From the lattice coloring scheme, we have $|L(A_i)| + |A_i| \leq 2(8kd)^k$. Hence, using Corollary 3.1 on exact unitary synthesis with geometrically-local circuits, we can implement $\hat{W}_{A_i}$ by a geometrically-local circuit with a circuit depth of

$$4(8kd)^k 4^{2(8kd)^k} \leq 4^{3(8kd)^k+1} \leq 4^{4(8kd)^k}. \tag{3.214}$$

Conditioned on Eq. (3.211) succeeds, the approximation error is bounded as follows:

$$\begin{aligned} \left\| \hat{W}_{A_i} - W_{A_i} \right\|_\infty &\leq 2 \left\| \frac{1}{2^{|A_i|}} \sum_{P \in \{I,X,Y,Z\}^{|A_i|}} \left(\hat{O}_{i,P_{A_i}} - U^\dagger P_{A_i} U \right) \otimes P_{A_i} \right\|_\infty \\ &\leq \frac{2}{2^{|A_i|}} \sum_{P \in \{I,X,Y,Z\}^{|A_i|}} \left\| \hat{O}_{i,P_{A_i}} - U^\dagger P_{A_i} U \right\|_\infty \\ &\leq \varepsilon_1. \end{aligned} \tag{3.215}$$

Here in the first line we use the same argument as in Eq. (3.123).

It remains to bound the time and query complexity to achieve the learning guarantee in Eq. (3.211). Given a randomized measurement dataset

$$\mathcal{T}_U(N) = \left\{ |\psi_\ell\rangle = \bigotimes_{i=1}^n |\psi_{\ell,i}\rangle, |\phi_\ell\rangle = \bigotimes_{i=1}^n |\phi_{\ell,i}\rangle \right\}_{\ell=1}^N, \tag{3.216}$$

for a Pauli operator $P \in \{I,X,Y,Z\}^{|A_i|}$ with weight $w \leq |A_i|$ (the weight of a Pauli operator is the number of non-identity elements), let

$$v_\ell^{U^\dagger P_{A_i} U} := 3^w \langle \phi_{\ell,A_i} | P | \phi_{\ell,A_i} \rangle, \tag{3.217}$$

where we let $|\phi_{\ell,A_i}\rangle := \otimes_{j \in A_i} |\phi_{\ell,j}\rangle$. The same argument in Lemma 3.12 shows that

$$\mathbb{E}\left[v_\ell^{U^\dagger P_{A_i} U} \right] = \langle \psi_\ell | U^\dagger P_{A_i} U | \psi_\ell \rangle. \tag{3.218}$$

Let $m := \max_i |L(A_i)| \leq (8kd)^k$ be the maximum support of the operators $U^\dagger P_{A_i} U$. Using Lemma 3.11, with a dataset size of

$$N = \frac{2^{\mathcal{O}(m)} \log(n/\delta)}{\varepsilon_1^2}, \tag{3.219}$$

Eq. (3.211) is achieved with success probability at least $1-\delta$.

Step 2: Compiling. Given a classical description of $\hat{W}_{A_i}$ as unitary acting on $L(A_i) \cup A_i'$, which can be implemented with a circuit depth of at most $4^{4(8kd)^k}$, we would like to find a depth-$(2d+1)$ circuit $\hat{W}'_{A_i}$ that is close to $\hat{W}_{A_i}$. To do this, we construct an ε-net for the circuit lightcone and perform a brute force search.

Definition 3.18 (ε-net for circuits)**.** *Consider a graph $G = (V, E)$. Let U be some unitary generated by d layers of 2-qubit gates where each gate is chosen from* SU(4) *and acts on an edge in E. An ε-net for circuits is a set of depth-d circuits defined on G, denoted as $\mathcal{N}_\varepsilon(G)$, such that for any choice of U, there exists $V \in \mathcal{N}_\varepsilon(G)$, such that $\|V - U\|_\infty \leq \varepsilon$.*

Lemma 3.19. *Let $G = (V, E)$ be a graph with $s = |V|$ vertices and maximum degree κ. An ε-net for depth-d circuits defined on G, denoted as $\mathcal{N}_\varepsilon(G)$, can be constructed with size at most $\left(\frac{\kappa s d}{\varepsilon}\right)^{\mathcal{O}(sd)}$ and in time $\left(\frac{\kappa s d}{\varepsilon}\right)^{\mathcal{O}(sd)}$.*

Proof. There are at most $sd/2$ 2-qubit gates in the circuit. We construct the ε-net by first enumerating all possible circuit architectures and then enumerate each 2-qubit gate using a $\frac{2\varepsilon}{sd}$-net for SU(4). In each layer, each qubit can interact with one of the κ neighboring qubits. This implies that the number of possible circuit architectures in one layer is at most κ^s. Therefore, the number of possible circuit architectures with depth d is at most κ^{sd}.

An ε_1-net for SU(4) can be constructed with $\left(\frac{c_0}{\varepsilon_1}\right)^{c_1}$ elements, where c_0, c_1 are absolute constants. Plugging in $\varepsilon_1 = \frac{2\varepsilon}{sd}$, the size of $\mathcal{N}_\varepsilon(G)$ is at most

$$\kappa^{sd} \cdot \left(\frac{\mathcal{O}(1) \cdot sd}{\varepsilon}\right)^{\mathcal{O}(1)\cdot sd} = \left(\frac{\kappa s d}{\varepsilon}\right)^{\mathcal{O}(sd)}. \tag{3.220}$$

This concludes the proof. □

Let $G_{L(A_i)}$ be the subgraph of k-dimensional lattice induced by vertices in $L(A_i)$. The lattice coloring scheme guarantees that the size of $L(A_i)$ is at most $(8kd)^k$. Let $\mathcal{N}_{\varepsilon_2}(G_{L(A_i)})$ be an ε_2-net for depth-d circuits acting on $L(A_i)$, which has size at most

$$\left(\frac{(8kd)^{k+1}}{\varepsilon_2}\right)^{\mathcal{O}(1)\cdot(8kd)^{k+1}}. \tag{3.221}$$

By definition, there is an element $V \in \mathcal{N}_{\varepsilon_2}(G_{L(A_i)})$ which is a depth-d circuit acting on $L(A_i)$, such that

$$\|(U^\dagger \otimes I) S_{A_i} (U \otimes I) - (V^\dagger \otimes I) S_{A_i} (V \otimes I)\|_\infty \leq 2\varepsilon_2, \tag{3.222}$$

which implies that

$$\|\hat{W}_{A_i} - (V^\dagger \otimes I)S_{A_i}(V \otimes I)\|_\infty \leq \varepsilon_1 + 2\varepsilon_2. \tag{3.223}$$

Therefore, enumerating over all elements in $\mathcal{N}_{\varepsilon_2}(G_{L(A_i)})$, we are guaranteed to find one element $\hat{V}$ that satisfies

$$\|\hat{W}_{A_i} - (\hat{V}^\dagger \otimes I)S_{A_i}(\hat{V} \otimes I)\|_\infty \leq \varepsilon_1 + 2\varepsilon_2. \tag{3.224}$$

Let $\varepsilon_2 = \varepsilon_1/2$ and define $\hat{W}'_{A_i} := (\hat{V}^\dagger \otimes I)S_{A_i}(\hat{V} \otimes I)$, we have $\|\hat{W}_{A_i} - \hat{W}'_{A_i}\|_\infty \leq 2\varepsilon_1$.

Putting everything together. To achieve diamond distance ε between the learned circuit $S\prod_{i=1}^{L}\hat{W}'_{A_i}$ and the true circuit $U \otimes U^\dagger$, it suffices to choose $\varepsilon_1 = \frac{\varepsilon}{3n}$. With probability at least $1-\delta$, we can learn all operators $\hat{W}_{A_i}$ within sufficient precision, using a dataset size of

$$N = \frac{2^{\mathcal{O}((8kd)^k)} n^2 \log(n/\delta)}{\varepsilon^2}. \tag{3.225}$$

Next, each $\hat{W}_{A_i}$ is classically compiled into a circuit, and they are combined together according to the order in Lemma 3.18, such that the learned circuit has total depth $(k+1)(2d+1)+1$. This classical postprocessing procedure takes a total time of

$$\mathcal{O}(nN) + (n/\varepsilon)^{\mathcal{O}(8kd)^{k+1})}, \tag{3.226}$$

which is polynomial in n and $1/\varepsilon$. If we do not compile $\hat{W}_{A_i}$ to the shorter-depth circuit $\hat{W}'_{A_i}$ and use $\hat{W}_{A_i}$ directly, then the classical postprocessing procedure only requires a computational time of

$$\mathcal{O}(nN), \tag{3.227}$$

but the learned circuit will have a total depth of $(k+1)4^{4(8kd)^k}+1$. This concludes the proof of the first part of Theorem 3.8.

Finite gate sets. The algorithm and the proof closely follow that of arbitrary SU(4) gates. When one considers a finite gate set with a constant size, a key simplification is the following: for any given $i \in \{1,\ldots,L\}$ and $P_{A_i} \in \{I,X,Y,Z\}^{|A_i|}$, $U^\dagger P_{A_i} U$ only takes on a constant number of options. Let $\varepsilon_{i,P_{A_i}} = \Omega(1)$ be the minimum distance in spectral norm between any pair of distinct $U^\dagger P_{A_i} U$.

From the same algorithm and proof in *Step 1: Learning*, we can ensure that

$$\left\|\hat{O}_{i,P_{A_i}} - U^\dagger P_{A_i} U\right\|_\infty \leq \frac{\varepsilon_{i,P_{A_i}}}{3}, \quad \forall i \in \{1,2,\ldots,L\}, \quad P_{A_i} \in \{I,X,Y,Z\}^{|A_i|}, \tag{3.228}$$

holds with probability at least $1-\delta$ using a sample complexity of

$$N = \mathcal{O}\left(\frac{\log(n/\delta)}{\varepsilon^2_{i,P_{A_i}}}\right) = \mathcal{O}\left(\log(n/\delta)\right). \tag{3.229}$$

From the definition of $\varepsilon_{i,P_{A_i}}$, we can identify $U^\dagger P_{A_i} U$ exactly from $\hat{O}_{i,P_{A_i}}$. This enables us to exactly reconstruct

$$W_{A_i} = \frac{1}{2^{|A_i|}} \sum_{P \in \{I,X,Y,Z\}^{|A_i|}} U^\dagger P U \otimes P = U^\dagger S_{A_i} U. \tag{3.230}$$

Because U is a quantum circuit of depth $d = \mathcal{O}(1)$ on a constant-dimensional lattice over a finite gate set of a constant size, we can perform a constant-time brute-force search to find a $(2d+1)$-depth circuit implementation for W_{A_i} instead of searching through the ε-net as in *Step 2: Compiling*. The computational time of the compiling step is improved from $(n/\varepsilon)^{\mathcal{O}(8kd)^{k+1}}$ to $\mathcal{O}(n)$. Following the rest of the proof for the case of SU(4) gates, we can learn U exactly with a learned circuit of depth $(k+1)(2d+1)+1$. The sample complexity is given in Eq. (3.229), and the computational time is dominated by reading the classical dataset, which is of $\mathcal{O}(nN) = \mathcal{O}(n \log(n/\delta))$. This concludes the proof of Theorem 3.8.

3.2.5 Learning shallow quantum circuits from quantum queries

We consider quantum learning algorithms that can access an unknown n-qubit unitary U through coherent quantum queries, which interleave the unitary U with quantum computation.

Definition 3.19 (Coherent quantum queries)**.** *The learning algorithm is a quantum algorithm with general coherent query access to the unknown unitary U. The quantum learning algorithm can interleave multiple accesses to the unknown unitary U with polynomial-size quantum circuits.*

We show the following result for learning geometrically-local shallow quantum circuits over finite gate sets with asymptotically optimal query complexity and time complexity. We only need to consider proving the matching upper bounds. The matching lower bounds to the query and time complexity are trivial: learning anything about U requires $\Omega(1)$ queries to U; writing down U requires $\Omega(n)$ time.

Theorem 3.9 (Learning geometrically-local shallow quantum circuits over a finite gate set)**.** *Given an unknown geometrically-local constant-depth n-qubit circuit U over a finite gate set. From*

$$N = \Theta(1) \tag{3.231}$$

queries to U, we can learn an n-qubit quantum channel $\hat{\mathcal{E}}$ that can be implemented by a geometrically-local constant-depth $2n$-qubit circuit, such that

$$\hat{\mathcal{E}} = \mathcal{U}, \tag{3.232}$$

with probability 1*. The computational time to learn $\hat{\mathcal{E}}$ is $\Theta(n)$.*

3.2.5.1 Learning local inversion using coherent quantum queries

When there is only a finite choice of possible unitaries, we can find the local inversion perfectly with $\mathcal{O}(1)$ queries, even if there is incoherent noise coming from the environment. This lemma is useful for showing the $\mathcal{O}(1)$ query complexity for learning n-qubit shallow quantum circuits with a finite gate set and a fixed geometric structure. The idea is to store multiple output quantum states in a quantum memory and utilize entangled quantum data processing. The formal statement is given below. We use the subscript on identity I or $\mathcal{I}$ to denote the number of qubits the identity acts on.

Lemma 3.20 (Perfect local inversion among finite choices)**.** *Consider $k, l, m = \mathcal{O}(1)$, unitaries $U_1, \ldots, U_m$ over k qubits, and unitaries $W_1, \ldots W_m$ over $(k-1)+l$ qubits. Let CPTP maps $\mathcal{E}_x$ from k to $k+l$ qubits be*

$$\mathcal{E}_x(\rho) := (\mathcal{I}_1 \otimes \mathcal{W}_x)(\mathcal{U}_x \otimes \mathcal{I}_l)(\rho \otimes I/2^l), \quad \forall x = 1, \ldots, m. \tag{3.233}$$

Given an unknown $\mathcal{E}_x$. Using $\mathcal{O}(1)$ queries to $\mathcal{E}_x$, we can find a perfect local inversion V_x of U_x on the first qubit. Furthermore, $V_x = U_i^\dagger$ for some i.

In order to prove the above lemma, we use a perfect local identity check for two choices given in Lemma 3.21. The proof of Lemma 3.20 is given after the proof of Lemma 3.21.

Lemma 3.21 (Perfect local identity check among two choices)**.** *Consider $k, l \geq 1$, two unitaries U_1, U_2 over k qubits, and two unitaries V_1, V_2 over $k+l-1$ qubits. Given CPTP maps from k qubits to $k+l$ qubits,*

$$\mathcal{E}_x(\rho) := (\mathcal{I}_1 \otimes \mathcal{V}_x)(\mathcal{U}_x \otimes \mathcal{I}_l)(\rho \otimes I/2^l), \quad \forall x = 1, 2. \tag{3.234}$$

Assume that k, l are constants, U_1 acts as identity on the first qubit $U_1 = I_1 \otimes \tilde{U}_1$, and U_2 is constant far from CPTP maps that act as an identity on the first qubit,

$$c := \min_{\mathcal{E}} \left\| \mathcal{U}_2 - \mathcal{I}_1 \otimes \mathcal{E} \right\|_\diamond = \Omega(1). \tag{3.235}$$

Given an unknown $\mathcal{E}_x$. Using $\mathcal{O}(1)$ queries to $\mathcal{E}_x$, we can perfectly distinguish between $\mathcal{E}_1$ and $\mathcal{E}_2$.

Proof. Let $|\Omega_k\rangle$ be the maximally entangled state over two copies of a k-qubit system. We define the following density matrices over $(k+l)+k$ qubits,

$$\rho_x := (\mathcal{I}_k \otimes \mathcal{E}_x)(|\Omega_k\rangle\!\langle\Omega_k|), \quad \forall x = 1, 2. \tag{3.236}$$

The support of a density matrix ρ is defined as

$$\mathrm{supp}(\rho) := \left\{ |\psi\rangle \,\middle|\, \langle\psi|\, \rho \,|\psi\rangle > 0 \right\}. \tag{3.237}$$

From the definition of ρ_x, we have

$$\mathrm{supp}(\rho_x) = \{(I_{k+1} \otimes V_x)(I_k \otimes U_x \otimes I_l)(|\Omega_k\rangle \otimes |\psi\rangle),\ \forall\, |\psi\rangle\}\,. \tag{3.238}$$

The maximal fidelity between two density matrices is defined as

$$\tilde{F}(\rho_1, \rho_2) := \max\big(|\,\langle\phi_1|\phi_2|\phi_1|\phi_2\rangle\,|\ \big|\ \ |\phi_x\rangle \in \mathrm{supp}(\rho_x),\ x = 1, 2\big)\,. \tag{3.239}$$

The maximal fidelity behaves similarly to fidelity and is multiplicative under tensor product

$$\tilde{F}(\rho_1 \otimes \sigma_1, \rho_2 \otimes\ \ \sigma_2) = \tilde{F}(\rho_1, \rho_2)\tilde{F}(\sigma_2, \sigma_2). \tag{3.240}$$

From the above definition, we see that there exists $|\psi_1\rangle\,, |\psi_2\rangle$ such that

$$\tilde{F}(\rho_1, \rho_2)^2 = \left|(\langle\Omega_k| \otimes \langle\psi_1|)(I_k \otimes U_2^\dagger \otimes I_l)(I_{k+1} \otimes (V_2^\dagger V_1(\tilde{U}_1 \otimes I_l)))(|\Omega_k\rangle \otimes |\psi_2\rangle)\right|^2. \tag{3.241}$$

We now consider two states associated with the above,

$$\sigma_1 := (I_{k+1} \otimes (V_2^\dagger V_1(\tilde{U}_1 \otimes I_l)))(|\Omega_k\rangle\!\langle\Omega_k| \otimes |\psi_2\rangle\!\langle\psi_2|)(I_{k+1} \otimes ((\tilde{U}_1^\dagger \otimes I_l)V_1^\dagger V_2)) \tag{3.242}$$

$$\sigma_2 := (I_k \otimes U_2 \otimes I_l)(|\Omega_k\rangle\!\langle\Omega_k| \otimes |\psi_1\rangle\!\langle\psi_1|)(I_k \otimes U_2^\dagger \otimes I_l) \tag{3.243}$$

The Fuchs–van de Graaf inequalities show that $\tilde{F}(\rho_1, \rho_2)^2 = \mathrm{Tr}(\sigma_1\sigma_2) \leq 1 - \frac{1}{4}\|\sigma_1 - \sigma_2\|_1^2$. We now consider a lower bound of the trace norm $\|\sigma_1 - \sigma_2\|_1$ by tracing out the last l qubits,

$$\|\sigma_1 - \sigma_2\|_1 \geq \|(\mathcal{I}_k \otimes \mathcal{I}_1 \otimes \mathcal{E})(|\Omega_k\rangle\!\langle\Omega_k|) - (\mathcal{I}_k \otimes \mathcal{U}_2)(|\Omega_k\rangle\!\langle\Omega_k|)\|_1, \tag{3.244}$$

where $\mathcal{E}$ is a CPTP map that acts on the last $k-1$ qubits. Recall that the 1-norm distance in the Choi states upper bounds the diamond distance in the CPTP maps up to the dimension factor $1/2^k$. From the definition of c in Eq. (3.235), we have the following inequality,

$$\|\sigma_1 - \sigma_2\|_1 \geq \frac{1}{2^k}\|\mathcal{I}_1 \otimes \mathcal{E} - \mathcal{U}_2\|_\diamond \geq \frac{c}{2^k}. \tag{3.245}$$

Therefore, we have

$$\tilde{F}(\rho_1, \rho_2) \leq \sqrt{1 - (c/2^{k+2})^2} < 1, \tag{3.246}$$

which is a key result that will be used later.

We need to consider another pair of states. Consider the Pauli decomposition of U_2 on the first qubit,

$$U_2 = \sum_{P \in \{I, X, Y, Z\}} P \otimes \tilde{U}_{2,P}, \tag{3.247}$$

where $\tilde{U}_{2,P}$ is a complex matrix of dimension 2^{k-1}. Because U_2 does not act as identity on the first qubit, we have $c' := \sum_{P \neq I} \mathrm{Tr}\Big(\tilde{U}_{2,P}^\dagger \tilde{U}_{2,P}\Big) > 0$ is a positive constant. Consider the following matrix,

$$M := \sum_{P \in \{X, Y, Z\}} P \otimes \tilde{U}_{2,P}, \tag{3.248}$$

and define two $2k$-qubit pure states,

$$|\psi_1\rangle := |\Omega_k\rangle\,, \tag{3.249}$$

$$|\psi_2\rangle := I_k \otimes \left(U_2^\dagger \frac{M}{\sqrt{\mathrm{Tr}(M^\dagger M)/2^k}} \right) |\Omega_k\rangle\,. \tag{3.250}$$

By the definition of c' and M, we have $\mathrm{Tr}\big(M^\dagger M\big) = 2c' > 0$ and

$$\tilde{F}(|\psi_1\rangle\!\langle\psi_1|\,, |\psi_2\rangle\!\langle\psi_2|) = |\, \langle \psi_1|\psi_2|\psi_1|\psi_2\rangle\,|^2 = 2c'/2^k > 0. \tag{3.251}$$

Furthermore, the overlap between $\mathcal{E}_x(|\psi_x\rangle\!\langle\psi_x|)$ satisfies

$$\begin{aligned}
&\mathrm{Tr}\left(\mathcal{E}_1(|\psi_1\rangle\!\langle\psi_1|)\mathcal{E}_2(|\psi_2\rangle\!\langle\psi_2|)\right) = \frac{1}{2c'/2^k}\cdot\frac{1}{2^k}\cdot\frac{1}{2^k}\cdot \\
&\sum_{P,Q\in\{X,Y,Z\}} \mathrm{Tr}\Big(\mathrm{Tr}_{\leq k}(P\otimes((\tilde{U}_{2,P}^\dagger\otimes I_l)V_2^\dagger V_1(\tilde{U}_1\otimes I_l)))\,\mathrm{Tr}_{\leq k}(Q\otimes((\tilde{U}_1^\dagger\otimes I_l)V_1^\dagger V_2(\tilde{U}_{2,Q}\otimes I_l)))\Big) \\
&= 0,
\end{aligned} \tag{3.252}$$

which implies that there exists a two-outcome projective measurement $\mathcal{M}$ that could perfectly distinguish between the two states $\mathcal{E}_1(|\psi_1\rangle\!\langle\psi_1|)$ and $\mathcal{E}_2(|\psi_2\rangle\!\langle\psi_2|)$.

Consider N queries to $\mathcal{E}_x$ to obtain $\rho_x^{\otimes N}$, where the number of queries is

$$N := \max\left(1, \left\lceil \frac{\log\left((2c'/2^k)\right)}{\log\left(\sqrt{1-(c/2^{k+2})^2}\right)} \right\rceil\right) = \mathcal{O}(1). \tag{3.253}$$

Using Eq. (3.240), (3.246), and (3.251), we have

$$\tilde{F}(\rho_1^{\otimes N}, \rho_2^{\otimes N}) = \tilde{F}(\rho_1,\rho_2)^N \leq \sqrt{1-(c/2^{k+2})^2}^N \leq (2c'/2^k) = \tilde{F}(|\psi_1\rangle\!\langle\psi_1|\,, |\psi_2\rangle\!\langle\psi_2|). \tag{3.254}$$

From Lemma 1 of [205], there exists a CPTP map $\mathcal{T}$ that takes ρ_x to $|\psi_x\rangle\!\langle\psi_x|$ for $x = 1, 2$. We apply $\mathcal{T}$ to ρ_x. And we evoke one additional query to $\mathcal{E}_x$ to obtain $\mathcal{E}_x(|\psi_x\rangle\!\langle\psi_x|)$. Finally, we perform the two-outcome projective measurement $\mathcal{M}$ to perfectly distinguish between $\mathcal{E}_1(|\psi_1\rangle\!\langle\psi_1|)$ and $\mathcal{E}_2(|\psi_2\rangle\!\langle\psi_2|)$. Together, with $N + 1 = \mathcal{O}(1)$ queries to $\mathcal{E}_x$, we can perfectly distinguish between $\mathcal{E}_1$ and $\mathcal{E}_2$. □

We are now ready to prove Lemma 3.20. The central idea is a bipartite tournament with a potential local inversion on one side and all possible non-local inversion on the other side.

Proof of Lemma 3.20. Each query to $\mathcal{E}_x$ allows us to create 1 query to any one of the following CPTP maps,

$$\mathcal{E}_{x,i} = (\mathcal{E}_x \circ \mathcal{U}_i^\dagger),\ \forall i = 1,\ldots,m. \tag{3.255}$$

The algorithm proceeds by going through all of i one by one. For each i, the algorithm creates two sets,

$$S_i := \left\{ y \in \{1, \ldots, m\} \mid U_y U_i^\dagger \text{ acts as identity on the first qubit} \right\}, \tag{3.256}$$

$$T_i := \{1, \ldots, m\} \setminus S_i. \tag{3.257}$$

Note that by definition, $i \in S_i$ and $i \notin T_i$. For each $y \in T_i$, the algorithm uses the algorithm given in the proof of Lemma 3.21 to test whether $\mathcal{E}_{x,i}$ is equal to $\mathcal{E}_{y,i}$ or $\mathcal{E}_{i,i}$. If $\mathcal{E}_{x,i}$ is indeed equal to one of them, then the algorithm in Lemma 3.21 is guaranteed to output the one that is equal to $\mathcal{E}_{x,i}$. If not, then the algorithm in Lemma 3.21 will output $\mathcal{E}_{y,i}$ or $\mathcal{E}_{i,i}$ arbitrarily. After going through all $y \in T_i$, if between $\mathcal{E}_{y,i}$ and $\mathcal{E}_{i,i}$, $\mathcal{E}_{i,i}$ is always chosen for all $y \in T_i$, then the algorithm sets $i^* := i$ and terminates the for-loop over i. The algorithm outputs $U_{i^*}^\dagger$ as the claimed perfect local inversion of U_x on the first qubit.

By construction, the total number of queries to $\mathcal{E}_x$ in the above algorithm is a constant. We now prove that (a) i^* can always be found by the above algorithm and (b) $U_{i^*}^\dagger$ is a perfect local inversion of U_x on the first qubit. The proof is separated into the following two paragraphs addressing each claim.

i^* can always be found. When $i = x$, for each $y \in T_i$, we are testing whether $\mathcal{E}_{x,x}$ is equal to $\mathcal{E}_{y,x}$ or $\mathcal{E}_{x,x}$. Because $U_y U_x^\dagger$ does not act as identity on the first qubit by definition of T_x, Lemma 3.21 shows that the algorithm will always return $\mathcal{E}_{x,x}$ when deciding between $\mathcal{E}_{y,x}$ and $\mathcal{E}_{x,x}$. Hence when $i = x$, the algorithm will set $i^* := i$ and terminate the for-loop over i. The algorithm could also terminate earlier for some $i < x$ but will always terminate when $i = x$. Therefore, i^*, as defined by the algorithm previously, can always be found.

$U_{i^*}^\dagger$ is a perfect local inversion of U_x on the first qubit. We first show by contradiction that $x \notin T_{i^*}$. Suppose that $x \in T_{i^*}$. For $y = x \in T_{i^*}$, we would be testing whether $\mathcal{E}_{x,i^*}$ is equal to $\mathcal{E}_{x,i^*}$ or $\mathcal{E}_{i^*,i^*}$. Recall that $i^* \notin T_{i^*}$, thus $x \neq i^*$. Lemma 3.21 thus implies that the algorithm will always return $\mathcal{E}_{x,i^*}$ when deciding between $\mathcal{E}_{x,i^*}$ and $\mathcal{E}_{i^*,i^*}$. As a result, the condition defining i^* is not satisfied, which is a contradiction. Because $S_{i^*} \cup T_{i^*} = \{1, \ldots, m\}$, we have $x \in S_{i^*}$. which means have $U_x U_{i^*}^\dagger$ acts as identity on the first qubit. As a result, $U_{i^*}^\dagger$ is a perfect local inversion of U_x on the first qubit. □

3.2.5.2 Learning geometrically-local shallow circuits over a finite gate set (Proof of Theorem 3.9)

We present the algorithm for learning an unknown geometrically-local shallow quantum circuit U over a finite gate set. Let the geometry over n qubits be represented by a graph $G = (V, E)$ with degree $\kappa = \mathcal{O}(1)$, the depth of U be $d = \mathcal{O}(1)$, and the finite gate set be $\mathcal{G}$ with $|\mathcal{G}| = \mathcal{O}(1)$. This algorithm requires coherent quantum queries to the unknown unitary U. The key ideas are constructing n CPTP maps $\mathcal{E}_i^U, \forall i \in \{1, \ldots, n\}$ from $\mathcal{O}(1)$ queries to U, utilizing Lemma 3.20 to find perfect local inversion among finite choices, and using Definition 3.13 and Lemma 3.7 to sew the local inversion unitaries together.

We consider the lightcone $L_d(i)$ of the geometry for qubit i under the unknown depth-d geometrically-local circuit U in Definition 3.10 and the properties of the lightcones given in Lemma 3.16.

For each qubit i in the n-qubit system, we can always decompose the depth-d geometrically-local quantum circuit U as the following,

$$U = \left(I_i \otimes W^{(i)} \otimes I_{\notin L_{2d}(i)}\right) \left(U^{(i)} \otimes \tilde{W}^{(i)}\right), \tag{3.258}$$

where $U^{(i)}$ acts on qubits in the set $L_d(i)$, $\tilde{W}^{(i)}$ acts on qubits not in the set $L_d(i)$, $W^{(i)}$ acts on qubits in the set $L_{2d}(i) \setminus \{i\}$, and $I_i, I_{\notin L_{3d}(i)}$ are identity matrices acting on qubit i and qubits not in $L_{3d}(i)$, respectively. Furthermore, $U^{(i)}, W^{(i)}, \tilde{W}^{(i)}$ are all subcircuits (circuits containing a subset of gates) of the unknown depth-d geometrically-local circuits U. We define the CPTP map $\mathcal{E}_i^U$,

$$\mathcal{E}_i^U(\rho) := \mathrm{Tr}_{\notin L_{2d}(i)} \left(U \left(\rho \otimes \frac{I_{\notin L_d(i)}}{2^{n-|L_d(i)|}} \right) U^\dagger \right) \tag{3.259}$$

$$= \left(\mathcal{I}_i \otimes \mathcal{W}^{(i)}\right) \left(\mathcal{U}^{(i)} \otimes \mathcal{I}_{L_{2d}(i) \setminus L_d(i)}\right) \left(\rho \otimes \frac{I_{L_{2d}(i) \setminus L_d(i)}}{2^{|L_{2d}(i)|-|L_d(i)|}} \right), \tag{3.260}$$

where ρ is a density matrix for qubits in $L_d(i)$, $I_{\notin L_d(i)}$ is the identity matrix over qubits not in $L_d(i)$, $I_{\notin L_d(i)}/2^{n-|L_d(i)|}$ is the maximally mixed state for qubits not in $L_d(i)$, and $\mathrm{Tr}_{\notin L_{2d}(i)}$ traces out all qubits not in $L_{2d}(i)$. Because $\mathcal{E}_i^U(\rho)$ uses a single query to U, naively, one would expect that to obtain a query to $\mathcal{E}_i^U$ for every qubit i requires n queries to U. The following lemma shows that we can do much more efficiently than what one would naively expect.

Lemma 3.22 (Queries to every $\mathcal{E}_i^U$ from only $\mathcal{O}(1)$ queries to U)**.** *We can construct a query to every $\mathcal{E}_i^U, 1 \leq i \leq n$ from only $\mathcal{O}(1)$ queries to the unknown constant-depth geometrically-local circuit U.*

Proof. Let $d = \mathcal{O}(1)$ be the depth of the circuit U. We consider a graph $G^{(3d)}$ over n qubits, where each pair of qubits is connected by an edge if their distance in G is at most $3d$. The degree of $G^{(3d)}$ is at most $(\kappa+1)^{3d} = \mathcal{O}(1)$. The graph only has $\mathcal{O}(n)$ edges and can be constructed as an adjacency list in time $\mathcal{O}(n)$. Let us define a coloring of the graph $G^{(3d)}$. By the standard greedy coloring algorithm, we can find a color $c^{(3d)}(i)$ for each qubit i in graph $G^{(3d)}$, where no adjacent vertices can have the same color, and there are only $\chi^{(3d)}$ distinct colors with

$$\chi^{(3d)} \leq (\kappa+1)^{3d} + 1 = \mathcal{O}(1). \tag{3.261}$$

The greedy coloring algorithm runs in time linear in the number of edges in $G^{(3d)}$, which is linear in the number n of qubits.

For each color $c = 1, \ldots, \chi^{(3d)}$, we consider the set of qubits with color c. We can construct one query to every $\mathcal{E}_i^U$ for qubits i with color $c^{(3d)}(i) = c$ from only one query to U. By the construction of the graph coloring, for two distinct qubits $i \neq j$ with the same color c, $L_{3d}(i) \cap L_{3d}(j) = \varnothing$. We now define the following sets of qubits for the color c,

$$A(c) := \left\{ i \in \{1, \ldots, n\} \,\middle|\, c^{(3d)}(i) = c \right\}, \quad B_q(c) := \bigcup_{i : c^{(3d)}(i) = c} L_q(i), \tag{3.262}$$

for any integer $q \geq 1$. Given the definition of $U^{(i)}, W^{(i)}$ in Eq. (3.258) for each qubit i. We can further decompose the shallow circuit U as

$$U = \left[\left(I_{A(c)} \otimes \bigotimes_{i : c^{(3d)}(i) = c} W^{(i)} \right) \otimes I_{\notin B_{2d}(c)} \right] \left[\left(\bigotimes_{i : c^{(3d)}(i) = c} U^{(i)} \right) \otimes \tilde{W}^{(c)} \right], \tag{3.263}$$

where $\tilde{W}^{(c)}$ acts on qubits not in $B_d(c)$. Consider initializing the qubits not in $B_d(c)$ as the maximally mixed state, evolving under U, and tracing out any qubits not in $B_{2d}(c)$. The resulting CPTP map $\mathcal{E}_c^U$ from qubits in $B_d(c)$ to qubits in $B_{2d}(c)$ can be written as

$$\mathcal{E}_c^U(\rho) = \left(\mathcal{I}_{A(c)} \otimes \bigotimes_{i : c^{(3d)}(i) = c} \mathcal{W}^{(i)} \right) \left(\bigotimes_{i : c^{(3d)}(i) = c} \mathcal{U}^{(i)} \otimes \mathcal{I}_{B_{2d}(i) \setminus B_d(i)} \right) \left(\rho \otimes \frac{I_{B_{2d}(c) \setminus B_d(c)}}{2^{|B_{2d}(c)| - |B_d(c)|}} \right), \tag{3.264}$$

where ρ is a density matrix over qubits in $B_d(c)$. It is not hard to see that

$$\mathcal{E}_c^U = \bigotimes_{i : c^{(3d)}(i) = c} \mathcal{E}_i^U. \tag{3.265}$$

Because $\mathcal{E}_c^U$ only requires one query to U, we can create $\mathcal{E}_i^U$ for all qubit i with color c from one query to U. Since there is only $\chi^{(3d)} = \mathcal{O}(1)$ colors, we can create a query to every $\mathcal{E}_i^U, 1 \leq i \leq n$ from only $\mathcal{O}(1)$ queries to the unknown circuit U. □

Because U is over a finite gate set with size $\mathcal{O}(1)$, we have $U^{(i)}$ and $W^{(i)}$ only have a constant number of choices. Furthermore, both $U^{(i)}$ and $W^{(i)}$ act on a constant number of qubits because $|L_d(i)| = \mathcal{O}(1), |L_{2d}(i)| = \mathcal{O}(1)$ for a constant depth d. From Lemma 3.20, for each qubit i, through $\mathcal{O}(1)$ queries to $\mathcal{E}_i^U$, we can learn a perfect local inversion V_i of $U^{(i)}$ on qubit i with no failure probability. The local inversion unitary V_i is the inverse of one of the possible choices for $U^{(i)}$. Hence, V_i is a geometrically-local depth-d circuit that only acts on qubits in $L_d(i)$. Combining with Lemma 3.22, from only $\mathcal{O}(1)$ queries to U, we can learn $V^{(i)}, \forall i = 1, \ldots, n$, such that

$$\mathcal{U}^{(i)} \mathcal{V}_i = \mathcal{I}^{(i)} \otimes \mathcal{E}_{\neq i}^{\mathcal{U}^{(i)} \mathcal{V}_i}, \tag{3.266}$$

where $\mathcal{I}^{(i)}$ is the identity map on qubit i and $\mathcal{E}_{\neq i}^{\mathcal{U}^{(i)} \mathcal{V}_i}$ is the reduced channel of $\mathcal{U}^{(i)} \mathcal{V}_i$ with qubit i removed. The quantum computational time is given by $\mathcal{O}(n)$. We now show that V_i is also the perfect local inversion unitary for U on qubit i. To see this, recall the decomposition in Eq. (3.258), we have

$$\mathcal{U}\mathcal{V}_i = \left(\mathcal{I}_i \otimes \mathcal{W}^{(i)} \otimes \mathcal{I}_{\notin L_{2d}(i)}\right)\left(\mathcal{U}^{(i)}\mathcal{V}_i \otimes \tilde{\mathcal{W}}^{(i)}\right) \tag{3.267}$$

$$= \mathcal{I}^{(i)} \otimes \left(\left(\mathcal{W}^{(i)} \otimes \mathcal{I}_{\notin L_{2d}(i)}\right)\left(\mathcal{E}_{\neq i}^{\mathcal{U}^{(i)}\mathcal{V}_i} \otimes \tilde{\mathcal{W}}^{(i)}\right)\right) \tag{3.268}$$

$$= \mathcal{I}^{(i)} \otimes \mathcal{E}_{\neq i}^{\mathcal{U}\mathcal{V}_i}. \tag{3.269}$$

We can now use Definition 3.13 and Lemma 3.7 to sew the perfect local inversion unitaries together. This gives the following $2n$-qubit unitary,

$$U_{\mathrm{sew}}(V_1, \ldots, V_n) = S\left[\prod_{i=1}^{n}\left(V_i^{(1)}\right) S_i \left(V_i^{(1)}\right)^{\dagger}\right] = U \otimes U^{\dagger}, \tag{3.270}$$

where $V_i^{(1)}$ is the unitary V_i acting on the first set of n qubits.

We now show that there exists a sewing ordering such that $U_{\mathrm{sew}}(V_1, \ldots, V_n)$ is a constant-depth geometrically-local circuit. Given the geometry over n qubits represented by a graph $G = (V, E)$. Consider a graph $G^{(2d)}$ over n qubits, where each pair (i, j) of qubits are connected by an edge if i, j is of distance at most $2d$ in the geometric graph G. Hence, equivalently, for all (i, j) not connected by an edge in $G^{(2d)}$, we have

$$L_d(i) \cap L_d(j) = \varnothing. \tag{3.271}$$

The degree of $G^{(2d)}$ is bounded above by $(\kappa + 1)^{2d}$. And $G^{(2d)}$ can be constructed as an adjacency list in time $\mathcal{O}(n)$. Because the graph has a constant degree, we can use a $\mathcal{O}(n)$-time greedy graph coloring algorithm to color the n-qubit graph $G^{(2d)}$ using only a constant number of colors. For each node/qubit i, we consider $c(i)$ to be the color. The sewing order for the n local inversion unitaries V_i is given by the greedy graph coloring, where we order from the smallest color to the largest color. By the definition of graph coloring, for any pair i, j of qubits with the same color, we have $L_d(i) \cap L_d(j) = \varnothing$. Furthermore, V_i is a constant-depth geometrically-local circuit that only acts on a constant number of qubits. Therefore, for any color c', we can find an implementation of the $2n$-qubit unitary

$$\prod_{i: c(i) = c'}\left(V_i^{(1)}\right) S_i \left(V_i^{(1)}\right)^{\dagger} \tag{3.272}$$

with a constant-depth geometrically-local quantum circuit in time $\mathcal{O}(n)$. Since there is only a constant number of colors, the $2n$-qubit unitary $U_{\mathrm{sew}}(V_1, \ldots, V_n)$ in Eq. (3.270) with the color-based ordering can be implemented with a constant-depth geometrically-local quantum circuit in time $\mathcal{O}(n)$. Finally, define an n-qubit channel $\hat{\mathcal{E}}$ as follows,

$$\hat{\mathcal{E}}(\rho) := \mathrm{Tr}_{>n}\left(\mathcal{U}_{\mathrm{sew}}(V_1, \ldots, V_n)(\rho \otimes |0^n\rangle\langle 0^n|)\right), \tag{3.273}$$

which can be implemented as a geometrically-local constant-depth quantum circuit over $2n$ qubits. Because $U_{\text{sew}}(V_1, \ldots, V_n) = U \otimes U^\dagger$ from Eq. (3.270), we have

$$\mathcal{E} = \mathcal{U} \tag{3.274}$$

with probability one. This concludes the proof of Theorem 3.9.

3.2.6 Hardness for learning log-depth quantum circuits

We have seen from the previous appendices that learning general constant-depth quantum circuits can be done efficiently. A natural follow-up question is whether one could efficiently learn log-depth quantum circuits. In the following, we show that learning log-depth quantum circuits to a constant diamond distance is exponentially hard, even when we allow coherent quantum queries to U. Hence, the problem of learning quantum circuits transitions from being polynomially easy to exponentially hard when we go from $\mathcal{O}(1)$-depth to $\mathcal{O}(\log n)$-depth.

Proposition 3.3 (Hardness for learning log-depth circuits)**.** *Consider an unknown n-qubit unitary U generated by a $\mathcal{O}(\log n)$-depth circuit over arbitrary two-qubit gates with n ancilla qubits. We have*

- *Learning U to $1/3$ diamond distance with high probability requires $\exp(\Omega(n))$ queries.*
- *Distinguishing whether U equals to the identity I or is $1/3$-far from the identity I in diamond distance with high probability requires $\exp(\Omega(n))$ queries.*

Proof. Without loss of generality, we consider n to be 2^k for an integer k. Consider the unknown unitary U to be I or one of $U_x, \forall x \in \{0,1\}^n$. The unitary U_x is defined to be

$$U_x \left|y\right\rangle = \begin{cases} 1, & x = y, \\ -1, & x \neq y, \end{cases} \tag{3.275}$$

for any $y \in \{0,1\}^n$. The n-qubit unitary U_x can be constructed as follows,

$$U_x = \left(\prod_{\substack{1 \leq i \leq n \\ x_i = 0}} X_i \right) C^n Z \left(\prod_{\substack{1 \leq i \leq n \\ x_i = 0}} X_i \right), \tag{3.276}$$

where X_i is the X gate on the i-th qubit, and $C^n Z$ is a controlled-Z gate controlled on all qubits. The circuit $\prod_{i:\, x_i = 0} X_i$ can be implemented in one layer. We can implement $C^n Z$ using n ancilla qubits in depth $\mathcal{O}(\log n)$. To see this, we first construct a $(2^k + 2^k - 1)$-qubit unitary V recursively as follows:

1. Set the $n = 2^k$ qubits to be the first set of control qubits. Set $j \leftarrow k$.

2. Consider the 2^j control qubits as 2^{j-1} pairs of two control qubits. Include 2^{j-1} new ancilla qubits initialized at $|0\rangle^n$.
3. For each pair of control qubits, implement a CCX gate on each newly added ancilla qubit controlled on the two control qubits.
4. Set the new 2^{j-1} ancilla qubits as the set of control qubits. Set $j \leftarrow j - 1$.
5. If $j > 0$, repeat Step 2.

We can compile the CCX gate acting on three qubits to be a sequence with a constant number of two-qubit gates. The depth of V is $\mathcal{O}(\log n)$. The unitary V computes whether all n qubits are one and stores the result in the $2n-1$ qubit. We can implement the n-qubit unitary $C^n Z$ using a $2n$-qubit $\mathcal{O}(\log n)$-depth circuit with n ancilla qubits,

$$C^n Z \otimes |0^n\rangle = (V \otimes I)^\dagger \, X_{2n} \, \mathrm{CZ}_{2n-1,2n} \, X_{2n} \, (V \otimes I) \, (I_n \otimes |0^n\rangle), \tag{3.277}$$

where X_{2n} is the NOT gate on the one ancilla qubit not acted by V, I is a single-qubit identity, I_n is an n-qubit identity, and $\mathrm{CZ}_{2n-1,2n}$ is controlled on the last ancilla qubit added in the recursive construction of V and acts on the one ancilla qubit not acted by V.

If one could learn U up to 1/3 error in the diamond distance with high probability or if one could distinguish whether U equals to the identity I or is 1/3-far from the identity I in the diamond distance with high probability, then one could successfully distinguish between the identity map I and the unitary U_x. Distinguishing I or one of $U_x, \forall x \in \{0,1\}^n$ is the well-known Grover search problem. Hence, from the well-known Grover lower bound [213], we have the number of queries must be at least $\Omega(2^{n/2}) = \exp(\Omega(n))$. This concludes the proof. □

3.3 Learning quantum states prepared by shallow quantum circuits

In this section we give efficient algorithms for learning quantum states prepared by shallow quantum circuits. The first algorithm presented below works for general finite dimensional lattices. The second algorithm presented in Section 3.3.4 is specialized to 2D lattices and uses no ancilla qubits for finite gate sets.

Theorem 3.10 (Restatement, simplified version of Theorem 3.13)**.** *There is an algorithm that, given copies of an unknown state $|\psi\rangle$, with the promise that $|\psi\rangle = U|0^n\rangle$ where U is an unknown depth-d circuit acting on a k-dimensional lattice (using arbitrary 2-qubit gates), outputs a depth-$(2k+1)d$ circuit W that prepares $|\psi\rangle$ up to 0.01 trace distance, with success probability 0.99. The algorithm uses M copies of $|\psi\rangle$ and runs in time T, where*

$$M = \tilde{O}(n^4) \cdot 2^{O(c)}, \quad T = \tilde{O}(n^4) \cdot 2^{O(c)} + (nkd \cdot c)^{O(d \cdot c)}. \tag{3.278}$$

Here, $c = O((3k)^{k+2}d)^k$, and W uses $r \cdot n$ ancilla qubits where $r > 0$ can be chosen to be an arbitrarily small constant.

Note that the running time is polynomial when $d = O(1)$, and quasi-polynomial when $d = \text{polylog}(n)$. In addition, the dominating term in the running time (second term in T) can be significantly improved when assuming a discrete gate set.

For quantum states prepared by shallow circuits, it is known that learning (sufficiently large) local reduced density matrices suffices to information-theoretically reconstruct the state [162, 201]. The question is whether the reconstruction can be computationally efficient. A naive approach finds small circuits for different local regions and stitch them together into a global circuit by checking local consistency, but this runs into a seemingly hard constraint satisfaction problem (in two and higher dimensions). In Section 3.3.4 we develop an efficient algorithm in 2D by showing that to learn a 2D state it suffices to solve a 1D constraint satisfaction problem which is efficient; however this approach runs into the same issue at three and higher dimensions. Here we develop new techniques that do not rely on solving any consistency problem; this enables the algorithm to work at arbitrary dimensions.

Testing quantum circuit complexity. As an application of our result, we also give an algorithm to test whether an unknown state on a k-dimensional lattice has low or high quantum circuit complexity.

Corollary 3.3 (Simplified version of Theorem 3.14)**.** *Fix some constant $L > 0$. Given copies of an unknown state $|\psi\rangle$ on a k-dimensional lattice, with the promise that one of the following two cases hold:*

- ***Case 1: Low complexity.*** *$|\psi\rangle = U|0^n\rangle$ where the depth of U is at most L;*
- ***Case 2: High complexity.*** *Any state prepared by a constant depth circuit using $O(n)$ ancilla qubits is at least 0.01-far from $|\psi\rangle$ in trace distance.*

There is an algorithm that decides which is the case, with success probability at least 0.99, with polynomial sample and time complexity.

Learning phases of matter. Quantum systems defined on finite-dimensional lattices are a central subject in condensed matter physics, where quantum states are classified into different "phases of matter" [214]. Quantum circuit complexity plays an important role in the definition of phases of matter: the "trivial" phase is typically defined as quantum states prepared by a constant (or $\text{polylog}(n)$) depth circuit acting on a k-dimensional lattice (e.g. [215, 216]), while quantum states in a topologically ordered phase have high circuit complexity [217]. Our result therefore shows that:

- Quantum states in the "trivial" phase can be learned in polynomial (or quasi-polynomial) time.

- Given an arbitrary quantum state, there is an efficient algorithm to test whether it is in the "trivial" phase or some other high-complexity phase.

Discussion. An interesting question is whether our algorithm can be generalized to other geometries (or interaction graphs) beyond finite-dimensional lattices. In fact, we show that our algorithm works for any geometry which has a property called "covering scheme" (Definition 3.23), and we construct covering schemes for finite-dimensional lattices. It is interesting to study what other geometries can have this covering scheme.

3.3.1 Learning algorithm

Notations. Our goal is to learn a quantum state $|\psi\rangle$, with the promise that $|\psi\rangle = U|0^n\rangle$ where U is an unknown depth-d circuit acting on a k-dimensional lattice. We do not assume knowledge of the circuit architecture: each layer of the circuit consists of non-overlapping 2-qubit gates, where each qubit could interact with any of its neighbors. The following concepts are used throughout the argument.

- Ball: $\mathcal{B}(A,r)$ denotes the radius-r neighborhood on the lattice for a set of vertices A (including A). For example, the dotted region in Fig. 3.3b shows a ball around region A on the 2D lattice.
- Lightcone: $\mathcal{L}(A,d)$ denotes the volume of locations that can be reached by a depth d circuit starting from region A. In particular, the support of the lightcone at top layer equals $\mathcal{B}(A,d)$. For example, the green region in Fig. 3.3a denotes the lightcone of the leftmost qubit for a circuit defined on 1D lattice.

A lightcone can be viewed as propagating or spreading causal influence in a circuit from input to output. Later we will define a dual notion of *backward* lightcone, which spreads from output to input.

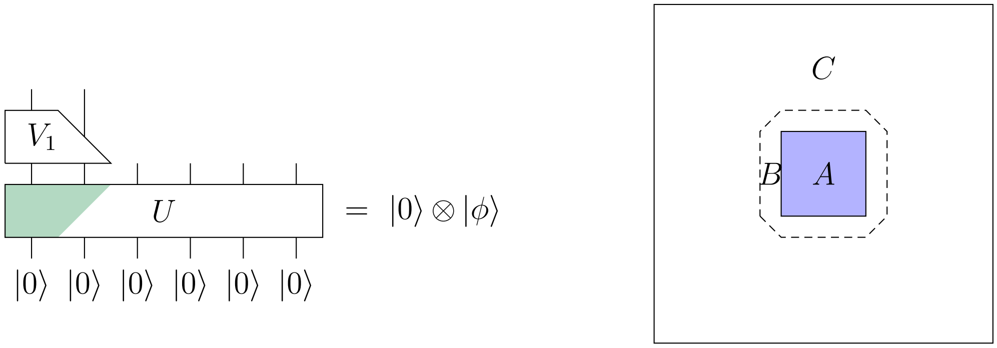
3.3.1.1 Overview of technical challenges and new ideas

We first give an overview of the technical challenges in developing a learning algorithm, and our new ideas to overcome these challenges. We start by defining a key concept of local inversion.

Definition 3.20 (Local inversion). *Given a state $|\psi\rangle$ and a subset $A \subseteq [n]$ of qubits, a unitary operator V is called a* local inversion *of A if V acts on a ball of A and $V|\psi\rangle = |0\rangle_A \otimes |\phi\rangle$ for some arbitrary pure state $|\phi\rangle$ on $n-|A|$ qubits.*

Fact 3.6. *Suppose $|\psi\rangle = U|0^n\rangle$ where U is a depth-d circuit acting on a k-dimensional lattice. Then for any $A \subseteq [n]$, there exists a local inversion V of A satisfying:*

1. *V is supported on $\mathcal{B}(A,d)$;*


V_1
U
$= |0\rangle \otimes |\phi\rangle$
$|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$
C
B
A

(a) Local inversion in 1D

(b) Local inversion in 2D

Figure 3.3: Local inversions for quantum states prepared by shallow circuits. (a) 1D lattice; (b) 2D lattice, where a local inversion of A can be constructed by applying a depth-d circuit on AB.

2. V is a depth d circuit, whose inverse shape is contained within the lightcone $\mathcal{L}(A, d)$.

This fact follows from *undoing* the gates in the lightcone of A. For example, Fig. 3.3a shows a state prepared by a circuit acting on a 1D lattice, and the green region denotes the lightcone of the leftmost qubit. There exists a local inversion V_1 (for example, by inverting the gates in the lightcone) which disentangles the leftmost qubit. Note that the shape of V_1 is the inverse of the shape of the lightcone. Fig. 3.3b shows a 2D lattice, where the dotted region AB equals $\mathcal{B}(A, d)$. There exists a local inversion of A which is a depth-d circuit acting on AB.

Learning local inversions. Local inversions provide the basic tool for a learning algorithm because they are easy to learn. For example, consider a constant-size region A in Fig. 3.3b. We first perform quantum state tomography to learn the reduced density matrix on AB, and then brute force search over all depth-d circuits acting on AB. For each circuit, we can efficiently check whether it correctly inverts the region A to $|0\rangle_A$. In this way we can find possibly a set of local inversion operators of A.

For the remainder of Section 3.3.1, we will assume that we have access to (exact) local inversion operators. In reality, we can only learn *approximate* local inversion operators. This issue is addressed in Section 3.3.2.

Our algorithm has a simple framework: (1) Quantum learning: learn local reduced density matrices; (2) Classical processing: find local inversion operators for local regions, and combine them into a circuit. Below we review the challenges in realizing this framework.

Challenges. The first issue is that there are multiple local inversion operators for a given local region. A naive approach to find a circuit is the following: pick a local inversion operator for each local region, such that local inversions for neighboring regions are *consistent*.

Consistency demands that two quantum circuits share the same gates where they overlap, so that they can be merged together as a bigger circuit. This is precisely a constraint satisfaction problem that is hard in general in 2D and higher dimensions.

The result of Section 3.3.4 addresses this problem in 2D, by showing that one can efficiently solve such a constraint satisfaction problem in 1D, to find 1D circuits that disentangle the 2D state into many 1D stripes (Fig. 3.7), and argue that the remaining 1D stripes are easy to learn. This approach fails at three and higher dimensions as the disentangling step still requires solving a hard constraint satisfaction problem.

The key challenge in solving the learning problem on finite dimensional lattices is to find a way to avoid any constraint satisfaction problem. In this section we give such a way.

Here is a basic idea: by definition, applying a local inversion V to a state $|\psi\rangle$ gives $V|\psi\rangle = |0\rangle_A \otimes |\phi\rangle$, where $|\phi\rangle$ is a *smaller* state on $n - |A|$ qubits. Can we repeat the process by further removing qubits from $|\phi\rangle$? The issue here is that now the state has been *disturbed*. In particular, the local inversion V we applied may not be the "ground truth" which undoes the gates in the lightcone of A, and V may just be an arbitray depth-d circuit that happens to invert A. Therefore the state $|\phi\rangle$ suffers from a potential quantum circuit complexity blow-up: we no longer have the guarantee that $|\phi\rangle$ is prepared by a depth-d circuit. Repeating this process will further increase the blow-up.

Key ideas. Overcoming these obstacles requires two insights. The first is to observe that we can *undo* the local inversion after applying it, so the state $|\psi\rangle$ is not disturbed. The utility of this observation is that in rewriting the state in this way, we have replaced part of the unknown state with a partial piece of known operations. The second insight is to realize that with a careful choice of local inversions based on the geometry of the lattice, these partial pieces can be layered together in a way that the *backward* lightcone of the final state is covered completely by these partial pieces and does not depend on the unknown initial state. The result is the learned constant depth circuit consisting of parts of local inversions and their inverses.

3.3.1.2 Replacement process

As discussed above, our first key idea is the following, which at first glance seems useless: apply a local inversion, and then undo it. We formally define this as a replacement process.

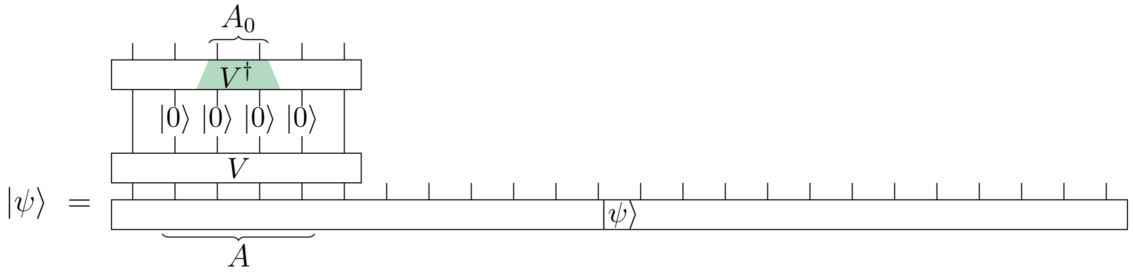
Definition 3.21 (Replacement process)**.** *Given a state $|\psi\rangle$ and a region A, define the A-*replacement process *as follows: take a local inversion V of region A, then perform the following operations on $|\psi\rangle$:*

1. *apply V,*

2. *trace out the qubits in A, replace each qubit with $|0\rangle$,*

3. *apply $V^\dagger$.*

Note that step 2 is in fact not doing anything (since step 1 already inverted the region A to $|0\rangle_A$) and is included here to help the illustration of the argument. Next, since step 2 is effectively the identity operation, step 1 and step 3 cancel with each other, and therefore the state $|\psi\rangle$ remains unchanged. This is illustrated as follows (for simplicity, here we draw local inversions as boxes without the wedges).



(3.279)

Fact 3.7. *The state $|\psi\rangle$ is invariant under any A-replacement process for any region A.*

Next we give an intuitive explanation of why the replacement process may be helpful for learning. First, we informally introduce the new concept of backward lightcone (green region in Eq. (3.279)). The formal definition is given in Definition 3.25.

Definition 3.22 (Backward lightcone, informal)**.** *The backward lightcone of a subset of qubits S at the output of a quantum circuit is the minimal part of the circuit diagram that determines the reduced density matrix of S.*

For example, in Eq. (3.279) the green region starts from A_0 at the output of the circuit and keeps growing backwards (which looks like the inverse of a (forward) lightcone), until it hits a region of $|0\rangle$. There is no need to grow further because the input is completely determined. Now, note that the reduced density matrix of A_0 at the output of the circuit (which equals the reduced density matrix of A_0 in $|\psi\rangle$) is determined by its backward lightcone: start from a region of $|0\rangle$ which is larger than A_0, apply the green circuit, and trace out the qubits not in A_0. All quantum gates not in the green region are irrelevant since removing them does not affect the reduced density matrix of A_0.

Here is an interesting observation about the A-replacement process: suppose we choose A to be a (sufficiently large) ball of some smaller region A_0 as in Eq. (3.279), then the backward lightcone of A_0 ends at the freshly initialized qubits in step 2 of Definition 3.21. In particular, the backward lightcone does not reach the unknown state $|\psi\rangle$, which allows us to reconstruct the reduced density matrix of $|\psi\rangle$ on A_0 by a *known* circuit. And yet, due to the invariance of $|\psi\rangle$ (Fact 3.7) we can pretend that nothing has happened to $|\psi\rangle$ and repeat this process. In other words,

> *Key observation: We have replaced part of the state with known operations, without disturbing the state.*

This observation suggests an approach to learning a circuit for $|\psi\rangle$: repeatedly apply A-replacement processes for different small regions A, and hope to have the backward lightcone of more and more output qubits be contained entirely within the replacement process, and hope that eventually this holds true for all of the output qubits. If we can manage this, it means we must have generated $|\psi\rangle$ solely from the collection of replacement processes (which are constructed by quantum circuits that are known to us) and thus we can extract from them a circuit that can generate $|\psi\rangle$, simply via the backward lightcone of all output qubits.

It is not obvious that this can work, because we need to apply replacement processes not only in parallel, but also on top of each other, since a single layer cannot cover all output qubits (e.g. Fig. 3.4a). The issue is that applying replacement processes on top of each other changes the lightcone structure: for example, the backward lightcone of A_0 may be much larger than in Eq. (3.279) if there are additional layers on top, because the backward lightcone starts from the output which is at the very top of the circuit.

In the next section we pin down the exact conditions for this approach to work.

3.3.1.3 Covering schemes and reconstruction circuits

It turns out that we can make this approach work if we can find a collection of small regions that satisfy the following conditions which we call a covering scheme. Our plan is:

1. In this section we show that a covering scheme implies a learning algorithm.

2. In Section 3.3.1.4 we show how to construct good covering schemes for k-dimensional lattices.

Definition 3.23 (Covering scheme)**.** *A (ℓ, c, d) covering scheme is a collection of subsets of qubits $S^i_j \subseteq [n]$*

$$S^1_1, S^1_2, \ldots, S^1_{m_1}, S^2_1, S^2_2, \ldots, S^2_{m_2}, \ldots, S^\ell_1, S^\ell_2, \ldots, S^\ell_{m_\ell}$$

which satisfy the following conditions.

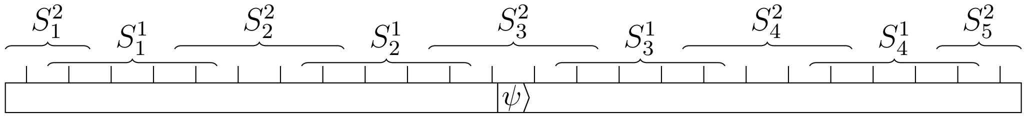
1. *The size of each $\mathcal{B}(S^i_j, d)$ is upper bounded by c.*

2. *For every fixed i, the sets $\mathcal{B}(S^i_j, d)$ are pairwise disjoint for $1 \leq j \leq m_i$.*

3. *For each qubit $v \in [n]$, there exists a S^i_j such that $\mathcal{B}(\{v\}, (2\ell - 1)d) \subseteq S^i_j$.*

We can think of these subsets as being divided into ℓ different layers: in each layer $1 \leq i \leq \ell$, there are subsets $S^i_1, S^i_2, \ldots, S^i_{m_i}$. Condition 1 says that each of them is small (even after being enlarged by a radius of d). Condition 2 says that the subsets in the same layer are disjoint, even after each of them is enlarged by a radius of d. Condition 3 says that for each qubit, a ball around that qubit (of radius $(2\ell - 1)d$) is entirely contained within some subset.

Examples in 1D. The example below shows a covering scheme in 1D with $\ell = 2$ layers.



(3.280)

Note that in each layer, the sets have some distance between each other, because we require that they remain disjoint even after being enlarged (Condition 2). In addition, the first layer and the second layer have a lot of overlap. This ensures that any qubit must lie within the interior of some set, which implies Condition 3.

Why is this useful? Recall that our idea is to repeatedly apply replacement processes. The reason to introduce covering schemes is the following:

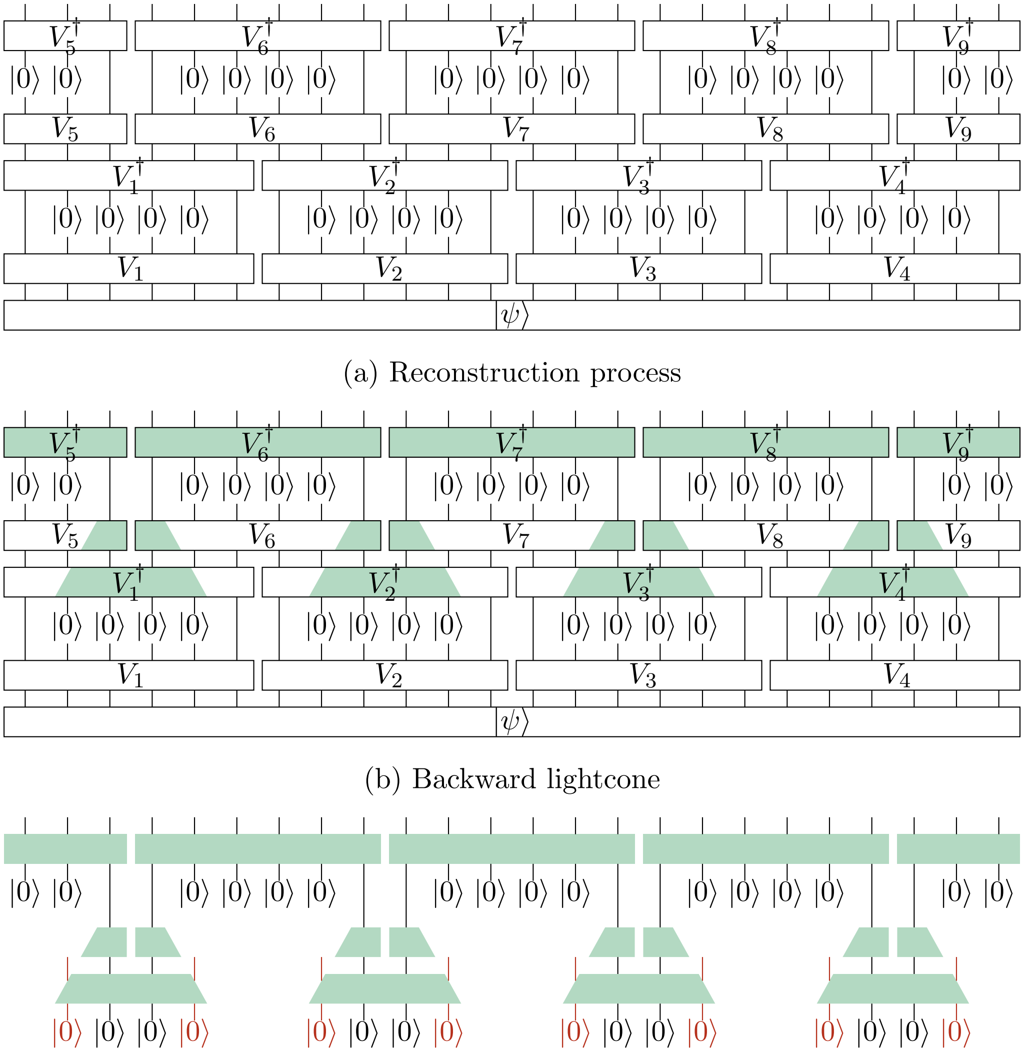
> *Key idea: Applying replacement processes for a covering scheme allows us to reconstruct the entire state.*

We first illustrate this idea in 1D in Fig. 3.4, and then give a formal argument in Section 3.3.1.3.

- **Reconstruction process.** Suppose we apply S_1^1-replacement process, which is supported on $\mathcal{B}(S_1^1, d)$ and looks exactly the same as Eq. (3.279). Note that all replacement processes corresponding to the first layer in the covering scheme can be implemented in parallel, due to Condition 2 in Definition 3.23. Next, we apply all replacement processes corresponding to the second layer in the covering scheme. We call the resulting diagram a reconstruction process, shown in Fig. 3.4a.

- **Backward lightcone.** Now we construct the backward lightcone for all output wires of the reconstruction process (Fig. 3.4b). The way to do this is to color all gates at the top layer in green, and then "spread" the green color backwards, until hitting a region of $|0\rangle$. Note that some of the spreading stops at the second layer of $|0\rangle$, but inevitably some of the spreading goes beyond the second layer and enters the first layer. However, all of the spreading stops at the first layer of $|0\rangle$ and never touches the bottom unknown state $|\psi\rangle$.

- **Reconstructed circuit.** By Fact 3.7, the state $|\psi\rangle$ remains invariant under the reconstruction process, and therefore the output state at the top layer equals $|\psi\rangle$. The backward lightcone consists entirely of known quantum circuits, and therefore it gives a reconstructed circuit that prepares $|\psi\rangle$ (Fig. 3.4c). This circuit has the following features: it has depth $3d$ and acts on the all-$|0\rangle$ input state, where we view the red qubits as ancilla qubits. After the circuit is applied, the entire state equals $|\psi\rangle \otimes |\text{junk}\rangle$, where the wires at the top correspond to $|\psi\rangle$, and the red wires (ancilla qubits) correspond to $|\text{junk}\rangle$.



(a) Reconstruction process

(b) Backward lightcone

(c) Reconstructed circuit

Figure 3.4: Illustration of the learning algorithm in 1D.

Reconstruction theorem. Next we give a rigorous argument for how to reconstruct the state in general. We start with a formal definition of the reconstruction process.

Definition 3.24 (Reconstruction process)**.** *The reconstruction process for* $|\psi\rangle$ *is defined as follows. Let* $V_{i,j}$ *be a local inversion for* S^i_j*. By Fact 3.6,* $V_{i,j}$ *is a depth-d circuit acting on* $\mathcal{B}(S^i_j, d)$*. The reconstruction process is defined as follows:*

> *For each* $1 \le i \le \ell$*, apply the* S^i_j*-replacement process using* $V_{i,j}$ *for all* j *in parallel.*

More specifically, for each fixed i we do the following:

1. *Apply $V_{i,j}$ for all $1 \leq j \leq m_i$ in parallel;*
2. *Trace out all qubits in S^i_j for each $1 \leq j \leq m_i$ and replace with $|0\rangle$;*
3. *Apply $V^\dagger_{i,j}$ for all $1 \leq j \leq m_i$ in parallel.*

As shown in the 1D example in Fig. 3.4, a two-step argument is used to show that a circuit for preparing the unknown state $|\psi\rangle$ can be extracted from the reconstruction process:

1. The state $|\psi\rangle$ is invariant under each S^i_j-replacement process (Fact 3.7), and therefore is invariant under the entire reconstruction process.
2. We can reconstruct the output state of the reconstruction process (which equals $|\psi\rangle$) via its backward lightcone, because the backward lightcone consists of known quantum circuits and in particular does not touch the unknown input state $|\psi\rangle$ at the bottom.

Below we elaborate on the second point. At this point a more precise definition of the backward lightcone is needed. Note that this definition needs to be applicable to slightly more general quantum circuits with reset gates.

Definition 3.25 (Backward lightcone)**.** *Let $|\phi\rangle = W|0^n\rangle$ where W is a quantum circuit consists of 2-qubit unitary gates and 1-qubit reset gates (a reset gate traces out the input and initializes a $|0\rangle$ state). Let $A \subseteq [n]$ be a subset of qubits. The backward lightcone of A is a circuit diagram which is part of the circuit diagram of W, defined as the collection of green gates acting on all-0 inputs, constructed as follows:*

1. *Color the output wires of W corresponding to A in blue.*
2. *Repeat the following process until no changes happen:*
   If there exists a 2-qubit gate G with a blue wire on top, color this gate in green. Moreover, consider each of the two wires at the bottom of G. If it is not connected to a reset gate, color the wire in blue.

In other words, the backward lightcone consists of all quantum gates that could influence the reduced density matrix of $|\phi\rangle\langle\phi|$ on A. The running time to construct the backward lightcone is linear in the size of W.

The process to construct the backward lightcone can be viewed as spreading the green color from top to bottom, as shown in Fig. 3.4b. The desired property for the backward lightcone, when applied to a reconstruction process, is that it stops entirely at intermediate reset gates and does not reach the bottom.

To help further illustrate this concept, we introduce a dual thought experiment. A different way of formulating our desired property is that we do not want the influence of the bottom input state to reach any top wire. The propagation of the influence can be viewed as forward lightcone spreading of the input state, that is, spreading the white color in Fig. 3.4b from bottom to top. Note that indeed, the white color stops entirely at intermediate reset gates and does not reach the top.

Now we are ready to prove that the desired property is guaranteed by the covering scheme.

Theorem 3.11 (Covering scheme implies learning algorithm)**.** *Suppose $|\psi\rangle = U|0^n\rangle$ where U is a depth-d circuit acting on a k-dimensional lattice. Let (ℓ, c, d) be a covering scheme for the k-dimensional lattice, and let W be the reconstruction process for this covering scheme, which satisfies $W|\psi\rangle = |\psi\rangle$. Then the backward lightcone of all output wires in $W|\psi\rangle$ is a depth-$(2\ell-1)d$ circuit C which is part of W.*

The circuit C can be used to prepare $|\psi\rangle$ in the following sense: it acts on n qubits as well as $m = O(n)$ ancilla qubits, and

$$C|0^n\rangle \otimes |0^m\rangle = |\psi\rangle \otimes |\text{junk}\rangle . \tag{3.281}$$

Proof. To prove that the backward lightcone of all output wires in $W|\psi\rangle$ is part of W and does not reach the input state $|\psi\rangle$, it suffices to show that the backward lightcone of each individual output wire in $W|\psi\rangle$ must stop at some regions of reset gates in some replacement processes.

Keeping track of the backward lightcone can be tricky: the backward spreading process can stop at various different layers such as in Fig. 3.4b. However, here we give a simple and pessimistic argument which suffices for our purpose.

Focus on a single output wire $v \in [n]$ of $W|\psi\rangle$ and imagine its backward spreading process. Part of the spreading may stop early at some reset gates, while other parts of the spreading may continue further. Instead of working with W, suppose we consider a new quantum circuit W' where all reset gates in W is replaced by the identity gate, with one exception: according to Condition 3 of Definition 3.23, there exists a S^i_j such that $\mathcal{B}(\{v\}, (2\ell-1)d) \subseteq S^i_j$; we keep the reset gates in W' that correspond to the S^i_j-replacement process.

Clearly, the backward spreading process in W is entirely contained within the backward spreading process in W', since W' removed some reset gates relative to W which can only help the spreading. Now, observe the following: *the backward spreading process in W' must stop at the reset gates in the S^i_j-replacement process.* This is because S^i_j contains a ball around v with radius $(2\ell-1)d$. By the time the spreading process reaches those reset gates, the process cannot spread further than a distance of $(2\ell-1)d$ and therefore is entirely covered by those reset gates. The number $(2\ell-1)d$ comes from a worst case estimate: suppose S^i_j is within the very bottom layer of the covering scheme, then the spreading process has gone through the top $\ell-1$ layers of depth $2d$ as well as $V^\dagger_{i,j}$ of depth d, with a total depth of $(2\ell-1)d$. This implies that the backward spreading process of v in W must be contained within W: if it cannot reach the bottom even in W', then it also cannot reach the bottom in W.

This concludes the proof that C is a (at most) depth-$(2\ell-1)d$ circuit and is part of W. Finally, note that C is a unitary quantum circuit which outputs a pure state. In addition, the reduced density matrix on the n output wires equals $|\psi\rangle\langle\psi|$. This implies that the system and ancilla must be tensor product pure states as in Eq. (3.281). □

To conclude this section we give a recap about the role of different parameters of a (ℓ, c, d) covering scheme in the learning algorithm.

- d is the promised depth of the unknown circuit that prepares $|\psi\rangle$.
- ℓ determines the depth of the learned circuit, which equals $(2\ell - 1)d$.
- c determines the running time of the learning algorithm: the algorithm needs to find local inversions acting on a region of size c, which takes time exponential in c and d.

Clearly, these parameters play an important role and are in tension with each other.

3.3.1.4 Good covering schemes

In this section we construct good covering schemes for k-dimensional lattices.

Theorem 3.12 (Lattice covering scheme)**.** *For k-dimensional lattice there exists a $(k+1, c, d)$ covering scheme for any integer $d > 0$, where*

$$c \leq \left((8k^2 + 14k + 2)d\right)^k . \tag{3.282}$$

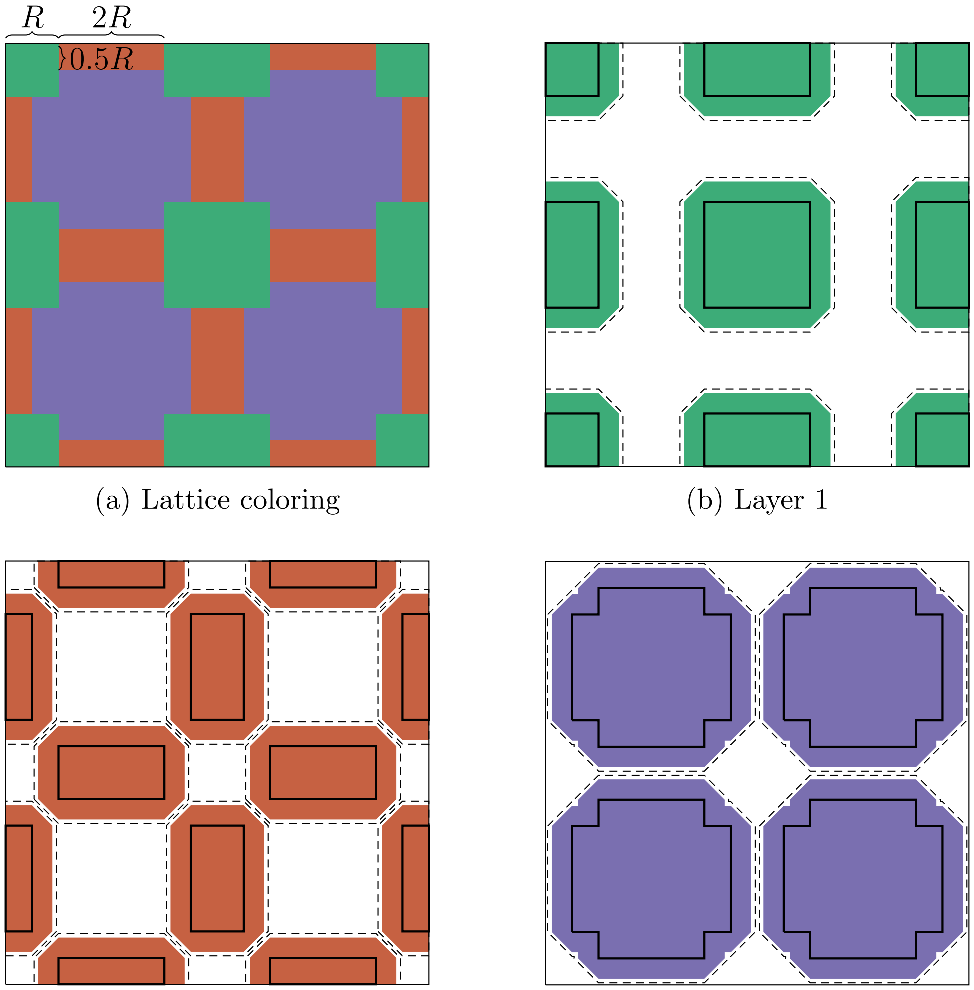
Before giving the proof, we first discuss an example in 2D (Fig. 3.5). Our starting point is a new concept called lattice coloring.

Definition 3.26 (Lattice coloring)**.** *A (w, c, R) lattice coloring for the k-dimensional lattice is defined as follows. Suppose the lattice is divided into disjoint subsets where each subset is connected. Each subset is assigned a color. We demand the following properties:*

1. *There are w different colors.*
2. *Each subset has size at most c.*
3. *Two different subsets with the same color must have distance at least R.*

Fig. 3.5a shows a $(3, 16R^2, R)$ coloring for the 2D lattice ($16R^2$ is an overestimate). Here R is a free parameter. Similar coloring schemes have been widely used in the physics literature (e.g. [212]). Now, we use this coloring to construct a covering scheme for the 2D lattice.

Lemma 3.23. *For 2D lattice there exists a $(3, 3844d^2, d)$ covering scheme for any integer $d > 0$.*



(a) Lattice coloring

(b) Layer 1

(c) Layer 2

(d) Layer 3

Figure 3.5: A covering scheme for the 2D lattice, which is derived from a lattice coloring. Here $R = 13d$.

This covering scheme is shown in Figs. 3.5b to 3.5d. Below we explain it in detail.

The idea is the following: each color in the lattice coloring corresponds to a layer in the covering scheme (so $\ell = 3$). We choose R to be large enough, and then

- Fix a color in the lattice coloring (say red regions in Fig. 3.5a), then for each small colored subset A in Fig. 3.5a, we assign a subset $S = \mathcal{B}(A, 5d)$ to the covering scheme (red regions in Fig. 3.5c).

Now, we choose R to make sure that Condition 2 in Definition 3.23 is satisfied. Say we consider two red regions A and B in Fig. 3.5a that are separated by distance R. The corresponding subsets in the covering scheme are $S_1 = \mathcal{B}(A, 5d)$ and $S_2 = \mathcal{B}(A, 5d)$. Condition 2 in Definition 3.23 demands that S_1 and S_2 are disjoint even after being enlarged further by distance d (dashed regions in Fig. 3.5c). That is, $\mathcal{B}(A, 6d)$ and $\mathcal{B}(B, 6d)$ must be disjoint. We therefore choose $R = 13d$ to ensure this property.

Next, we show that Condition 3 in Definition 3.23 is satisfied. Take any qubit $v \in [n]$, then it must lie within *some* colored region in Fig. 3.5a. Suppose the region is red and call it A. Then $\mathcal{B}(\{v\}, 5d) \subseteq \mathcal{B}(A, 5d)$ which is one of the subsets in the covering scheme in Fig. 3.5c.

Finally, note that each small colored region in Fig. 3.5a is contained with in a box of $4R \times 4R$. The length scale of each subset in the covering scheme is at most $4R + 2 \times 5d = 62d$. Therefore, all subsets in the 2D covering scheme has size at most $62d \times 62d = 3844d^2$.

Proof of Theorem 3.12. The proof essentially repeats the above example. We first quote a coloring scheme for the k-dimensional lattice (see Definition 3.17).

Lemma 3.24 (k-dimensional lattice coloring)**.** *For k-dimensional lattice there exists a*

$$\big(k+1, (2kR)^k, R\big)$$

lattice coloring. Each small colored region is contained in a k-dimensional box of length scale $2kR$.

Each color in the lattice coloring corresponds to a layer in the covering scheme (so $\ell = k + 1$). We choose R to be large enough, and then for each small colored subset A in the lattice coloring, we assign a subset $S = \mathcal{B}(A, (2k+1)d)$ to the corresponding layer of the covering scheme.

- The depth of learned circuit is $(2\ell - 1)d = (2k + 1)d$.
- To ensure Condition 2 in Definition 3.23 is satisfied, we choose $R = 2 \times (2k+2)d + d = (4k + 5)d$.
- The length scale of each subset in the covering scheme is at most $2kR + 2 \times (2k+1)d = (8k^2 + 14k + 2)d$. Therefore, c is at most $\left((8k^2 + 14k + 2)d\right)^k$.

□

Remark 3.8. *In the proof of Theorem 3.12 we have chosen $R = (4k + 5)d$ which leads to the stated scaling of c. Note that choosing R to be any number larger than that also gives a valid covering scheme, with a larger c. This is useful for the discussions in Section 3.3.2.3.*

3.3.2 Detailed analysis

In this section we give a detailed analysis which leads to the following main result.

Theorem 3.13 (Main result)**.** *There is an algorithm that, given copies of an unknown state $|\psi\rangle$, with the promise that $|\psi\rangle = U|0^n\rangle$ where U is an unknown depth-d circuit acting on a k-dimensional lattice (using arbitrary 2-qubit gates), outputs a depth-$(2k+1)d$ circuit W that prepares $|\psi\rangle$ up to ε trace distance, with success probability $1-\delta$. The algorithm uses M copies of $|\psi\rangle$ and runs in time T, where*

$$M = \frac{n^4 \cdot 2^{O(c)}}{\varepsilon^4}\log\frac{n}{\delta}, \quad T = \frac{n^4 \cdot 2^{O(c)}}{\varepsilon^4}\log\frac{n}{\delta} + \left(\frac{nkd\cdot c}{\varepsilon}\right)^{O(d\cdot c)}. \tag{3.283}$$

Here, $c = O((3k)^{k+2}d)^k$, and W uses $r\cdot n$ ancilla qubits where $r>0$ can be chosen to be an arbitrarily small constant.

3.3.2.1 The reconstruction process is robust

We first show that the final error is controlled, if we replace local inversions in the reconstruction process with approximate local inversions.

Definition 3.27 (ε-approximate local inversion)**.** *Let V be a unitary acting on AB, and denote the remaining system as C (Fig. 3.3b). Let $|\psi\rangle$ be a state defined on ABC. V is an ε-approximate local inversion of the region A for $|\psi\rangle$ if*

$$\langle 0|_A \operatorname{Tr}_{BC}\left(V|\psi\rangle\langle\psi|V^\dagger\right)|0\rangle_A \geq 1-\varepsilon. \tag{3.284}$$

Lemma 3.25. *Let V be an ε-approximate local inversion of the region A for $|\psi\rangle$. The corresponding A-replacement process is given by*

$$\mathcal{V}^\dagger \circ \mathcal{R}_A \circ \mathcal{V}, \tag{3.285}$$

where calligraphic letters denote channels, and $\mathcal{R}_A$ is the reset channel acting on A which traces out the input and prepares $|0\rangle\langle 0|_A$. Then we have

$$\left\|\mathcal{V}^\dagger \circ \mathcal{R}_A \circ \mathcal{V}(|\psi\rangle\langle\psi|) - |\psi\rangle\langle\psi|\right\|_1 \leq 4\sqrt{\varepsilon}. \tag{3.286}$$

Proof. First, note that due to the unitary invariance of the trace distance,

$$\begin{aligned}\left\|\mathcal{V}^\dagger \circ \mathcal{R}_A \circ \mathcal{V}(|\psi\rangle\langle\psi|) - |\psi\rangle\langle\psi|\right\|_1 &= \left\|\mathcal{V}^\dagger \circ \mathcal{R}_A \circ \mathcal{V}(|\psi\rangle\langle\psi|) - \mathcal{V}^\dagger \circ \mathcal{V}(|\psi\rangle\langle\psi|)\right\|_1 \\ &= \left\|\mathcal{R}_A \circ \mathcal{V}(|\psi\rangle\langle\psi|) - \mathcal{V}(|\psi\rangle\langle\psi|)\right\|_1.\end{aligned} \tag{3.287}$$

Next, we can write $V|\psi\rangle$ as

$$V|\psi\rangle = \sqrt{1-\varepsilon'}\,|0\rangle_A|\phi\rangle_{BC} + \sqrt{\varepsilon'}\,|\text{else}\rangle_{ABC}, \tag{3.288}$$

where $\varepsilon' \leq \varepsilon$ and $\langle 0|_A|\text{else}\rangle_{ABC} = 0$. Using the relationship between fidelity and trace distance,

$$\left\|\mathcal{V}(|\psi\rangle\langle\psi|) - |0\rangle\langle 0|_A \otimes |\phi\rangle\langle\phi|_{BC}\right\|_1 = 2\sqrt{1-\left|\langle\psi|V^\dagger|0\rangle_A|\phi\rangle_{BC}\right|^2} = 2\sqrt{\varepsilon'}. \tag{3.289}$$

Finally,

$$\begin{aligned}
&\|\mathcal{R}_A \circ \mathcal{V}(|\psi\rangle\langle\psi|) - \mathcal{V}(|\psi\rangle\langle\psi|)\|_1 \\
&\leq \|\mathcal{R}_A \circ \mathcal{V}(|\psi\rangle\langle\psi|) - |0\rangle\langle 0|_A \otimes |\phi\rangle\langle\phi|_{BC}\|_1 + \||0\rangle\langle 0|_A \otimes |\phi\rangle\langle\phi|_{BC} - \mathcal{V}(|\psi\rangle\langle\psi|)\|_1 \\
&= \|\mathcal{R}_A \circ \mathcal{V}(|\psi\rangle\langle\psi|) - \mathcal{R}_A\left(|0\rangle\langle 0|_A \otimes |\phi\rangle\langle\phi|_{BC}\right)\|_1 + \||0\rangle\langle 0|_A \otimes |\phi\rangle\langle\phi|_{BC} - \mathcal{V}(|\psi\rangle\langle\psi|)\|_1 \\
&\leq \|\mathcal{V}(|\psi\rangle\langle\psi|) - |0\rangle\langle 0|_A \otimes |\phi\rangle\langle\phi|_{BC}\|_1 + \||0\rangle\langle 0|_A \otimes |\phi\rangle\langle\phi|_{BC} - \mathcal{V}(|\psi\rangle\langle\psi|)\|_1 \\
&\leq 4\sqrt{\varepsilon}.
\end{aligned} \tag{3.290}$$

Here, the second line is by triangle inequality; the third line is because $|0\rangle\langle 0|_A \otimes |\phi\rangle\langle\phi|_{BC}$ is invariant under $\mathcal{R}_A$; the fourth line is by data-processing inequality. □

Lemma 3.26. *Let $\Phi_1, \Phi_2, \ldots, \Phi_T$ be arbitrary replacement processes using ε-approximate local inversions of $|\psi\rangle$, where each Φ_i has the form of Eq. (3.285) where V is an ε-approximate local inversion for some arbitrary region. Then*

$$\|\Phi_T \circ \cdots \circ \Phi_2 \circ \Phi_1(|\psi\rangle\langle\psi|) - |\psi\rangle\langle\psi|\|_1 \leq 4T\sqrt{\varepsilon}. \tag{3.291}$$

Proof. Note that

$$\begin{aligned}
&\|\Phi_T \circ \cdots \circ \Phi_2 \circ \Phi_1(|\psi\rangle\langle\psi|) - |\psi\rangle\langle\psi|\|_1 \\
&\leq \|\Phi_T \circ \cdots \circ \Phi_2 \circ \Phi_1(|\psi\rangle\langle\psi|) - \Phi_T(|\psi\rangle\langle\psi|)\|_1 + \|\Phi_T(|\psi\rangle\langle\psi|) - |\psi\rangle\langle\psi|\|_1 \\
&\leq \|\Phi_{T-1} \cdots \circ \Phi_2 \circ \Phi_1(|\psi\rangle\langle\psi|) - |\psi\rangle\langle\psi|\|_1 + 4\sqrt{\varepsilon},
\end{aligned} \tag{3.292}$$

where the second line is by triangle inequality, the third line is by data-processing inequality and Lemma 3.25. The claim follows from induction. □

3.3.2.2 Analysis of the learning algorithm

Next we put everything together and give a detailed analysis of the learning algorithm. The algorithm is given copies of an unknown quantum state $|\psi\rangle$, with the promise that $|\psi\rangle = U|0^n\rangle$ where U is a depth-d circuit acting on a k-dimensional lattice.

Step 1: build a covering scheme. We build a $(k+1, c, d)$ covering scheme (here $c = O(k^2 d)^k$ or larger, depending on the choice of R, see Remark 3.8) for the k-dimensional lattice according to Theorem 3.12. This covering scheme is divided into $\ell = k+1$ layers. Note that there are less than n subsets in total.

Step 2: learn reduced density matrices. To find local inversions for the covering scheme it suffices to learn reduced density matrices on a radius-d ball around each subset (dashed regions in Figs. 3.5b to 3.5d). Each reduced density matrix has size which has the same scaling as c and there are less than n of them.

Suppose we would like to learn all of them within ε_1 trace distance, with δ failure probability. Here we use a simple existing result (see Lemma 3.27): for each copy of $|\psi\rangle$, we measure each qubit in a random Pauli basis. The collection of measurement outcomes is known as classical shadows or randomized measurement dataset [103, 195]. The desired result can be achieved by processing this dataset, which uses

$$M = \frac{2^{O(c)}}{\varepsilon_1^2} \log \frac{n}{\delta} \tag{3.293}$$

copies of $|\psi\rangle$. The running time to process this dataset scales similarly. From now on everything is classical processing on the learned reduced density matrices, and we assume that all learned reduced density matrices are within ε_1 error (this happens with probability at least $1-\delta$).

Step 3: find approximate local inversions. For each reduced density matrix, suppose it looks like the region AB in Fig. 3.3b. We will brute force search over an ε_1-net on depth-d circuits acting on AB to find a depth-d circuit that approximately inverts the region A. In this way we can find a $3\varepsilon_1$-approximate local inversion of A for $|\psi\rangle$ (see Proof of first claim of Theorem 3.15 for a proof). The running time scales as the size of the ε_1-net, which equals

$$\left(\frac{kd\cdot c}{\varepsilon_1}\right)^{O(d\cdot c)}. \tag{3.294}$$

See Lemma 3.19 for a proof.

Note that this can be significantly improved if we assume a discrete gate set, which we do not discuss here.

Step 4: find approximate reconstruction circuit. We have found a $3\varepsilon_1$-approximate local inversion for every subset in the covering scheme. Construct a reconstruction process (Definition 3.24) using these approximate local inversions, denote as Φ. By Lemma 3.26, we have

$$\|\Phi(|\psi\rangle\langle\psi|) - |\psi\rangle\langle\psi|\|_1 \leq 4n\sqrt{3\varepsilon_1}. \tag{3.295}$$

By Theorem 3.11, the covering scheme guarantees that the backward lightcone C of the output of Φ is contained entirely within Φ (in fact, the output state of Φ is independent of its input state). C is a depth-$(2k+1)d$ circuit. Our final learned state $\hat{\rho}$ is thus equal to $\Phi(|\psi\rangle\langle\psi|)$, which can be prepared by running C on all-$|0\rangle$ input and trace out those qubits that do not belong to the output of Φ. To achieve ε-closeness in trace distance, it suffices to choose $\varepsilon_1 = \frac{\varepsilon^2}{48n^2}$.

This concludes the proof of the complexity statements in the main result.

3.3.2.3 Optimizing the number of ancilla qubits

Here we explicitly calculate (a rough upper bound of) the number of ancilla qubits needed for reconstructing $|\psi\rangle$, which corresponds to the red qubits in Fig. 3.4c. In Figs. 3.5b to 3.5d, the ancilla qubits live in the colored regions outside solid black boxes.

We start from a lattice coloring (e.g. Fig. 3.5a). Note that each small colored region is contained in a k-dimensional box of length scale $2kR$; meanwhile each small colored region at least contains a k-dimensional box of length scale R (see Definition 3.17). The number of colored regions is at most n/R^k.

Next we upper bound the number of ancilla qubits that can be associated with a colored region. The length scale of a subset in a covering scheme is at most $2kR+2\times(2k+1)d \leq 3kR$, since we will choose R to be at least $(4k+5)d$. The ancilla qubits live at $2k$ different $k-1$ dimensional surfaces with thickness $(2k+1)d \leq 3kd$; the total volume is at most $(2k)\times(3kR)^{k-1}\times(3kd)$. Overall, the total number of ancilla qubits in the entire circuit is at most

$$\frac{n}{R^k}\times(2k)\times(3kR)^{k-1}\times(3kd) \leq \frac{n}{R}\times(3k)^{k+1}d. \tag{3.296}$$

Choosing $R = L\cdot(3k)^{k+1}d$ for some sufficiently large constant L, the total number of ancilla qubits is an arbitrarily small constant times n. In fact we could also afford to choose $L = \omega(1)$ to be some small function (e.g. $\log\log\log n$) so that the total number of ancilla qubits is sublinear. Below we just consider L as being a constant.

The complexity scales with c as shown in Section 3.3.2.2, which is given by

$$c \leq (3kR)^k = O((3k)^{k+2}d)^k. \tag{3.297}$$

3.3.3 Testing quantum circuit complexity

The following result is a stronger version than stated in Corollary 3.3.

Theorem 3.14. *Given copies of an unknown state $|\psi\rangle$ on a k-dimensional lattice, with the promise that one of the following two cases hold:*

- ***Case 1: Low complexity.*** *$|\psi\rangle = U|0^n\rangle$ where the depth of U is at most d;*
- ***Case 2: High complexity.*** *Let ρ be any n-qubit state prepared by a depth at most $(2k+1)d$ circuit using $r\cdot n$ ancilla qubits, where r is some small constant. Then $\|\rho - |\psi\rangle\langle\psi|\|_1 > \varepsilon$.*

There is an algorithm that decides which is the case, with success probability at least $1-\delta$, where the sample complexity and running time is the same as in Theorem 3.13.

Proof. We perform Step 1 and 2 as prescribed by Section 3.3.2.2 (note that $\varepsilon_1 = \frac{\varepsilon^2}{48n^2}$). Then, we declare "Case 1: Low complexity" if the search procedure to find all desired approximate local inversions in Step 3 all succeed. Otherwise, declare "Case 2: High complexity".

We assume that the reduced density matrices in Step 2 are all within ε_1 error, which happens with probability at least $1-\delta$. Conditioned on this event, the algorithm is always correct, because of the following reasoning.

- In Case 1, all search procedures to find the desired approximate local inversions are guaranteed to succeed, so the algorithm must output "Case 1".

- In Case 2, suppose by contradiction that all search procedures succeed. Then, by the arguments of the main result, this actually gives a *proof* that $|\psi\rangle$ can be prepared within ε error, using a quantum circuit of depth $(2k+1)d$ and a small amount of ancilla qubits, which contradicts the assumption of Case 2. Therefore, some of the search procedures must fail, and the algorithm must output "Case 2".

□

3.3.4 Alternative algorithm for learning quantum states prepared by shallow circuits in 2D

Given copies of an unknown quantum state $|\psi\rangle = U|0^n\rangle$, with the promise that U is a depth-d circuit acting on a 2-dimensional lattice. In this section, we present an algorithm to learn a description of a shallow circuit that prepares $|\psi\rangle$ up to a desired precision. The algorithm can be viewed as first collecting a sufficiently large randomized measurement dataset [103, 195] from the unknown state and then classically reconstructing the circuit based on the dataset.

Definition 3.28 (Randomized measurement dataset for an unknown state)**.** *The learning algorithm accesses the unknown state via a randomized measurement dataset of the following form,*

$$\mathcal{T}_{|\psi\rangle}(N) = \left\{ |\phi_\ell\rangle = \bigotimes_{i=1}^{n} |\phi_{\ell,i}\rangle \right\}_{\ell=1}^{N}. \tag{3.298}$$

A randomized measurement dataset of size N is constructed by obtaining N samples from the unknown state $|\psi\rangle$. One sample is obtained from one experiment given as follows: measure every qubit of $|\psi\rangle$ under a random Pauli basis. The measurement collapses the state $|\psi\rangle$ to a state $|\phi_\ell\rangle = \bigotimes_{i=1}^{n} |\phi_{\ell,i}\rangle$, where $|\phi_{\ell,i}\rangle$ is a single-qubit stabilizer state in stab_1.

Together, N copies of $|\psi\rangle$ construct a dataset $\mathcal{T}_{|\psi\rangle}(N)$ with N samples. The dataset can be represented efficiently on a classical computer with $\mathcal{O}(Nn)$ bits.

Theorem 3.15 (Learning quantum states generated by shallow circuits in 2D)**.** *Given copies of an unknown state $|\psi\rangle$, with the promise that $|\psi\rangle = U|0^n\rangle$ for an unknown n-qubit circuit U with circuit depth d acting on a 2-dimensional lattice, then the following holds.*

1. *Suppose each two-qubit gate in U is chosen from* $\mathrm{SU}(4)$. *With a randomized measurement dataset $\mathcal{T}_{|\psi\rangle}(N)$ of size*

$$N = \frac{2^{\mathcal{O}(d^2)} n^{50}}{\varepsilon^{64}} \log \frac{n}{\delta}, \tag{3.299}$$

we can learn a quantum circuit V with depth $3d$ acting on $n+m$ qubits on an extended 2-dimensional lattice, such that

$$\frac{1}{2}\left\| \mathrm{Tr}_B\left(V |0^n\rangle\!\langle 0^n|_A \otimes |0^m\rangle\!\langle 0^m|_B V^\dagger \right) - |\psi\rangle\!\langle\psi| \right\|_1 \leq \varepsilon, \tag{3.300}$$

with probability at least $1-\delta$. The computational time to learn V is $\left(\frac{nd^3}{\varepsilon}\right)^{\mathcal{O}(d^3)}$. The number of ancilla qubits can be chosen as $m = tn$ for an arbitrarily small constant $t > 0$.

2. *In addition, if each two-qubit gate in U is chosen from a finite gateset of constant size and $d = \mathcal{O}(1)$, then there is an algorithm that learns an exact preparation circuit V with depth $3d$ acting on $n+m$ qubits, such that $V\left|0^n\right\rangle_A\left|0^m\right\rangle_B = \left|\psi\right\rangle_A \otimes \left|\mathrm{junk}\right\rangle_B$ with probability $1-\delta$, with sample complexity $N = \mathcal{O}(\log(n/\delta))$ and time complexity $\mathcal{O}(n\log(n/\delta))$. The number of ancilla qubits can be chosen as $m = tn$ for an arbitrarily small constant $t > 0$.*

3. *In addition, if each two-qubit gate in U is chosen from a finite gateset of constant size and $d = \mathcal{O}(1)$, then there is an algorithm that learns a circuit V with depth $2^{c\cdot d^2}$ (for some universal constant c) acting on n qubits (without using any ancilla), such that $\left|\left\langle 0^n\right| V^\dagger \left|\psi\right\rangle\right|^2 \geq 1-\varepsilon$ with probability $1-\delta$, with query complexity $N = \mathcal{O}(\log(n/\delta))$ and time complexity $(n/\varepsilon)^{\mathcal{O}(1)}$.*

Remark 3.9. *The first claim in Theorem 3.15 holds for any gateset and any circuit depth d (which may not be a constant), while the second and third claims are specialized to the simpler setting of finite gateset and constant depth.*

In particular, the first claim implies that when $d = \mathrm{polylog}(n)$, the state $\left|\psi\right\rangle$ can be learned within ε trace distance with sample complexity $N = \frac{2^{\mathrm{polylog}(n)}}{\varepsilon^{\mathcal{O}(1)}}\log\frac{n}{\delta}$, in time $(n/\varepsilon)^{\mathrm{polylog}(n)}$.

We prove Theorem 3.15 in the remainder of this section. We start by assuming a finite gate set, and address general SU(4) gates in Section 3.3.4.4.

3.3.4.1 Learning 1D states by solving a constraint satisfaction problem

We start by assuming U is a depth-d circuit acting on a 1D lattice, for some constant $d = \mathcal{O}(1)$. The learning problem is equivalent to finding a low-depth circuit V such that $V\left|\psi\right\rangle = \left|0^n\right\rangle$. Consider Fig. 3.6 where A, B, C are contiguous regions of size $3d$. Suppose we want to locally invert the qubits in region A back to $\left|0\right\rangle_A$. We can do so by undoing the gates within the lightcone of A, i.e. apply a depth-d circuit of the blue shape (that acts on $4d$ qubits) on top of $\left|\psi\right\rangle$. As we do not know what is the correct circuit to apply, we enumerate over all possible circuits of the blue shape (we can do it because its size is small). There are $2^{\mathcal{O}(d^2)}$ such circuits in total, and for each circuit we apply it to $\left|\psi\right\rangle$ and test if the state on A actually equals to $\left|0\right\rangle_A$ (we can do it by measuring many copies, and seeing the outcome all-0 with high probability). For now we assume that all local inversion circuits can be found exactly; this is addressed in more detail later.

At the end of this procedure, we end up with a list of candidate circuits $\mathcal{C}_A$ of the blue shape, such that each of them is a valid local inversion of A, i.e., for all $V_A \in \mathcal{C}_A$ we have $V_A\left|\psi\right\rangle = \left|0\right\rangle_A \otimes \left|\psi'\right\rangle$. The inverse of the lightcone of A in the unknown circuit U is among

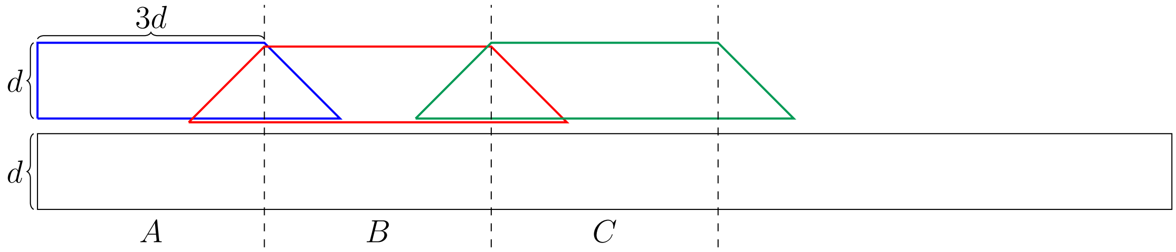


Figure 3.6: Efficient learning of quantum states generated by a shallow circuit in 1D. For each local region $A, B, C, \ldots$ we find a list of local inversion circuits, and merge them together by solving a constraint satisfaction problem.

them, but we don't know which one. We repeat the same procedure for each region A, B, C, ... and get a list of candidate local inversions $\mathcal{C}_A$, $\mathcal{C}_B$, $\mathcal{C}_C$, ... for each region.

Note that in this construction shown in Fig. 3.6, only the local inversions acting on neighboring regions could overlap. For example, the blue and green circuit does not overlap because A and C are separated by distance $3d$, and each circuit could "spread" into region B for distance at most d.

The next observation is that there are certain blue circuits in $\mathcal{C}_A$ that share the same overlapping region with certain red circuits in $\mathcal{C}_B$, i.e. they share the same gates in the overlapping triangle of blue and red. For example, the inverse of the lightcone of A in U and the inverse of the lightcone of B in U share the same overlap. We call such circuits "consistent" with each other. Note that if two circuits are consistent, they can be merged into a bigger one. For example, take a blue circuit and a red circuit that are consistent, then they can be merged by considering the union of the gates, and applying the merged circuit to $|\psi\rangle$ will simultaneously invert both regions A and B. If we can find a local inversion for each region such that all nearest neighbors are consistent, then they can be merged into a depth-d circuit V that satisfies $V|\psi\rangle = |0^n\rangle$.

Now the task can be viewed as a constraint satisfaction problem: for each region, find a local inversion circuit among all candidate local inversions (there are at most $2^{\mathcal{O}(d^2)}$ choices), such that each pair of nearest neighbor circuits are consistent. This can be solved efficiently by a simple dynamic programming algorithm in time $n \cdot 2^{\mathcal{O}(d^2)}$.

To be more specific, suppose the system is divided into $L = \frac{n}{3d}$ regions of size $3d$ as in Fig. 3.6, and suppose we have found at most $M = 2^{\mathcal{O}(d^2)}$ local inversions for each region. These circuits are stored in an array C, where $C[i][j]$ denotes the jth local inversion circuit for the ith region. Define an arrays $cost$, where $cost[i][j] = 0$ if there exists a consistent assignment at locations $1, 2, \ldots, i$ where $C[i][j]$ is used at location i; and $cost[i][j] \geq 1$ otherwise (let $cost[0][j] = 0$ for all j). Also define an array $prev$, where $prev[i][j]$ is an index k, such that there exists a consistent assignment at locations $1, 2, \ldots, i$ where $C[i][j]$ is used at location i and $C[i-1][k]$ is used at location $i-1$. $prev[i][j]$ is not defined when $cost[i][j] \geq 1$.

Once these arrays are constructed, we can take any circuit j such that $\mathit{cost}[L][j] = 0$, and construct a consistent assignment by tracing back through the *prev* array. Let *temp* be an array of size M. The following pseudocode shows how to construct these arrays in time $\mathcal{O}(LM^2)$.

```
for $i = 1, 2, \ldots, L$ do
    for $j = 1, 2, \ldots, M$ do
        for $k = 1, 2, \ldots, M$ do
            $\mathit{temp}[k] = \mathit{cost}[i-1][k] + 1\,[C[i][j] \text{ is not consistent with } C[i-1][k]]$
        end for
        $\mathit{cost}[i][j] = \min_k \mathit{temp}[k]$
        if $\mathit{cost}[i][j] = 0$ then
            $\mathit{prev}[i][j] = \arg\min_k \mathit{temp}[k]$
        end if
    end for
end for
```

Finally, note that the above procedure can be implemented by a two-step process:

1. Learn reduced density matrices of $|\psi\rangle$ supported on the lightcone of each small region $A, B, C, \ldots$.

2. Find local inversions classically using the learned classical descriptions of the reduced density matrices, and then solve the constraint satisfaction problem.

This is because to find local inversions, say for the B region, we only need access to the reduced density matrix of $|\psi\rangle$ on the lightcone of B, which has $5d$ qubits, since the local inversion only acts on the reduced density matrix.

We need to learn $\frac{n}{3d}$ reduced density matrices of size at most $5d$. The following general lemma shows the complexity for learning reduced density matrices which we use throughout this section.

Lemma 3.27 (Learning reduced density matrices)**.** *Let ρ be an unknown n-qubit mixed state. Suppose we would like to learn its reduced density matrices $\rho_{A_1}, \ldots, \rho_{A_m}$ where A_i are subsystems of size at most k. Given a randomized measurement dataset $\mathcal{T}_\rho(N)$ of size $N = \frac{2^{\mathcal{O}(k)}}{\varepsilon^2}\log\frac{m}{\delta}$, we can learn a list of Hermitian matrices (not necessarily density matrices) $\{\sigma_{A_i}\}$ such that with probability at least $1-\delta$, we have $\|\rho_{A_i} - \sigma_{A_i}\|_1 \leq \varepsilon$ for all i.*

Proof. Fix some i, we can write $\rho_{A_i} = \sum_{P\in\{I,X,Y,Z\}^{|A_i|}} \alpha_P P$. It suffices to learn the Pauli coefficients $\alpha_P = \frac{1}{2^{|A_i|}}\operatorname{Tr}(\rho_{A_i}P) = \frac{1}{2^{|A_i|}}\operatorname{Tr}(\rho P)$. Suppose we have learned these coefficients (denote as $\{\beta_P\}$) to within ε_1 precision. Let $\sigma_{A_i} := \sum_{P\in\{I,X,Y,Z\}^{|A_i|}} \beta_P P$, then

$$\|\rho_{A_i} - \sigma_{A_i}\|_1^2 \leq 2^{|A_i|}\operatorname{Tr}(\rho-\sigma)^2 = 2^{2|A_i|}\sum_P(\alpha_P - \beta_P)^2 \leq 2^{4k}\varepsilon_1^2, \tag{3.301}$$

which gives $\|\rho_{A_i} - \sigma_{A_i}\|_1 \leq 2^{2k}\varepsilon_1$. Thus to achieve $\|\rho_{A_i} - \sigma_{A_i}\|_1 \leq \varepsilon$ it suffices to learn $\{\mathrm{Tr}(\rho P)\}$ within accuracy $\varepsilon/2^k$; there are at most $m \cdot 4^k$ k-local Pauli operators that we need to learn.

By the main result of [103], given a randomized measurement dataset of size

$$N = \frac{2^{\mathcal{O}(k)}}{\varepsilon^2} \log \frac{m}{\delta}, \tag{3.302}$$

with probability at least $1 - \delta$, we can learn all observables $\mathrm{Tr}(\rho P)$ for the $m \cdot 4^k$ k-local Pauli operators within accuracy $\varepsilon/2^k$; this is sufficient to obtain Hermitian matrices $\{\sigma_{A_i}\}$ that satisfy $\|\rho_{A_i} - \sigma_{A_i}\|_1 \leq \varepsilon$ for all i. □

Note that when the gates in the unknown circuit are assumed to come from a constant-size gate set, the reduced density matrices only have $2^{\mathcal{O}(d^2)} = \mathcal{O}(1)$ choices. Therefore, choosing ε to be some small constant in Lemma 3.27 suffices to learn all the reduced density matrices *exactly*. This allows us to find the exact local inversions by classically processing the reduced density matrices.

In summary, we have shown an algorithm that learns a depth-d circuit V that satisfies $|\psi\rangle = V^\dagger |0^n\rangle$ with success probability $1 - \delta$, using a randomized measurement dataset of size $N = \mathcal{O}(\log(n/\delta))$, in time $\mathcal{O}(n)$.

3.3.4.2 Disentangling a 2D state

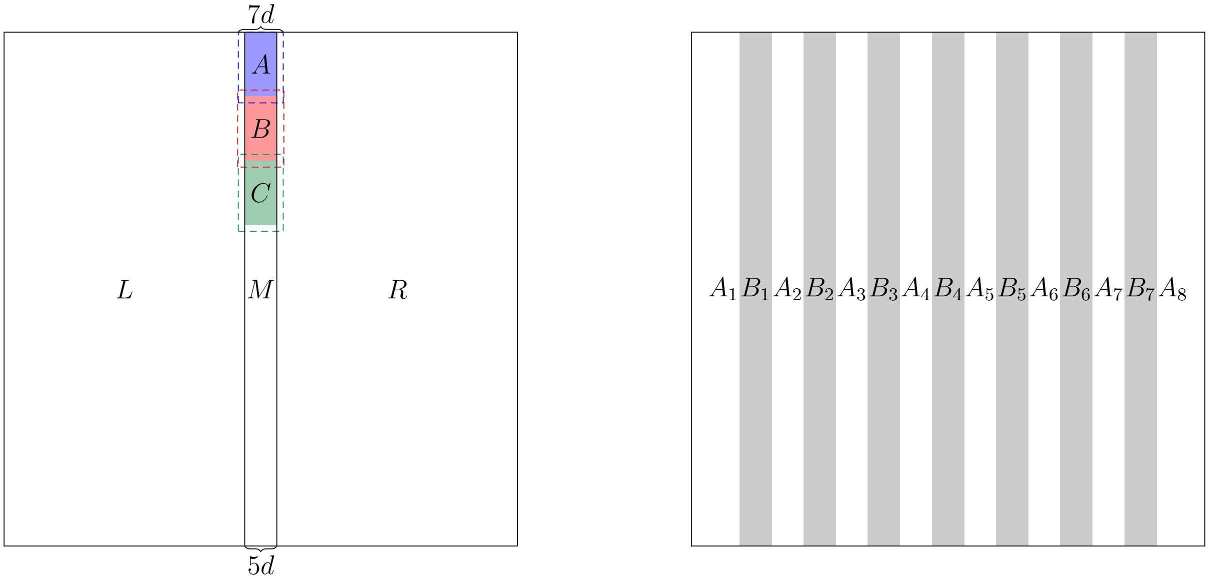
Next we use the 1D techniques developed above to disentangle a state $|\psi\rangle = U|0^n\rangle$, where U is a depth-d circuit acting on a 2D lattice, for some constant $d = \mathcal{O}(1)$.

For this purpose we need to introduce a general property for quantum states generated by low depth circuits, that is they have finite correlation length.

Lemma 3.28 (Finite correlation length)**.** *Let $|\psi\rangle$ be a state generated by a depth-d geometrically-local circuit (Definition 3.9). Let A, B be two regions that are separated by distance at least $2d$ in the connectivity graph. Then $I(A : B)_\psi = 0$. In other words, let ρ_{AB}, ρ_A, ρ_B be the reduced density matrices of $|\psi\rangle$ on AB, A and B, then $\rho_{AB} = \rho_A \otimes \rho_B$.*

Proof. As A and B are separated by distance $2d$, their lightcones $L(A)$ and $L(B)$ are disjoint. $\rho_{AB} = \rho_A \otimes \rho_B$ follows from the fact that ρ_{AB} is generated by the gates in $L(AB)$, which is a tensor product between $L(A)$ and $L(B)$. □

Fig. 3.7 (a) shows a quantum state $|\psi\rangle$ (let $\rho = |\psi\rangle\langle\psi|$) prepared by a depth-$d$ circuit on a 2D lattice, divided into three regions L, M, R. Since L and R are separated by distance $5d$, Lemma 3.28 implies that $\rho_{LR} = \rho_L \otimes \rho_R$. Although subsystems L and R are not entangled with each other, they both could be entangled with M. Therefore we develop an argument to invert the qubits in M, so that the state on L and R could become a tensor product of pure states.


7d
A
B
C
L
M
R
5d
A_1 B_1 A_2 B_2 A_3 B_3 A_4 B_4 A_5 B_5 A_6 B_6 A_7 B_7 A_8

(a) (b)

Figure 3.7: Learning to disentangle a quantum state generated by a shallow circuit in 2D. (a) The middle region M can be inverted by solving a similar 1D constraint satisfaction problem as in Fig. 3.6. (b) After inverting all the gray B_i regions, the remaining white A_i regions are disentangled into a tensor product of pure states.

Note that M is a 1D-like region. Our goal is to find a depth-d circuit V acting on a slightly wider strip (of width $7d$) around M, such that $V\left|\psi\right\rangle = \left|0\right\rangle_M \otimes \left|\psi'\right\rangle$. Such a circuit exists since we can undo the lightcone of M, and we can find such a circuit using the same argument as in the previous section. In Fig. 3.7 (a), the blue, red and green regions play the same role as in Fig. 3.6. For example, we can find a set of local inversions $\mathcal{C}_A$ for the shaded blue region A, by first learning the reduced density matrix on the dotted blue region, and then enumerating over all depth-d circuits acting on the dotted blue region. After learning a set of local inversions for each local region, we can find a desired depth-d circuit that inverts M by solving a 1D constraint satisfaction problem.

Now, we have effectively reduced the problem of learning $\left|\psi\right\rangle$ to the following problem: given copies of a state $\left|\psi_1\right\rangle$ with the promise that

1. it is prepared by a depth-$2d$ circuit (defined on a 2D lattice) acting on $\left|0^n\right\rangle$;
2. its reduced density matrix on M equals $\left|0\right\rangle\!\left\langle 0\right|_M$.

The goal is to learn the state $\left|\psi_1\right\rangle$. Note that in this new state $\sigma = \left|\psi_1\right\rangle\!\left\langle\psi_1\right|$, even though its circuit depth has increased from d to $2d$, the reduced state on L and R is still in tensor product, i.e. $\sigma_{LR} = \sigma_L \otimes \sigma_R$, due to the fact that M (with width $5d$) is sufficiently wide.

The main purpose of inverting the M region is that now σ_L and σ_R are guaranteed to be pure states, as shown by the following.

Lemma 3.29. *Let ρ_{ABC} be a pure state such that the following two properties hold:*

1. $\rho_B = |0\rangle\langle 0|_B$,
2. $\rho_{AC} = \rho_A \otimes \rho_C$.

Then ρ_A and ρ_C are both pure states.

Proof. This is a special case of Lemma 3.33. □

Next, we apply the above argument across the entire system. In Fig. 3.7 (b), the system is divided into many vertical strips of width $5d$. By repeating the above argument, we can learn a inverting circuit V_i for each shaded B_i region. Note that each V_i acts on a width-$7d$ strip around B_i and therefore different V_is do not overlap. By combining these different inverting circuits, overall we have learned a depth-d circuit V such that $V|\psi\rangle = |0\rangle_B \otimes |\psi'\rangle$ where B denotes the union of B_i.

Finally, by repeatedly applying Lemma 3.29, we know that the reduced density matrix of $V|\psi\rangle$ on each region A_i is a pure state. This means that overall the state can be written as $V|\psi\rangle = |0\rangle_B \otimes (\otimes_i |\phi\rangle_{A_i})$ for some pure states $|\phi\rangle_{A_i}$.

Now, we have disentangled the state $|\psi\rangle$ into a tensor product of many 1D-like pure states, and the problem of learning $|\psi\rangle$ is reduced to the following problem:

Problem 1. We are given copies of a state $|\psi_2\rangle$ with the promise that

1. it is prepared by a depth-$2d$ circuit (defined on a 2D lattice) acting on $|0^n\rangle$;
2. its reduced density matrix on each of the B_i regions in Fig. 3.7 (b) equals $|0\rangle\langle 0|_{B_i}$; in particular, this implies that $|\psi_2\rangle = |0\rangle_B \otimes (\otimes_i |\phi\rangle_{A_i})$ for some pure states $|\phi\rangle_{A_i}$.

The goal is to learn the state $|\psi_2\rangle$, and it suffices to learn each of the individual states $|\phi\rangle_{A_i}$.

3.3.4.3 Learning finite correlated states in 1D

Next we show how to learn a state $|\phi\rangle$ (abbreviating the subscript A_i) on a specific region A_i that came from Problem 1. Besides the fact that $|\phi\rangle$ is a pure state, the learning algorithm heavily relies on the property that $|\phi\rangle$ is part of a larger state that is prepared by a depth-$2d$ circuit. Note that this does not imply that $|\phi\rangle$ itself can be prepared by a depth-$2d$ circuit acting on A_i. Instead, we will use this property to derive useful facts about $|\phi\rangle$, presented as two different viewpoints. Each of them leads to a learning algorithm that is similar to the approach in Section 3.3.4.1.

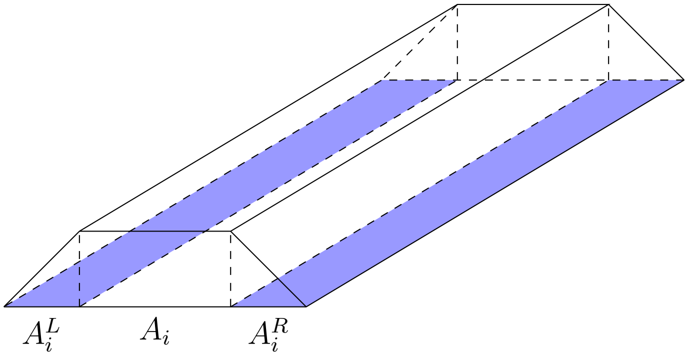


Figure 3.8: Each of the states on the white A_i regions in Fig. 3.7 (b) can be viewed as being prepared by a depth-$2d$ circuit acting on A_i (white) as well as ancilla qubits A_i^L and A_i^R (blue).

Viewpoint 1. By Lemma 3.28, the state $|\phi\rangle$ is a finite correlated state with correlation length $\ell = 4d$. That is, let $\sigma = |\phi\rangle\langle\phi|$ and let $R_1, R_2 \subseteq A_i$ be two regions that are separated by distance at least $4d$, then $\sigma_{R_1R_2} = \sigma_{R_1} \otimes \sigma_{R_2}$.

Viewpoint 2. $|\phi\rangle$ can be prepared by a depth-$2d$ circuit acting on A_i as well as some ancilla qubits A_i^L and A_i^R, shown in Fig. 3.8. To see this, recall that $|\phi\rangle$ is part of a state that is prepared by a depth-$2d$ circuit. Now, imagine that we *undo* all the gates in that circuit, except for those in the *backward lightcone* of A_i. This procedure does not affect the state on A_i, and the resulting circuit (denote as W_i) has exactly the same shape as in Fig. 3.8, where A_i^L, A_i^R both has width $2d$. Moreover, since $|\phi\rangle$ is a pure state, it is disentangled with the ancilla qubits, which means

$$W_i \left|0\right\rangle_{A_i^L} \left|0\right\rangle_{A_i} \left|0\right\rangle_{A_i^R} = \left|\text{junk}\right\rangle_{A_i^L} \otimes \left|\phi\right\rangle \otimes \left|\text{junk}'\right\rangle_{A_i^R} . \tag{3.303}$$

Clearly, Viewpoint 2 is a much stronger characterization of $|\phi\rangle$ and derives Viewpoint 1 as a corollary; however, it involves additional ancilla qubits. In the following, we show that each of these Viewpoints itself is sufficient to derive a learning algorithm; in particular,

- Using Viewpoint 1, we show that the state $|\phi\rangle$ can be prepared by a depth-$2^{\mathcal{O}(d^2)}$ circuit acting on A_i (without ancilla), therefore it can be learned using the techniques in Section 3.3.4.1.
- Using Viewpoint 2, we show how to learn a depth-$2d$ circuit W_i that prepares the state $|\phi\rangle$ using ancilla qubits, according to Eq. (3.303).

Central to both of these results is a technique that allows us to disentangle a finite correlated state in 1D. For simplicity, below we present this technique for a 1D system on a line with no width.

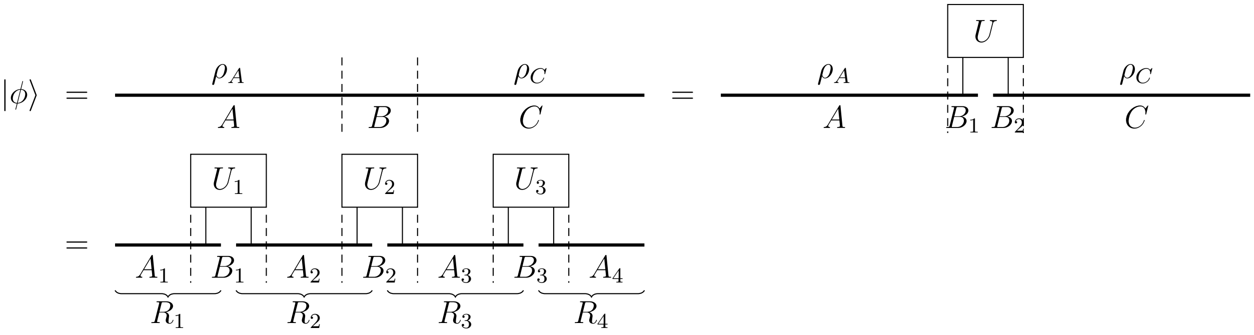


Figure 3.9: Disentangling a finite correlated state in 1D.

Lemma 3.30 (Disentangling finite correlated states in 1D)**.** *Let $|\phi\rangle$ be a state defined on a line with correlation length ℓ, that is, every two regions R_1, R_2 that are separated by distance at least ℓ have zero mutual information, i.e. $\rho_{R_1R_2} = \rho_{R_1} \otimes \rho_{R_2}$, where $\rho = |\phi\rangle\langle\phi|$. Divide the 1D line into contiguous regions of size ℓ, denote as $A_1, B_1, A_2, B_2, \ldots, B_{L-1}, A_L$ (Fig. 3.9). Then for each i there exists a unitary U_i acting on the B_i region, such that $\prod_{i=1}^{L-1} U_i |\phi\rangle$ is a tensor product of L pure states.*

Proof. We start with three subsystems A, B, C (first line of Fig. 3.9), where B has size ℓ. Then we have

$$\operatorname{rank}(\rho_B) = \operatorname{rank}(\rho_{AC}) = \operatorname{rank}(\rho_A \otimes \rho_C) = \operatorname{rank}(\rho_A) \cdot \operatorname{rank}(\rho_C) \le \dim(B). \tag{3.304}$$

Purifying the state ρ_A (ρ_C) requires an ancilla system with dimension $\operatorname{rank}(\rho_A)$ ($\operatorname{rank}(\rho_C)$). Therefore we can partition B into two systems B_1, B_2, such that there exists pure states $|\phi_1\rangle_{AB_1}$ and $|\phi_2\rangle_{B_2C}$, such that $|\phi_1\rangle_{AB_1}$ is a purification of ρ_A, and $|\phi_2\rangle_{B_2C}$ is a purification of ρ_C. This implies that $|\phi_1\rangle_{AB_1} \otimes |\phi_2\rangle_{B_2C}$ is a purification of ρ_{AC}. Since $|\phi\rangle_{ABC}$ is also a purification of ρ_{AC}, by Uhlmann's theorem there exists a unitary U_B such that $|\phi\rangle_{ABC} = U_B |\phi_1\rangle_{AB_1} \otimes |\phi_2\rangle_{B_2C}$.

Applying this argument independently at different B_i regions (bottom line of Fig. 3.9), we have that for each $i = 1, 2, \ldots, L-1$, there exists a partition of the system B_i as two systems B_i^L and B_i^R, as well as a unitary U_i acting on $B_i = B_i^L \cup B_i^R$, such that

$$|\phi\rangle = U_i |\phi_1\rangle_{A_1 \ldots B_i^L} \otimes |\phi_2\rangle_{B_i^R A_{i+1} \cdots A_L}, \tag{3.305}$$

or equivalently, $U_i^\dagger |\phi\rangle = |\phi_1\rangle_{A_1 \ldots B_i^L} \otimes |\phi_2\rangle_{B_i^R A_{i+1} \cdots A_L}$, for some pure states $|\phi_1\rangle$ and $|\phi_2\rangle$. Next, we relabel the systems according to

$$R_i := B_{i-1}^R \cup A_i \cup B_i^L. \tag{3.306}$$

Intuitively, after applying all $U_i^\dagger$s, the system must be disentangled across all the R_i regions. To prove this we use a simple argument based on the strong subadditivity of quantum entropy (Lemma 3.31).

Let $\sigma := \left(\prod_{i=1}^{L-1} U_i^\dagger\right) |\phi\rangle\langle\phi| \left(\prod_{i=1}^{L-1} U_i\right)$ be the final (pure) state. Fix some i, our goal is to prove that σ_{R_i} is pure, i.e., $S(\sigma_{R_i}) = 0$. The strong subadditivity of quantum entropy gives

$$S(\sigma_{R_i}) \leq S(\sigma_{R_1 \ldots R_i}) + S(\sigma_{R_i \ldots R_L}) - S(\sigma) = S(\sigma_{R_1 \ldots R_i}) + S(\sigma_{R_i \ldots R_L}). \tag{3.307}$$

Note that when calculating $S(\sigma_{R_1 \ldots R_i})$ we can *undo* all the unitaries $U_j^\dagger$ for $j < i$ due to the invariance of entropy under unitary. Then $S(\sigma_{R_1 \ldots R_i}) = 0$ immediately follows from Eq. (3.305), and a similar argument shows $S(\sigma_{R_i \ldots R_L}) = 0$, which concludes the proof. □

Lemma 3.31 (Strong subadditivity of quantum entropy [218])**.** *Let ρ be a mixed state defined on three systems A, B, C. Let $S(\rho) := -\operatorname{Tr}(\rho \log \rho)$ be the von Neumann entropy. Then we have*

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}). \tag{3.308}$$

Learning under Viewpoint 1. A corollary of Lemma 3.30 is that any finite correlated state in 1D can be prepared by a low-depth circuit, because each of the small pure state on the R_i regions in the bottom line of Fig. 3.9 can be prepared by a local unitary acting on $\mathcal{O}(\ell)$ qubits. Applying this argument to the state $|\phi\rangle_{A_i}$ shown in Fig. 3.8, we conclude that it can be prepared by two layers of unitaries acting on $\mathcal{O}(d^2)$ qubits, acting on the A_i region only. This implies that the state $|\phi\rangle_{A_i}$ can be prepared by a depth-$2^{\mathcal{O}(d^2)}$ circuit acting on A_i, and thus can be learned by applying the argument in Section 3.3.4.1.

Learning under Viewpoint 2. The main drawback of the above argument is that the learned circuit depth has an exponential blowup. To reduce this blowup we use additional structure of the state $|\phi\rangle_{A_i}$, described in Viewpoint 2 and Fig. 3.8. Note that there is a key difference between learning the state $|\phi\rangle_{A_i}$ and learning 1D states discussed in Section 3.3.4.1. Here, while the state $|\phi\rangle_{A_i}$ has a low-depth property shown in Fig. 3.8, this property relies on ancilla qubits (the $|\text{junk}\rangle$ states in Eq. (3.303)) that *we do not have access to.* Therefore we cannot directly apply the techniques in Section 3.3.4.1, which requires access to all qubits prepared by the low-depth circuit.

The main idea is to learn a mixed state ρ that is *locally consistent* with the state $|\phi\rangle\langle\phi|$, i.e., they have the same local reduced density matrices, and then show that this forces the two states to be globally the same.

The argument is illustrated in Fig. 3.10, where we learn to locally *prepare* the state instead of *invert* the state. Consider the state $|\phi\rangle$ on the A_i region shown in Fig. 3.10, and suppose we have learned its reduced density matrix ρ_{blue} on the solid blue region. Due to the fact that $|\phi\rangle$ is prepared by a depth-$2d$ circuit acting on A_i^L, A_i, A_i^R, we know that there exists a depth-$2d$ circuit acting on the dotted blue region that prepares ρ_{blue} (the circuit looks like a small piece of Fig. 3.8), by undoing all the gates except for those in the backward lightcone of the solid blue region. We can perform a brute force search over all depth-$2d$ circuits acting

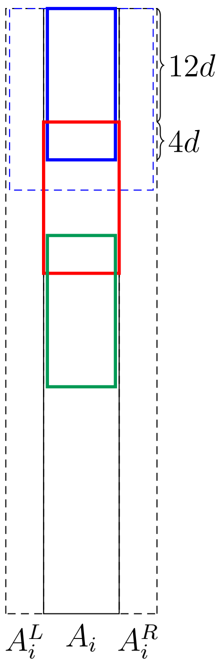


Figure 3.10: Learning a quantum state generated by a depth-$2d$ circuit with ancilla.

on the dotted blue region, and for each of them we can test whether it prepares ρ_{blue}. In this way we obtain a list of depth-$2d$ circuits acting on the dotted blue region that prepares ρ_{blue}.

By repeating the above procedure we can obtain a list of local preparation circuits for each of the solid colored regions. A key point here is that the neighboring colored regions overlap by distance $4d$. Moreover, the local preparation circuits for the blue and green regions do not overlap, since the red region is sufficiently big. This enables us to solve a constraint satisfaction problem of the same nature as in Section 3.3.4.1, where we can choose a local preparation circuit for each region, such that neighboring circuits are consistent and can be merged together. Overall we have learned a depth-$2d$ circuit W acting on A_i^L, A_i, A_i^R, that simultaneously prepares all the local reduced density matrices.

Let $\rho := \mathrm{Tr}_{A_i^L A_i^R}(W\,|0\rangle\!\langle 0|_{A_i^L A_i A_i^R}\,W^\dagger)$ be the learned density matrix on A_i. At this point we know that ρ and $|\phi\rangle\!\langle\phi|$ are locally the same on the solid blue, red, and green regions (and so on), but this does not directly imply that $\rho = |\phi\rangle\!\langle\phi|$. For example, a Haar random pure state and the maximally mixed state are locally very close but globally very far. Next, we show that the finite correlation property forces ρ and $|\phi\rangle\!\langle\phi|$ to be globally equal.

Lemma 3.32 (Local consistency implies global consistency)**.** *Let $|\psi\rangle$ be a state defined on a 1D line with correlation length ℓ and let $\sigma = |\psi\rangle\!\langle\psi|$. Suppose the system is partitioned into contiguous regions $A_1, \ldots, A_L$ where $|A_i| \geq \ell$. Suppose ρ is a mixed state that satisfies $\rho_{A_iA_{i+1}} = \sigma_{A_iA_{i+1}}$ for all i, then $\rho = \sigma$.*

Proof. We show this for 3 subsystems; generalizing to more subsystems is straightforward. Let ρ be a mixed state satisfying $\rho_{A_1A_2} = \sigma_{A_1A_2}$ and $\rho_{A_2A_3} = \sigma_{A_2A_3}$. Following the proof of Lemma 3.30, there exists a unitary U acting on A_2 such that

$$U_{A_2} |\psi\rangle_{A_1A_2A_3} = |\phi_1\rangle_{A_1A_{21}} \otimes |\phi_2\rangle_{A_{22}A_3}\,, \tag{3.309}$$

where A_{21}, A_{22} is a partition of A_2, and $|\phi_1\rangle_{A_1A_{21}}$, $|\phi_2\rangle_{A_{22}A_3}$ are some pure states. Equivalently, we have

$$U_{A_2}\sigma U_{A_2}^\dagger = |\phi_1\rangle\!\langle\phi_1|_{A_1A_{21}} \otimes |\phi_2\rangle\!\langle\phi_2|_{A_{22}A_3}\,. \tag{3.310}$$

Let $\tau := U_{A_2}\rho U_{A_2}^\dagger$, we will show that $\tau = |\phi_1\rangle\!\langle\phi_1|_{A_1A_{21}} \otimes |\phi_2\rangle\!\langle\phi_2|_{A_{22}A_3}$, which implies $\rho = \sigma$.

First, taking the partial trace over A_3 on both sides of Eq. (3.310), we have

$$U\sigma_{A_1A_2}U^\dagger = |\phi_1\rangle\!\langle\phi_1|_{A_1A_{21}} \otimes \mathrm{Tr}_{A_3} |\phi_2\rangle\!\langle\phi_2|\,. \tag{3.311}$$

Then, notice that

$$\tau_{A_1A_2} = U\rho_{A_1A_2}U^\dagger = U\sigma_{A_1A_2}U^\dagger = |\phi_1\rangle\!\langle\phi_1|_{A_1A_{21}} \otimes \mathrm{Tr}_{A_3} |\phi_2\rangle\!\langle\phi_2|\,. \tag{3.312}$$

Tracing out A_{22} on both sides, we have $\tau_{A_1A_{21}} = |\phi_1\rangle\!\langle\phi_1|_{A_1A_{21}}$; similarly, $\tau_{A_{22}A_3} = |\phi_2\rangle\!\langle\phi_2|_{A_{22}A_3}$. Since $\tau_{A_1A_{21}}$ and $\tau_{A_{22}A_3}$ are both pure states, this implies that the global state τ is a tensor product

$$\tau = \tau_{A_1A_{21}} \otimes \tau_{A_{22}A_3} = |\phi_1\rangle\!\langle\phi_1|_{A_1A_{21}} \otimes |\phi_2\rangle\!\langle\phi_2|_{A_{22}A_3}\,. \tag{3.313}$$

Thus we have $\tau = U\sigma U^\dagger$, which implies $\rho = \sigma$. □

Summary of our progress so far. So far we have developed all technical ingredients for learning a quantum state $|\psi\rangle = U|0^n\rangle$, under the simplified setting that U is a depth $d = \mathcal{O}(1)$ circuit acting on a 2D lattice, and each gate in U is from a constant size gate set.

Note that all the above arguments can be viewed as first learning the local reduced density matrices of $|\psi\rangle$ followed by classically reconstructing the circuit. As we have discussed before in Section 3.3.4.1, a reduced density matrix of constant size can be learned *exactly* as it only has a constant number of choices. In the disentangling step shown in Fig. 3.7, we can learn $\mathcal{O}(n)$ reduced density matrices on the dotted regions of size $\mathcal{O}(d^2)$, and then classically reconstruct a depth-d circuit V in time $\mathcal{O}(n)$, such that $V|\psi\rangle = |0\rangle_B \otimes (\otimes_i |\phi\rangle_{A_i})$ where the pure states $|\phi\rangle_{A_i}$ live on the white regions of Fig. 3.7 (b).

Proof of second claim of Theorem 3.15. Next, we start with Viewpoint 2. As shown in Fig. 3.10, learning a state $|\phi\rangle_{A_i}$ requires learning its reduced density matrices of size $5d \times 16d$. This can be achieved by experimentally applying V to $|\psi\rangle$ and then learning the reduced density matrices. Equivalently, say we want to learn the reduced density matrix of $|\phi\rangle_{A_i}$ on a region M of size $5d \times 16d$, then it suffices to learn a reduced density matrix of $|\psi\rangle$ of size $7d \times 18d$ on a region surrounding M, then *classically* apply the gates of V within the backward lightcone of M, and then classically trace out the qubits outside M. In other words, the reduced density matrices of $|\phi\rangle_{A_i}$ can be simulated by slightly larger reduced density matrices of $|\psi\rangle$. Using these reduced density matrices, for each i we can learn a depth-$2d$ circuit W_i such that

$$W_i\,|0\rangle_{A_i^L A_i A_i^R} = |\phi\rangle_{A_i} \otimes |\mathrm{junk}\rangle_{A_i^L A_i^R}\,, \tag{3.314}$$

which takes total time $\mathcal{O}(n)$. The entire process requires $\mathcal{O}(n)$ reduced density matrices of $|\psi\rangle$ of size $\mathcal{O}(d^2)$, which can be learned exactly with probability at least $1-\delta$, using a randomized measurement dataset of size $N = \mathcal{O}(\log(n/\delta))$.

The state $|\psi\rangle$ can be prepared as follows:

1. Initialize registers A_i, B_i, A_i^L, A_i^R in the state $|0\rangle$. Let $A = \cup_i A_i$ and $B = \cup_i B_i$.
2. For each i, apply the depth-$2d$ circuit W_i to $A_i^L A_i A_i^R$.
3. Apply the depth-d circuit $V^\dagger$ to AB, and the state $|\psi\rangle$ lives on AB.

Overall the learned circuit has depth $3d$ and can be implemented on an extended 2D lattice, where the qubits in A_i can interact with its ancilla qubits A_i^L, A_i^R as well as neighboring B_i regions.

In Fig. 3.10 we have chosen the width of A_i to be $5d$. Note that the width of A_i^L and A_i^R are both $2d$, regardless of the width of A_i. In fact we could have chosen the width of A_i to be Cd for some large constant C, and the number of ancilla qubits is at most $n/(Cd)\cdot 4d = \frac{4}{C}n$, which can be made arbitrarily small.

Proof of third claim of Theorem 3.15. Using Viewpoint 1, the state $|\phi\rangle_{A_i}$ can be prepared by a depth-$2^{\mathcal{O}(d^2)}$ circuit acting on A_i, and thus can be learned by applying the argument in Section 3.3.4.1. Let $|\phi\rangle_{A_i} = W\,|0\rangle_{A_i}$ for some depth-$2^{\mathcal{O}(d^2)}$ circuit W acting on A_i. A technical issue here is that we no longer have the guarantee that W consists of gates from a finite gate set as in U, because the existence of W comes from the disentangling argument in Lemma 3.30, instead of coming from the original circuit U as in Viewpoint 2. Below we discuss how to find this circuit W.

Let $d' = 2^{\mathcal{O}(d^2)}$ be the circuit depth of W. Following Section 3.3.4.1, we can learn reduced density matrices of $\sigma := |\phi\rangle\!\langle\phi|_{A_i}$ of size $5d \times 5d'$ (which can be done exactly, as discussed above) and then classically find local inversions for regions of size $5d\times 3d'$. Following Fig. 3.6, let A be a region of size $5d\times 3d'$, and let AA_1 be the lightcone of A with size $5d\times 4d'$. Then there is a depth-d' circuit W_{AA_1} acting on AA_1 such that

$$\mathrm{Tr}_{A_1}\left(W_{AA_1}\sigma_{AA_1}W_{AA_1}^\dagger\right) = |0\rangle\!\langle 0|_A\,. \tag{3.315}$$

To find the local inversion W_{AA_1} we use an ε_0-net over depth-d' circuits acting on AA_1, denoted as $\mathcal{N}_{\varepsilon_0}(AA_1)$ (see Definition 3.18 and Lemma 3.19), which has size at most

$$S = \left(\frac{d'^3}{\varepsilon_0}\right)^{\mathcal{O}(d'^3)}. \tag{3.316}$$

By definition, there exists $\hat{W}_{AA_1} \in \mathcal{N}_{\varepsilon_0}(AA_1)$ such that $\|\hat{W}_{AA_1} - W_{AA_1}\|_\infty \leq \varepsilon_0$, which gives

$$\langle 0_A | \mathrm{Tr}_{A_1}\left(\hat{W}_{AA_1} \sigma_{AA_1} \hat{W}^\dagger_{AA_1}\right) |0_A\rangle \geq 1 - 2\varepsilon_0. \tag{3.317}$$

By enumerating over every element in $\mathcal{N}_{\varepsilon_0}(AA_1)$, we can find a list of circuits which satisfy the above equation. Following the argument in Section 3.3.4.1, we repeat the same procedure for each local region and merge the local circuits into a global depth-d' circuit $\hat{W}_i$, which approximately inverts each local region up to $1 - 2\varepsilon_0$ fidelity. By union bound, we have

$$\left| \langle 0_{A_i} | \, \hat{W}_i \, |\phi\rangle_{A_i} \right|^2 \geq 1 - 2\sqrt{n}\varepsilon_0. \tag{3.318}$$

After learning each region A_i, the state $|\psi\rangle$ can be approximately prepared as follows:

1. Initialize registers A_i, B_i in the state $|0\rangle$. Let $A = \cup_i A_i$ and $B = \cup_i B_i$.
2. For each i, apply the depth-d' circuit $\hat{W}_i^\dagger$ to A_i.
3. Apply the depth-d circuit $V^\dagger$ to AB, and the state on AB, which is $\left|\hat{\psi}\right\rangle = V^\dagger(\otimes_i \hat{W}_i^\dagger)\,|0^n\rangle$, approximately equals to $|\psi\rangle$.

We bound the approximation error as follows.

$$\left|\left\langle \hat{\psi}|\psi \middle| \hat{\psi}|\psi \right\rangle\right|^2 = \left| \langle 0^n | \, (\otimes_i \hat{W}_i) V \, |\psi\rangle \right|^2 = \prod_i \left| \langle 0_{A_i} | \, \hat{W}_i \, |\phi\rangle_{A_i} \right|^2 \geq 1 - 2n\varepsilon_0. \tag{3.319}$$

Therefore to achieve $1 - \varepsilon$ fidelity it suffices to choose $\varepsilon_0 = \frac{\varepsilon}{2n}$, which gives total running time $n \cdot S = (n/\varepsilon)^{\mathcal{O}(1)}$.

3.3.4.4 Robustness to imprecision

In the previous sections we have been focusing on a finite gateset, which allows us to learn reduced density matrices exactly, and therefore the disentangling procedure in Fig. 3.7 can be performed exactly. However, it's not clear that this argument still works for general SU(4) gates, because in this case each step can only be performed *approximately*. In particular, we can only approximately disentangle the state using the procedure in Fig. 3.7, and learning the remaining 1D states poses new technical challenges as they are no longer pure.

In this section we address this issue. In the following we first outline the argument and develop key technical lemmas, before going into the full proof of the first claim in Theorem 3.15.

We start with the disentangling step in Fig. 3.7. Here, instead of exhaustively enumerating small circuits acting on local regions, we can only enumerate over an ε-net of the circuit. Therefore, we are only able to find circuits that approximately invert each B_i region shown in Fig. 3.7 (b). This means that after the disentangling step, the reduced density matrix on B will be *close* to $|0\rangle\langle 0|_B$, instead of being *exactly equal* to $|0\rangle\langle 0|_B$.

Now the question is what happens to the remaining A_i regions. Note that the state is still in tensor product across different A_i regions due to the finite correlation length property, but the reduced density matrices on each A_i region will not be pure. The following lemma shows that these states are *approximately pure*.

Lemma 3.33. *Let $\rho_{A_1A_2\ldots A_LB}$ be a pure state such that the following two properties hold:*

1. $\langle 0_B|\rho_B|0_B\rangle \geq 1-\varepsilon$,
2. $\rho_{A_1A_2\ldots A_L} = \rho_{A_1} \otimes \cdots \otimes \rho_{A_L}$.

Then for each $i = 1, \ldots, L$ there exists a pure state $|\phi\rangle_{A_i}$ such that $\langle \phi_{A_i}|\rho_{A_i}|\phi_{A_i}\rangle \geq 1-\varepsilon$.

Proof. Consider the operator norm $\|\rho\|_\infty := \lambda_{\max}(\rho) = \max_{|\psi\rangle} \langle\psi|\rho|\psi\rangle$. Condition 1 gives $\|\rho_B\|_\infty \geq 1-\varepsilon$. Using condition 2 we have

$$\|\rho_B\|_\infty = \|\rho_{A_1\ldots A_L}\|_\infty = \|\rho_{A_1} \otimes \cdots \otimes \rho_{A_L}\|_\infty = \prod_{i=1}^{L} \|\rho_{A_i}\|_\infty \geq 1-\varepsilon, \tag{3.320}$$

which implies that $\lambda_{\max}(\rho_{A_i}) \geq 1-\varepsilon$ for any i. □

Next, we discuss how to learn these states $\{\rho_{A_i}\}$ that are approximately pure. Again, we still have the property that each ρ_{A_i} is a 1D-like state with finite correlation length. However, our previous techniques developed in Section 3.3.4.3 only work for *exactly* pure states. We develop new techniques by examining the robustness of the key technical lemma developed in Section 3.3.4.3, Lemma 3.30.

There are two key ingredients in the proof of Lemma 3.30:

1. The use of Uhlmann's theorem to prove the existence of a local disentangling unitary;
2. The use to entropy inequalities (in particular, strong subadditivity) to prove that the state is disentangled into many local pieces after applying Uhlmann's unitaries across the entire system.

Fortunately, both ingredients are robust. First, Uhlmann's theorem says that if two mixed states are close, then there exists a unitary (acting on the purifying system) that approximately maps between their purifications. Second, entropy inequalities are robust, thanks to the continuity of entropy given below.

Lemma 3.34 (Fannes–Audenaert inequality)**.** *Let ρ, σ be two n-qubit density matrices, and let $\varepsilon := \frac{1}{2}\|\rho-\sigma\|_1$. Then*

$$|S(\rho) - S(\sigma)| \leq n\varepsilon + h(\varepsilon), \tag{3.321}$$

where $h(\cdot)$ is the binary entropy function and can be upper bounded as $h(\varepsilon) \leq 2\sqrt{\varepsilon}$.

We formalize the above intuitions as the following main technical lemma, which is a robust version of Lemma 3.32.

Lemma 3.35. *Let ρ be an n-qubit mixed state defined on systems $A_1, \ldots, A_L$, with the following properties:*

1. *there exists an n-qubit pure state $|\psi\rangle$, such that $\langle\psi|\rho|\psi\rangle \geq 1 - \varepsilon$.*
2. *for any $i = 2, 3, \ldots, L-1$, it holds that $I(A_1 \cdots A_{i-1} : A_{i+1} \cdots A_L)_\rho = 0$.*

For simplicity we assume that L is odd. Let σ be another n-qubit mixed state that satisfies

$$\frac{1}{2}\|\sigma_{A_{2i}A_{2i+1}A_{2i+2}} - \rho_{A_{2i}A_{2i+1}A_{2i+2}}\|_1 \leq \delta, \quad \forall i = 0, 1, \ldots, (L-1)/2, \tag{3.322}$$

Then

$$\frac{1}{2}\left\|\sigma - \rho\right\|_1 \leq 13n\varepsilon^{1/16} + 4n\delta^{1/4}. \tag{3.323}$$

Proof. The above condition says that ρ and σ are close on local regions

$$A_1A_2, A_2A_3A_4, A_4A_5A_6, \ldots, A_{L-1}A_L.$$

The goal is to prove that they are globally close.

Let $\tau := |\psi\rangle\langle\psi|$ denote the density matrix of $|\psi\rangle$. For any $j \in \{1, 2, \ldots, (L-1)/2\}$, define three regions $L^{(j)} := A_{\leq 2j-1}$, $M^{(j)} := A_{2j}$, $R^{(j)} := A_{\geq 2j+1}$ (the superscript (j) is abbreviated when there is no confusion).

Note that for any subsystem W, we have

$$\frac{1}{2}\|\tau_W - \rho_W\|_1 \leq \frac{1}{2}\|\tau - \rho\|_1 \leq \sqrt{1 - \langle\psi|\rho|\psi\rangle} \leq \sqrt{\varepsilon}. \tag{3.324}$$

Therefore,

$$\begin{aligned} \|\tau_{LR} - \tau_L \otimes \tau_R\|_1 &\leq \|\tau_{LR} - \rho_{LR}\|_1 + \|\rho_{LR} - \rho_L \otimes \rho_R\|_1 + \|\rho_L \otimes \rho_R - \tau_L \otimes \tau_R\|_1 \\ &\leq \|\tau_{LR} - \rho_{LR}\|_1 + \|\rho_L - \tau_L\|_1 + \|\rho_R - \tau_R\|_1 \\ &\leq \varepsilon_1 \end{aligned} \tag{3.325}$$

where we let $\varepsilon_1 := 6\sqrt{\varepsilon}$. Then, the relationship between fidelity and trace distance implies that

$$F(\tau_{LR}, \tau_L \otimes \tau_R) \geq 1 - \|\tau_{LR} - \tau_L \otimes \tau_R\|_1 \geq 1 - \varepsilon_1. \tag{3.326}$$

Let $|\phi_1\rangle_{LM_1^{(j)}}$ be a purification of τ_L, and let $|\phi_2\rangle_{M_2^{(j)}R}$ be a purification of τ_R. Note that $\dim(M_1^{(j)}) \leq \dim(L)$ and $\dim(M_2^{(j)}) \leq \dim(R)$. Let $M'^{(j)}$ be an ancilla space with dimension $\dim(M_1^{(j)})\dim(M_2^{(j)})/\dim(M^{(j)})$. Here $M'^{(j)}$ is needed in case $M^{(j)}$ is smaller than $M_1^{(j)}M_2^{(j)}$. Now, $|\psi\rangle_{LMR}|0\rangle_{M'^{(j)}}$ is a purification of the state τ_{LR}, while $|\phi_1\rangle_{LM_1^{(j)}} \otimes |\phi_2\rangle_{M_2^{(j)}R}$ is a purification of the state $\tau_L \otimes \tau_R$, and they have the same dimension. Then by Uhlmann's theorem, there exists a unitary $U^{(j)} : M^{(j)}M'^{(j)} \to M_1^{(j)}M_2^{(j)}$, such that

$$U^{(j)}_{M^{(j)}M'^{(j)}}|\psi\rangle_{LM^{(j)}R}|0\rangle_{M'^{(j)}} \approx_{\varepsilon_1} |\phi_1\rangle_{LM_1^{(j)}} \otimes |\phi_2\rangle_{M_2^{(j)}R}. \tag{3.327}$$

Here, $|u\rangle \approx_\varepsilon |v\rangle$ means $|\langle u|v|u|v\rangle|^2 \geq 1-\varepsilon$.

The above argument shows the existence of a unitary $U^{(j)}$ acting on $M^{(j)} = A_{2j}$ (as well as an ancilla system $M'^{(j)}$), that approximately disentangles the state $|\psi\rangle$ into a tensor product between $LM_1^{(j)}$ and $M_2^{(j)}R$, where $M_1^{(j)}$, $M_2^{(j)}$ are ancilla systems associated with A_{2j}. We apply all such unitaries $U^{(j)}$ ($j \in \{1,2,\dots,(L-1)/2\}$) to $|\psi\rangle$, and obtain

$$\eta := \left(\prod_{j=1}^{(L-1)/2} U^{(j)}\right) |\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|_{M'} \left(\prod_{j=1}^{(L-1)/2} U^{(j)\dagger}\right), \tag{3.328}$$

where M' represents the union of all $M'^{(j)}$. Note that η supports on $A_1, A_3, A_5, \dots, A_L$ as well as $M_1^{(j)}, M_2^{(j)}$ for $j \in \{1,2,\dots,(L-1)/2\}$. Now, we relabel the systems according to

$$B_j := M_2^{(j-1)} \cup A_{2j-1} \cup M_1^{(j)}, \quad j \in \{1,2,\dots,(L+1)/2\}, \tag{3.329}$$

and the state η supports on B_j, $j \in \{1,2,\dots,(L+1)/2\}$, and we want to prove that it is approximately a tensor product across all B_j regions via upper bounding the relative entropy

$$D(\eta|| \otimes_j \eta_{B_j}) = \sum_j S(\eta_{B_j}) - S(\eta) = \sum_j S(\eta_{B_j}). \tag{3.330}$$

By the strong subadditivity of quantum entropy,

$$S(\eta_{B_j}) \leq S(\eta_{B_{\leq j}}) + S(\eta_{B_{\geq j}}) - S(\eta) = S(\eta_{B_{\leq j}}) + S(\eta_{B_{\geq j}}). \tag{3.331}$$

Focusing on the entropy of $S(\eta_{B_{\leq j}})$, we can ignore the unitaries that are applied on regions other than A_{2j}. Note that Eq. (3.327) implies that

$$\frac{1}{2}\left\|\mathrm{Tr}_{M_2^{(j)}R}(U^{(j)}|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|_{M'^{(j)}} U^{(j)\dagger}) - |\phi_1\rangle\langle\phi_1|_{LM_1^{(j)}}\right\|_1 \leq \sqrt{\varepsilon_1}. \tag{3.332}$$

Therefore by the Fannes-Audenaert inequality,

$$S(\eta_{B_{\leq j}}) = S(\mathrm{Tr}_{M_2^{(j)}R}(U^{(j)}|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|_{M'^{(j)}} U^{(j)\dagger})) \leq 2|L|\sqrt{\varepsilon_1} + 2\varepsilon_1^{1/4} \leq 2n\sqrt{\varepsilon_1} + 2\varepsilon_1^{1/4}. \tag{3.333}$$

A similar argument holds for $S(\eta_{B_{\geq j}})$. Therefore we have

$$S(\eta_{B_j}) \leq 4n\sqrt{\varepsilon_1} + 4\varepsilon_1^{1/4}, \quad \forall j \in \{1,2,\dots,(L+1)/2\}. \tag{3.334}$$

Let

$$\omega := \left(\prod_{j=1}^{(L-1)/2} U^{(j)}\right) \sigma \otimes |0\rangle\langle 0|_{M'} \left(\prod_{j=1}^{(L-1)/2} U^{(j)\dagger}\right), \tag{3.335}$$

then $\|\sigma - |\psi\rangle\langle\psi|\|_1 = \|\omega - \eta\|_1$. Note that for any j, η_{B_j} only depends on the reduced density matrix $\tau_{A_{2j-2}A_{2j-1}A_{2j}}$; similarly, ω_{B_j} only depends on the reduced density matrix $\sigma_{A_{2j-2}A_{2j-1}A_{2j}}$. Therefore,

$$\begin{aligned} \left\|\omega_{B_j} - \eta_{B_j}\right\|_1 &\leq \left\|\sigma_{A_{2j-2}A_{2j-1}A_{2j}} - \tau_{A_{2j-2}A_{2j-1}A_{2j}}\right\|_1 \\ &\leq \left\|\sigma_{A_{2j-2}A_{2j-1}A_{2j}} - \rho_{A_{2j-2}A_{2j-1}A_{2j}}\right\|_1 + \left\|\rho_{A_{2j-2}A_{2j-1}A_{2j}} - \tau_{A_{2j-2}A_{2j-1}A_{2j}}\right\|_1 \\ &\leq 2\delta + 2\sqrt{\varepsilon}. \end{aligned} \tag{3.336}$$

Note that $|B_j| \leq 3n$, by the Fannes-Audenaert inequality,

$$S(\omega_{B_j}) \leq S(\eta_{B_j}) + 3n(\delta + \sqrt{\varepsilon}) + 2\sqrt{\delta + \sqrt{\varepsilon}}. \tag{3.337}$$

This implies that

$$\begin{aligned} D(\omega|| \otimes_j \omega_{B_j}) &= \sum_j S(\omega_{B_j}) - S(\omega) \\ &\leq \sum_j S(\omega_{B_j}) \\ &\leq \sum_j S(\eta_{B_j}) + 3n^2(\delta + \sqrt{\varepsilon}) + 2n\sqrt{\delta + \sqrt{\varepsilon}}. \end{aligned} \tag{3.338}$$

Then

$$\begin{aligned} \|\sigma - \rho\|_1 &\leq \|\sigma - \tau\|_1 + \|\tau - \rho\|_1 \\ &\leq \|\omega - \eta\|_1 + 2\sqrt{\varepsilon} \\ &\leq \left\|\omega - \otimes_j\omega_{B_j}\right\|_1 + \left\|\otimes_j\omega_{B_j} - \otimes_j\eta_{B_j}\right\|_1 + \left\|\otimes_j\eta_{B_j} - \eta\right\|_1 + 2\sqrt{\varepsilon} \\ &\leq \sqrt{2D(\omega|| \otimes_j \omega_{B_j})} + 2n\delta + 2n\sqrt{\varepsilon} + \sqrt{2D(\eta|| \otimes_j \eta_{B_j})} + 2\sqrt{\varepsilon} \\ &\leq \sqrt{8n(n\sqrt{\varepsilon_1} + \varepsilon_1^{1/4}) + 6n^2(\delta + \sqrt{\varepsilon}) + 4n\sqrt{\delta + \sqrt{\varepsilon}}} \\ &\quad + \sqrt{8n(n\sqrt{\varepsilon_1} + \varepsilon_1^{1/4})} + 2n\delta + 2(n+1)\sqrt{\varepsilon}. \end{aligned} \tag{3.339}$$

Here in the fourth line we use the quantum Pinsker inequality, which says that $\|\rho - \sigma\|_1 \leq \sqrt{2D(\rho\|\sigma)}$ for two density matrices ρ, σ. Using the fact that $\varepsilon_1 = 6\sqrt{\varepsilon}$, we have

$$\begin{aligned} \frac{1}{2}\|\sigma - \rho\|_1 &\leq n\delta + 2n\sqrt{\varepsilon} + \sqrt{8}n\varepsilon_1^{1/4} + \sqrt{8}\sqrt{n}\varepsilon_1^{1/8} + \frac{\sqrt{6}}{2}n\sqrt{\delta} + \frac{\sqrt{6}}{2}n\varepsilon^{1/4} + \sqrt{n}\delta^{1/4} + \sqrt{n}\varepsilon^{1/8} \\ &\leq \sqrt{8}n\varepsilon_1^{1/4} + \sqrt{8}\sqrt{n}\varepsilon_1^{1/8} + 5n\varepsilon^{1/8} + 4n\delta^{1/4} \\ &\leq 13n\varepsilon^{1/16} + 4n\delta^{1/4}. \end{aligned} \tag{3.340}$$

□

Finally, the next technical lemma bounds the distance between the learned state and the unknown state $|\psi\rangle$.

Lemma 3.36. *Let $|\psi\rangle_{A_1 \ldots A_L B}$ be a pure state, and let $\rho_{A_1 \ldots A_L B} = |\psi\rangle\langle\psi|_{A_1 \ldots A_L B}$. Suppose the following two properties hold:*

1. $\langle 0_B | \rho_B | 0_B \rangle = 1 - \varepsilon$,
2. $\rho_{A_1 \ldots A_L} = \rho_{A_1} \otimes \cdots \otimes \rho_{A_L}$.

Suppose $\{\sigma_{A_i}\}$ are density matrices that satisfies $\frac{1}{2}\left\|\rho_{A_i} - \sigma_{A_i}\right\|_1 \leq \delta$ for any i. Then

$$\frac{1}{2}\left\|(\otimes_{i=1}^L \sigma_{A_i}) \otimes |0\rangle\langle 0|_B - |\psi\rangle\langle\psi|\right\|_1 \leq \sqrt{2\varepsilon + L\delta}. \tag{3.341}$$

Proof. The state $|\psi\rangle_{A_1 \ldots A_L B}$ can be written as

$$|\psi\rangle_{A_1 \ldots A_L B} = \sqrt{1-\varepsilon}\, |0\rangle_B \, |\phi\rangle_{A_1 \ldots A_L} + \sqrt{\varepsilon}\, |\mathrm{else}\rangle_{A_1 \ldots A_L B}\,, \tag{3.342}$$

where $\langle 0|_B \, |\mathrm{else}\rangle_{A_1 \ldots A_L B} = 0$. This implies that

$$\rho_{A_1 \ldots A_L} = \mathrm{Tr}_B\, \rho_{A_1 \ldots A_L B} = (1-\varepsilon)\, |\phi\rangle\langle\phi|_{A_1 \ldots A_L} + \varepsilon\, \mathrm{Tr}_B\, |\mathrm{else}\rangle\langle\mathrm{else}|\,. \tag{3.343}$$

Note that

$$\begin{aligned} \frac{1}{2}\left\|\rho_{A_1 \ldots A_L} - \sigma_{A_1} \otimes \cdots \otimes \sigma_{A_L}\right\|_1 &= \frac{1}{2}\left\|\rho_{A_1} \otimes \cdots \otimes \rho_{A_L} - \sigma_{A_1} \otimes \cdots \otimes \sigma_{A_L}\right\|_1 \\ &\leq \frac{1}{2}\sum_{i=1}^{L}\left\|\rho_{A_i} - \sigma_{A_i}\right\|_1 \\ &\leq L\delta. \end{aligned} \tag{3.344}$$

Therefore,

$$\begin{aligned} \langle\psi|_{A_1 \ldots A_L B}\, \sigma_{A_1} \otimes \cdots \otimes \sigma_{A_L} \otimes |0\rangle\langle 0|_B \, |\psi\rangle_{A_1 \ldots A_L B} &\geq \langle\psi|\, \rho_{A_1 \ldots A_L} \otimes |0\rangle\langle 0|_B \, |\psi\rangle - L\delta \\ &\geq (1-\varepsilon)^2 - L\delta \\ &\geq 1 - 2\varepsilon - L\delta. \end{aligned} \tag{3.345}$$

This implies that

$$\frac{1}{2}\left\|(\otimes_{i=1}^L \sigma_{A_i}) \otimes |0\rangle\langle 0|_B - |\psi\rangle\langle\psi|\right\|_1 \leq \sqrt{2\varepsilon + L\delta}. \tag{3.346}$$

□

Proof of first claim of Theorem 3.15. Next we show how to use the above techniques to learn an unknown quantum state $|\psi\rangle = U|0^n\rangle$, with the promise that U is a depth-d circuit acting on a 2D lattice (here d is treated as a generic parameter which is not necessarily a constant) with arbitrary SU(4) gates.

We work with Viewpoint 2 described in Section 3.3.4.3. As discussed at the end of Section 3.3.4.3, the learning process requires $\mathcal{O}(n)$ reduced density matrices of $|\psi\rangle$ of size $\mathcal{O}(d^2)$. Suppose all of these reduced density matrices are learned to within ε_0 trace distance with probability $1-\delta$, then by Lemma 3.27 it suffices to take a randomized measurement dataset $\mathcal{T}_{|\psi\rangle}(N)$ of size

$$N = \frac{2^{\mathcal{O}(d^2)}}{\varepsilon_0^2} \log \frac{n}{\delta}. \tag{3.347}$$

Next we proceed with the disentangling step shown in Fig. 3.7. We have learned the reduced density matrices on the dotted regions shown in Fig. 3.7 (a) to within ε_0 trace distance. Denote the dotted blue region as AA_1 where A is the colored blue region, and let ρ_{AA_1} be the reduced density matrix of $|\psi\rangle$ on AA_1. We know that there exists a depth-$2d$ circuit V_{AA_1} such that

$$V_{AA_1}\rho_{AA_1}V_{AA_1}^\dagger = |0\rangle\langle 0|_A \otimes \sigma_{A_1} \tag{3.348}$$

for some density matrix σ_{A_1}. We have learned a density matrix $\hat{\rho}_{AA_1}$ such that $\|\hat{\rho}_{AA_1} - \rho_{AA_1}\|_1 \leq \varepsilon_0$. To find an approximate local inversion for the region A, we perform a brute force search over an ε_0-net for depth-$2d$ circuits acting on AA_1, denoted as $\mathcal{N}_{\varepsilon_0}(AA_1)$, which is constructed by discretizing each SU(4) gate (see Definition 3.18 and Lemma 3.19), which has size at most

$$S = \left(\frac{d^3}{\varepsilon_0}\right)^{\mathcal{O}(d^3)}. \tag{3.349}$$

Note that Eq. (3.348) together with $\|\hat{\rho}_{AA_1} - \rho_{AA_1}\|_1 \leq \varepsilon_0$ implies that

$$\mathrm{Tr}\Big(\langle 0|_A\, V_{AA_1}\hat{\rho}_{AA_1}V_{AA_1}^\dagger\, |0\rangle_A\Big) \geq 1-\varepsilon_0. \tag{3.350}$$

By definition of ε_0-net, there exists a unitary $\hat{V}_{AA_1} \in \mathcal{N}_{\varepsilon_0}(AA_1)$ that satisfies $\|\hat{V}_{AA_1} - V_{AA_1}\|_\infty \leq \varepsilon_0$, which gives

$$\mathrm{Tr}\Big(\langle 0|_A\, \hat{V}_{AA_1}\hat{\rho}_{AA_1}\hat{V}_{AA_1}^\dagger\, |0\rangle_A\Big) \geq 1-2\varepsilon_0. \tag{3.351}$$

The algorithm is to enumerate over all elements in $\mathcal{N}_{\varepsilon_0}(AA_1)$ and find the ones which satisfy the above equation. Each of these circuits is an approximate local inversion in the sense that

$$\mathrm{Tr}\Big(\langle 0|_A\, \hat{V}_{AA_1}\rho_{AA_1}\hat{V}_{AA_1}^\dagger\, |0\rangle_A\Big) \geq 1-3\varepsilon_0. \tag{3.352}$$

Using the same argument as in Section 3.3.4.2, in Fig. 3.7 (a) we can find a depth-d circuit $\hat{V}$ acting on the width-$7d$ strip around M, such that Eq. (3.352) is satisfied for all local colored regions. There are at most $\sqrt{n}$ such regions. Let $\rho = |\psi\rangle\langle\psi|$, by union bound,

$$\mathrm{Tr}\Big(\langle 0|_M\, \hat{V}\rho\hat{V}^\dagger\, |0\rangle_M\Big) \geq 1-3\sqrt{n}\varepsilon_0. \tag{3.353}$$

Repeat the same procedure for all vertical B_i strips shown in Fig. 3.7 (b). There are at most $\sqrt{n}$ different vertical strips. Let $B = \cup_i B_i$, and let V denote the union of all learned inversion circuits across different regions, we have

$$\mathrm{Tr}\big(\langle 0|_B V \rho V^\dagger |0\rangle_B\big) \geq 1 - 3n\varepsilon_0. \tag{3.354}$$

Now, the problem reduces to learning the state $V|\psi\rangle$, which can be formulated as follows.

Problem 2. We are given copies of a state $\sigma = |\phi\rangle\langle\phi|$ with the promise that

1. it is prepared by a depth-$2d$ circuit (defined on a 2D lattice) acting on $|0^n\rangle$;
2. its reduced density matrix on each of the B_i regions in Fig. 3.7 (b) is close $|0\rangle\langle 0|_{B_i}$, i.e. $\langle 0_B|\sigma_B|0_B\rangle \geq 1 - \varepsilon_1$.

The goal is to (approximately) learn the state $|\phi\rangle$.

Let $|\phi\rangle := V|\psi\rangle$ and let $\varepsilon_1 := 3n\varepsilon_0$. Consider dividing the state $\sigma = |\phi\rangle\langle\phi|$ into regions $A_1, A_2, \ldots, A_L$ and $B = \cup_i B_i$ as in Fig. 3.7 (b). As the regions $\{A_i\}$ are sufficiently far from each other, the reduced density matrix on $A = \cup_i A_i$ is a tensor product across each region, i.e., $\sigma_{A_1 \ldots A_L} = \sigma_{A_1} \otimes \cdots \otimes \sigma_{A_L}$. By Eq. (3.354), we have $\langle 0_B|\sigma_B|0_B\rangle \geq 1 - \varepsilon_1$. By Lemma 3.33, for each $i = 1, \ldots, L$ there exists a pure state $|\phi\rangle_{A_i}$ such that $\langle \phi_{A_i}|\sigma_{A_i}|\phi_{A_i}\rangle \geq 1 - \varepsilon_1$.

Next we discuss how to learn the state σ_{A_i} for a fixed i. This is similar to the earlier situation in Viewpoint 2, but with the critical difference that here σ_{A_i} is no longer pure. So we list the updated Viewpoint below.

Viewpoint 2'. σ_{A_i} can be prepared by a depth-$2d$ circuit acting on A_i as well as some ancilla qubits A_i^L and A_i^R, shown in Fig. 3.8. To see this, recall that σ_{A_i} is part of a state that is prepared by a depth-$2d$ circuit. Now, imagine that we *undo* all the gates in that circuit, except for those in the *backward lightcone* of A_i. This procedure does not affect the state on A_i, and the resulting circuit (denote as W_i) has exactly the same shape as in Fig. 3.8, where A_i^L, A_i^R both has width $2d$. Note that here σ_{A_i} could be entangled with the ancilla qubits, and we have

$$\mathrm{Tr}_{A_i^L A_i^R}\left(W_i |0\rangle\langle 0|_{A_i^L A_i A_i^R} W_i^\dagger\right) = \sigma_{A_i}. \tag{3.355}$$

Using the same argument as the end of Section 3.3.4.3, the reduced density matrices of σ_{A_i} can be simulated by reduced density matrices of $|\psi\rangle\langle\psi|$ on slightly larger regions. Therefore we can obtain reduced density matrices of σ_{A_i} within trace distance ε_0. Let C be the solid blue region in Fig. 3.10, and let CC_1 be the dotted blue region. We have learned a reduced density matrix $\hat{\sigma}_C$ such that $\|\hat{\sigma}_C - \sigma_C\|_1 \leq \varepsilon_0$. From Viewpoint 2', we know that there is a depth-$2d$ circuit W_{CC_1} acting on CC_1, such that

$$\mathrm{Tr}_{C_1}\left(W_{CC_1} |0\rangle\langle 0|_{CC_1} W_{CC_1}^\dagger\right) = \sigma_C. \tag{3.356}$$

Consider an ε_0-net for depth-$2d$ circuits acting on CC_1, denoted as $\mathcal{N}_{\varepsilon_0}(CC_1)$. By definition, there exists a unitary $\hat{W}_{CC_1}$ that satisfies $\|\hat{W}_{CC_1} - W_{CC_1}\|_\infty \leq \varepsilon_0$, which means that

$$\begin{aligned}&\left\|\mathrm{Tr}_{C_1}\left(\hat{W}_{CC_1}\left|0\middle\rangle\!\middle\langle 0\right|_{CC_1}\hat{W}^\dagger_{CC_1}\right) - \hat{\sigma}_C\right\|_1\\ \leq&\left\|\mathrm{Tr}_{C_1}\left(\hat{W}_{CC_1}\left|0\middle\rangle\!\middle\langle 0\right|_{CC_1}\hat{W}^\dagger_{CC_1}\right) - \sigma_C\right\|_1 + \|\sigma_C - \hat{\sigma}_C\|_1\\ \leq& 2\varepsilon_0.\end{aligned} \tag{3.357}$$

By enumerating over every element in $\mathcal{N}_{\varepsilon_0}(CC_1)$, we can find a list of circuits $\{\hat{W}'_{CC_1}\}$ that satisfy $\left\|\mathrm{Tr}_{C_1}\left(\hat{W}'_{CC_1}\left|0\middle\rangle\!\middle\langle 0\right|_{CC_1}\hat{W}'^\dagger_{CC_1}\right) - \hat{\sigma}_C\right\|_1 \leq 2\varepsilon_0$. Any such circuit $\hat{W}'_{CC_1}$ will also satisfy

$$\left\|\mathrm{Tr}_{C_1}\left(\hat{W}'_{CC_1}\left|0\middle\rangle\!\middle\langle 0\right|_{CC_1}\hat{W}'^\dagger_{CC_1}\right) - \sigma_C\right\|_1 \leq 3\varepsilon_0. \tag{3.358}$$

Using the same argument as in Section 3.3.4.3, we can merge these learned local circuits into a global depth-$2d$ circuit $\hat{W}_i$. Let $\hat{\sigma}_{A_i} := \mathrm{Tr}_{A_i^L A_i^R}\left(\hat{W}_i\left|0\middle\rangle\!\middle\langle 0\right|_{A_i^L A_i A_i^R}\hat{W}_i^\dagger\right)$ be the learned reduced density matrix on A_i, then the local reduced density matrices of $\hat{\sigma}_{A_i}$ and σ_{A_i} are $3\varepsilon_0$ close in trace distance on solid colored regions in Fig. 3.10. This allows us to invoke the main technical lemma, Lemma 3.35, which gives

$$\frac{1}{2}\|\hat{\sigma}_{A_i} - \sigma_{A_i}\|_1 \leq 13n\varepsilon_1^{1/16} + 8n\varepsilon_0^{1/4} \leq 22n^{17/16}\varepsilon_0^{1/16}. \tag{3.359}$$

The state $|\psi\rangle$ can be approximately prepared as follows:

1. Initialize registers A_i, B_i, A_i^L, A_i^R in the state $|0\rangle$. Let $A = \cup_i A_i$ and $B = \cup_i B_i$.
2. For each i, apply the depth-$2d$ circuit $\hat{W}_i$ to $A_i^L A_i A_i^R$. The reduced density matrix on AB equals $(\otimes_i\hat{\sigma}_{A_i}) \otimes \left|0\middle\rangle\!\middle\langle 0\right|_B$
3. Apply the depth-d circuit $V^\dagger$ to AB, and the reduced density matrix on AB is $\hat{\rho} = V^\dagger(\otimes_i\hat{\sigma}_{A_i}) \otimes \left|0\middle\rangle\!\middle\langle 0\right|_B V$, which approximately equals to $\left|\psi\middle\rangle\!\middle\langle\psi\right|$.

Similar to the proof of second claim of Theorem 3.15 at the end of Section 3.3.4.3, we can choose the A_i regions to be sufficiently wide, such that the number of ancilla qubits equals to tn for an arbitrarily small constant t.

The final task is to bound the error between the learned density matrix and $\left|\psi\middle\rangle\!\middle\langle\psi\right|$. Using Lemma 3.36, the trace distance can be bounded as

$$\begin{aligned}\frac{1}{2}\left\|V^\dagger(\otimes_i\hat{\sigma}_{A_i}) \otimes \left|0\middle\rangle\!\middle\langle 0\right|_B V - \left|\psi\middle\rangle\!\middle\langle\psi\right|\right\|_1 &= \frac{1}{2}\left\|(\otimes_i\hat{\sigma}_{A_i}) \otimes \left|0\middle\rangle\!\middle\langle 0\right|_B - V\left|\psi\middle\rangle\!\middle\langle\psi\right|V^\dagger\right\|_1\\ &\leq \sqrt{2\cdot 3n\varepsilon_0 + \sqrt{n}\cdot 22n^{17/16}\varepsilon_0^{1/16}}\\ &\leq 6n^{25/32}\varepsilon_0^{1/32}.\end{aligned} \tag{3.360}$$

Therefore, to achieve trace distance ε, it suffices to choose $\varepsilon_0 = \mathcal{O}(\frac{\varepsilon^{32}}{n^{25}})$. The total sample complexity is

$$N = \frac{2^{\mathcal{O}(d^2)}}{\varepsilon_0^2} \log \frac{n}{\delta} = \frac{2^{\mathcal{O}(d^2)} n^{50}}{\varepsilon^{64}} \log \frac{n}{\delta}. \tag{3.361}$$

The total running time is

$$n \cdot S = \left(\frac{nd^3}{\varepsilon} \right)^{\mathcal{O}(d^3)}. \tag{3.362}$$

Chapter 4

Fault tolerance

This chapter studies fault tolerance techniques for shallow quantum circuits and their applications. In Section 4.1 we discuss the motivation for studying this direction and propose a key open question toward quantum computational advantage in the early fault tolerant era: constructing a family of shallow quantum circuits that is both fault tolerant and classically hard. The rest of this chapter presents partial progress toward addressing this question. In Section 4.2 we construct a family of local Hamiltonians which demonstrate quantum computational advantage with constant-temperature Gibbs sampling, which follows from a fault tolerance scheme for shallow IQP circuits against input noise. This is based on joint work with Thiago Bergamaschi and Chi-Fang Chen [219]. In Section 4.3 we prove that encoded logical states of arbitrary quantum LDPC codes can be fault tolerantly prepared in a single shot: a constant depth noisy quantum circuit followed by a single round of measurement and classical feedforward, based on joint work with Thiago Bergamaschi [220]. Both results develop techniques for achieving fault tolerance within noisy shallow quantum circuits.

4.1 Toward quantum computational advantage in the early fault tolerant era

Recent developments in quantum hardware have ushered the transition of quantum computing from the NISQ era to the early fault tolerant era. This motivates the question of what is the next step regarding achieving quantum computational advantage using current or near-term quantum devices.

In Section 2.1 we gave an overview of the exciting developments in achieving quantum computational advantage in the NISQ era via random circuit sampling. While these experiments require a large resource cost to classically spoof, their main drawback is the lack of scalable computational hardness in the asymptotic regime of constant noise per gate, which can be viewed as a consequence of the lack of error correction. As we discussed in Section 2.1, a key question now is to formulate the next motivating target that drives our field. Here we discuss such a proposal: demonstrate quantum computational advantage with provable hardness using fault tolerance techniques within noisy shallow quantum circuits.

Noisy shallow quantum circuits and fault tolerance. Recall that a noisy shallow quantum circuit is a low-depth quantum circuit which consists of noisy initial states, noisy quantum gates and noisy measurements. This has been a standard model of quantum computation in the NISQ era, which includes Google's random circuit sampling experiments [7, 21].

In 2023, Harvard/QuEra performed an experiment which marks the beginning of the early fault tolerant era [221]. While the experiment is still within the model of noisy shallow quantum circuits, it demonstrated fundamentally new capabilities: *long-range connectivity* and *transversal gates on encoded qubits*. This demonstrated a new potential of noisy shallow circuits: they can achieve some limited form of fault tolerance, even though the circuit is shallow and does not have mid-circuit error correction.

While the experiment is not scalable due to the exponential sampling overhead in post-selected error detection, it makes an important step beyond NISQ. A concrete goal is therefore to achieve scalable quantum computational advantage with provable hardness, using fault tolerance techniques within noisy shallow circuits. This goal aims to achieve the best case scenario, where what we currently have in terms of hardware capabilities already suffice for scalable computational hardness. This goal is a specific question in a broader context of finding plausible architectures for early fault tolerant devices that can demonstrate computational hardness in the next few years.

The key technical challenge is therefore to develop a family of shallow quantum circuits that are both fault tolerant and classically hard, in the sense that sampling from the output distribution is classically hard to simulate even in the presence of constant noise per gate.

Overhead and hardware requirements of fault tolerance. Why is the above question technically non-trivial? The difficulty lies in the circuit depth overhead to implement quantum fault tolerance. Suppose we would like to encode an n-qubit, constant depth quantum circuit (which is already classically hard in the absence of noise) into a fault tolerant circuit. The quantum fault tolerance threshold theorem [222] based on recursive concatenation gives a quantum circuit that acts on $n \cdot \text{polylog}(n)$ qubits, with depth $\text{polylog}(n)$, and requires fresh qubit initialization. Our goal is to reduce the circuit depth to $O(\log n)$ or even less, which likely requires new techniques (the significance of $O(\log n)$ depth circuits is discussed below).

An alternative model of quantum fault tolerance assumes mid-circuit error correction: measure a subset of qubits (mid-circuit measurement), perform a classical computation based on the measurement outcome (decoding), and apply subsequent quantum gates based on the result of the computation (feedforward). The assumption here is that the decoding time is negligible. It is known that the quantum circuit depth overhead for fault tolerance can be reduced to constant in this model using single-shot error correction [223].

Here we would like to achieve low circuit depth overhead without assuming mid-circuit error correction, for the following two reasons. The first is to reduce the hardware requirement for demonstrating quantum computational advantage. A mid-circuit decoder requires specialized hardware and has not been demonstrated in large scale experiments. As we discussed earlier, removing this assumption would imply that our current hardware capabilities already suffice for scalable quantum computational advantage in the early fault tolerant era. The second is to understand the fundamental question of whether a noisy quantum computer can demonstrate computational hardness *by itself*, without the help of a classical computer, even at low depth. Analogous to the fundamental notion of self-correcting quantum memory which requires a noisy quantum computer to store a quantum state by itself, here we are asking for "self-correcting quantum advantage" at low depth.

Finally, we comment on the significance of $O(\log n)$ depth quantum circuits, which is related to the requirement of fresh qubit initialization (initializing a noisy ancilla in the $|0\rangle$ state at any time) in fault tolerance. Without fresh qubit initialization, it is known that any noisy quantum circuit output becomes trivial at $\omega(\log n)$ depth due to the accumulation of noise [36], which implies the fundamental importance of fresh qubit initialization in fault tolerance. Conversely, it is also known [36] that any noisy quantum circuit of $O(\log n)$ depth that uses fresh qubits can be converted to one without fresh qubits, with the same depth and polynomial blow-up in system size. Therefore, our goal addresses the fundamental question of whether quantum computational advantage exists in a model of computation without fresh qubits or mid-circuit error correction.

Concrete direction: fault tolerant IQP sampling. A concrete direction to address the above question is to study low-overhead fault tolerant implementation of IQP sampling. The definition of an IQP circuit is shown in Fig. 4.1, which is known to be hard to simulate classically even at constant depth [118, 119]. From the fault tolerance perspective, IQP circuits are the most natural family of classically-hard circuits, because conceiving a blueprint for implementing such a quantum circuit fault tolerantly is straightforward: just implement each component in Fig. 4.1 using logical ones with a suitable quantum LDPC code. This involves (1) fault tolerant preparation of the logical $|+\rangle$ state, (2) transversal diagonal gates, and (3) logical X measurement. In particular, there is no need to perform magic state distillation or code switching, which are typically required for universal fault tolerant quantum computation, because we only need transversal diagonal gates (instead of universal gates) for IQP sampling. Examining this blueprint, the key challenge in fact lies in the first step: how to prepare the logical state in low depth.

Ref. [114] showed that a constant depth Clifford circuit can be implemented fault tolerantly in constant depth (without classical feedforward), using single-shot logical state preparation (see below and Section 4.3) for the surface code. Instead of correcting the Pauli error in state preparation using classical feedforward, the error is propagated through the circuit and corrected at the end. However, this approach runs into issues when applied to non-Clifford gates, as Pauli errors no longer propagate as Pauli errors. Reducing the over-

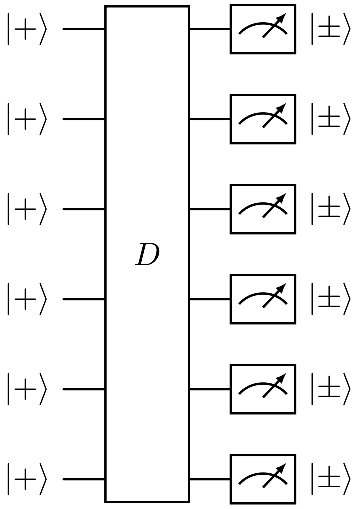


Figure 4.1: IQP sampling. An IQP circuit starts with every qubit in the $|+\rangle$ state, followed by a circuit of diagonal gates, and measurement in the X ($\{|+\rangle, |-\rangle\}$) basis.

head for logical state preparation seems to be a fundamental challenge for fault tolerantly implementing quantum circuits with non-Clifford gates.

Overview of Section 4.2 and Section 4.3. The remainder of this chapter aims to develop techniques and achieve a deeper understanding towards addressing the key question discussed above. Meanwhile, the two results presented in Section 4.2 and Section 4.3 also address fundamental questions in quantum complexity theory and fault tolerance, respectively.

In Section 4.2 we construct a family of local Hamiltonians which demonstrate quantum computational advantage with constant-temperature Gibbs sampling: the Gibbs state can be efficiently prepared quantumly by a rapidly-mixing quantum Markov chain but hard to sample from using a classical computer, addressing a key question about the complexity of quantum Gibbs sampling. Interestingly, the key idea is to reduce this question to constructing a fault tolerance scheme for shallow IQP circuits, under a special noise model where only input qubits are subject to noise. Our construction gives a fault tolerant circuit under this noise model with $O(\log\log n)$ depth. This makes partial progress toward the goal. Meanwhile, it also hints that achieving fault tolerant IQP sampling in constant depth with the standard noise model may be difficult, as it already seems difficult in this weaker noise model.

In Section 4.3 we prove that single-shot logical state preparation can be achieved for arbitrary quantum LDPC codes. This allows constant-depth fault tolerant preparation of encoded $|0\rangle$ or $|+\rangle$ states in any LDPC code with classical feedforward. As discussed above, it remains to be seen whether this result can help achieve low overhead fault tolerant IQP sampling.

4.2 Quantum computational advantage with constant-temperature Gibbs sampling

A major goal of today's quantum computing efforts is to realize quantum computational advantage in realistic physical setups. One such setup is open system thermalization, where a quantum many-body system is specified by a Hamiltonian H and then coupled to a bath at finite (constant) temperature β, and the system converges to the Gibbs state $\rho_\beta \propto e^{-\beta H}$. Under physical assumptions,[1] this thermalization process can be described by a *thermal Lindbladian* (a continuous-time quantum Markov chain), most notably the Davies generator [224] and its variants (e.g. [225]). This setup is especially relevant for physical platforms in which implementing digital quantum circuits is difficult. However, there has been no complexity-theoretic evidence showing that quantum computational advantage can be achieved in this model (see Section 4.2.1 for a discussion).

In this section, we provide such evidence by showing that quantum computational advantage can be achieved for the task of sampling from the measurement outcome distribution of Gibbs states at constant temperatures. In particular, we construct a family of commuting local Hamiltonians and show that its thermalization process (described by the Davies generator) is rapidly mixing. Meanwhile, its Gibbs state is classically intractable to sample from.

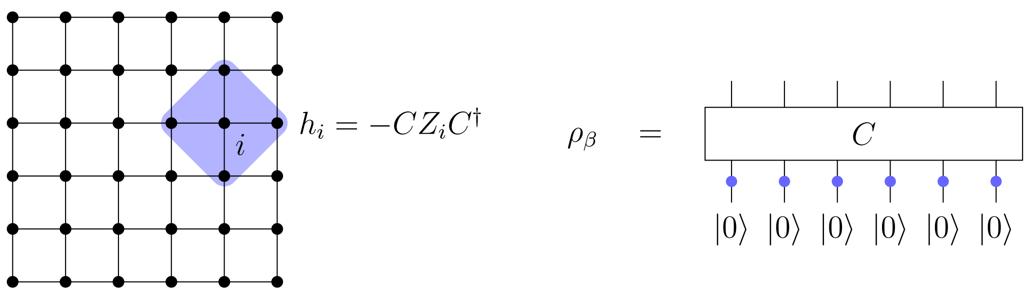
Theorem 4.1 (Main result)**.** *For any constant inverse-temperature $\beta = \Theta(1)$, there exists a family of n-qubit commuting $O(1)$-local Hamiltonians, such that the n-qubit Gibbs state ρ_β is both*

1. Rapidly Thermalizing. *It can be prepared within small trace distance by the Davies generator (a quantum Markov chain describing thermalization), in time $n^{o(1)}$. In addition, this process can be simulated on a quantum computer in time $n^{1+o(1)}$. And yet,*

2. Classically Intractable. *Under certain complexity-theoretic assumptions, there is no polynomial time classical algorithm to sample from the measurement outcome distribution $p(x) = \langle x| \rho_\beta |x\rangle$ within small total variation distance.*

The classical hardness is based on the hardness of approximate sampling from the output distribution of ideal shallow quantum circuits. The main result, therefore, places the hardness of rapidly mixing thermalization to the same level as ideal sampling-based quantum supremacy experiments (see Section 4.2.8 for more details).

[1] The bath is Markovian and the coupling is weak.


$h_i = -CZ_iC^\dagger$
i
$\rho_\beta =$
C
$|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$

(a) A Local Hamiltonian

(b) The Noise Model

Figure 4.2: (a) We consider local Hamiltonians which are parent Hamiltonians of shallow quantum circuits. (b) The Gibbs states of these Hamiltonians $\rho_\beta \propto e^{-\beta H}$ are equivalent to the output state of C, where the input qubits are subject to bit-flip errors (blue dots) of rate $(1 + e^{2\beta})^{-1}$.

A more general version of the result (Theorem 4.8) is given in Section 4.2.10, where we generalize the above and show how to trade-off locality for mixing time, including a family of $O(\log \log n)$-local Hamiltonians which thermalizes in polylog(n) time.[2]

Our approach. The family of Hamiltonians we consider is the class of "parent" Hamiltonians of shallow quantum circuits (Fig. 4.2a). Starting from a trivial, non-interacting Hamiltonian $H_{\mathsf{NI}} = -\sum_i Z_i$ consisting of single-qubit Pauli-Z terms, we consider the family of Hamiltonians that are related to H_{NI} by a low depth circuit,

$$\mathscr{H} = \left\{ H : \exists \text{ low-depth circuit } C,\ H = CH_{\mathsf{NI}}C^\dagger \right\}. \tag{4.1}$$

Each $H \in \mathscr{H}$ is local, commuting, and it encodes the computation C in the sense that its ground state is the output of the circuit $C\,|0^n\rangle$. The reason that these Hamiltonians are good candidates for quantum advantage at constant temperatures lies in the following key observation:

The Gibbs state of each $H \in \mathscr{H}$ is a noisy version of the underlying computation, where random bit-flip errors are applied to the input qubits (Fig. 4.2b).

This is a clean example of the general intuition that constant-temperature Gibbs states are very noisy and far from ground states. To encode computational hardness into the Gibbs states of $H \in \mathscr{H}$, it then suffices to design a shallow quantum circuit which is classically intractable to simulate even under input noise. Our main result then follows from two key technical contributions:

[2] In the initial posting of this work, we stated this latter construction as our main result. We thank James Watson and Joel Rajakumar for the observation that under appropriate parameter choices, our construction in fact has constant locality (see Section 4.2.4.2).

1. **A construction of classically-hard shallow quantum circuits that are fault-tolerant against input noise.** Standard techniques in quantum fault-tolerance blow up the circuit depth, and in turn, the locality of the parent Hamiltonian[3]. We start from a specific family of classically-hard shallow circuits (namely, IQP circuits [118, 119]), and then design a low-overhead fault-tolerance scheme tailored to IQP circuits and the input noise model.

2. **A proof that these Hamiltonians thermalize rapidly, via a modified log-Sobolev inequality.** We prove a rapid mixing bound for Hamiltonians in $\mathscr{H}$ which leverages the structure of the thermal Lindbladian (the quantum Markov chain describing thermalization), in combination with a carefully constructed lightcone argument for shallow quantum circuits.

4.2.1 Related work

Complexity of Gibbs states. Establishing quantum computational advantage with constant-temperature Gibbs sampling faces inherent difficulties. After all, at high enough temperatures, Gibbs states are expected to be essentially classical objects; in particular, sampling from these Gibbs states is efficient to simulate on a classical computer[4] [189, 226]. On the other hand, in the low temperature regime, preparing Gibbs states is expected to be hard in general even for a quantum computer;[5] in particular, the thermalization process may take exponential time.

Nevertheless, a path exists to circumvent these issues, by embedding some classically hard quantum computation into a local Hamiltonian. It is reasonable to hope that the nature of this embedding ensures that producing the Gibbs state is still tractable for quantum computers[6] (e.g. [228, 229]), and one can further hope that the Gibbs state is classically hard. But there is yet another issue: standard means to embed quantum circuits into Hamiltonians [230] typically encode the quantum computation into its ground state. However, Gibbs states at constant temperatures are understood to be very noisy, and far from the ground state. In this manner, to argue that this noisy version of the ground state remains classically hard, there must be an inherent *fault-tolerance* to the circuit-to-Hamiltonian mapping. Our approach can be viewed as a clean example that satisfies all of the above criteria.

Gibbs samplers and rapid mixing. Preparing Gibbs states (or Gibbs sampling) is a candidate application of quantum computers as well as an important quantum algorithmic

[3] Our interest in decreasing the locality stems both from the practical challenges behind engineering systems with many-body interactions, and a complexity-theoretic understanding of the role of locality in the hardness of Gibbs sampling.

[4] This does not contradict our result which holds for arbitrary constant temperature, due to the order of quantifiers; see Remark 4.1.

[5] Indeed, NP-hard due to the classical PCP theorem [227].

[6] In fact, we desire something even stronger: that the Hamiltonian is rapidly thermalizing.

primitive. While there are many proposed quantum Gibbs samplers, recent developments have focused on an approach of simulating open system (Lindbladian) dynamics, in particular the Davies generator and its variants which mimic thermalization in nature [231, 232, 233].

The key missing ingredient to the efficiency of these quantum simulation algorithms is a bound on the mixing time of the underlying quantum Markov chain. The standard approach, via a bound on the spectral gap, gives a mixing time that has intrinsic polynomial dependence in n [234]. A much stronger approach known as (quantum) *log-Sobolev inequalities* consists of a decay of the relative entropy, and results in only polylog(n) mixing time, a phenomenon known as *rapid mixing*. These stronger inequalities are notoriously hard to prove: examples have only been shown for certain commuting systems, in 1D [235, 236] or on lattices above a threshold temperature [237]. Our rapid mixing bound uses the lightcone structure of shallow quantum circuits, and does not require geometric locality or a temperature threshold.

Shallow quantum circuits and fault-tolerance. Shallow quantum circuits are widely used in quantum algorithms for near-term devices and quantum supremacy experiments. The hardness of sampling from the output distribution of shallow quantum circuits provides the complexity foundation for these experiments (see [8] for a review). We focus on constant-depth *instantaneous quantum polynomial time* (IQP) circuits $C = H^{\otimes n} D H^{\otimes n}$ where D is a constant-depth diagonal unitary, which provides hardness due to the universality of measurement-based quantum computation [118, 119]. However, these circuits are not noise-robust and become classically simulable under noise [40, 238]; fault-tolerance techniques are therefore necessary for classical hardness in our context.

There is a tension between shallow quantum circuits and the overhead of quantum fault-tolerance.[7] Standard techniques encode a constant depth quantum circuit into a fault-tolerant circuit of polylog(n) depth [222], and fault-tolerance with constant circuit depth overhead is only known for shallow Clifford circuits [114]. Ref. [40] devised a fault-tolerance scheme specialized to IQP circuits and the input noise model, and we design a new scheme in this setting which achieves a significantly smaller overhead.

4.2.2 Our Contributions

4.2.2.1 Efficient quantum Gibbs sampling via rapid mixing

Our first result is a quantum algorithm for preparing the Gibbs states of $H \in \mathscr{H}$, given only a description of its local terms $H = \sum_i h_i$ (as $2^\ell \times 2^\ell$ matrices).[8]

Lemma 4.1 (Gibbs State Preparation). *Fix $\beta > 0$, and let $H \in \mathscr{H}$ be the parent Hamiltonian of a quantum circuit on n qubits, of depth d and lightcone size ℓ. Then, there exists*

[7] Note that some models of fault-tolerance assume instant classical computation and feedforward within a quantum circuit [239, 240]. This is not allowed in our setting: all operations must be realized by quantum gates.

[8] Although H has a simple structure by definition, the underlying global structure (the low-depth circuit C) is hidden among the local terms, and is not directly accessible. See Remark 4.2 for a discussion.

a quantum algorithm which can prepare the Gibbs state of H at inverse-temperature β up to an error ε in trace distance in time $O(2^{4\ell} \cdot 2^d \cdot e^{2\beta} \cdot n \cdot \mathrm{poly}(\log \frac{n}{\varepsilon}, \ell, \beta))$.

In general, the lightcone size ℓ is upper bounded by $\ell \leq 2^d$. We emphasize we do not make any assumptions on the temperature or geometric locality. This is important as our fault-tolerant circuits (Lemma 4.2) are not naturally defined on a lattice.

The algorithm in Lemma 4.1 follows from a two-step argument. The first step is the design and analysis of a particular family of Davies generators [224], a family of dissipative Lindbladians whose local jumps (or transitions) are engineered to resemble the connectivity of the Hamiltonian. In Lemma 4.8, we prove that the mixing time of our Lindbladians is $t_{mix} = O(4^\ell \log n)$ via a modified log-Sobolev inequality. In principle, this step is already a *thermal algorithm*, in the sense that "placing the system in a fridge" would drive it to the Gibbs state in time $t_{mix} \cdot \log(1/\varepsilon)$.

The second step is the simulation of the dissipative (non-unitary) dynamics on a quantum computer. We employ the block-encoding framework of [232] which we significantly simplify as our family of Hamiltonians is commuting and has integer spectra. The quantum simulation adds a factor of n to the running time, which may be hard to improve due to the absence of geometric locality.

4.2.2.2 Fault-tolerance of shallow IQP circuits

The key ingredient for the classical hardness of sampling from quantum Gibbs states is to produce a shallow quantum circuit which is hard to sample from even under input noise. For this purpose, we design a fault-tolerance scheme for shallow IQP circuits [118, 119] since their gate set works nicely with fault tolerance techniques. Our result ensures that any IQP circuit can be made robust to input noise with only a small additive blow-up to the circuit depth (see Lemma 4.14 for a more general statement).

Lemma 4.2. *Let $p < \frac{1}{2}$ be a constant bit-flip error rate, and let C be an n qubit IQP circuit of depth d. Then, there exists an $O(n \log \frac{n}{\varepsilon})$ qubit circuit $\tilde{C}$ of depth $d + o(\log \frac{n}{\varepsilon})$, such that a sample from $\tilde{C}$ under input noise (Fig. 4.2b) can be efficiently post-processed into a sample within ε total variation distance to the output distribution of C.*

This result significantly reduces the blow-up in circuit depth compared to a prior fault-tolerance scheme of [40]. Moreover, the locality of the resulting parent Hamiltonian $H = -\sum_i \tilde{C} Z_i \tilde{C}^\dagger$ is only a constant. Our key idea is a non-adaptive state distillation scheme, drawing inspiration from magic state distillation [241]: distilling a near-perfect initial state from noisy initial states, up to a known but uncorrected Pauli error. The error is propagated through the circuit and corrected in post-processing, similar to [114]. Propagating Pauli errors through non-Clifford circuits is hard in general, but here it works thanks to the structure of IQP circuits.

4.2.2.3 Applications

BQP Completeness under adaptive single-qubit measurements. In addition to quantum advantage, using our techniques we can show that constant-temperature Gibbs states do have some inherent form of universality for quantum computation. In Section 4.2.11 we prove that there exist local Hamiltonians whose Gibbs states are universal resource states for quantum computation, in the sense that they can be used for universal measurement-based quantum computation.

Theorem 4.2. *Fix an inverse-temperature $\beta = \Theta(1)$. Then, there exists an n-qubit, $O(1)$-local commuting Hamiltonian, whose Gibbs state at inverse-temperature β is a universal resource state for quantum computation and is efficiently preparable on a quantum computer.*

Theorem 4.2 is based on the universality of cluster-states for measurement-based quantum computation. That is to say, any quantum computation of bounded size can be implemented using adaptive single-qubit measurements on top of a fixed 2D cluster-state (e.g. [242]). We design a Hamiltonian whose Gibbs state resembles a noisy version of a cluster-state, such that under adaptive single-qubit operations, one can nevertheless correct and distill out computation.

Gibbs sampling under measurement errors. An interesting question is whether the thermal quantum advantage demonstrated in this section, is itself robust to noise. That is, in realistic physical platforms, we expect imperfect state preparation, noisy system-bath couplings, and erroneous measurements. As a starting point to this problem, we consider a model where the Gibbs state preparation is ideal, but there are random bit-flip errors in the measurement outcome.

We show that the quantum advantage survives in this model, albeit at a higher locality.

Theorem 4.3. *Fix an inverse temperature $\beta = \Theta(1)$, and a measurement error rate $p < \frac{1}{2}$. There exists a family of n-qubit, $O(\log n)$-local Hamiltonians, such that sampling from their Gibbs state at inverse-temperature β, under measurement errors of rate p, is classically intractable under certain complexity-theoretic assumptions. Moreover, there exists a* $\mathrm{poly}(n)$ *time quantum algorithm to produce said Gibbs state.*

The Hamiltonians of Theorem 4.3 are similar to that of Theorem 4.1, in the sense that they are parent Hamiltonians of fault-tolerant IQP circuits. However, to ensure classical hardness under measurement errors, our quantum circuits now need to be fault-tolerant against both input and output errors. (Recall that the "input errors" come from temperature, while "output errors" come from actual physical noise in measurements.) To do so, in Section 4.2.12 we appeal to an optimized construction of a prior fault-tolerance scheme by [40], at the cost of an increase to the locality of the Hamiltonians, which also changes the mixing time from $n^{o(1)}$ to $\mathrm{poly}(n)$.

4.2.3 Discussion

We conclude by discussing two future directions, broadly related to the complexity of Gibbs sampling. The first of which concerns the BQP Completeness of Gibbs sampling (without adaptivity).

Question 4.1 (BQP Completeness of Gibbs Sampling). *For every n qubit,* $\mathrm{poly}(n)$ *depth quantum circuit C, does there exist a Hamiltonian H and a constant inverse-temperature $\beta > 0$ such that by sampling from its Gibbs state one can recover the output of the quantum computation C?*

Partial progress on this question has recently been made by [229], albeit, only at very low temperatures where the Gibbs state approximates the ground state. In particular, they showed how to embed an arbitrary quantum computation into a (modified) Feynman-Kitaev circuit-to-Hamiltonian mapping, which could be efficiently prepared by a Lindbladian evolution. Whether similar ideas could work at constant temperatures remains an open problem.

Another interesting direction lies in the time overhead for fault-tolerance, and for quantum advantage using shallow circuits which are robust to noise.

Question 4.2 (Quantum Advantage in Noisy Shallow Circuits). *Does there exist a family of constant depth quantum circuits (using only quantum gates) which is classically hard to sample from in the presence of depolarizing noise on each gate?*

Its main motivation lies in the design of quantum advantage experiments, which can be implemented on near-term devices. Depolarizing noise on each gate of the circuit, however, is naturally a significantly more general noise model than input noise. Nevertheless, the same question with input noise remains open as well.

4.2.4 Technical Overview

In this section we give a sketch of our two main technical contributions: (1) A proof of a modified log-Sobolev inequality for a family of Davies generators, via a lightcone argument (Section 4.2.4.1); and (2) A fault-tolerance scheme for shallow IQP circuits against input noise, via non-adaptive state distillation (Section 4.2.4.2). We begin by presenting some basic notation and background on thermal Linbladians.

4.2.4.1 Gibbs state preparation via rapid mixing

Fix a Hamiltonian $H \in \mathscr{H}$. By definition, there exists a shallow circuit C such that

$$H = \sum_{i\in[n]} h_i, \quad \text{where} \quad h_i = C\left(|1\rangle\langle 1|_i \otimes \mathbb{I}_{[n]\setminus i}\right) C^\dagger, \tag{4.2}$$

and each $|1\rangle\langle 1|_i$ is a single-qubit projection. Note that Eq. (4.2) is equivalent to Eq. (4.1) up to a shift. The eigenstates of H of energy $k \in [n]$ are all the states $C\,|x\rangle$, where $x \in \{0,1\}^n$

has hamming weight $|x| = k$. We denote the projection Π_k onto the eigenspace of H of energy k as

$$\Pi_k = C\bigg(\sum_{|x|=k} |x\rangle\langle x|\bigg)C^\dagger. \tag{4.3}$$

We consider two notions of locality for C and H respectively:

- The circuit lightcone. The *lightcone* L_i of qubit i is the set of qubits that can be reached by i via gates in C, and we define the *lightcone size* as $\ell = \max_i |\mathsf{L}_i|$.
- The Hamiltonian locality. Let $\mathsf{S}_i = \text{supp}(h_i) = \text{supp}(CZ_iC^\dagger)$ be the set of qubits that h_i acts nontrivially on. The *locality* of the Hamiltonian H is defined as $r = \max_i |\mathsf{S}_i|$.

Note that S_i is related to the propagation of Z_i under C, and thus by definition we have $\mathsf{S}_i \subseteq \mathsf{L}_i$. In fact, $r \ll \ell$ for the family of circuits we consider.

Our Davies generators. We determine our family of thermal Linbladians, or Davies generators, by specifying two ingredients: a set of jump operators, and transition weights. Technically, general thermal Lindbladians for noncommuting Hamiltonians need not take the Davies' form (see, e.g, [232, 233]), but for commuting Hamiltonians, the Davies' generator is nonetheless sufficient for all our discussions.

- **Jump Operators.** To generate the transitions, we consider the set of jump operators which are local, ℓ-qubit Pauli operators on the support of each lightcone L_i,[9]

  $$\{A^a\}_{a\in\mathcal{A}} = 2^{-\ell}\cdot\bigg\{P_{\mathsf{L}_i}\otimes\mathbb{I}_{[n]\setminus\mathsf{L}_i} : i\in[n], P\in\mathcal{P}_\ell\bigg\}, \tag{4.4}$$

  where $\mathcal{P}_\ell = \{I, X, Y, Z\}^{\otimes\ell}$.[10] In contrast to classical Markov Chain transitions, these quantum jumps will change the energy of the system in superposition. Thereby, it will be convenient to decompose the jump operators into the energy basis:

  $$A^a_\nu := \sum_{k\in[n]} \Pi_{k+\nu}A^a\Pi_k \quad \text{such that} \quad \sum_{\nu\in[-n,n]} A^a_\nu = A^a. \tag{4.5}$$

- **Transition Weights.** The transition weight is selected to be the Glauber dynamics weight, $\gamma(\nu) = 1/(1+e^{-\beta\nu})$ for all $\nu\in[-n,n]$.

[9] This set of jump operators which "drive" the transition can be essentially arbitrary, however, this choice resembling the connectivity of the underlying Hamiltonian will play an important role in our analysis.

[10] Note that there are $|\mathcal{A}| = n\cdot 4^\ell$ jump operators.

Put together, the associated family of Davies generators $\mathcal{L}$ can be written down as[11]

$$\mathcal{L}[\rho] = \sum_{a\in\mathcal{A}}\sum_{\nu}\gamma(\nu)\left(A^a_\nu\rho(A^a_\nu)^\dagger - \frac{1}{2}\left\{(A^a_\nu)^\dagger A^a_\nu, \rho\right\}\right). \tag{4.6}$$

This construction satisfies the *quantum detailed balance* condition, which implies that the desired Gibbs state is a fixed point $\mathcal{L}[\rho_\beta] = 0$ of the evolution (see e.g. [243], or Fact 4.1). It remains to show that the Lindblad dynamics, governed by the exponential map

$$\frac{d}{dt}\rho = \mathcal{L}[\rho] \Rightarrow \rho(t) = e^{\mathcal{L}t}[\rho_0], \tag{4.7}$$

converges quickly to ρ_β. This is achieved by presenting a bound on the mixing time of $\mathcal{L}$, which is the shortest time t_{mix} such that

$$\left\|e^{\mathcal{L}t_{mix}}[\rho - \sigma]\right\|_1 \leq \frac{1}{2}\left\|\rho - \sigma\right\|_1, \quad \text{for all density matrices } \rho, \sigma. \tag{4.8}$$

A lightcone argument for the modified log-Sobolev inequality. To study the mixing time of our algorithm, our starting point is first to study the trivial non-interacting Hamiltonian $H_{\mathsf{NI}} = \sum_{i\in[n]} |1\rangle\langle 1|_i \otimes \mathbb{I}_{[n]\setminus\{i\}}$, and prove a rapid mixing bound for the associated Davies generator $\mathcal{L}_{\mathsf{NI}}$. Subsequently, we argue that the mixing time of $\mathcal{L}$ can be *compared* with that of $\mathcal{L}_{\mathsf{NI}}$. This is achieved by leveraging the lightcone structure of shallow quantum circuits. We begin by presenting basic definitions of Log-Sobolev bounds.

Mixing time bounds via log-Sobolev inequalities. There are two general purpose methods to bound the mixing time of Lindbladian evolution. The first of which consists of a bound on the spectral gap of $\mathcal{L}$. Unfortunately, a spectral gap bound comes at an inherent polynomial overhead to the mixing time, see Section 4.2.5. Instead, we make use of a much sharper notion of convergence known as a modified log-Sobolev inequality (MLSI) [234]. Informally, a MLSI quantifies the rate of decay of the relative entropy,[12] by relating it to the relative entropy itself:

$$\left.\frac{d}{dt}\right|_{t=0} D\big(e^{t\mathcal{L}}[\rho]||\rho_\beta\big) \leq -\alpha\cdot D\big(\rho||\rho_\beta\big) \quad \text{for every density matrix } \rho, \tag{MLSI}$$

where α is known as the MLSI constant. This clearly implies an exponential decay of $D(e^{\mathcal{L}t}[\rho]||\rho_\beta) \leq e^{-\alpha t}\cdot D(\rho||\rho_\beta)$. Which, in turn, tells us the mixing time is bounded by $t_{mix} \leq \alpha^{-1}\cdot O(\log n)$ (Pinsker's inequality). This logarithmic mixing time bound is known as *rapid mixing*, and proving good lower bounds on the constant α has proven to be quite challenging in the quantum setting.

[11] Where $\{A, B\} := AB + BA$ is the anti-commutator.

[12] The quantum relative entropy between two density matrices ρ, σ is given by $D(\rho||\sigma) = \mathrm{Tr}\big[\rho\cdot(\log\rho - \log\sigma)\big]$.

The non-interacting Lindbladian. The simplest Hamiltonian in the family $\mathscr{H}$ is the non-interacting system H_{NI}. Its Gibbs state is the tensor product state $\sigma_\beta \propto \left(e^{-\beta|1\rangle\langle 1|}\right)^{\otimes n}$. Under our framework (described in Eq. (4.6)), its associated Linbladian $\mathcal{L}_{\mathsf{NI}}$ has the same form as $\mathcal{L}$, except that the circuit C has been replaced by the identity. In this manner, $\mathcal{L}_{\mathsf{NI}}$ itself can also be written as a sum of non-interacting, single-qubit components:

$$\mathcal{L}_{\mathsf{NI}} = \sum_{i\in[n]} \mathcal{L}^i_{single} \otimes \mathbb{I}_{[n]\setminus\{i\}} \tag{4.9}$$

Since each single qubit Lindbladian $\mathcal{L}_{single}$ is highly explicit (it acts on 2×2 matrices), in Section 4.2.5, following now standard techniques, we are able to prove simple bounds on its MLSI constant.

Claim 4.1. *$\mathcal{L}_{\mathsf{NI}}$ satisfies a MLSI with constant $\Omega(e^{-\beta})$.*

The convex combination argument. The main technical challenge in our analysis lies in relating $\mathcal{L}_{\mathsf{NI}}$ with our family of Davies generators $\mathcal{L}$ from Eq. (4.6), in order to inherit the rapid mixing properties from the former. The crux of our proof lies in analyzing $\mathcal{L}$ in a basis rotated by C, to show that the rotated Davies generator is a *convex combination* of $\mathcal{L}_{\mathsf{NI}}$ and some other Davies generator. This involves a delicate lightcone argument shown in Fig. 4.3, and discussed shortly.

Claim 4.2. *In a basis rotated by C, the Lindbladian $\mathcal{L}$ from Eq. (4.6) can be written as a convex combination*

$$\tilde{\mathcal{L}} \equiv C^\dagger \mathcal{L}[C \cdot C^\dagger]C = q \cdot \mathcal{L}_{\mathsf{NI}}[\cdot] + (1-q) \cdot \mathcal{L}_{rest}[\cdot], \tag{4.10}$$

where both $\mathcal{L}_{\mathsf{NI}}, \mathcal{L}_{rest}$ share the fixed point σ_β, and $q = 4^{1-\ell}$.

We emphasize that the parameter $q \in [0,1]$ only depends on the lightcone size of C. Moreover, while $\mathcal{L}_{\mathsf{NI}}$ is the well-understood non-interacting system discussed previously, $\mathcal{L}_{rest}$ may apriori be arbitrary. However, at the very least we know it shares a fixed point with $\mathcal{L}_{\mathsf{NI}}$.

Neverthless, in Section 4.2.4.1 we show that convexity is precisely enough[13] to exhibit an MLSI for $\mathcal{L}$, with a constant which is only a multiplicative factor of q off of that of $\mathcal{L}_{\mathsf{NI}}$.

Lemma 4.3. *$\mathcal{L}$ satisfies a MLSI with constant $\Omega(4^{-\ell}e^{-\beta})$.*

[13] This is inspired by [244], who leveraged the concavity of the spectral gap to prove mixing properties of stochastic Hamiltonians.

The lightcone argument. We conclude by discussing the key technical step: a proof of Claim 4.2 via a lightcone argument. The starting point is to examine the Davies generator $\mathcal{L}$ (Eq. (4.6)) and its rotated version $\tilde{\mathcal{L}} = C^\dagger \mathcal{L}[C \cdot C^\dagger]C$. The goal is to show that "a fraction of" $\tilde{\mathcal{L}}$ equals the Davies generator of the non-interacting Hamiltonian H_{NI}, that is, within that fraction, the effect of C is erased.

To begin, note that the Davies generator $\mathcal{L}$ (Eq. (4.6)) only depends on the circuit C through the jump operators decomposed into the frequency basis, A^a_ν, which we recollect can be written as

$$A^a_\nu = \sum_k \Pi_{k+\nu} A^a \Pi_k = \sum_k C\Bigg(\sum_{|y|=k+\nu} |y\rangle\langle y|\Bigg) C^\dagger A^a C \Bigg(\sum_{|x|=k} |x\rangle\langle x|\Bigg) C^\dagger. \tag{4.11}$$

Crucially, due to the rotation of $\mathcal{L}$ to $\tilde{\mathcal{L}}$,[14] we observe that the dependence of C within $\tilde{\mathcal{L}}$ is only through second-moment operators of the form

$$\mathop{\mathbb{E}}_{P\sim\mathcal{P}_\ell}[C^\dagger P C \otimes C^\dagger P C], \tag{4.12}$$

where we consider the sum of all jump operators that act on a specific lightcone L_i of size ℓ, and recall that the jump operators are ℓ-qubit Pauli operators. It remains to express this operator as a convex combination, as shown in Fig. 4.3. A sketch of the argument follows:

- Step (i): Uses the identity $\mathbb{E}_{P\sim\mathcal{P}_\ell}[P \otimes P] = \frac{1}{2^\ell} \cdot \mathrm{SWAP}$, and linearity of expectation.
- Step (ii): Since C is a low-depth circuit, one can cancel the quantum gates within the lightcone of qubit i with their inverse.
- Step (iii): Uses the identity $\mathbb{E}_{P\sim\mathcal{P}_\ell}[P \otimes P] = \frac{1}{2^\ell} \cdot \mathrm{SWAP}$ again, but in the other direction.
- Step (iv): Re-writes the expectation into two parts: the first part is over the 4 single-qubit Paulis that act only on qubit i, and the second part, is all the remaining ℓ-qubit Paulis.

The crux of the argument lies in noting that in the first part of Step (iv), the ith qubit has been completely disentangled from the remaining circuit. Thereby, the single-qubit Pauli acts on a disentangled wire, and all remaining gates cancel with each other.

This gives the desired convex combination, where the first term corresponds to the non-interacting Hamiltonian with single-qubit jump operators. Finally, note that our choice of the jump operators (as ℓ-qubit Paulis acting on each lightcone) is crucial for this argument, and it is unclear if an arbitrary choice of jump operators would suffice.

[14] And the fact our jump operators are Pauli operators.

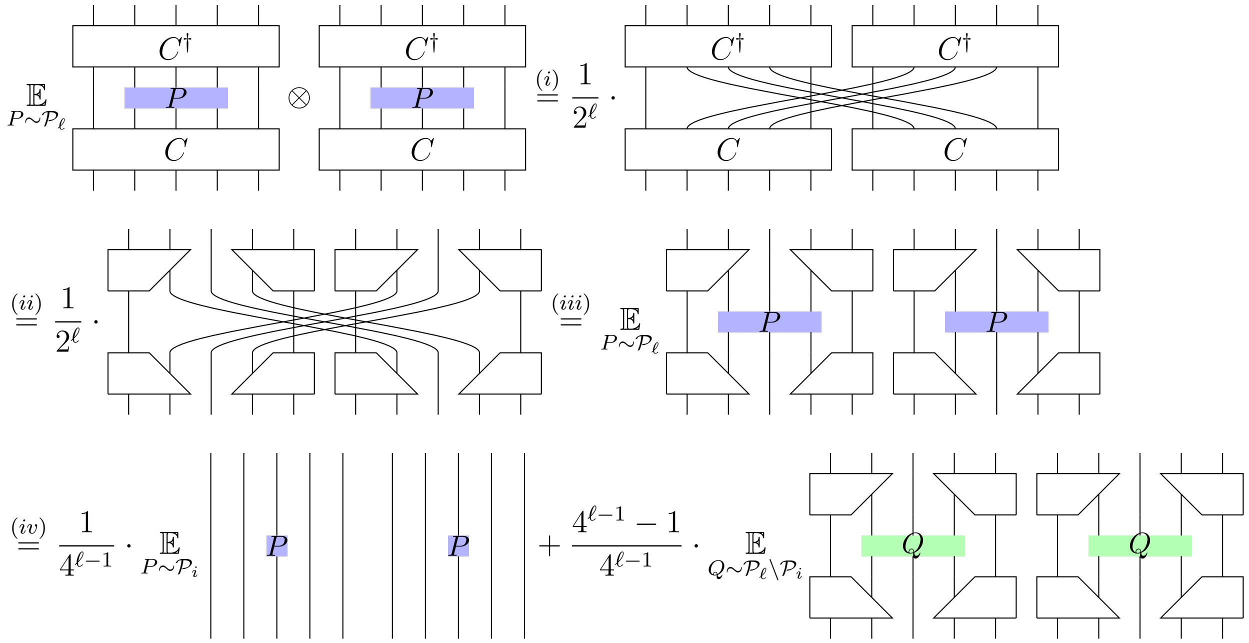

$\mathbb{E}_{P\sim\mathcal{P}_\ell}$
$C^\dagger$
P
C
$\otimes$
$\overset{(i)}{=}\ \frac{1}{2^\ell}\cdot$
$\overset{(ii)}{=}\ \frac{1}{2^\ell}\cdot$
$\overset{(iii)}{=}\ \mathbb{E}_{P\sim\mathcal{P}_\ell}$
$\overset{(iv)}{=}\ \frac{1}{4^{\ell-1}}\cdot\mathbb{E}_{P\sim\mathcal{P}_i}$
$+\ \frac{4^{\ell-1}-1}{4^{\ell-1}}\cdot\mathbb{E}_{Q\sim\mathcal{P}_\ell\backslash\mathcal{P}_i}$
Q

Figure 4.3: A lightcone argument for proving the modified log-Sobolev inequality.

Efficient implementation on a quantum computer. Rapid mixing of the Davies generator implies that the Gibbs state can be efficiently prepared in the thermal model of computation, described by coupling the quantum system to a thermal bath [245]. Next, we briefly discuss how to simulate the dissipative Lindbladian evolution $e^{\mathcal{L}t}$ on a quantum computer.

We leverage the "continuous-time quantum Gibbs sampler" framework of [232]. They show that to implement the map $e^{\mathcal{L}t}$ one requires $\tilde{O}(t)$ black-box invocations to a *unitary block-encoding* of the Lindblad operators ([232], Theorem I.1). In turn, to implement such a block-encoding for Hamiltonians H of integer spectra, it suffices to design quantum circuits which implement the Hamiltonian simulation of H, a block-encoding for the jump operators A^a, as well as a certain"frequency filter" which implements the Glauber dynamics weight. In Section 4.2.6, we discuss circuit implementations of all these ingredients, summarized in the following Lemma.

Lemma 4.4 (Dissipative Lindbladian Implementation)**.** *Fix parameters $t \geq 1$ and $\varepsilon \leq \frac{1}{2}$. Let $\mathcal{L}$ denote the Lindbladian of Eq. (4.6), defined by a quantum circuit C on n qubits of depth d and lightcone size ℓ. Then, we can simulate the map $e^{t\mathcal{L}}$ to error ε in diamond norm using a quantum circuit of depth $O(t \cdot n \cdot 4^\ell \cdot 2^d \cdot \mathrm{poly}(\ell, \log n, \log \frac{1}{\varepsilon}, \log t))$.*

Put together with our bound on the mixing time, we arrive at our main statement on Gibbs state preparation in Lemma 4.1.

4.2.4.2 Classical hardness of gibbs sampling

As discussed earlier, to obtain the classical intractability of quantum Gibbs sampling it suffices to construct a family of low depth quantum circuits which are hard to sample from even in the presence of input errors (Fig. 4.2b). The reason this imposes a challenge is two-fold. First, it is known that many classically hard shallow quantum circuits actually become classically simulable under input noise [40], thereby suggesting a need for fault-tolerance techniques. However, standard fault-tolerance techniques [222] often come with a prohibitive circuit depth overhead, which blows up the locality of the parent Hamiltonian. We address these challenges by designing a fault-tolerance scheme tailored to the input noise model with small overhead.

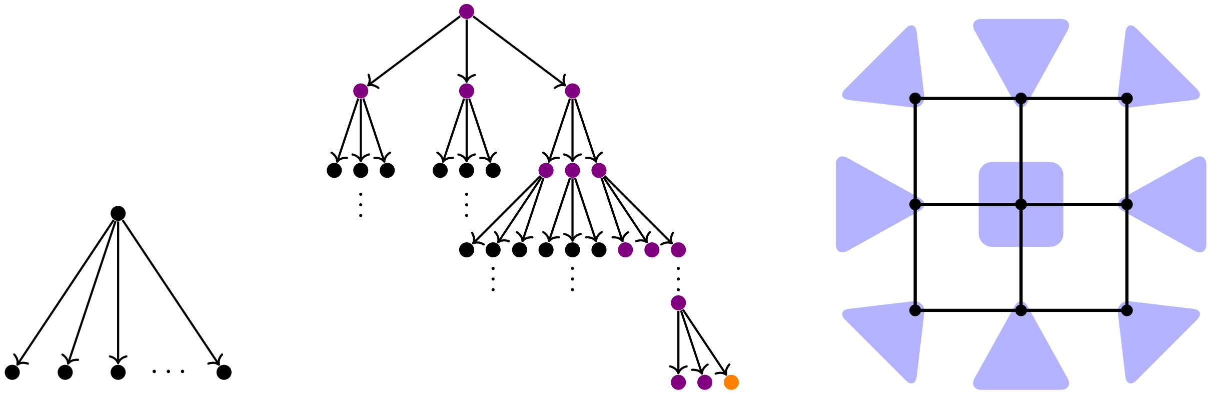
Our plan is to focus on IQP circuits, which are known to be already classically hard at constant depth. We show that their commuting structure plays an important role in our fault-tolerance techniques at low overhead.

Quantum computational advantage with shallow IQP circuits. Recall that IQP circuits can be written as $C = H^{\otimes n} D H^{\otimes n}$, where D is a diagonal unitary. The induced probability distribution $p(x) = |\langle x|C|0^n\rangle|^2$ is hard to sample from classically in general [246]. While any family of constant-depth and classically-hard IQP circuits suffices for our purpose, here we use the concrete example of cluster states on regular lattices composed with random Z-rotations[15], which have become the basis for various proposals of sampling-based quantum supremacy using low-depth circuits [246, 40, 118, 119, 247, 248].

We present the structure of these circuits in more detail in Section 4.2.8, where we additionally present a comprehensive discussion on the foundations of their hardness. As a brief overview, note that 2D cluster states with single-qubit Z rotations is a universal resource state for measurement-based quantum computation (MBQC) [242]. This implies that *exactly* sampling from their output distribution is hard in the worst-case [24]. The hardness of approximate sampling from these architectures are based on further assumptions [249, 118, 247], which we rigorously define in Section 4.2.8. The following theorem thus provides the complexity-theoretic basis of our hardness arguments.

Theorem 4.4 (Complexity of constant-depth IQP sampling [118, 119])**.** *There exists a constant $\delta > 0$, and a family of constant depth IQP circuits $\{C_n\}_{n\geq 1}$ on n qubits, such that no randomized classical polynomial-time algorithm can sample from the output distribution of C_n up to additive error δ in total variation distance, assuming the average-case hardness of computing a fixed family of partition functions (Conjecture 4.1), and the non-collapse of the Polynomial Hierarchy.*

[15] In the literature, these circuits are also known as the "evolution (quench) of an nearest-neighbor, translationally invariant (NNTI) Hamiltonian".



(a) Repetition gadget (b) Recursive concatenation (c) Fault-tolerant circuit

Figure 4.4: Fault-tolerance via state distillation gadgets. (a) The repetition code gadget. (b) A B-Tree and the recursive concatenation scheme. Arrows denote the direction of CNOT gates. (c) Pre-processing the circuit using distillation gadgets.

To establish classical hardness of the Gibbs sampling task, it suffices to map the above circuit C to a fault-tolerant circuit $\tilde{C}$, such that a sample from the output distribution of $\tilde{C}$ under input noise can be efficiently post-processed into an ideal sample from C. The key challenge is to reduce the fault-tolerance overhead in $\tilde{C}$, so that the corresponding parent Hamiltonian has small locality.

Fault-tolerance of IQP circuits against input noise. The starting point in our approach is the observation that it suffices to *error-detect* the random inputs bits, instead of correcting them, to preserve the hardness-of-sampling of C. Indeed, bit-flip errors (which are Pauli-X errors) on the input of IQP circuits, become phase-flip errors after the first layer of Hadamard gates, and thus commute with the entire IQP circuit. In this manner, they are equivalent to bit-flip errors on the measured output string. Therefore, if we could identify the computational basis state $|r\rangle = \otimes_i |r_i\rangle$ fed into the IQP circuit, we would be able to correct the measured output sample by simply subtracting $r \in \{0,1\}^n$. Indeed, we emphasize we don't intend to correct the input error within the quantum circuit at all, as this would require decoding and feedforward, and potentially a much deeper circuit. Instead, we correct the error only during classical post-processing (that is, after all qubits are measured).

The crux of our approach is the design of a "distillation" gadget, which independently pre-processes each input bit r_i into k others in such a manner which enables us to reconstruct r_i (with high probability) given only the other $k-1$ noisy bits. We illustrate this task with a simple example, based on the repetition code.

A distillation gadget based on the repetition code. Recall that all input bits are noisy: each of them is flipped from 0 to 1 with probability q. Given k bits drawn from

$s \leftarrow \mathsf{Bern}^k(q)$, suppose we designate the k-th bit as the "root" and apply a CNOT gate from it to the other $k-1$ bits (Fig. 4.4a). During the decoding stage, we would like to reconstruct the root bit given the other $k-1$ bits. To do this we simply compute the majority of the "leaves":

$$\begin{aligned} \text{Gadget}(s_1, s_2, \cdots, s_k) &= (s_1 \oplus s_k, s_2 \oplus s_k, \cdots, s_{k-1} \oplus s_k, s_k), \\ \hat{s}_k &= \mathsf{Maj}(s_1 \oplus s_k, s_2 \oplus s_k, \cdots, s_{k-1} \oplus s_k). \end{aligned} \tag{4.13}$$

We show that the probability of failure (when $\hat{s}_k \neq s_k$) equals $\delta = q^{\Omega(k)}$.

To highlight how these gadgets can be used for fault-tolerance, given an n-qubit IQP circuit C, we begin by pre-processing each of n input bits independently into a distillation gadget of size k, resulting in a circuit on $n \cdot k$ bits. Each of the n "root" bits are then fed into C (Fig. 4.4c). Note that the $n \cdot (k-1)$ remaining bits are untouched by C. In the end, after all qubits are measured, we can use the $n \cdot (k-1)$ ancilla bits to infer if an error had happened on each of the "root" bits fed into the circuit. As argued earlier, if an error did happen, it can be corrected by simply flipping the measurement outcome since the error commutes with the circuit. If we choose $k = \Theta(\log n)$, then the entire error correction process succeeds with high probability.

Recursive concatenation and B-Trees. In effect, the scheme above distills the "root" bit s_k with an effective bit-flip error rate $q^{\Omega(k)}$, using $k-1$ redundant "syndrome" bits of error rate q. Note that it used no information about the distribution of s_k, only that of the "leaves" $s_1, \cdots, s_{k-1}$.

To improve on this example, we bootstrap the above technique by recursively preparing "syndrome" bits of better and better fidelity.[16] Suppose we organize k bits into a tree of arity B and depth 2, such $k = 1 + B + B^2$. Moreover, apply the repetition code gadget on each layer, from leaves to root of the tree, by applying a CNOT gate from each parent bit to their respective children bits in the tree. In doing so, by the previous analysis we can identify each bit at the middle layer, just using the bits at the leaves, with error probability $q^{\Omega(B)}$. By performing majority again at the middle layer, we are now able to identify the bit at the root of this two-layer tree with error rate $(q^{\Omega(B)})^{\Omega(B)} = q^{\Omega(B^2)}$. By recursively applying this approach on a B-tree of depth d, the error probability at the root of the tree scales doubly-exponentially with the depth d, $q^{B^{\Omega(d)}}$.

At face value, it may seem that we haven't gained anything over the repetition code, as the error probability still only decays exponentially with the size of the gadget. The advantage lies instead in the *locality* of the gadget. Indeed, consider the lightcone of the orange qubit u at the leaf of the tree in Fig. 4.4b. By examining the causal influence of this qubit, we conclude that only the qubits in the neighborhood of its path to the root (the purple nodes in Fig. 4.4b) can lie in its lightcone. That is, if

$$u = u_0 \to u_1 \to \cdots \to u_d \equiv \text{ root} \tag{4.14}$$

[16] This construction is largely inspired by recursive magic state distillation schemes.

denotes the path from leaf to root, then the *lightcone* of u is contained the union of the neighborhoods $\mathsf{L}_u \equiv \cup_i^d N(u_i)$. Therefore, $|\mathsf{L}_u| \leq O(B \cdot d)$, which is a linear function of the depth of the tree. By further studying the propagation of Z Pauli's through the gadget, we analogously show that that the locality of the parent Hamiltonian of the distillation circuit is $|\mathsf{S}_u| \leq d$; precisely the nodes on the path from leaf to root. Lemma 4.2 then follows from a careful choice of B and d.

Organization. We organize the rest of this work as follows. In Section 4.2.5, we prove our rapid mixing bounds for Davies Generators, and in Section 4.2.6 discuss their simulation on a quantum computer. In Section 4.2.7, we prove that the constant temperature Gibbs states of the Hamiltonians in $\mathscr{H}$, can be interpreted as the output of noisy circuits. In Section 4.2.8, we present an overview of the computational complexity of shallow IQP sampling. In Section 4.2.9, we present our fault-tolerance scheme based on state distillation. Finally, in Section 4.2.10, we put everything together and prove our main result (Theorem 4.1). In Section 4.2.12, we present our results on Gibbs sampling with measurement errors, and in Section 4.2.11, we discuss the BQP completeness of Gibbs sampling with adaptive single-qubit measurements.

4.2.5 Rapid Mixing and Efficient Gibbs State Preparation

We dedicate this section to a proof of the rapid convergence of our dissipative Lindbladians. We defer a discussion on its implementation using quantum circuits to Section 4.2.6. For simplicity, henceforth we re-scale the class of parent Hamiltonians,[17]

$$H = \sum_{i\in[n]} h_i = \sum_i C\big(\,|1\rangle\!\langle 1| \otimes \mathbb{I}_{[n]\setminus i}\big)C^\dagger \tag{4.15}$$

to ensure frustration-freeness and positive integer spectra $[n] = \{0, \cdots, n\}$. Recall this Hamiltonian is commuting, and its eigenstates are given by $\{C\,|x\rangle : x \in \{0,1\}^n\}$. Let ℓ be the lightcone size of C. We refer the reader to Section 4.2.4.1, Eq. (4.6) for a description of our Lindbladian.

4.2.5.1 Preliminaries on thermal Lindbladians and their convergence

We dedicate this subsection to background on the evolution and convergence of open quantum systems described by a Lindbladian. Recall, a general Lindbladian is a continuous-time Markov chain acting on density operators

$$\mathcal{L}[\rho] = \sum_j J_j \rho J_j^\dagger - \frac{1}{2}\{J_j J_j^\dagger, \rho\} \quad \text{for some set of Lindblad operators} \quad \{J_j\}_j, \tag{4.16}$$

[17] Note that we are simply applying an affine transformation, $H = \frac{1}{2}\big(n\cdot\mathbb{I} + C\sum Z_i C^\dagger\big)$, such that the Gibbs state of H at temperature β is the same as that of $C\sum Z_i C^\dagger$ at temperature $\beta/2$.

which generates a family of completely positive and trace-preserving map

$$e^{\mathcal{L}t}[\rho] \quad \text{for each} \quad t \geq 0. \tag{4.17}$$

Our Linbladians of interest satisfy a particular property known as detailed balance.

Definition 4.1 (s-Inner Product). *Fix a full rank density matrix σ and $s \in [0,1]$. We define the weighted Hilbert-Schmidt inner product:*

$$\langle A, B\rangle_s = \langle A, \sigma^{1-s}B\sigma^s\rangle = \mathrm{Tr}\big[A^\dagger\sigma^{1-s}B\sigma^s\big] \quad \textit{for each} \quad A, B. \tag{4.18}$$

Definition 4.2 (s-Detailed Balance). *A Lindbladian $\mathcal{L}$ is s-*Detailed Balance *with respect to σ if $\mathcal{L}^\dagger$ is self-adjoint with respect to $\langle\cdot,\cdot\rangle_s$:*

$$\forall A, B : \langle A, \mathcal{L}^\dagger[B]\rangle_s = \langle \mathcal{L}^\dagger[A], B\rangle_s \tag{4.19}$$

There are two important structural consequences of this detailed balance condition. The first is that the density operator σ is a fixed point of Lindbladian evolution:

$$\mathcal{L}[\sigma] = 0 \tag{4.20}$$

The second, as discussed shortly, is that it implies a powerful means to understand the convergence of the mixing process. For the reader most familiar with classical Markov chains, the detailed balance condition above is an analog to its classical counterpart, however, with an additional degree of freedom $0 \leq s \leq 1$ which arises due to non-commutativity.

Two special cases of the above are the GNS (where $s = 1$) and KMS ($s = 1/2$) detailed balance conditions. Fortunately, under minor constraints on the family of Lindbladians (which our Lindbladian satisfies), all these definitions collapse. We refer the reader back to Eq. (4.6) for the definition of the family of Linbladians we consider, Davies Generators.

Fact 4.1 (Davies' generators are detailed balanced). *Consider the Davies generator $\mathcal{L}$ described in Eq. (4.6), subject to constraint that the transition weights satisfy $\forall\nu : \gamma(\nu)/\gamma(-\nu) = e^{-\beta\nu}$, and the jump operators contain their adjoints $\{A_a\} = \{A_a^\dagger\}$. Then, $\mathcal{L}$ satisfies s-DB $\forall s \in [0,1]$ w.r.t. the Gibbs state $\rho_\beta \propto e^{-\beta H}$.*

In this manner, the Gibbs state $\rho_\beta \propto e^{-\beta H}$ is a fixed point of the Davies generator we designed in Section 4.2.4.1. However, it may not be the unique stationary state, nor may its evolution converge rapidly. To understand the rate of convergence of this process, we need a bound on its mixing time $t_{mix}(\mathcal{L})$. Physically, the mixing time provides an estimate for the thermalization time of the system.

Definition 4.3 (Mixing time). *The mixing time $t_{mix}(\mathcal{L})$ of a Lindbladian $\mathcal{L}$ is the smallest time $t \geq 0$ for which*

$$\|e^{t\mathcal{L}}(\rho_1 - \rho_2)\|_1 \leq \frac{1}{2}\|\rho_1 - \rho_2\|_1 \quad \textit{for any two states } \rho_1, \rho_2. \tag{4.21}$$

In what remains of this subsection, we describe two means to analyze t_{mix}. The first of which consists of a bound on the spectral gap of $\mathcal{L}$. Apriori, however, the super-operator $\mathcal{L}$ is not even Hermitian, and its spectral gap may not even be well-defined. Fortunately, under an appropriate similarity transformation, we can appeal to a related Hermitian quantity known as *the discriminant*:

Definition 4.4 (Quantum discriminant)**.** *Fix $s \in [0,1]$ and a full-rank density matrix σ. The discriminant $\mathcal{K}_s$ of $\mathcal{L}$ consists of the super-operator*

$$\mathcal{K}_s(\cdot) = \sigma^{-\frac{1-s}{2}} \mathcal{L}\left(\sigma^{\frac{1-s}{2}} \cdot \sigma^{\frac{s}{2}}\right) \sigma^{-\frac{s}{2}}. \tag{4.22}$$

Lemma 4.5 ([243], Lemma 5 and 7)**.** *The discriminant $\mathcal{K}_s$ of $\mathcal{L}$ satisfies the following properties*

1. *$\mathcal{L}$ satisfies s-DB if and only if $\mathcal{K}_s$ is Hermitian.*
2. *If $\mathcal{L}$ satisfies s-DB, then the eigenvalues of $\mathcal{L}$ are the same as that of $\mathcal{K}_s$, which are real.*
3. *If $\mathcal{L}$ is a Davies generator satisfying the constraints of Fact 4.1, then $\mathcal{K} \equiv \mathcal{K}_s$ is independent of $s \in [0,1]$.*

The spectral gap $\Delta(\mathcal{L}) = \Delta(\mathcal{K}_s)$ of a given Lindbladian is defined to be that of the associated discriminant. Analyzing this gap can be a challenging task, and concrete bounds are often case-dependent. Nevertheless, it provides a powerful means to control the convergence of the time-evolution.

Lemma 4.6 (Mixing time from the Spectral Gap, [234])**.** *If a Lindbladian $\mathcal{L}$ satisfies KMS reversibility with fixed point σ, then*

$$t_{mix}(\mathcal{L}) \leq \frac{\log\big(2\|\sigma^{-1/2}\|\big)}{\Delta(\mathcal{L})}, \tag{4.23}$$

We remark that the dependence on $\log\|\sigma^{-1/2}\| \approx O(\beta n)$ often-times incurs a polynomial overhead to the mixing time. The notion of a (modified) Log Sobolev inequality provides a significantly stronger means of analyzing the mixing time. To formalize this method, we first require the definition of the conditional expectation of an operator X, $\mathcal{E}[X] = \lim_{t\to\infty} e^{t\mathcal{L}}[X]$.

Definition 4.5 (Modified Logarithmic Sobolev inequality)**.** *The Markov semigroup $(e^{t\mathcal{L}})_{t\geq 0}$ satisfies a* Modified Logarithmic Sobolev inequality *(MSLI) with constant α if*

$$\frac{d}{dt}D\bigg(e^{t\mathcal{L}}[\rho]||\mathcal{E}[\rho]\bigg)\bigg|_{t=0} = \mathrm{Tr}\,\mathcal{L}[\rho](\log\rho - \log\mathcal{E}[\rho]) \leq -\alpha\cdot D(\rho||\mathcal{E}[\rho]) \quad \textit{for each} \quad \rho, \tag{4.24}$$

where $D(\rho||\sigma) = \mathrm{Tr}\,\rho(\log\rho - \log\sigma)$ is the quantum relative entropy.

In other words, a MLSI quantifies the decay of the relative entropy, which converts to a bound on the mixing time through Pinsker's inequality.

Lemma 4.7 (Mixing time from MLSI, [234])**.** *If a Lindbladian $\mathcal{L}$ satisfies KMS-detailed balance with fixed point σ and a MLSI with constant α, then*

$$t_{mix}(\mathcal{L}) \leq \frac{2 \cdot \log(4 \cdot \log \|\sigma^{-1}\|)}{\alpha} \tag{4.25}$$

This *polylogarithmic* overhead in system size is known as *rapid mixing.* Moreover, if given an additional entangled reference system R the semigroup $(e^{t\mathcal{L}} \otimes \mathbb{I}_R)_{t\geq 0}$ satisfies an MSLI, then $\mathcal{L}$ is said to satisfy a *complete* modified logarithmic Sobolev inequality (CMLSI).

4.2.5.2 Analysis

The main result of this subsection is a bound on the mixing time of our family of Lindbladians.

Lemma 4.8. *The mixing time of our family of Lindbladians $\mathcal{L}$ defined in Eq. (4.6) is bounded by*

$$t_{mix}(\mathcal{L}) = O(4^\ell \cdot e^\beta \cdot \log n). \tag{4.26}$$

The starting point of our analysis is based on that of a much simpler Lindbladian, namely, that corresponding to the trivial circuit $C = \mathbb{I}$. In this setting, both the associated parent Hamiltonian, and the associated Lindbladian, are a sum over non-interacting, single-qubit terms:

$$H_{\mathsf{NI}} = \sum_i |1\rangle\langle 1|_i \text{ and } \mathcal{L}_{\mathsf{NI}} = \sum_{i\in[n]} \mathcal{L}^i_{single}, \quad \text{where} \quad \mathcal{L}^i_{single}[\sigma^i_\beta] = 0 \quad \text{and} \quad \sigma^i_\beta \propto e^{-\beta|1\rangle\langle 1|_i}. \tag{4.27}$$

The jump operators of $\mathcal{L}^i_{single}$ are simply single-qubit Pauli operators, and the single-qubit Gibbs state σ^i_β is its fixed point. Using now standard techniques, one can prove that this non-interacting Lindbladian is both gapped and mixes rapidly:

Claim 4.3 (The Non-Interacting Lindbladian is rapidly mixing)**.** *The non-interacting Lindbladian $\mathcal{L}_{\mathsf{NI}}$ has a constant spectral gap $\Delta(\mathcal{L}_{\mathsf{NI}}) \geq 4^{-1}$ and satisfies a MSLI with constant $\alpha_{\mathsf{NI}} = \Omega(e^{-\beta})$.*

The *unique* fixed point of $\mathcal{L}_{NI}$ is thus the tensor product state $\sigma_\beta = \otimes_i \sigma^i_\beta \propto e^{-\beta H_{\mathsf{NI}}}$. We defer a proof of Claim 4.3 to the next subsection. In the rest of this subsection, we show how to relate our Lindbladian $\mathcal{L}$ of Eq. (4.6) (implicitly defined by the quantum circuit C), to $\mathcal{L}_{\mathsf{NI}}$, and moreover how to inherit its rapid mixing properties.

Claim 4.4 (A Convex Combination of Lindbladians)**.** *In a basis rotated by C, the Lindbladian $\mathcal{L}$ can be written as the convex combination*

$$C^{\dagger}\mathcal{L}[C\cdot C^{\dagger}]C = q\cdot\mathcal{L}_{\mathsf{NI}}[\cdot] + (1-q)\cdot\mathcal{L}_{rest}[\cdot], \tag{4.28}$$

of two Lindbladians $\mathcal{L}_{\mathsf{NI}}, \mathcal{L}_{rest}$ which share the fixed point $\sigma_\beta = \otimes_i\sigma_\beta^i$. Moreover, the parameter $q = 4^{1-\ell}$ depends only on the lightcone size of C.

A proof of which we also defer to a future subsection. The convex combination claim above is the heart of our analysis, as it enables us to inherit the gap and mixing properties of $\mathcal{L}_{NI}$, without knowing properties of $\mathcal{L}_{rest}$ except for its (common) fixed point. To conclude this subsection, we present a proof of the MLSI of $\mathcal{L}$:

Claim 4.5 (The Modified Log-Sobolev Inequality)**.** *The Lindbladian $\mathcal{L}$ satisfies a MSLI with constant $\alpha \geq q\cdot\alpha_{NI} = \Omega(4^{-\ell}\cdot e^{-\beta})$.*

Proof. [of Claim 4.5] From Claim 4.4, we can write our Lindbladian $\mathcal{L}$ in a basis rotated by the circuit C as a convex combination

$$\tilde{\mathcal{L}} = C^{\dagger}\mathcal{L}[C\cdot C^{\dagger}]C = q\cdot\mathcal{L}_{\mathsf{NI}}[\cdot] + (1-q)\cdot\mathcal{L}_{rest}[\cdot] \tag{4.29}$$

Since relative entropy is basis independent, proving a MLSI for $\tilde{\mathcal{L}}$ similarly implies one for $\mathcal{L}$ with the same constant. To do so, we begin by expressing the "entropy production rate" as a convex combination.

$$\begin{aligned}\mathsf{EP}_{\tilde{\mathcal{L}}}(\rho) &= \mathrm{Tr}\Big[\tilde{\mathcal{L}}[\rho](\log\rho - \log\sigma_\beta)\Big] = \qquad (4.30)\\ &= q\,\mathrm{Tr}[\mathcal{L}_{\mathsf{NI}}[\rho](\log\rho-\log\sigma_\beta)] + (1-q)\,\mathrm{Tr}[\mathcal{L}_{rest}[\rho](\log\rho-\log\sigma_\beta)] \qquad (4.31)\end{aligned}$$

To the first term on the RHS above, we can simply apply the MLSI for the non-interacting Linbladian Claim 4.3:

$$\mathrm{Tr}[\mathcal{L}_{\mathsf{NI}}[\rho](\log\rho-\log\sigma_\beta)] \leq -\alpha_{\mathsf{NI}}\cdot D(\rho||\sigma_\beta). \tag{4.32}$$

In turn, we claim that the second term on the RHS above is non-positive. Indeed, note that by Claim 4.4, σ_β is a fixed point of $\mathcal{L}_{rest}$. The Data-processing inequality for the relative entropy then tells us that

$$\mathrm{Tr}[\mathcal{L}_{rest}[\rho](\log\rho-\log\sigma_\beta)] = \frac{d}{dt}D(e^{t\mathcal{L}_{rest}}[\rho]||\sigma_\beta)\bigg|_{t=0} = \frac{d}{dt}D(e^{t\mathcal{L}_{rest}}[\rho]||e^{t\mathcal{L}_{rest}}[\sigma_\beta])\bigg|_{t=0} \leq 0. \tag{4.33}$$

Put together, we conclude $\mathsf{EP}_{\tilde{\mathcal{L}}}(\rho) \leq -q\cdot\alpha_{\mathsf{NI}}\cdot D(\rho||\sigma_\beta)$. □

4.2.5.3 The non-interacting Lindbladian is gapped (Claim 4.3)

We dedicate this subsection to an analysis of the non-interacting Lindbladian $\mathcal{L}_{NI}$ (Claim 4.3).

Lemma 4.9. *The spectral gap of the single-qubit Lindbladian $\mathcal{L}_{single}$ is $\Delta(\mathcal{L}_{single}) \geq 4^{-1}$.*

To understand this spectral gap, we revisit the (Hermitian) Discriminant super-operator $\mathcal{K}$ defined in Definition 4.4. Recall, from Lemma 4.5, that (under detailed balance) this super-operator has the same eigenvalues of $\mathcal{L}$. In turn, to understand the spectral gap of $\mathcal{K}$, we vectorize this super-operator (on 2×2 matrices) into an operator (a 4×4 matrix).

$$\mathcal{K}[\cdot] = \sum_j A_j[\cdot]B_j \to \mathbf{K} = \sum_j A_j \otimes B_j^T. \tag{4.34}$$

Proof. To analyze the gap, we consider the discriminant $\mathcal{K}_{single}$ of the Lindbladian $\mathcal{L}_{single}$, and in particular its vectorization:

$$\mathbf{K}_{single} = \sum_{a\in\mathcal{A}}\sum_{\nu\in[-n,n]} -\sqrt{\gamma(\nu)\gamma(-\nu)}\cdot A^a_\nu \otimes (A^a_\nu)^* + \frac{\gamma(\nu)}{2}\Big((A^a_\nu)^\dagger A^a_\nu \otimes \mathbb{I} + \mathbb{I}\otimes (A^a_\nu)^T(A^a_\nu)^*\Big) \tag{4.35}$$

Which is PSD, frustration free, and preserves the eigenvalues of $\mathcal{L}_{single}$ (up to a factor of -1) Lemma 4.5. Moreover, via detailed balance, the purified Gibbs state $|\sqrt{\sigma_\beta}\rangle \propto |00\rangle + e^{-\beta/2}|11\rangle$ is a ground state of $\mathbf{K}_{single}$. Since the jump operators are single qubit Pauli operators $\{\mathbb{I}, X, Y, Z\}$, they can be written in the energy basis as

$$A_0^{\mathbb{I}} \propto \mathbb{I} \quad \text{and} \quad A_0^Z \propto Z_i \quad \text{and} \quad A_1^X = (-i)A_1^Y \propto \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \tag{4.36}$$

and that the conjugates can be inferred from the identity $A^a_\nu = A^{a\dagger}_{-\nu}$. The 4×4 vectorized discriminant can therefore be written as

$$\mathbf{K}_{single} = \frac{1}{2}\begin{bmatrix} \gamma(1) & 0 & 0 & -\sqrt{\gamma(1)\gamma(-1)} \\ 0 & \frac{\gamma(1)+\gamma(-1)}{2}+\gamma(0) & 0 & 0 \\ 0 & 0 & \frac{\gamma(1)+\gamma(-1)}{2}+\gamma(0) & 0 \\ -\sqrt{\gamma(1)\gamma(-1)} & 0 & 0 & \gamma(-1) \end{bmatrix}. \tag{4.37}$$

Which we identify to be frustration free and have spectral gap $\frac{\gamma(1)}{2}\cdot\min(1+e^\beta, \frac{1+e^\beta}{2}+\frac{\gamma(0)}{\gamma(1)}) = \frac{\gamma(1)}{2} \geq \frac{1}{4}$ under Glauber Dynamics, where $\gamma(\nu) = (1+e^{-\beta\nu})^{-1}$.

□

The positivity of the spectral gap can be used to show a complete MLSI, as shown by [250]. This conversion comes at the cost of factors of the local dimension of the Lindbladian - which in the case of $\mathcal{L}_{single}$, is just 2.

Theorem 4.5 (CMLSI from the Spectral Gap, [250] Theorem 4.3)**.** *Suppose a Lindbladian $\mathcal{G}$, acting on a D-dimensional Hilbert space, is GNS-symmetric w.r.t a fixed state $\sigma > 0$. Then, it satisfies a CMSLI with constant*

$$\alpha_c \geq \Delta(\mathcal{G}) \cdot \frac{\|\sigma^{-1}\|^{-1}}{D^2}. \tag{4.38}$$

In this manner, $\mathcal{L}_{single}$ satisfies a CMSLI with constant $\alpha_{single} = \frac{1}{16\cdot(1+e^{\beta})}$. We are now in a position to prove the MLSI for $\mathcal{L}_{NI}$.

Proof. [of Claim 4.3] We begin by leveraging the Complete MLSI, on the local Lindbladians (Theorem 4.5 and Lemma 4.9):

$$\begin{aligned}
\mathrm{Tr}[\mathcal{L}_{NI}[\rho](\log(\rho)-\log(\sigma_\beta))] &= \sum_{i\in[n]} \mathrm{Tr}\big[\mathcal{L}^i_{single}[\rho](\log(\rho)-\log(\sigma_\beta))\big] &(4.39)\\
&= \sum_{i\in[n]} \mathrm{Tr}\big[\mathcal{L}^i_{single}[\rho](\log(\rho)-\log(\mathcal{E}_i[\rho]))\big] &(4.40)\\
&\leq -\alpha_{single}\sum_{i\in[n]} D(\rho||\mathcal{E}_i[\rho]) &(4.41)
\end{aligned}$$

Where $\mathcal{E}_i$ is the conditional expectation of the ith semigroup $e^{t\mathcal{L}^i_{single}}$. Next, we leverage the strong subadditivity of non-interacting conditional expectations [250] (eq. 5), see also [251]:

$$\sum_{i\in[n]} D(\rho||\mathcal{E}_i[\rho]) \geq D(\rho||\prod_i \mathcal{E}_i[\rho]). \tag{4.42}$$

To conclude, observe that for any ρ, the collection of conditional expectation $\prod_{i\in[n]}\mathcal{E}_i[\rho] = \sigma_\beta$ maps to the unique stationary state. □

4.2.5.4 The convex combination claim (Claim 4.4)

Recall that the parent Hamiltonian H has solvable eigenstates given by $\{C|x\rangle : x\in\{0,1\}^n\}$, with energies given by the hamming weight $|x|\in[n]$.

Proof. We begin by explicitly writing down each jump operator in the frequency basis:

$$\begin{aligned}
A^a_\nu = \sum_k \Pi_{k+\mu}A^a\Pi_k &= C\Bigg(\sum_{k\in[n]}\sum_{\substack{|x|=k\\|y|=k+\nu}} |y\rangle\langle x|\cdot\langle y|C^\dagger A^a C|x\rangle\Bigg)C^\dagger &(4.43)\\
&=: C\Bigg(\sum_{k\in[n]}\sum_{\substack{|x|=k\\|y|=k+\nu}} |y\rangle\langle x|\cdot N^a_{x,y}\Bigg)C^\dagger &(4.44)
\end{aligned}$$

For conciseness, we have denoted the coefficient by $N^a_{x,y} = \langle y| C^\dagger A^a C |x\rangle$. Our Lindbladian $\mathcal{L}$ of Eq. (4.6) can thus be written in a basis rotated by the circuit C, in terms of the second moment of these coefficients:

$$C^\dagger \mathcal{L}[C\rho C^\dagger]C = \sum_\nu \gamma_\nu \cdot \sum_a \Bigg(\sum_{k,k'\in[n]} \sum_{\substack{|x|=k \\ |y|=k+\nu}} \sum_{\substack{|x'|=k' \\ |y'|=k'+\nu}} N^a_{x,y} \cdot (N^a_{x',y'})^* \cdot |y\rangle \langle x| \rho |x'\rangle \langle y'| \tag{4.45}$$

$$-\frac{1}{2} \sum_{k\in[n]} \sum_{\substack{|x|,|x'|=k \\ |y|=k+\nu}} N^a_{x,y} \cdot (N^a_{x',y})^* \cdot \Big\{ |x'\rangle \langle x| , \rho \Big\} \Bigg). \tag{4.46}$$

For $i \in [n]$, consider the subset of jump operators $\mathcal{A}_i$, centered around the i-th lightcone L_i:

$$\mathcal{A}_i = 2^{-\ell} \cdot \big\{ P_{\mathsf{L}_i} \otimes \mathbb{I}_{[n]\setminus \mathsf{L}_i} : P \in \mathcal{P}_\ell \big\} \tag{4.47}$$

By definition, these subsets are disjoint, and form a partition $\cup_i \mathcal{A}_i = \mathcal{A}$. We claim that we can rotate the jump operators in each subset, by substituting

$$\mathcal{A}_i \to \mathcal{A}_i' = U_i \mathcal{A}_i U_i^\dagger \tag{4.48}$$

for an *arbitrary* choice of unitary U_i of support contained in L_i, while keeping the Lindbladian $\mathcal{L}$ invariant. Essentially, this is because the Lindbladian is only defined by the second moments of the jump operators, and that the second moment of random Pauli operators is Haar random (via the 1-design property) and thus invariant under unitary conjugation:

$$\sum_{a\in\mathcal{A}_i} A^a[\cdot]A^a = \frac{1}{2^\ell} \operatorname{tr}_{\mathsf{L}_i}[\cdot] = \sum_{a\in\mathcal{A}_i} U_i^\dagger A^a U_i [\cdot] U_i^\dagger A^a U_i \quad \text{for any } U_i \text{ supported on } \mathsf{L}_i. \tag{4.49}$$

Indeed, for every choice of basis elements $x, x', y, y' \in \{0,1\}^n$, the pre-factor

$$\sum_{a\in\mathcal{A}_i} N^a_{x,y} \cdot (N^a_{x',y'})^* = \sum_{a\in\mathcal{A}_i} \langle y| C^\dagger A_a C |x\rangle \langle x'| C^\dagger A_a C |y'\rangle \tag{4.50}$$

$$= \sum_{a\in\mathcal{A}_i} \langle y| C^\dagger U_i A_a U_i^\dagger C |x\rangle \langle x'| C^\dagger U_i A_a U_i^\dagger C |y'\rangle = \sum_{a'\in\mathcal{A}_i'} N^{a'}_{x,y} \cdot (N^{a'}_{x',y'})^* \tag{4.51}$$

is preserved, whether in $\mathcal{A}_i$ or $\mathcal{A}_i'$.

Finally, let U_i be the gates in C contained in the lightcone of the ith qubit. The sum over jump operators $a \in \mathcal{A}$ can be written as an expectation over random ℓ-qubit Paulis, $P \in \mathcal{P}_\ell$, and then an expectation over the center $i \in [n]$ in which to place P. With probability $4/4^\ell$, P is a single qubit Pauli P_i centered at i. Moreover, for a single-qubit Pauli P_i centered at

i, the choice of U_i *exactly cancels with the circuit* $C^\dagger$:

$$\langle y| C^\dagger U_i \bigg(P_i \otimes \mathbb{I}_{[n]\setminus\{i\}} \bigg) U_i^\dagger C |x\rangle = \langle y| \bigg(P_i \otimes \mathbb{I}_{n\setminus\{i\}} \bigg) |x\rangle \,, \tag{4.52}$$

$$\text{for each} \quad x, y \in \{0,1\}^n \quad \text{and} \quad P_i \in \{\mathbb{I}, X, Y, Z\}_i. \tag{4.53}$$

We note that these are precisely the jump operators we expect in the non-interacting case $\mathcal{L}_{\mathsf{NI}}$, where the circuit is replaced by the trivial circuit $C = \mathbb{I}$. In this manner, we conclude that the rotated Lindbladian can be written a convex combination:

$$C^\dagger \mathcal{L}[C \cdot C^\dagger] C = 2^{2(1-\ell)} \cdot \mathcal{L}_{\mathsf{NI}}[\cdot] + (1 - 2^{2(1-\ell)}) \cdot \mathcal{L}_{rest}[\cdot], \tag{4.54}$$

where both $\mathcal{L}_{\mathsf{NI}}[\cdot]$, $\mathcal{L}_{rest}[\cdot]$ are Davies' generators defined on disjoint sets of jump operators. Both of them satisfy detailed balance and share the Gibbs state as the stationary state. □

4.2.6 Circuit Implementation of the Dissipative Lindbladian

The main claim of this section is an efficient implementation of the Lindbladian time-evolution using a quantum circuit. Put together with our bound on the mixing time of our Lindbladians, this all but concludes the proof of the preparation of Gibbs states of the parent Hamiltonians of quantum circuits.

Lemma 4.10 (Dissipative Lindbladian Implementation)**.** *Fix parameters $t \geq 1$ and $\varepsilon \leq \frac{1}{2}$. Let $\mathcal{L}$ denote the Lindbladian of Eq. (4.6), defined by a quantum circuit on n qubits, of lightcone size ℓ and depth d. Then, we can simulate the map $e^{t\mathcal{L}}$ to error ε in diamond norm using a quantum algorithm of depth $O(t \cdot n \cdot 2^{2\ell} \cdot 2^d \cdot \mathrm{poly}(\ell, \log n, \log \frac{1}{\varepsilon}, \log t))$.*

We dedicate Section 4.2.6.1 to presenting the required background results on implementing Lindbladian evolution using quantum circuits. In the ensuing section Section 4.2.6.2, we discuss optimizations both particular to our systems, and generic, to the runtime of our algorithms.

4.2.6.1 Preliminaries on simulating Lindbladian evolution

Our implementation of the map $e^{t\mathcal{L}}$ follows the framework of [232], in reducing the task to constructing a block-encoding of the Lindblad operators. To implement their scheme it is suitable to renormalize the time-scale and Lindblad operators $\{L_j\}_{j\in\mathcal{A}}$ such that the resulting Lindbladian has norm 1:

$$t \to t \cdot \| \sum_j L_j^\dagger L_j \| \text{ and } L_j \to L_j \cdot \| \sum_j L_j^\dagger L_j \|^{-1/2} \tag{4.55}$$

Given the choice of jump operators from Eq. (4.6), under this normalization we have $t \to n \cdot t$.

Definition 4.6 (Unitary block encoding for Lindblad Operators [232, Definition I.2])**.** *Given a purely irreversible Lindbladian determined by the Lindblad operators $\{L_j\}_{j\in\mathcal{A}}$, a unitary U is said to be block-encoding of the Lindblad operators if*

$$\big(\langle 0|^b \otimes \mathbb{I}\big)U\big(|0\rangle^c \otimes \mathbb{I}\big) = \sum_{a\in\mathcal{A}} |a\rangle \otimes L_j \text{ for } b, c \in \mathbb{N} \tag{4.56}$$

Given a black-box circuit corresponding to a block-encoding of $\mathcal{L}$, the following theorem stipulates that one can simulate the corresponding Lindbladian evolution for time t using just $\tilde{O}(t)$ invocations of the black-box:

Theorem 4.6 (Theorem I.2, [232])**.** *Suppose U is a unitary block-encoding of the Lindbladian $\mathcal{L}$ as in Definition 4.6. Let time $t \geq 1$ and error $\varepsilon \leq \frac{1}{2}$, then we can simulate the map $e^{t\mathcal{L}}$ to error ε in diamond norm using*

1. *$O((c + \log\frac{t}{\varepsilon})\log\frac{t}{\varepsilon}))$ resettable ancilla qubits,*
2. *$\tilde{O}(t)$ controlled uses of U and $U^\dagger$, and*
3. *$\tilde{O}(t + c)$ other 2-qubit gates.*

We remark that since our Hamiltonian has a integer spectra $[n]$, one can exactly implement the projection of the jump operators $\{A^a\}$ onto the energy eigenbasis by performing an operator fourier transform with uniform weights:

$$A^a_\nu \propto \sum_{\bar{t}\in S_{\pi/n}} e^{i\nu\bar{t}} e^{iH\bar{t}} A^a e^{-iH\bar{t}} \text{ where } S_{\pi/n} = \frac{\pi}{n}\cdot\{-n, -(n-1), \cdots, -1, 0, 1, \cdots, n\} \tag{4.57}$$

In this setting, we can now apply a lemma on the efficient implementation of block-encodings from [232], simplified to the context of integer spectra Hamiltonians.

Lemma 4.11 (Lemma I.1, [232])**.** *In the setting of Theorem 4.6, a unitary block encoding for the Lindblad operators corresponding to a Hamiltonian H of integer spectra $[n]$ can be created using $O(n + \log|A|)$ ancilla qubits, as well as one query to*

1. *The controlled Hamiltonian simulation: $\sum_{\bar{t}\in S_{\pi/n}} |\bar{t}\rangle\langle\bar{t}| \otimes e^{\pm iH\bar{t}}$,*
2. *A block-encoding of the jump operators: $\sum_{a\in\mathcal{A}} |a\rangle \otimes A^a$,*
3. *$O(\log n)$ qubit Quantum Fourier transform: $|\bar{t}\rangle \to (2n)^{-1/2}\sum_{\omega\in[-n,\cdots n]} e^{-i\omega\bar{t}}|\omega\rangle$*
4. *And a controlled filter for the Boltzmann factors:*

$$W = \sum_{\omega\in[-n,\cdots,n]} \begin{bmatrix} \sqrt{\gamma(\omega)} & -\sqrt{1-\gamma(\omega)} \\ \sqrt{1-\gamma(\omega)} & \sqrt{\gamma(\omega)} \end{bmatrix} \otimes |\omega\rangle\langle\omega| \tag{4.58}$$

4.2.6.2 Optimizing the circuit implementation

In light of Theorem 4.6 and Lemma 4.11, in what remains of this section, we describe how to implement the controlled Hamiltonian simulation (Claim 4.6), the block-encoding of the jump operators (Claim 4.7), and the controlled Boltzmann filter (Claim 4.8), in circuit depth $O(4^\ell \cdot 2^d \cdot \text{poly}(\log n, \log \frac{1}{\varepsilon}, \ell))$. While the first two optimizations are particular to our family of Hamiltonians, the latter may find independent application to the framework of [232].

We begin with a simple lemma which "colors" the interaction graph of the Hamiltonian, partitioning the interactions into disjoint subsets $S_1, S_2 \cdots S_\Delta \subset [n]$ such that no two terms h_i, h_j of the same subset have overlapping support.

Lemma 4.12. *Any parent Hamiltonian $H \in \mathscr{H}$ defined by a quantum circuit of depth d and lightcone size ℓ can be Δ-colored with $\Delta \leq \ell \cdot 2^d + 1$ colors.*

Proof. Two interactions h_i, h_j overlap at a qubit only if their lightcones intersect in the underlying circuit C, which determines H. Let ℓ_r denote the maximum "reverse lightcone" size of the circuit, that is, the maximum number of qubits which have a given qubit in their lightcone. Since the Hamiltonian is at most ℓ-local, any interaction h_i overlaps with at most $\ell \cdot \ell_r$ other terms, which in turn tells us the interactions can then be partitioned using $\Delta \leq \ell \cdot \ell_r + 1$ different colors. If the depth of the circuit C as measured by layers of 2-qubit gates is d, then $\ell_r \leq \min(n, 2^d)$. □

Claim 4.6. *The controlled time-evolution of an n qubit parent Hamiltonian of a quantum circuit with lightcone size ℓ, can be implemented using a quantum circuit of depth $O(4^\ell \cdot \Delta \cdot \log n)$ and size $O(4^\ell \cdot \Delta \cdot n \cdot \log n)$.*

At a high level, the circuit of Claim 4.6 partitions the terms of the (commuting) Hamiltonian into disjoint subsets of non-overlapping terms, which can be implemented in parallel. However, since we need to implement the *controlled* Hamiltonian simulation, all of these Hamiltonian terms need to act conditioned on the time-register, which is a sequential bottleneck to the circuit depth. In order to further compress the depth, we parallelized the access to the time-register by encoding it into a GHZ state.

Proof. Since the Hamiltonian is commuting, let us restrict our attention to a fixed subset S in the partition guaranteed by Lemma 4.12. It suffices to prove how to implement the controlled time evolution of each subset of non-overlapping terms $H_S = \sum_{i\in S} h_i$. For this purpose, we begin by parallelizing the access to the clock register: $|\bar{t}\rangle \rightarrow |\bar{t}\rangle^{\otimes n}$, using a $O(\log n)$ depth circuit of CNOT gates, of size $O(n \log n)$.

Next, controlled on the jth clock register, we apply the time evolution of the jth interaction. Although this gate acts on $O(\log n) + \ell$ qubits, it can be implemented via a sequence of $O(\log n)$ gates acting only on $\ell + 1$ qubits, by applying a binary expansion of the time register $\bar{t} = \frac{\pi}{n} \cdot \sum_k 2^k \bar{t}_k$:

$$\sum_{\bar{t}} |\bar{t}\rangle\langle\bar{t}| \otimes e^{i\bar{t}h_j} = \prod_k \mathbb{I}_{\setminus k} \otimes \sum_{\bar{t}_k \in \{0,1\}} |\bar{t}_k\rangle\langle\bar{t}_k| \otimes \exp\Big[i\frac{\pi}{n} \cdot 2^k \bar{t}_k \cdot h_j\Big] \tag{4.59}$$

In turn, each unitary on $\ell+1$ qubits can generically be implemented in $O(4^\ell)$ size and depth. After all the colors have concluded, we revert the copies of the clock register. □

Claim 4.7. *A block encoding of the jump operators can be implemented using a quantum circuit of depth $O(\Delta\cdot(\ell+\log n))$ and size $O(n\cdot\Delta\cdot(\ell+\log n))$.*

Proof. There are $|A| = n\cdot 4^\ell$ jump operators, which we represent by indexing them using a pair $a=(i,P)$ in terms of the center of its support $i\in[n]$, using $\log n$ qubits, as well as the ℓ-local Pauli P, using $2\cdot\ell$ qubits. Next we proceed using similar techniques to Claim 4.6. We begin by partitioning the jump operators into Δ disjoint subsets using Lemma 4.12, where any two jump operators in the same subset either have the same center, or do not intersect. Our implementation proceeds by addressing each color $c\in[\Delta]$ independently.

First, we create copies of the control register $|a\rangle\to|a\rangle^{\otimes n}$, to parallelize access to it. Suppose all the jump operators centered at j have been colored c. Our goal is to coherently apply all the controlled jump operators of the form (j,P) by acting only on the support (centered at j) and the jth control register $|a=(i,Q)\rangle_j$. For this purpose, we first check whether $i=j$, and controlled on the one-qubit outcome, we apply the Pauli Q. The check can be implemented using $O(\log n)$ size and depth, and the controlled Pauli in $O(\ell)$ size and depth. We conclude by inverting the checking and copying steps. □

Claim 4.8. *The controlled filter W can be implemented up to error ε in spectral norm using a circuit of size $O(\text{polylog}(\frac{n}{\varepsilon}))$ 2-qubit gates.*

Proof. Let us denote $n_\delta = \beta^{-1}\log\frac{1}{\delta}$. Then, the Glauber dynamics weight $\gamma(\nu) = (1+e^{-\beta\nu})^{-1}$ satisfies

$$\gamma(\nu)\leq\delta \qquad \text{if}\quad \nu\leq -n_\delta, \tag{4.60}$$

$$\text{and } \gamma(\nu)\geq 1-\delta \quad \text{if}\quad \nu\geq n_\delta. \tag{4.61}$$

We claim that the W gate can be replaced by a truncation W_δ,

$$W_\delta = \sum_{\omega\in[-n,\cdots,n]}\begin{bmatrix}\sqrt{\tilde\gamma(\omega)} & -\sqrt{1-\tilde\gamma(\omega)}\\ \sqrt{1-\tilde\gamma(\omega)} & \sqrt{\tilde\gamma(\omega)}\end{bmatrix}\otimes|\omega\rangle\langle\omega|\,,\quad \tilde\gamma(\omega):=\begin{cases}\gamma(\omega) & \text{if } \omega\in[-n_\delta,n_\delta]\\ 1 & \text{if } \omega>n_\delta\\ 0 & \text{if } \omega<-n_\delta\end{cases}. \tag{4.62}$$

Indeed, the truncation error is controlled by

$$\|W-W_\delta\|\leq\sum_{j\in[-n,-n_\delta]}\|\begin{bmatrix}\sqrt{\gamma(\omega)} & 1-\sqrt{1-\gamma(\omega)}\\ \sqrt{1-\gamma(\omega)}-1 & \sqrt{\gamma(\omega)}\end{bmatrix}\| \tag{4.63}$$

$$+\sum_{j\in[n_\delta,n]}\|\begin{bmatrix}\sqrt{\gamma(\omega)}-1 & -\sqrt{1-\gamma(\omega)}\\ \sqrt{1-\gamma(\omega)} & \sqrt{\gamma(\omega)}-1\end{bmatrix}\| \tag{4.64}$$

$$\leq 2n\cdot(2\sqrt{\delta}+2\delta)\leq 8n\cdot\sqrt{\delta}, \tag{4.65}$$

where the last line uses that $1-\sqrt{1-x}\leq x$ when $x\in[0,1/2]$.

It only remains now analyze the gate complexity of implementing W_δ. Following [232] (pg. 25, footnote 33), the W_δ filter for the Glauber weight between $[-n_\delta, n_\delta]$ can be implemented using the QSVT up to error ε using $\tilde{O}((1+\beta n_\delta)\text{polylog}\frac{1}{\varepsilon})$ 2-qubit gates. With the choice $\delta = O(\frac{\varepsilon}{n})$, we arrive at the advertised bounds by combining with the trivial cases $\omega\notin[-n_\delta, n_\delta]$. □

We remark that this error in spectral norm between unitaries is equivalent to the channel diamond norm distance, up to a constant: $\|U-V\|_\diamond\leq 2\cdot\|U-V\|$.

Put together, Claim 4.6, Claim 4.7 and Claim 4.8 imply Lemma 4.10.

4.2.7 The Input Noise Model and Gibbs States of Quantum Circuits

In this section, we show that the Gibbs states of parent Hamiltonians of quantum circuits correspond to noisy versions of the output of the quantum circuit, under a certain input noise model. To begin, let us recollect the noise model. Fix a noise rate $p\in(0,1)$. The single-qubit bit-flip error channel consists of the superoperator

$$\mathcal{D}_p(\sigma) = (1-p)\cdot\sigma + p\cdot X\sigma X. \tag{4.66}$$

Given a quantum circuit C on n qubits, the input noise model consists of independent applications of the bit-flip error channel on the input wires of C. In particular, the mixed state given by the output of the noisy circuit is:

$$\rho = C\bigg(\mathcal{D}_p(|0\rangle\langle 0|)\bigg)^{\otimes n} C^\dagger \tag{4.67}$$

For a fixed n qubit quantum circuit C, recall that we refer to the parent Hamiltonian of C as

$$H_C = C\bigg(\sum_{i\in[n]} |1\rangle\langle 1|_i\otimes\mathbb{I}_{[n]\setminus i}\bigg)C^\dagger \tag{4.68}$$

Lemma 4.13. *Fix $\beta>0$, and let H_C be the parent Hamiltonian of a quantum circuit C. The Gibbs state of H_C at inverse-temperature β is given by the output of the circuit C under input level noise with probability $p=(1+e^\beta)^{-1}$:*

$$\rho_\beta = \frac{e^{-\beta H_C}}{\operatorname{Tr} e^{-\beta H_C}} = C\bigg(\mathcal{D}_p(|0\rangle\langle 0|)\bigg)^{\otimes n} C^\dagger \tag{4.69}$$

Proof. It suffices to consider the Gibbs state σ_β of the Hamiltonian $H = \sum_{i\in[n]} |1\rangle\langle 1|_i$, as $\rho_\beta = C\sigma_\beta C^\dagger$. Since H is commuting, the partition function can be written as:

$$\operatorname{Tr} e^{-\beta H_C} = \operatorname{Tr} e^{-\beta H} = \sum_{x\in\{0,1\}^n} \prod_i^n \langle x_i|\, e^{-\beta|1\rangle\langle 1|_i}\, |x_i\rangle = \tag{4.70}$$

$$= \prod_i^n \sum_{x_i\in\{0,1\}} \langle x_i|\, e^{-\beta|1\rangle\langle 1|_i}\, |x_i\rangle = (1+e^{-\beta})^n. \tag{4.71}$$

Therefore, the Gibbs state of H can be expressed as the outcome of the depolarizing channel:

$$\sigma_\beta = (1+e^{-\beta})^{-n} \cdot e^{-\beta H} = \bigotimes_i^n \left(\frac{|0\rangle\langle 0|}{1+e^{-\beta}} + \frac{|1\rangle\langle 1|}{1+e^{\beta}} \right) = \Big(\mathcal{D}_p(|0\rangle\langle 0|)\Big)^{\otimes n}, \tag{4.72}$$

with $p = (1+e^\beta)^{-1}$.

□

4.2.8 Computational Complexity of Shallow IQP Sampling

In recent years several architectures have been proposed for achieving a quantum speedup, based on quantum processes which resemble or are equivalent to the IQP Circuit Sampling task discussed in Section 4.2.4.2. The basis for these speedups is on standard complexity-theoretic conjectures, including the non-collapse of the Polynomial Hierarchy, often in addition to strong assumptions on the hardness of computing permanents or partition functions. We dedicate this section to a discussion on the background behind Theorem 4.4, as well as a comparison to related statements in the literature.

To begin, let us recollect the circuit described in Section 4.2.4.2, comprised of a 2D cluster state and random phase gates [119, 118].

1. Prepare an n qubit cluster state on a 2D rectangular lattice using a layer of Hadamard gates and 4 layers of CZ gates.

2. Sample a random string $b \in [7]^n$, and apply powers of T gates to each qubit:

$$\bigotimes_{i\in[n]} T^{b_i} \left(\prod_{<i,j>} CZ_{i,j} \,|+\rangle^{\otimes n} \right) = C_b\, |0\rangle^{\otimes n}, \text{ where } T = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}. \tag{4.73}$$

3. Finally, measure the output in the X basis.

Figure 4.5: A family of random IQP circuits, $\{C_b\}$.

If instead of *random* powers of single-qubit T gates, the powers were chosen adaptively given partial measurements of the circuit, this scheme would implement measurement-based quantum computation [242]. The universality of MBQC under adaptivity (or post-selection) implies the hardness of *exactly* sampling from the output distribution, unless the polynomial hierarchy collapses to the third level [246, 118]. To reproduce their argument, universality implies

$$\mathsf{PostIQP} \underbrace{=}_{[246]} \mathsf{PostBQP} \underbrace{=}_{[252]} \mathsf{PP}. \tag{4.74}$$

If we now assume there existed a classical algorithm to exactly sample from arbitrary IQP circuits, that would imply $\mathsf{PP} = \mathsf{PostIQP} \subseteq \mathsf{PostBPP}$, which in turn gives us a collapse of the Polynomial Heirarchy (henceforth, PH):

$$\mathsf{PH} \underbrace{=}_{\text{Toda's Theorem}} P^{\mathsf{PP}} \underbrace{=}_{\text{By assumption}} P^{\mathsf{PostBPP}} = \Sigma_3. \tag{4.75}$$

In fact, by similar reasoning [246] (Theorem 2) showed that no classical algorithm can even *weakly* approximately sample from IQP circuits - i.e. up to some fixed multiplicative error. To extend these hardness results to approximate sampling (up to some additive error) in total variation distance, we require stronger assumptions.

[249] were the first to show that, assuming an additional complexity-theoretic conjecture on the average-case hardness of computing partition functions, approximately sampling from the output of IQP circuits remains classically intractable even up to small total variation distance. They noted that the output distribution of IQP circuits,

$$p_x = |\langle x| C |0\rangle^{\otimes n}|^2 = 2^{-n} \cdot \left|\mathcal{Z}_x\right|^2, \tag{4.76}$$

precisely resembles a complex-valued partition function, defined by $w_{u,v}, w_u$ real-valued edge and vertex weights on some underlying architecture graph G:

$$\mathcal{Z}_x = \sum_{z\in\{\pm 1\}^n} \exp\left[i\bigg(\sum_{<u,v>} w_{uv} z_u z_v + \sum_u (\pi \cdot x_u + w_u) z_u\right]. \tag{4.77}$$

They prove that approximating $\left|\mathcal{Z}_x\right|^2$, and therefore p_x, up to multiplicative error is $\#P$ hard in the worst-case, and pose as a conjecture its hardness in the average case over x. Under this conjecture, [249] show that the existence of an efficient classical algorithm to approximately sample from $\{p_x\}$, even up to constant TVD, would imply a collapse of the polynomial heirarchy.

However, the original results of [249] referred to a complete graph G, which, roughly speaking, correspond to IQP circuits of some polynomial depth. In follow up work by the same authors [40], they reduced the circuit depth to logarithmic under a sparsified version of the graph G. It was only in [118] and [119] that the $\#P$ hardness of approximately computing p_x on 2D circuit architectures was established (in the worst-case), corresponding to constant depth IQP circuits in 2D. Their analogous average-case conjecture for approximately computing p_x on 2D circuits, is reproduced below:

Conjecture 4.1 ([118]). *There exists a choice of vertex and edge weights $\{w_{uv}, w_u\}_{u,v\in[n]}$ on a 2D lattice G, and constants ε, δ, such that approximating the measurement distribution $\{p_x\}$ to the following mixture of multiplicative and additive errors*

$$|\tilde{p}_x - p_x| \leq \frac{1}{\mathrm{poly}(n)} \cdot p_x + \frac{\varepsilon}{\delta \cdot 2^n} \tag{4.78}$$

is #P hard for any $1-\delta$ fraction of instances x.

[118] show that Conjecture 4.1 implies Theorem 4.4:

Theorem 4.7 ([118], restatement of Theorem 4.4). *Assuming Conjecture 4.1, simulating the distribution $\{p_x\}$ up to ε total variation distance is classically intractable, assuming* PH *doesn't collapse.*

A related result was shown by [119]. They start from the (weaker) conjecture that computing $p(x)$ up to a multiplicative factor is hard-on-average, and combine it with a further conjecture on the anti-concentration of the output distribution of random linear-depth IQP circuits. Put together, they also arrive at Theorem 4.4.

4.2.9 Fault Tolerance of IQP Circuits under Input Noise

We dedicate this section to a proof of Lemma 4.14, on the fault tolerance of IQP circuits under input noise.

Lemma 4.14. *Fix an input noise rate $p < \frac{1}{2}$ and a positive integer D. Let C be an n qubit IQP circuit with depth d and lightcone size ℓ. Then, there exists another quantum circuit $\tilde{C}$, such that a sample from the output of $\tilde{C}$ under input bit-flip errors can be post-processed using an efficient classical algorithm into a sample ε-close to the output distribution of C. The circuit $\tilde{C}$*

1. *acts on $O(n \log \frac{n}{\varepsilon})$ qubits,*
2. *has lightcone size $\ell + O\left(D \log^{1/D}\left(\frac{n}{\varepsilon}\right)\right)$,*
3. *depth $d + O\left(D \log^{1/D}\left(\frac{n}{\varepsilon}\right)\right)$, and*
4. *the locality of its parent Hamiltonian is $\ell + O(D)$.*

At a high level, our approach is based on pre-processing each of the n input bits into "code-blocks" or gadgets of size $k = O(\log \frac{n}{\varepsilon})$ bits, where each gadget has a designated "root" bit. The n root bits are then input into the IQP circuit C. Since bit-flip errors commute with the IQP circuit, to be able to sample from the original output distribution of C, it suffices to *identify* these root bits. Indeed, we emphasize that we do not use the encoding to *correct* the errors within the circuit, as this would require adaptivity and an increase in circuit depth, and instead perform the correction only in post-processing.

4.2.9.1 The Distillation Gadget

We place the noisy bits into a tree of arity B (a "B-tree") of depth D. For notational convenience, let us partition the nodes in the tree into disjoint subsets, $L_1 \cup L_2 \cdots \cup L_D = [k]$, the "layers" of the tree. Moreover, for each node u in the tree, let the subset N_u denote its children or (downwards) neighbors in the tree. The encoding circuit proceeds over the layers from the leaves to the root, where at the ith layer L_i of the tree, a CNOT gate is applied from each parent bit to each of its children.

Note that the size of the tree k is implicitly defined by B and d: $k = \sum_{j=0}^{D-1} B^j = \Theta(B^D)$.

Algorithm 4 The Distillation Gadget U

Input: k qubits in the computational basis $|s\rangle$, where $s \leftarrow \mathrm{Ber}^k(p/2)$.

```
1: For each layer $i \in [2, \cdots, D]$ from leaves to root,
2: For each child $c \in N_p$ of a parent node $p$, apply a $\mathsf{CNOT}$ gate from $p$ to $c$.
```

$$\prod_{i\in[D]} \bigotimes_{p\in L_i} \left(\prod_{c\in N_p} \mathsf{CNOT}_{p,c} \right) |s\rangle \equiv U\,|s\rangle\,. \tag{4.79}$$

We emphasize that the ordering of operations, from leaves to root, matters crucially. In this manner, the ith layer acts as a "parity check syndrome" for the $(i+1)$st. When implemented using 2-qubit gates, the depth of the distillation circuit is $B \cdot D$, as the CNOT gates at the same layer but operating on different subtrees can be performed in parallel, but the B CNOT gates which act on the same parent must be performed sequentially.

4.2.9.2 The Decoding Algorithm

Next, suppose that all the qubits of $U\,|s\rangle$ except for that at the root of the tree have been measured, resulting in bits $b_2, \cdots b_k$. Can we reconstruct s_1, the bit at the root? The decoding algorithm below traverses the tree layer by layer, from leaves to root, attempting to reconstruct the bit s_p of the next layer.

Algorithm 5 The Decoding Algorithm

Input: $(k-1)$ bits $b_2, \cdots, b_k$, organized into a B-tree, where the root bit has been removed.
Output: A single bit $\tilde{s}_1$, a guess for the bit at the root.

```
At the leaves $L_1 \subset [k]$, let us denote $\tilde{b}_u = b_u$ for $u \in L_1$.
For each layer $i \in [2, \cdots, D]$, from leaves to root,
For each parent node $p \in L_i$, let $\tilde{s}_p = \mathrm{Maj}(\tilde{b}_c : c \in N_p)$ be the majority of its children bits.
If the root hasn't been reached, update $\tilde{b}_p \leftarrow \tilde{s}_p \oplus b_p$. Otherwise, output $\tilde{s}_1$.
```

The decoding algorithm above maintains the invariant that $\tilde{s}_u$ is a "guess" for the original noisy bit s_u input into the distillation gadget. Since U acts from leaves to root, the children in each layer contain (with high probability) the necessary information to reconstruct the parents' bit s_p. Together with the measurement outcome b_p - which reveals information about the layer above - we can continue the reconstruction up the tree.

4.2.9.3 Analysis

We divide the analysis into three claims, which consider the correctness, the lightcone size of the circuit, and the "Z-locality" of the distillation gadget which determines the locality of the parent Hamiltonian.

Claim 4.9 (Correctness)**.** *Fix any noise rate $p \leq \frac{1-\delta}{2}$ and let $B = \Omega(\delta^{-2})$. Then, the effective bit-flip error rate at the root of the depth d B-tree is $\leq 2^{-B^{\Omega(d)}}$.*

Proof. We prove inductively that the *effective* bit-flip error rate p_i at the ith layer, i.e.,

$$p_i \equiv \mathbb{P}_{s\leftarrow \mathsf{Bern}^k(p)}[s_u \neq \tilde{s}_u] \text{ for each node } u \in L_i, \tag{4.80}$$

decays doubly-exponentially with the layer index $i > 2$. As the base case, $p_1 = p$ is the probability of a bit-flip error on the leaves. Suppose $p = \frac{1-\delta}{2}$. Then, after the first layer, the probability the majority vote of the children bits is incorrect is

$$p_2 \leq \sum_{j=B/2}^{B} \binom{B}{j} \big(p_1\big)^j \big(1-p_1\big)^{B-j} = 2^{-B} \sum_{j=B/2}^{B} \binom{B}{j} \big(1-\delta\big)^j \big(1+\delta\big)^{B-j} \tag{4.81}$$

$$\leq \big(1-\delta\big)^{B/2} \cdot \big(1+\delta\big)^{B/2} \leq \big(1-\delta^2\big)^{B/2} \leq \frac{1}{16}, \tag{4.82}$$

so long as B is chosen to be $\Omega(\delta^{-2})$. For each layer $i \geq 2$, the effective bit-flip error rate on the $(i+1)$st layer is

$$p_{i+1} \leq \sum_{j=B/2}^{B} \binom{B}{j} \big(p_i\big)^j \cdot \big(1-p_i\big)^{B-j} \leq 2^B (p_i)^{B/2} \leq p_i^{B/4}. \tag{4.83}$$

In this manner, $p_{i+1} \leq 2^{-(B/4)^i}$ for $i \geq 1$.

□

Claim 4.10 (Circuit lightcone size). *The circuit lightcone size of the distillation scheme is $\leq B \cdot D$.*

Proof. The lightcone size of the quantum circuit U is upper bounded by the size of the lightcone of the qubits at the leaves of the tree. Crucially, we claim that if

$$u = u_1 \to u_2 \to u_3 \cdots \to u_D = \text{root} \tag{4.84}$$

denotes the path from a leaf $u \in L_1$ to the root, then only the children of these nodes can be in the lightcone of u. Indeed, this is since the CNOT gates in Section 4.2.9.1 are applied layer by layer in increasing order, so the only nodes which are causally connected to u in the circuit are its immediate ascendants or their neighbors. In turn, the size of this set is bounded by $B \cdot D$.

□

The last key claim makes reference to the locality of the parent Hamiltonian of the distillation circuit, that is, the size of the support of the operator $U(Z_i \otimes \mathbb{I})U^\dagger$, maximized over bits i in the gadget.

Claim 4.11 (Parent Hamiltonian Locality). *The locality of parent Hamiltonian of the distillation circuit is $\leq D$.*

Proof. The following two circuit identities describe how Pauli Z operators propagate through CNOT gates.

$$\mathsf{CNOT}_{i,j}(Z_i \otimes \mathbb{I})\mathsf{CNOT}_{i,j} = Z_i \otimes \mathbb{I} \tag{4.85}$$

$$\mathsf{CNOT}_{i,j}(\mathbb{I} \otimes Z_j)\mathsf{CNOT}_{i,j} = Z_i \otimes Z_j \tag{4.86}$$

Crucially, the locality only increases (or propagates) from the target qubit to the control qubit. Applied to our gadget in Section 4.2.9.1, we conclude that the qubits in the Z-lightcone of any qubit i in the tree, are precisely the ancestors of i. Thus, $|\text{supp}(U(Z_i \otimes \mathbb{I})U)| \leq D$, the depth of the tree. □

We are now in a position to conclude the proof of Lemma 4.14.

Proof. [of Lemma 4.14] By Claim 4.9, if $p \leq \frac{1}{2}(1-\delta)$, then, so long as

$$B = \max\left(\Theta(\delta^{-2}), \log^{1/D}\left(\frac{n}{\varepsilon}\right)\right) \tag{4.87}$$

the probability the decoding algorithm incorrectly outputs the bit at the root of the tree is $\leq \varepsilon n^{-1}$. By a union bound, all the gadgets succeed with probability $\geq 1-\varepsilon$. Conditioned on this event, the output distribution of $\tilde{C}$ corrected by the output of the n decoding algorithms is exactly that of C, which implies the bound on the TV distance. To conclude, the locality parameters are then implied by Claim 4.10 and Claim 4.11 □

4.2.10 Quantum Advantage in Gibbs Sampling

We dedicate this section to combining all the aforementioned ingredients and concluding the proof of our main result in Theorem 4.1.

Theorem 4.8 (General version of Theorem 4.1)**.** *For any constant inverse-temperature $\beta = \Theta(1)$ and integer L, there exists a family of n-qubit commuting $O(L)$-local Hamiltonians, such that the n-qubit Gibbs state ρ_β is both*

1. Rapidly Thermalizing. *It can be prepared within small trace distance by the Davies generator (Eq. (4.6)) which has mixing time $e^{O(L\cdot\log^{1/L}(n))}$. In addition, this process can be simulated on a quantum computer in time $n\cdot e^{O(L\cdot\log^{1/L}(n))}$. And yet,*
2. Classically Intractable. *Under Conjecture 4.1, there is no polynomial time classical algorithm to sample from the measurement outcome distribution $p(x) = \langle x|\,\rho_\beta\,|x\rangle$ within small constant total variation distance.*

In particular, the choice of a sufficiently large constant L recovers our main result of Theorem 4.1. When $L = \log\log n$, we obtain a mixing time of $\mathrm{polylog}(n)$.

Proof of Theorem 4.8. To begin our proof, let us fix an inverse-temperature $\beta = \Theta(1)$, and consider the equivalent bit-flip error rate

$$p = (1+e^{\beta})^{-1} < \frac{1}{2}, \tag{4.88}$$

as guaranteed by Lemma 4.13.

Classical Intractability. Consider the family of constant-depth, classically intractable, n-qubit IQP circuits C guaranteed by Theorem 4.4 (Conjecture 4.1). Using Lemma 4.2, let us fix a depth parameter L, and embed each circuit in said family into a new circuit $\tilde{C}$, which is fault tolerant to input noise of rate $p = \frac{1}{2}(1-\Theta(1))$. $\tilde{C}$ now has Z-locality $O(L)$, circuit depth and lightcone size $O(L\log^{1/L}(\frac{n}{\varepsilon}))$; and a noisy sample from $\tilde{C}$ can be efficiently classically post-processed into a sample ε-close in trace distance to an ideal sample from C.

Now, consider the family of parent Hamiltonians defined by the family of Fault-Tolerant circuits $\tilde{C}$,

$$H = \sum_i \tilde{C}\bigg(Z_i\otimes \mathbb{I}_{[n]\setminus i}\bigg)\tilde{C}^\dagger. \tag{4.89}$$

The support size of each term is given by the Z-locality of the fault-tolerant circuit $\tilde{C}$, which is $O(L)$.

If, by assumption, there was a polynomial time classical algorithm $\mathcal{A}$ to sample from the Gibbs state of H at inverse-temperature β, then we could construct a polynomial time classical algorithm to sample from a distribution ε-close to the ideal distribution of C, as follows: First, construct $\tilde{C}$ and thus the local terms of H from C. Then, leverage $\mathcal{A}$ to sample from $\propto e^{-\beta H}$. Finally, process the output sample using the post-processing algorithm from the fault-tolerance statement of Lemma 4.2.

Rapid Thermalization. To conclude, via Lemma 4.1, the Gibbs state of H can be prepared using the Davies generator of Eq. (4.6) of mixing time exponential in the circuit lightcone size, $\log n \cdot \exp\Big(O(L \cdot \log^{1/L}(n))\Big) = \exp\Big(O(L \cdot \log^{1/L}(n))\Big)$. To simulate this process on a quantum computer, the overall runtime $n \cdot \exp\Big(O(L \cdot \log^{1/L}(n))\Big)$ has an additional quasi-linear overhead. □

Remark 4.1. *Theorem 4.1 asserts that for every constant temperature, there exists a Hamiltonian H which is classically hard-to-sample from. Conversely, results by [189] and [226] show that every local Hamiltonian (of fixed degree) has a critical temperature, such that above said threshold one can efficiently classically sample from their Gibbs state. The resolution to this apparent contradiction lies in the order of quantifiers. The degree/locality of our Hamiltonians increases with the temperature, see Section 4.2.9 for their dependence on the noise rate.*

Remark 4.2. *Since the Gibbs state is determined by a low depth quantum circuit C, with access to a description of C, one could trivially produce it on a quantum computer. However, if given access only to the local Hamiltonian terms $\{h_i\}_i = \{-CZ_iC^\dagger\}_i$, we don't believe it to be computationally efficient to recover the global structure of C, in general. While this is not rigorous statement, we only know how to do so for 1D circuits, via dynamic programming. It is worthwhile to contrast this to the Feynman-Kitaev circuit-to-Hamiltonian mapping [230], wherein the gates of the circuit can be exactly read-off from the local Hamiltonian interactions.*

4.2.11 BQP Completeness with Adaptive Single-Qubit Measurements

We dedicate this section to a proof of Theorem 4.2, on the BQP completeness of Gibbs Sampling with adaptive measurements.

Theorem 4.9. *Fix an inverse-temperature $\beta = \Theta(1)$. Then, there exists an n-qubit $O(1)$-local Hamiltonian, whose Gibbs state at inverse-temperature β is a universal resource state for quantum computation and is efficiently preparable on a quantum computer.*

This result is all but a corollary of our fault tolerance techniques for IQP circuits, applied to measurement-based quantum computation. Indeed, it is well known that 2D cluster states, in addition to single-qubit measurents in adaptively chosen basis on the $X - Y$ plane, is universal for quantum computation. The following lemma shows that one can produce said cluster state out of the Gibbs state of a local Hamiltonian, so long as we are allowed to measure a subset of the qubits, and subsequently apply a Pauli correction to "distill" out the cluster state.

Lemma 4.15. *There exists a n-qubit, $O(1)$-local commuting Hamiltonian, whose Gibbs state at inverse-temperature β can be used to prepare a cluster state. That is, by measuring a subset of the qubits of the Gibbs state, and then with 1 round of adaptive Pauli correction, one can produce a 2D cluster state on $O(n/\log\frac{n}{\varepsilon})$ qubits with probability $1-\varepsilon$.*

Proof. Let C be the circuit which prepares a 2D cluster state on m qubits, comprised of Hadamard gates and CZ gates. Let $\tilde{C}$ be the $n = \Theta(m\log\frac{m}{\varepsilon})$ qubit circuit defined by the fault tolerance scheme of Lemma 4.14, which is robust to input errors of finite probability $< \frac{1}{2}$. Then, consider the parent Hamiltonian H of $\tilde{C}$, on n qubits and with locality $O(1)$.

By construction, its Gibbs state is a quantum-classical state, of classical bits lying in the fault-tolerance gadget of Lemma 4.14, and qubits comprising a cluster-state under input noise. Again, recall that input bit-flip errors are equivalent to output Z errors, due to the gate structure of C. From Lemma 4.14, by measuring the classical bits of the fault-tolerance gadget, one can recover the output Z error with probability $1-\varepsilon$. □

We remark that the adaptively chosen $X-Y$ measurements can be performed simultaneously with the Pauli corrections. In this manner, after producing the desired resource Gibbs state, it suffices to perform adaptively chosen single-qubit measurements to achieve universal measurement based quantum computation.

4.2.12 Addressing Output Measurement Errors

In this section, we prove Theorem 4.3 on sampling from finite-temperature Gibbs states subject to measurement errors.

Lemma 4.16. *Fix an inverse temperature $\beta = \Theta(1)$, and a measurement error rate $p < \frac{1}{2}$. There exists a family of n-qubit, $O(\log n)$-local Hamiltonians, such that sampling from their Gibbs state at inverse-temperature β, under measurement errors of rate p, is classically intractable under Theorem 4.4. Moreover, there exists a* $\mathrm{poly}(n)$ *time quantum algorithm to produce said Gibbs state.*

Our construction of Lemma 4.16 is similarly based on the parent Hamiltonians of fault-tolerant IQP circuits, which are hard-to-sample from in the ideal case. We note that the distribution defined by sampling from the Gibbs state of the parent Hamiltonian of a quantum circuit C, given measurement errors, corresponds exactly to sampling from C under both input and output noise, albeit with different noise rates. Unfortunately, to address this mixed noise model, we do need to appropriately modify our fault-tolerance scheme. For this purpose, we appeal to prior work by [40], at the cost of a higher locality.

4.2.12.1 Overview

To model the noise in this section, recall the definition of the bit-flip error channel $\mathcal{D}_p$ in Eq. (4.66). Given a quantum circuit C on n qubits, and fixed noise rates $p_{in}, p_{out} \in [0, \frac{1}{2})$, the noisy output distribution of C given input and output noise is given by

$$p_{C,p_{in},p_{out}}(x) = \mathrm{Tr}\left[|x\rangle\langle x| \cdot \mathcal{D}_{p_{out}}^{\otimes n} \circ C\left(\mathcal{D}_{p_{in}} \circ (|0\rangle\langle 0|)\right)^{\otimes n} C^{\dagger}\right] \tag{4.90}$$

If $\mathcal{A} : \{0,1\}^n \to \{0,1\}^{n'}$ is a deterministic classical post-processing algorithm, we denote as $\mathcal{A} \circ p$ the distribution given by sampling $x \leftarrow p$ and outputting $\mathcal{A}(x)$. The following lemma is a fault-tolerance statement for IQP circuits against this input/output noise model.

Lemma 4.17. *Let C be an n qubit IQP circuit of depth d and lightcone size ℓ, and fix input and output bit-flip error rates $p_{in}, p_{out} \in [0, \frac{1}{2})$. Then, for every $r \in \mathbb{N}$ there exists a quantum circuit C_r and a deterministic, $O(n_r)$-time decoding algorithm $\mathcal{A}_r : \{0,1\}^{n_r} \to \{0,1\}^n$, such that in the presence of input and output noise, the statistical distance*

$$\|\mathcal{A}_r \circ p_{C_r,p_{in},p_{out}} - p_{C,0,0}\|_1 \leq n \cdot (4q(1-q))^{r/2}, \text{ where } q = p_{in}(1-p_{out}) + p_{out}(1-p_{in}) < \frac{1}{2}. \tag{4.91}$$

Moreover, C_r is defined on $n_r = n \cdot r$ qubits, has depth $d_r = d \cdot \log r$ and lightcone size $\leq \ell \cdot r$.

In other words, noisy samples from C_r can be post-processed into nearly-ideal samples from C. Note that $q < \frac{1}{2}$ implies the total variation distance above decays exponentially with r.

Corollary 4.1. *Fix input and output bit-flip error rates $< \frac{1}{2}$. Then, any IQP circuit on n qubits and constant depth can be efficiently transformed into a quantum circuit of $O(\log \log n)$ depth and $O(\log n)$ lightcone size, robust to input and output noise with error $n^{-\Omega(1)}$.*

Starting from the hard-to-sample IQP circuits ensured by Theorem 4.4, we can construct circuits fault-tolerant to input and output noise via the Corollary above. In turn, these fault-tolerant circuits define a parent Hamiltonian, which is rapidly thermalizing (via Lemma 4.1), and yet, classically hard to sample from. Put together, we prove Lemma 4.16.

4.2.12.2 Analysis

We remark that if the circuit C itself is an IQP circuit, then the bit-flip noise model $\mathcal{B}_p$ commutes with the circuit, and thus the input/output noise is equivalent to input noise at a higher rate: $p_{C,p_{in},p_{out}}(x) = p_{C,q,0}(x)$, with

$$q = p_{in}(1 - p_{out}) + p_{out}(1 - p_{in}) < \frac{1}{2} \tag{4.92}$$

To leverage this equivalence, however, we need to design a fault-tolerant circuit which itself is an IQP circuit. Fortunately, here we can appeal to [40], who achieved precisely that. To summarize their construction, their circuit embedding leverages the following property of IQP circuits. The diagonal part D of any IQP circuit can be expressed as a matrix-exponential of a polynomial of Z Pauli matrices:

$$D = \exp\left[i\sum_{j\in[m]}\theta_j\bigotimes_{i\in[n]}Z_i^{M_{ji}}\right], \text{ for real coefficients } \{\theta_j\}, \text{ and a boolean matrix } M\in\mathbb{F}_2^{m\times n}. \tag{4.93}$$

If D is comprised of 2-qubit gates, then the weight of any row of M is ≤ 2. Now, suppose $G\in\mathbb{F}_2^{(n\cdot r)\times n}$ is the generator matrix of a repetition code, on $n' = n\cdot r$ bits and rate $n/n' = \frac{1}{r}$. [40] observe that the new IQP circuit defined by mapping $M\rightarrow\tilde{M} = M\cdot G^T$ is robust to input noise, up to (roughly) the random-error-correction capacity of G. Indeed, this follows from the fact that

$$\left\langle G^Tx\right| D\left|G^Tx\right\rangle = \langle x|\,\tilde{D}\,|x\rangle\,, \forall x\in\{0,1\}^{n'}. \tag{4.94}$$

Therefore, the output distribution of the new circuit $\tilde{C}$ under input (or output) noise is the same as sampling $y\in\{0,1\}^n$ from C, encoding y into the code $\tilde{y} = Gy\in\{0,1\}^{n'}$, and finally flipping each entry of $\tilde{y}$ independently with probability q. If the repetition code can tolerate random bit-flip errors with rate q, then one can approximately sample from C using noisy samples from $\tilde{C}$.

The caveat in their approach is that the resulting IQP circuits maybe polynomially larger. Indeed, each two qubit gate in the original circuit C, is mapped to a $2\cdot r$ multi-qubit gate in $\tilde{C}$:

$$e^{i\theta Z_a\otimes Z_b}\rightarrow e^{i\theta Z_a^1\otimes Z_a^2\cdots Z_a^r\otimes Z_b^1\cdots Z_b^r} \tag{4.95}$$

which is complex to implement using only diagonal operations. Instead, we dispense with the requirement that the intermediate gates in the circuit be diagonal (and thus the circuit is not an IQP circuit), however, globally it is equivalent to the same (IQP) unitary operation.

Definition 4.7. *A k-local* Pauli rotation gate *is the k qubit unitary U defined by an angle $\theta\in[0,2\pi]$ and a k-qubit Pauli P where $U = e^{i\theta P}$.*

Of particular note to us are multi-controlled Z rotations, where $P = Z_1\otimes Z_2\cdots Z_k$.

Claim 4.12. *Any k-local Pauli rotation gate can be implemented using an $\leq\log k$ depth circuit on a fully connected architecture of 2-qubit gates.*

For simplicity, we prove the above for multi-qubit Z Paulis, as the general case is analogous.

Proof. Let U be a k-local Z rotation gate, and V be any unitary. Then, the identity $Ve^{i\theta P}V^\dagger = e^{i\theta VPV^\dagger}$ tells us that it suffices to find a depth $d\leq\log k$ Clifford circuit V such that $V(\otimes_i^k Z_i)V^\dagger = Z_1\otimes\mathbb{I}_{[k]\setminus 1}$. We claim that this can be done recursively, where each layer of V halves the weight of the remaining Z's. Indeed, since $(\mathbb{I}\otimes Z) = \mathsf{CNOT}(Z\otimes Z)\mathsf{CNOT}^\dagger$, layers of CNOT gates on a matching of the remaining Z's will suffice. □

To prove our statement, we instantiate the lemma below with our implementation of multi-controlled Z gates.

Lemma 4.18 ([40])**.** *Let C be an n qubit IQP circuit of depth d. Then, for every $r \in \mathbb{N}$, there exists a deterministic, $O(n \cdot r)$-time decoding algorithm $\mathcal{A}_r : \{0,1\}^{n \cdot r} \to \{0,1\}^n$, and a quantum circuit C_r on $n_r = n \cdot r$ qubits, comprised only of Hadamard gates and $O(d)$ layers of $\leq 2r$-local Z rotation gates, satisfying*

1. *In the absence of noise, the distribution $\mathcal{A}_r \circ p_{C_r,0,0}$ given by sampling $y \leftarrow p_{\tilde{C}_r,0,0}$ from the output of C_r, and outputting $\mathcal{A}_r(y)$, is the same as sampling from C.*

2. *In the presence of input-level noise with probability q, the statistical distance*

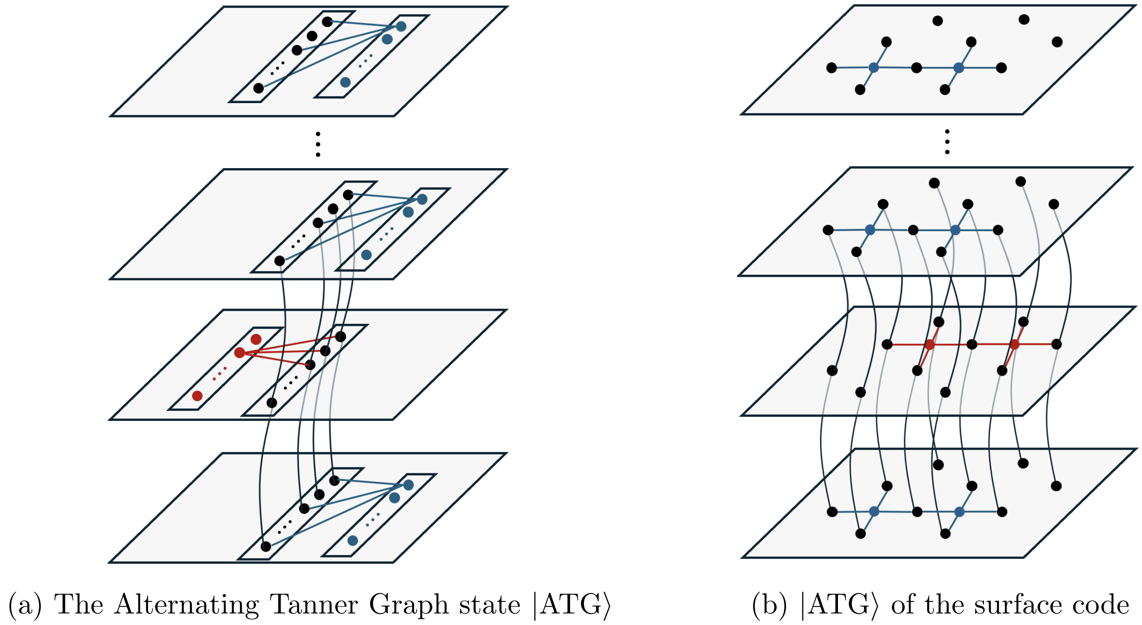
$$\|\mathcal{A}_r \circ p_{\tilde{C}_r,q,0} - p_{C,0,0}\|_1 \leq n \cdot (4 \cdot q \cdot (1-q))^{r/2}. \tag{4.96}$$

4.3 Single-shot logical state preparation for arbitrary quantum LDPC codes

Single-shot logical state preparation is a procedure in which a code-state (such as the logical $|0\rangle$ or $|+\rangle$ state) of a quantum error correcting code is prepared by a constant depth quantum circuit, followed by one round of single qubit measurements, and an adaptive Pauli correction using classical feedforward. In this section, we prove that every CSS quantum LDPC code admits a single-shot state preparation procedure with logarithmic space overhead that is fault tolerant against local stochastic noise, generalizing prior work by [253, 114] for the surface code. Our proof is based on a clustering-of-errors argument by [254] and [255].

The intuition behind our construction is to realize *repeated measurement* via *measurement-based quantum computation*. A standard approach to initialize a quantum LDPC code in the logical $|+\rangle$ state is as follows: first initialize all physical qubits in a code block in the $|+\rangle$ state, then perform repeated Z and X syndrome measurements, and finally perform a Pauli error correction using classical feedforward based on the syndrome measurement outcomes. A key observation here is that the repeated measurements are non-adaptive, thus we can attempt to simulate this process using a cluster state, achieving a space-time trade-off.

Our construction, which we refer to as the Alternating Tanner Graph state $|\mathrm{ATG}\rangle$ (Fig. 4.6) formalizes this intuition: there are $2T+1$ copies of the code block. A copy of the Z Tanner graph is placed on each odd layer (simulating Z syndrome measurements), and a copy of the X Tanner graph is placed on each even layer (simulating X syndrome measurements). Vertical connections are added between code qubits of neighboring layers, which is used to propagate quantum computation in the time direction. To prepare the cluster state in constant depth, we initialize all qubits in the $|+\rangle$ state and apply a CZ gate on each edge. Note that this is equivalent to the Raussendorf-Bravyi-Harrington (RBH) cluster state when applied to the surface code [253].



(a) The Alternating Tanner Graph state $|\mathrm{ATG}\rangle$

(b) $|\mathrm{ATG}\rangle$ of the surface code

Figure 4.6: (a) The ATG is a graph state, defined on a vertex set comprised of layers of copies of the code block and the Z (resp. X) Tanner graph (in blue, resp. red) of the qLDPC code. (b) The RBH cluster state [253], which prepares code states of the surface code in a 3D cubic lattice arrangement, is a special case of this construction.

4.3.1 Single-shot logical state preparation

4.3.1.1 Basic definitions and notation

The set of Pauli operators on a set of n qubits is denoted as $\mathsf{Pauli}(n)$. A Pauli error $E \in \mathsf{Pauli}(n)$ is said to be a local stochastic error of noise rate $p \in [0,1]$, or, $E \leftarrow N(p)$ if

$$\forall S \subset [n], \quad \mathbb{P}_{E\leftarrow N(p)}\big[S \subset \mathrm{supp}(E)\big] \leq p^{|S|}. \tag{4.97}$$

Definition 4.8. *An $[[n,k,d]]$ stabilizer code Q is said to be an ℓ-LDPC code if there exists a choice of generators $\in \mathsf{Pauli}(n)$ for Q with Pauli weight $\leq \ell$ and such that the number of generators acting non-trivially on any fixed qubit is $\leq \ell$.*

We will henceforth assume that Q is a CSS code, specified by X and Z parity check matrices (H^x, H^z) satisfying $H^x(H^z)^\dagger = 0$. Let $H^x \in \mathbb{F}_2^{m_x\times n}$ and $H^z \in \mathbb{F}_2^{m_z\times n}$ respectively, where m_x, m_z determine the number of X and Z parity checks. We say $i \sim c$ if i is in the support of the check c.

Definition 4.9 (Tanner Graph)**.** *The tanner graph of a parity check matrix $H \in \mathbb{F}_2^{m\times n}$ is the bipartite graph on the vertex set $[m] \cup [n]$, such that there exists an edge (c,i) for $c \in [m]$ and $i \in [n]$ if $H_{c,i} = 1$.*

We refer to the tanner graphs of H^X, H^Z as G^X, G^Z.

Here we overview our fault-tolerant state preparation algorithm for any LDPC CSS code. Our algorithm is based on that of [114] and is comprised of three general steps, summarized below.

Single-Shot State Preparation.

1. Using a constant-depth circuit W, we prepare a graph state $|\text{ATG}\rangle = W\left|0^V\right\rangle$. $G = (V, E)$ is defined on "bulk" qubits $\mathcal{B}$ and "boundary" qubits $\partial = V \setminus \mathcal{B}$.
2. Measure all the bulk qubits $\mathcal{B}$ in the Hadamard basis $|\pm\rangle$, resulting in a string s.
3. Using s, compute a Pauli correction $\mathsf{Rec}(s) \in \mathsf{Pauli}(\partial)$, and adaptively apply it to the boundary qubits ∂. The resulting (unnormalized) state is given by:

$$\Big(|\pm_s\rangle\langle\pm_s|_{\mathcal{B}} \otimes \mathsf{Rec}(s)_\partial \Big) |\text{ATG}\rangle \tag{4.98}$$

In the absence of any errors, we design $\mathsf{Rec}(s)$ to ensure that the resulting state is $\left|\bar{\Phi}\right\rangle = \frac{1}{\sqrt{2}}(|\bar{0}\rangle|\bar{0}\rangle + |\bar{1}\rangle|\bar{1}\rangle)$, a logical EPR pair across two copies of the LDPC code – "the boundaries". More generally, the resulting state is $\left|\bar{\Phi}\right\rangle^{\otimes k}$ if the LDPC code has k logical qubits. For simplicity, below we state the result for one of the k logical qubits. To ensure that this state-preparation algorithm is fault-tolerant, we stipulate that even in the presence of random errors during the execution of W and the measurements, the resulting output state is statistically close to $\left|\bar{\Phi}\right\rangle$ with local stochastic noise.

Definition 4.10 (cf. [114])**.** *A family of stabilizer codes Q admits a* Single-Shot State Preparation *procedure if there exists a constant depth circuit W on a set of qubits $\partial \cup \mathcal{B}$, and deterministic recovery and repair functions,*

$$\mathsf{Rec} : \{0,1\}^{\mathcal{B}} \to \mathsf{Pauli}(\partial) \tag{4.99}$$

$$\mathsf{Rep} : \mathsf{Pauli}(\mathcal{B}) \to \mathsf{Pauli}(\partial), \tag{4.100}$$

such that, for any error on the Bulk $P \in \mathsf{Pauli}(\mathcal{B})$ and measurement outcome $s \in \{0,1\}^{\mathcal{B}}$,

$$\Big(|\pm_s\rangle\langle\pm_s| \otimes \mathsf{Rec}(s) \Big) PW\,|0\rangle^{\mathcal{B}} \otimes |0\rangle^{\partial} = \gamma_s\,|\pm_s\rangle \otimes \mathsf{Rep}(P)\left|\bar{\Phi}\right\rangle, \tag{4.101}$$

where $\gamma_s \in \mathbb{C}$. Moreover, if $P \leftarrow \mathcal{N}(p)$ is a local stochastic error on the Bulk, then $\mathsf{Rep}(P) \leftarrow \mathcal{N}(c_1 \cdot p^{c_2})$ is a local stochastic Pauli error on ∂, for two constants c_1, c_2.

Naturally, in the presence of measurement errors, one cannot hope to perfectly prepare the encoded logical state. Instead, the function Rep quantifies the residual error on the ideal state. The state preparation procedure is then fault-tolerant if it manages to convert local stochastic noise during the state preparation circuit into local stochastic noise on the output state.

The main result of this work is the following theorem:

Theorem 4.10. *Fix an integer $T \geq 1$. Let Q be any $[[n, k, d]]$ CSS code, which is ℓ-LDPC. Then, Q admits a single-shot state preparation procedure (for the encoded logical state $\left|\bar{\Phi}\right\rangle$) using a circuit W on $O(\ell \cdot n \cdot T)$ qubits and of depth $O(\ell^2)$, satisfying the following guarantees:*

1. *There exists a constant $p^*(\ell) \in (0, 1)$, such that if W is subject to local-stochastic noise $\mathcal{N}(p)$ of rate $p < p^*$, the state preparation procedure succeeds with probability at least $1 - n \cdot (\frac{p}{p^*})^{\Omega(\min(T,d))}$.*
2. *Conditioning on this event, the resulting state is subject to a local stochastic noise of rate $\mathcal{N}(O(p^{1/2}))$.*

Theorem 4.10 can be interpreted as a repeated measurement state preparation scheme, realized in measurement based quantum computation. In this sense, it "tradeoffs space for time", suppressing errors exponentially in $\min(T, d)$, at the cost of a $O(T)$ multiplicative space overhead. Its proof is based on an error-clustering argument, inspired by the results of [254] and [255].

4.3.1.2 The Alternating Tanner Graph state $|\text{ATG}\rangle$

Using the LDPC code Q, we will define the cluster state $|\text{ATG}\rangle$ by specifying a graph $G = (V, E)$. So long as the degree of G is bounded, the resulting state-preparation circuit is low-depth.

$$|\text{ATG}\rangle = \left(\prod_{(u,v)\in E} \mathsf{CZ}_{u,v} \right) |+\rangle^{\otimes |V|} \tag{4.102}$$

Fix an integer $T \geq 1$. G will be defined on $(2T + 1)$ layers of copies of the LDPC code, which alternate X and Z checks. At each layer, we will place a copy of the tanner graph of H^x or H^z.

The Alternating Tanner Graph. Of the $2T+1$ layers, $\{1, 2, \cdots, 2T+1\}$

- Every layer has a copy of the n "code" qubits. Each code qubit $i \in [n]$ at layer t is connected to its copy above and below it $(i, t \pm 1)$.
- *Z Layers.* Odd layers have a copy of m_z "Z" parity check qubits. A copy of the Z tanner graph is placed between the n code qubits and the m_z Z parity check qubits.
- *X Layers.* Even layers have a copy of m_x "X" parity check qubits. A copy of the X tanner graph is placed between the n code qubits and the m_x X parity check qubits.

One should picture the copies of the code qubits stacked vertically in 1D layers, while the X and Z ancilla qubits lie to their left and right respectively. The "boundary" qubits ∂ referred to in the overview will simply be the $2n$ code qubits on the 1st and $(2T+1)$th layers. Henceforth we will index the vertices in G as tuples, (i, t) for any code qubit $i \in [n]$ at layer t, and (c, t) for check qubits $c \in [m_x]$ (or m_z). We highlight that the only vertical connections in E are between copies of the same code qubit.

4.3.1.3 The stabilizers of $|\text{ATG}\rangle$

The alternating Tanner graph state $|\text{ATG}\rangle$ is a stabilizer state. A complete set of stabilizer generators is defined as follows: for each $u \in V$, there is an associated stabilizer

$$G_u = X_u \bigotimes_{v:\,(u,v)\in E} Z_v. \tag{4.103}$$

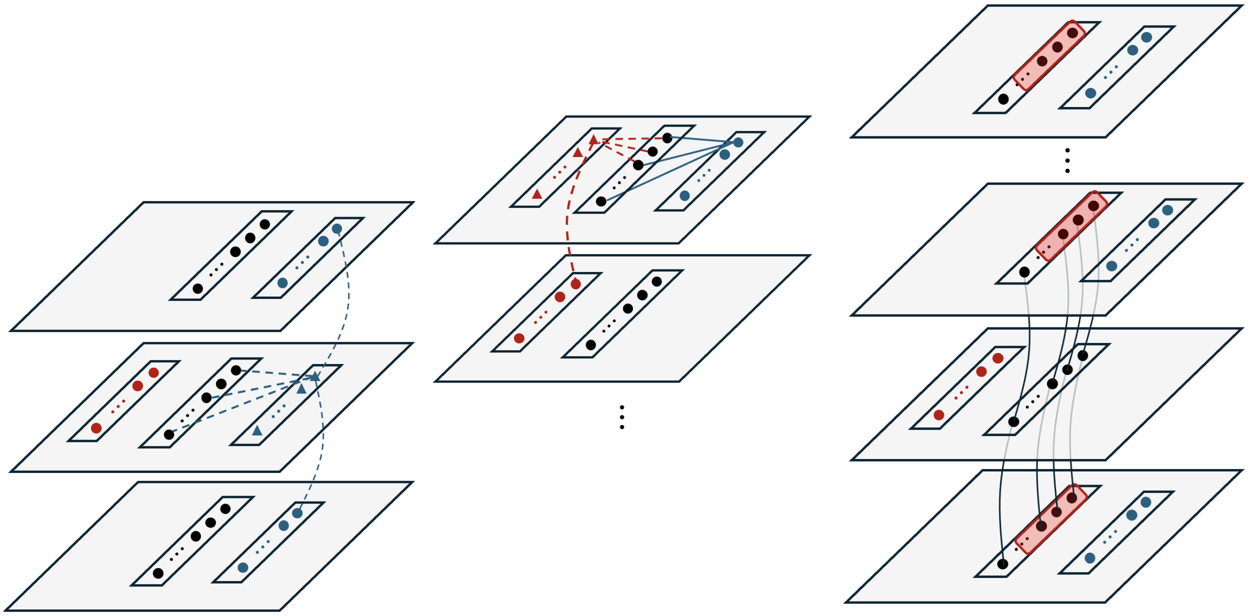
We will use this collection of graph-state stabilizers to define stabilizers of the post-measurement state. Following [114], we will identify two subgroups of the stabilizer group $\mathcal{S}$ of $|\text{ATG}\rangle$, $\mathcal{S}^0 \subset \mathcal{S}^1 \subset \mathcal{S}$, satisfying the following constraints:

(i) Any element of $\mathcal{S}^0$ can be written as $\mathbb{I}_\partial \otimes X(\alpha)_\mathcal{B}$ on some subset $\alpha \in \{0,1\}^\mathcal{B}$.

(ii) Any element of $\mathcal{S}^1$ can be written as $S_\partial \otimes X(\alpha)_\mathcal{B}$ on some subset $\alpha \in \{0,1\}^\mathcal{B}$, and for some stabilizer S of $|\bar{\Phi}\rangle$.

(iii) For every stabilizer S of $|\bar{\Phi}\rangle$, there exists an element of $\mathcal{S}^1$ of the form $S_\partial \otimes X(\alpha)_\mathcal{B}$.

The fact that these subgroups act only as Pauli X's on the Bulk implies that after an X basis measurement, they remain stabilizers of the post-measurement state, and we can recover their information from the measured string s. We will shortly describe how to use these stabilizers to define the Recover and Repair Functions $\mathsf{Rec}, \mathsf{Rep}$. In this section, we show how to define these stabilizers using the structure of the tanner graph state. We defer



(a) A Z meta-check, on an even layer (blue triangles).

(b) The stabilizers of the top boundary code.

(c) An Encoded $\bar{X} \otimes \bar{X}$ Logical Stabilizer.

Figure 4.7: The Stabilizers of the Graph State $|\text{ATG}\rangle$

to Section 4.3.3 rigorous proofs that these stabilizers factor as above. Let us begin with $\mathcal{S}^0$.

The Meta-Checks $\mathcal{S}^0$. $\mathcal{S}^0$ will consist of X or Z "meta-checks" which encode redundancies into the bulk qubits. Each meta-check will be centered around an "meta-vertex" - the support of such meta-checks will consist of copies of the X and Z tanner graphs in alternating layers (offset from those of E); together with vertical connections between copies of the same ancillas (Section 4.3.1.3).

For each even layer t and $c \in [m_z]$, we place a Z meta-check:

$$G_{(c,t-1)} \cdot G_{(c,t+1)} \cdot \prod_{i \sim c} G_{(i,t)} = X_{(c,t-1)} \otimes X_{(c,t+1)} \bigotimes_{i \sim c} X_{(i,t)} \tag{4.104}$$

For each odd layer t and $c \in [m_x]$, we place a X meta-check:

$$G_{(c,t-1)} \cdot G_{(c,t+1)} \cdot \prod_{i \sim c} G_{(i,t)} = X_{(c,t-1)} \otimes X_{(c,t+1)} \bigotimes_{i \sim c} X_{(i,t)} \tag{4.105}$$

The fact that these products of graph state factorize cleanly into products of X operators is non-trivial, and carefully leverages the fact that Q is a CSS code. While we defer a rigorous proof to Section 4.3.3, the underlying intuition is that each X ancilla qubit $d \in [m_x]$

which arises in the neighborhood of the support of any given Z meta-check $c \in [m_z]$, appears precisely an even number of times. This is since d and c on layer t are connected through a qubit i iff $H^Z_{c,i} \cdot H^X_{d,i} = 1$ (see Section 4.3.1.3 (b)), and therefore the number of appearances is

$$\sum_i H^Z_{c,i} \cdot H^X_{d,i} = \big(H^Z (H^X)^T\big)_{c,d} = 0 \mod 2 \tag{4.106}$$

The stabilizers in $\mathcal{S}^1$ arise in two types. Recall that $|\bar{\Phi}\rangle$ consists of an encoded maximally entangled state across two copies of the LDPC code Q. Then, within $\mathcal{S}^1$ there will be stabilizers of the individual boundary codes, and encoded stabilizers of the Bell state.

The Stabilizers of the Boundary Codes. We specify the X and Z stabilizers of the boundary codes as follows. To begin, let us consider the simpler case, consisting of the Z-type stabilizers. For each $c \in [m_z]$, there exists graph state stabilizers $\in \mathcal{S}$, satisfying the following decomposition

$$G_{(c,1)} = X_{(c,1)} \bigotimes_{i\sim c} Z_{(i,1)}, \quad G_{(c,2T+1)} = X_{(c,2T+1)} \bigotimes_{i\sim c} Z_{(i,2T+1)} \tag{4.107}$$

Which directly follows from the definition of the graph state stabilizers. Note that these operators act as Z Paulis on the boundary ∂ and X on $\mathcal{B}$.

The X-type stabilizers on the boundary codes are slightly more complicated, and require products of graph state stabilizers. For each X-type stabilizer $c \in [m_x]$ of Q, there exists products of graph state stabilizers $\in \mathcal{S}$, satisfying the decomposition

$$G_{(c,2)} \cdot \prod_{i\sim c} G_{(i,1)} = X_{(c,2)} \bigotimes_{i\sim c} X_{(i,1)}, \quad G_{(c,2T)} \cdot \prod_{i\sim c} G_{(i,2T+1)} = X_{(c,2T)} \bigotimes_{i\sim c} X_{(i,2T+1)}. \tag{4.108}$$

Which, we note, act as X stabilizers on the boundary codes ∂ in tensor product with an X Pauli on the Bulk, as desired.

The Encoded Logical Stabilizers. The encoded Bell pairs are stabilized by products $\bar{X}_1 \otimes \bar{X}_{2T+1}, \bar{Z}_1 \otimes \bar{Z}_{2T+1}$ of logical operators. We construct these operators using graph state stabilizers in G, and only X operators on the Bulk, via products of stabilizers in alternating layers of G.

To begin, let us consider the encoded $\bar{X}_1 \otimes \bar{X}_{2T+1}$ stabilizer. Let $\alpha_x \subset [n]$ denote the support of a logical $\bar{X}$ on Q. Then, we can write the $\bar{X}_1 \otimes \bar{X}_{2T+1}$ stabilizer as:

$$\prod_{\substack{i\in\alpha_x \\ t \text{ odd}}} G_{(i,t)} = \bigotimes_{\substack{i\in\alpha_x \\ t \text{ odd}}} X_{(i,t)} = \bar{X}_1 \otimes \bar{X}_{2T+1} \bigotimes_{\substack{i\in\alpha_x \\ t\in\{3,5,\cdots\}}} X_{(i,t)} \tag{4.109}$$

Similarly, let $\alpha_z \subset [n]$ denote the support of a logical $\bar{Z}$ on Q. Then, we can write the $\bar{Z}_1 \otimes \bar{Z}_{2T+1}$ stabilizer as:

$$\prod_{\substack{i\in\alpha_z \\ t \text{ even}}} G_{(i,t)} = \bar{Z}_1 \otimes \bar{Z}_{2T+1} \bigotimes_{\substack{i\in\alpha_z \\ t \text{ even}}} X_{(i,t)} \tag{4.110}$$

To argue that these operators factor as desired, we similarly apply the constraint that Q defines a CSS code. We refer the reader to Section 4.3.3 for the proofs.

4.3.1.4 The Recover and Repair functions

We are now in a position to define the Pauli frame correction Rec, and the residual Pauli noise Rep.

The Pauli Frame Correction Rec. The definition of Rec proceeds in two steps. At a high level, we begin by leveraging the meta-check information $\mathcal{S}^0$, to make a guess $Z(\beta)$ for the error which occurs on the Bulk $\mathcal{B}$.[18] Then, we pick Rec to be an arbitrary Pauli error supported on the boundary ∂, which is consistent with the information from $\mathcal{S}^1$, and the inferred error $Z(\beta)$.

To understand the role of the meta-checks in this sketch, let us concretely show how the measurement outcome string $s \in \{0,1\}^{\mathcal{B}}$ allows us to extract partial information about any Z-type Pauli error on the Bulk $P = Z(\eta)_{\mathcal{B}}$. For any stabilizer $\mathbb{I}_\partial \otimes X(\alpha)_{\mathcal{B}} \in \mathcal{S}^0$, we have

$$\begin{aligned} &(-1)^{s\cdot\alpha}\bigg(|\pm_s\rangle\langle\pm_s| \otimes \mathbb{I}_\partial\bigg) Z(\eta)_{\mathcal{B}} \otimes \mathbb{I}_\partial \,|\text{ATG}\rangle \\ &= \bigg(|\pm_s\rangle\langle\pm_s| \otimes \mathbb{I}_\partial\bigg) X(\alpha) Z(\eta)_{\mathcal{B}} \otimes \mathbb{I}_\partial \,|\text{ATG}\rangle \\ &= (-1)^{\alpha\cdot\eta}\bigg(|\pm_s\rangle\langle\pm_s| \otimes \mathbb{I}_\partial\bigg) Z(\eta)_{\mathcal{B}} \otimes \mathbb{I}_\partial \,|\text{ATG}\rangle\,, \end{aligned} \tag{4.111}$$

therefore revealing the "syndrome" information $\eta \cdot \alpha = s \cdot \alpha \mod 2$. By collecting this syndrome information, we can make a guess for the error on the Bulk, as described in Step 2 of Fig. 4.8. An analogous calculation can be reproduced for the stabilizers $S_\partial \otimes X(\alpha)_{\mathcal{B}} \in \mathcal{S}^1$:

$$\begin{aligned} &S_\partial \cdot \bigg(|\pm_s\rangle\langle\pm_s| \otimes \mathbb{I}_\partial\bigg) Z(\eta)_{\mathcal{B}} \otimes \mathbb{I}_\partial \,|\text{ATG}\rangle \\ &= (-1)^{s\cdot\alpha}\bigg(|\pm_s\rangle\langle\pm_s| \otimes \mathbb{I}_\partial\bigg)\bigg(X(\alpha) Z(\eta)_{\mathcal{B}} \otimes S_\partial\bigg) |\text{ATG}\rangle \\ &= (-1)^{\alpha\cdot\eta+\alpha\cdot s}\bigg(|\pm_s\rangle\langle\pm_s| \otimes \mathbb{I}_\partial\bigg) Z(\eta)_{\mathcal{B}} \otimes \mathbb{I}_\partial \,|\text{ATG}\rangle\,. \end{aligned} \tag{4.112}$$

In this manner, if we were to perform an ideal syndrome measurement of S_∂ on the post-measurement state, the value readoff would be $\alpha\cdot\eta + \alpha\cdot s$. The goal is to correct all of them to 0 (or +1 for the stabilizer measurement).

[18] Note that we can restrict ourselves to Z errors on $\mathcal{B}$, since these qubits are measured in the X basis.

Recover $\mathsf{Rec}(s)$**.** Given a measurement string $s \in \{0,1\}^{\mathcal{B}}$,

1. For each stabilizer $\mathbb{I}_\partial \otimes X(\alpha)_\mathcal{B} \in \mathcal{S}^0$, compute its syndrome $s_\alpha = s \cdot \alpha$.
2. Find the minimum-weight Z-type Pauli $Z(\beta)$ supported on $\mathcal{B}$, consistent with the syndromes s_α of $\mathcal{S}^0$.
3. For each stabilizer $\mathcal{S}_\partial \otimes X(\gamma)_\mathcal{B} \in \mathcal{S}^1$, compute the *corrected* syndrome
$$s'_\gamma = s \cdot \gamma \oplus \gamma \cdot \beta.$$
4. Let $\mathsf{Rec}(s)$ be an *arbitrary* Pauli supported on ∂, consistent with the computed corrected syndromes s' of $\mathcal{S}^1$. [a]

[a] Here, for simplicity we assume the CSS code is specified by full rank matrices H_X, H_Z, such that there always exists an error associated to every syndrome vector.

Figure 4.8: The Pauli Frame

Unfortunately, our decoder does not have access to these ideal values, and instead can only make a guess of them using s, and the inferred error $Z(\beta)$, as done in step 3. Note that this step is not necessarily efficient.

After $\mathsf{Rec}(s)$ is applied, the state equals $\Big(|\pm_s\rangle\langle\pm_s| \otimes \mathsf{Rec}(s)_\partial \Big) Z(\eta)_\mathcal{B} \otimes \mathbb{I}_\partial |\text{ATG}\rangle$ up to normalization. For each code stabilizer S_∂, suppose the corresponding cluster state stabilizer is $S_\partial \otimes X(\alpha)_\mathcal{B} \in \mathcal{S}^1$, then the *residual syndrome* is given by

$$\begin{aligned}
&S_\partial \cdot \Big(|\pm_s\rangle\langle\pm_s| \otimes \mathsf{Rec}(s)_\partial \Big) Z(\eta)_\mathcal{B} \otimes \mathbb{I}_\partial |\text{ATG}\rangle \\
&= (-1)^{\alpha\cdot\beta+\alpha\cdot s} \mathsf{Rec}(s)_\partial \cdot S_\partial \cdot \Big(|\pm_s\rangle\langle\pm_s| \otimes \mathbb{I}_\partial \Big) Z(\eta)_\mathcal{B} \otimes \mathbb{I}_\partial |\text{ATG}\rangle \\
&= (-1)^{\alpha\cdot\beta+\alpha\cdot s} \mathsf{Rec}(s)_\partial \cdot (-1)^{\alpha\cdot\eta+\alpha\cdot s} \Big(|\pm_s\rangle\langle\pm_s| \otimes \mathbb{I}_\partial \Big) Z(\eta)_\mathcal{B} \otimes \mathbb{I}_\partial |\text{ATG}\rangle \\
&= (-1)^{\alpha\cdot\beta+\alpha\cdot\eta} \Big(|\pm_s\rangle\langle\pm_s| \otimes \mathsf{Rec}(s)_\partial \Big) Z(\eta)_\mathcal{B} \otimes \mathbb{I}_\partial |\text{ATG}\rangle ,
\end{aligned} \tag{4.113}$$

that is, the residual syndrome is given by $\alpha \cdot \beta + \alpha \cdot \eta$. This constitutes a proof that the decoding process in $\mathsf{Rec}(s)$ (step 4 of Fig. 4.8) can be arbitrary, because the residual syndrome only depends on β.

The Residual Error Rep**.** Once the $\mathsf{Rec}(s)$ has been computed and applied, we pick Rep to be a carefully designed Pauli operator (residual error) on ∂ satisfying:

$$\bigg(|\pm_s\rangle\langle\pm_s| \otimes \mathsf{Rec}(s)\bigg) P\,|\mathrm{ATG}\rangle = \gamma_s\,|s\rangle \otimes \mathsf{Rep}(P)\,\big|\bar{\Phi}\big\rangle \tag{4.114}$$

Note that the decoder need not know what Rep is; however, should they be able to perform a *perfect* syndrome measurement after Rec is applied, then Rep can essentially be understood as the minimum weight operator consistent with that syndrome. To be more concrete, we define $\mathsf{Rep}(P)$ as a product of 4 terms:

$$\mathsf{Rep}(P) = \mathsf{Rep}_X(P) \cdot \mathsf{Rep}_Z(P) \cdot \mathsf{Rep}_{\bar{X}}(P) \cdot \mathsf{Rep}_{\bar{Z}}(P) \tag{4.115}$$

Each of the terms above will correspond to the residual correction of the syndrome of a given stabilizer of $\bar{\Phi}$. If we partition the stabilizers $S_\partial \otimes X(\alpha) \in \mathcal{S}^1$, based on whether S_∂ is an X or Z stabilizer of the LDPC code Q, or whether it is a logical XX or ZZ stabilizer of the encoded Bell state, then

- $\mathsf{Rep}_X(P)$ is the minimum weight Z-type Pauli operator on ∂ consistent with the residual syndrome of the X stabilizers of Q;
- $\mathsf{Rep}_Z(P)$ is the minimum weight X-type Pauli operator on ∂ consistent with the residual syndrome of the Z stabilizers of Q;
- $\mathsf{Rep}_{\bar{X}}(P) \in \{\mathbb{I}, \bar{Z}\}$ is the residual Z logical error, which ensures $\mathsf{Rep}(P)$ and $Z(\beta)P$ have the same syndrome under the encoded XX stabilizer.
- $\mathsf{Rep}_{\bar{Z}}(P) \in \{\mathbb{I}, \bar{X}\}$ is the residual X logical error, which ensures $\mathsf{Rep}(P)$ and $Z(\beta)P$ have the same syndrome under the encoded ZZ stabilizer.

We will argue in the following section that so long as P is a local stochastic error, then $\mathsf{Rep}_X(P)$, $\mathsf{Rep}_Z(P)$ are local stochastic as well. Moreover, $\mathsf{Rep}_{\bar{X}}(P)$, $\mathsf{Rep}_{\bar{Z}}(P)$ are trivial (identity) with high probability.

4.3.2 Proof of fault-tolerance via clustering

This section is the main technical part, in which we show that if P is a local stochastic error, then the residual error $\mathsf{Rep}(P)$ is also local stochastic. Our proof approach follows closely that of [255] and [254], who argued that quantum LDPC codes can (information-theoretically) correct from random errors of rate less than a critical threshold $p \leq p^*$, even in the presence of syndrome measurement errors. Their key idea was a "clustering" argument, which reasons that stochastic errors on LDPC codes cluster into connected components on a certain low-degree graph. Using a percolation argument on this low-degree graph, they prove these errors are unlikely to accumulate into a logical error on the codespace.

The syndrome adjacency graphs. To setup notation, let $P_2, P_3, \cdots P_{2T} \in \{0,1\}^n$ denote the support of the physical Z errors that occur on each layer of the Bulk qubits, and $B_1, B_3, \cdots B_{2T+1} \in \{0,1\}^{m_z}, B_2, B_4, \cdots B_{2T} \in \{0,1\}^{m_x}$ denote the support of the Z errors that occur the ancilla qubits in G. Following the syntax of $\mathsf{Rec}, \mathsf{Rep}$, let $R_2, R_3, \cdots, R_{2T} \in \{0,1\}^n$ and $C_1, C_2, \cdots, C_{2T+1}$ denote the set of *deduced* Z-errors on the physical and ancilla qubits respectively, using the information from the meta-checks $\mathcal{S}^0$.[19] For $E \in \{0,1\}^n$, $\mathsf{Syn}_X(E) \in \{0,1\}^{m_x}$ is the X syndrome vector associated to E (when E is interpreted as a Pauli Z error), and $\mathsf{Syn}_Z(E)$ is defined analogously.

By definition, R, C and P, B are both consistent with syndromes inferred from $\mathcal{S}^0$, which implies the relation

$$B_{2t} \oplus B_{2t+2} \oplus \mathsf{Syn}_X(P_{2t+1}) = C_{2t} \oplus C_{2t+2} \oplus \mathsf{Syn}_X(R_{2t+1}), \quad \forall t = 1, 2, \ldots, T-1, \tag{4.116}$$

and similarly for the Z syndrome on even layers:

$$B_{2t-1} \oplus B_{2t+1} \oplus \mathsf{Syn}_Z(P_{2t}) = C_{2t-1} \oplus C_{2t+1} \oplus \mathsf{Syn}_Z(R_{2t}), \quad \forall t = 1, 2, \ldots, T. \tag{4.117}$$

We represent the decoding process on a pair of new graphs, coined the “syndrome adjacency graphs”, following minor modifications to [254]'s ideas. The X (resp, Z) syndrome adjacency graph is defined on $T+1$ layers, where in each layer we place n nodes, and the 1st and $T+1$st layers are referred to as the “boundary” nodes. For each timestep $t \in [T]$ and X parity check $c \in [m_x]$, we also place a node (c, t) *between* code layers t and $t+1$. Intuitively, the nodes in the X syndrome adjacency graph are in bijection with the nodes in the Z layers of the ATG. In turn, the edges in this graph represent the connectivity of the stabilizers of the ATG: we connect any two nodes if they both are both acted on by a X (resp. Z) meta-checks or X (resp. Z) boundary code stabilizers. If the code is ℓ-LDPC, the degree of this graph is $z \leq \ell(\ell-1) + 2\ell \leq \ell(\ell+1)$.

We connect code nodes (i, t) and (j, t) if $i, j \sim c$ are both in the support of some check $c \in [m_x]$, we connect (c, t) to both (i, t) and $(i, t+1)$, and finally we connect (c, t) to $(c, t+1)$.

To represent the decoding process on these new graphs, we will mark the vertices in which decoding has failed: for $t \in \{1, \cdots T-1\}$ we paint a code vertex (i, t) in the Bulk of the X syndrome adjacency graph iff the physical error differs from the inferred error on the associated vertex of the ATG: $P_{(i,2t+1)} \neq R_{(i,2t+1)}$; similarly, we paint a check vertex (c, t) of the syndrome adjacency graph iff $B_{(c,2t)} \neq C_{(c,2t)}$. Finally, we paint a vertex on the boundary of the X syndrome adjacency graph iff $\mathsf{Rep}_X(P)$ is non-zero on the associated qubit of ∂.

Perhaps the key observation in the approach of [255, 254] is to decompose the marked vertices in these adjacency graphs into (maximal) connected components[20]. By definition, the total weight of $R + C$ must be less than the total error on the Bulk $P + B$. What is non-trivial is that this is true even within each connected components of marked vertices, as formalized in Lemma 4.19 below.

[19] Note that R, C define $Z(\beta)$, in the notation of Section 4.3.1.4.

[20] A connected component K of marked vertices is maximal if there is no marked vertex adjacent to but not in K.

Lemma 4.19 (Mismatched errors form connected components)**.** *Consider a connected component K within the Bulk of the X syndrome adjacency graph. Then,*

$$\sum_{t\in[1,T-1]} \big|R_{2t+1}\big|_K + \sum_{t\in[1,T]} \big|C_{2t}\big|_K \leq \sum_{t\in[1,T-1]} \big|P_{2t+1}\big|_K + \sum_{t\in[1,T]} \big|B_{2t}\big|_K, \tag{4.118}$$

and similarly for the Z syndrome adjacency graph.

Proof. By definition, R, C are chosen to have minimal weight consistent with the syndrome $\mathcal{S}^0$ information. We claim that if on K, the weight of R, C are not minimal, then we could swap $R|_K, C|_K \leftrightarrow P|_K, B|_K$ on K, and decrease the overall error weight. Indeed, to see that performing such a swap still produces an operator R', C' consistent with the syndrome information, note that (1) all checks contained entirely within K are consistent, since so is $P|_K, B|_K$; (2) on the "closure" $\bar{K}$ of K (K and its boundary), $R'|_{\bar{K}}, C'|_{\bar{K}} = P|_{\bar{K}}, B|_{\bar{K}}$, since K is a *maximal* connected component. □

The clustering arguments in the next lemmas require a short technical fact.

Fact 4.2 ([255, 254])**.** *Consider a set T of t nodes in a graph G of degree $\leq z$. The number of sets S of nodes which contain T, of total size s, and which form a union of connected components in G, is $\leq z^{s-t}\cdot 4^s$.*

The first step in the clustering argument is to reason that there does not exist any connected component (in either X or Z syndrome adjacency graph) which connects the two boundary codes in ∂. This is the only location the third dimension/number of layers of the graph state, T, appears. Henceforth, we will refer to this non-connected boundary condition as the *clustering condition*, or $\mathsf{CC}_X, \mathsf{CC}_Z$. As we discuss shortly, conditioned on this event, the residual errors can be described as local stochastic errors.

Lemma 4.20 (The Boundaries aren't Connected)**.** *There exists $p_0 \in (0,1)$ s.t. $\forall p < p_0$, the probability there exists a connected component K in the Bulk of the X syndrome adjacency graph which spans the two boundaries is*

$$1 - \mathbb{P}[\mathsf{CC}_X] \equiv \mathbb{P}_{P\leftarrow N(p)}\bigg[\exists K \text{ which spans } \partial\bigg] \leq m_x \cdot \frac{(p/p_0)^{T/2}}{1-\sqrt{p/p_0}}, \tag{4.119}$$

where $p_0 \equiv (8z)^{-2}$. The Z clusters are analogous.

Proof. Let us fix a connected component of size $|K| = s$. We aim to find a lower bound on the number of Bulk Z errors which occur in K, as a function of s. For this purpose, note that the number of marked vertices satisfies:[21]

$$\begin{aligned} s = &\sum_{t\in[1,T-1]} \left|R_{2t+1}P_{2t+1}\right|_K + \sum_{t\in[1,T]} \left|C_{2t}B_{2t}\right|_K \\ \leq &\sum_{t\in[1,T-1]} \left(\left|R_{2t+1}\right|_K + \left|P_{2t+1}\right|_K\right) + \sum_{t\in[1,T]} \left(\left|C_{2t}\right|_K + \left|B_{2t}\right|_K\right). \end{aligned} \tag{4.120}$$

Next, we leverage the fact that the weight of the inferred error (R, C) is *minimal*, from Lemma 4.19. That is, within any connected component K,

$$\sum_{t\in[1,T-1]} \left|R_{2t+1}\right|_K + \sum_{t\in[T]} \left|C_{2t}\right|_K \leq \sum_{t\in[1,T-1]} \left|P_{2t+1}\right|_K + \sum_{t\in[T]} \left|B_{2t}\right|_K. \tag{4.121}$$

This implies that

$$\sum_{t\in[1,T-1]} \left|P_{2t+1}\right|_K + \sum_{t\in[T]} \left|B_{2t}\right|_K \geq \frac{s}{2}. \tag{4.122}$$

Therefore, there are at least $s/2$ physical errors (P, B) in K. Finally, if a connected component K spans the boundaries of ∂, it must have size $s \geq T$. By a union bound,

$$\begin{aligned} &\mathbb{P}_{P\leftarrow N(p)}\left[\exists K \text{ which spans } \partial\right] \\ =&\mathbb{P}_{P\leftarrow N(p)}\left[\exists K, \exists i, j \in [m_x] \text{ s.t. } (i,1),(j,T) \in K\right] \\ \leq&\sum_{s\geq T}\left(\#\text{ Clusters of Size } s \text{ incident on } \partial\right)\cdot\left(\#\text{ Error Patterns}\right)\cdot p^{s/2} \\ \leq&m_x\cdot\sum_{s\geq T}(4z)^s\cdot 2^s\cdot p^{s/2} \leq m_x\frac{(p/p_0)^{T/2}}{1-\sqrt{p/p_0}} \end{aligned} \tag{4.123}$$

where $p_0 \equiv (8z)^{-2}$. In the last inequality, we leveraged Fact 4.2 and the fact that the number of ways to pick $s/2$ locations out of a set of size s is $\leq 2^s$. □

We are now in a position to prove the residual errors $\mathsf{Rep}_Z, \mathsf{Rep}_X$ are local stochastic noise, so long as we condition on the disconnected boundaries condition CC of Lemma 4.20. The proof strategy is similar to that above, and that of [254]: We relate the size of the clusters to the number of true physical errors within it, and subsequently union bound over such configurations of clusters.

To proceed, we need another short lemma on the weight of clusters connected to *only one* of the boundary codes:

[21] Here, we let $|RP|$ denote the number of locations the vectors R, P differ, or alternatively, the weight of the operator associated to the product of R, P as Pauli Z operators.

Lemma 4.21. *Consider a connected component K in the X syndrome adjacency graph, incident on only one of the boundary codes. Let $G = \mathsf{Rep}_X(P)$ denote the residual Z error on ∂. Then,*

$$|G|_K \leq \sum_{t\in[1,T-1]} \big|P_{2t+1} \cdot R_{2t+1}\big|_K. \tag{4.124}$$

Proof. The crux of the proof lies in the following claim (which we prove shortly): If the connected component K is incident on only one of the boundaries of ∂, then the operators $G|_K \in \mathsf{Pauli}(n)$ and $\prod_{t\in[1,T-1]} P_{2t+1}|_K \cdot R_{2t+1}|_K \in \mathsf{Pauli}(n)$ have the same X syndrome. This tells us $G|_K$ and $\prod_{t\in[1,T-1]} P_{2t+1}|_K R_{2t+1}|_K$ differ only by a Z stabilizer. However, if G is *minimal*, then $G|_K$ must have weight less than that of $\prod_{t\in[1,T-1]} P_{2t+1}|_K R_{2t+1}|_K$. Otherwise, we could replace G with $G \cdot G|_K \cdot \prod_{t\in[1,T-1]} P_{2t+1}|_K R_{2t+1}|_K$, and decrease the overall weight of G, without changing the syndrome information. Thus,

$$|G|_K \leq \left| \prod_{t\in[1,T-1]} P_{2t+1}\big|_K R_{2t+1}\big|_K \right| \leq \sum_{t\in[1,T-1]} \big|P_{2t+1} \cdot R_{2t+1}\big|_K. \tag{4.125}$$

To prove the missing claim, we note two facts. Let us assume, WLOG, that K is incident on the boundary on the first layer. First, recall that the residual error $G|_K = \mathsf{Rep}_X(P)|_K$ is, by definition, the minimum weight Z error consistent with the *residual* X syndrome (restricted to the cluster K). By the definition of the Pauli frame in Fig. 4.8, we observe that this residual syndrome is simply $B_2|_K \oplus C_2|_K$, the mismatch on the first layer of X checks.

Therefore, it remains to show that $\mathsf{Syn}_X(\prod_{t\in[1,T-1]} P_{2t+1}|_K \cdot R_{2t+1}|_K) = B_2|_K \oplus C_2|_K$. By a telescoping argument using Eq. (4.116),

$$\begin{aligned}
\mathsf{Syn}_X\Big(\prod_{t\in[1,T-1]} P_{2t+1}|_K \cdot R_{2t+1}|_K\Big) &= \sum_{t\in[1,T-1]} \mathsf{Syn}_X(P_{2t+1}|_K \cdot R_{2t+1}|_K) \\
&= \sum_{t\in[1,T-1]} \Big(B_{2t}|_K \oplus B_{2t+2}|_K \oplus C_{2t}|_K \oplus C_{2t+2}|_K\Big) \\
&= B_2|_K \oplus C_2|_K \oplus B_{2t^*}|_K \oplus C_{2t^*}|_K,
\end{aligned} \tag{4.126}$$

where we assume the cluster K is entirely contained within layers $[1, t^* < T]$. However, note that the last layer of K must be a "code" layer, i.e. we must have $B_{2t^*}|_K \oplus C_{2t^*}|_K = 0$. As otherwise, by Eq. (4.116) and the meta-check connectivity, we must have at least one connected node at some layer $> 2t^*$, a contradiction to K being contained within $[1, t^* < T]$. □

Lemma 4.22 (The Residual Error is Stochastic)**.** *Let $S \subset \partial$ denote a subset of qubits on the boundary of size $|S| = a$. Then there exists $p_1 \in (0, 1)$ s.t. $\forall p < p_1$,*

$$\mathbb{P}_{P\leftarrow N(p)}\Big[S \subseteq \mathsf{Supp}(\mathsf{Rep}_X(P)) \Big| \mathsf{CC}_X\Big] \leq \frac{(p/p_1)^{a/2}}{1-(p/p_1)^{1/4}} \cdot \frac{1}{\mathbb{P}[\mathsf{CC}_X]}, \tag{4.127}$$

Where $p_1 = (8z)^{-4}$*, and* $\mathbb{P}[\mathsf{CC}_X]$ *is defined in Lemma 4.20.* $\mathsf{Rep}_Z(P)$ *is analogous.*

Proof. We wish to bound the probability that the residual error $G = \mathsf{Rep}_X(P)$ is supported on a subset S of size $|S| = a$. Let us once again decompose the X syndrome adjacency graph into clusters, and sum up the total size of clusters connected to (and including) G, let this size be s. Assuming none of these clusters span the two boundary codes, we have from Lemma 4.21,

$$2a \leq 2 \cdot \sum_K |G|_K \leq \sum_K |G|_K + \sum_{t\in[1,T-1]} \big|P_{2t+1}R_{2t+1}\big|_K \leq s. \tag{4.128}$$

Moreover, since the weight of R, C is minimal, they must have weight less than that of P, B on each cluster (Lemma 4.19):

$$\begin{aligned} s &= \sum_K \Bigg(|G|_K + \sum_{t\in[1,T-1]} \big|R_{2t+1}P_{2t+1}\big|_K + \sum_{t\in[1,T]} \big|C_{2t}B_{2t}\big|_K\Bigg) \\ &\leq 4 \cdot \sum_K \Bigg(\sum_{t\in[1,T-1]} \big|P_{2t+1}\big|_K + \sum_{t\in[1,T]} \big|B_{2t}\big|_K\Bigg). \end{aligned} \tag{4.129}$$

In other words, there must be at least $s/4$ real errors "connected" to the residual error G. We can now apply a union bound over the cluster configurations:

$$\begin{aligned} &\mathbb{P}_{P\leftarrow N(p)}\Big[S \subseteq \mathsf{Supp}(\mathsf{Rep}_X(P)) \text{ and } \mathsf{CC}_X\Big] \\ &\leq \sum_{s\geq 2a} \Big(\#\text{ Clusters of Size } s\Big) \cdot \Big(\#\text{ Error Patterns}\Big) \cdot p^{s/4} \\ &\leq \sum_{s\geq 2a} \big(4^s \cdot z^{s-a}\big) \cdot \big(2^s\big) \cdot p^{s/4} \leq \big(2^6 z\sqrt{p}\big)^a \cdot \sum_{i\geq 0} (2^3 z p^{1/4})^i \leq \frac{(p/p_1)^{a/2}}{1-(p/p_1)^{1/4}}. \end{aligned} \tag{4.130}$$

So long as $p \leq p_1 \equiv (8z)^{-4}$. In the above, we leverage Fact 4.2. Bayes rule with $\mathbb{P}[\mathsf{CC}_X]$ concludes the proof. The Z errors are analogous. □

It only remains to consider the logical stabilizers of $\bar{\Phi}$. As described in Section 4.3.1.3, $\mathsf{Rep}_{\bar{X}} \in \{\mathbb{I}, \bar{Z}_1\}$ quantifies the necessary logical correction operation on ∂, to ensure the resulting state is $\bar{\Phi}$. The lemma below stipulates that except with exponentially small probability, this correction operator is simply identity.

Lemma 4.23 (There are no Logical Errors)**.** *The probability the* X *logical correction* $\mathsf{Rep}_{\bar{X}}(P)$ *is non-trivial is*

$$\mathbb{P}_{P\leftarrow N(p)}\Big[\mathsf{Rep}_{\bar{X}}(P) \neq \mathbb{I}\Big] \leq \frac{(p/p_2)^{d/4}}{1-(p/p_2)^{1/4}} \cdot \frac{1}{\mathbb{P}[\mathsf{CC}_X]}, \tag{4.131}$$

where $p_2 = (8z)^{-4}$*, and* $\mathbb{P}[\mathsf{CC}_X]$ *was defined in Lemma 4.20. The* Z *correction* $\mathsf{Rep}_{\bar{Z}}(P)$ *is analogous.*

Proof. Suppose, after we apply $\mathsf{Rep}_X(P) = \mathsf{Rep}_X(P)_1 \otimes \mathsf{Rep}_X(P)_{2T+1}$ to the boundary ∂ of the post-measurement state, we were able to perform a perfect (noiseless) syndrome measurement of an encoded $\bar{X}_1 \otimes \bar{X}_{2T+1}$ stabilizer. By definition of the associated encoded logical stabilizers from Section 4.3.1.3, the resulting syndrome outcome $s_{\bar{X}} \in \{0, 1\}$ is

$$s_{\bar{X}} = \mathsf{Syn}_{\bar{X}_1 \bar{X}_{2T+1}}\left(\mathsf{Rep}_X(P)_1 \otimes \mathsf{Rep}_X(P)_{2T+1} \bigotimes_{t\in[1,T-1]} P_{2t+1} \cdot R_{2t+1}\right). \tag{4.132}$$

If we let the support of $\bar{X}$ be $\alpha_x \subset [n]$, this syndrome outcome is equal to the parity of the following operator $Z(\gamma), \gamma \subset [n]$ on α_x:

$$Z(\gamma) \equiv \mathsf{Rep}_X(P)_1 \cdot \mathsf{Rep}_X(P)_{2T+1} \cdot \prod_{t\in[1,T-1]} P_{2t+1} \cdot R_{2t+1}, \quad s_{\bar{X}} = |\gamma \cap \alpha_x| \mod 2 \tag{4.133}$$

We claim that this parity is 0 with high probability, such that no logical correction $\mathsf{Rep}_{\bar{X}}(P)_1$ is needed. To see this, we follow the procedure from the previous proofs, and decompose the X syndrome adjacency graph into clusters. Recall via the proof of Lemma 4.21 that each cluster defines a logical (or trivial) operator of Q, since

$$\mathsf{Syn}_X\left(\mathsf{Rep}_X(P)|_K \cdot \prod_{t\in[1,T-1]} P_{2t+1}|_K R_{2t+1}|_K\right) = 0. \tag{4.134}$$

Moreover, if this product of operators on K is an X stabilizer of Q, then by definition it has 0 logical $\bar{X}_1\bar{X}_{2T+1}$ syndrome. The only remaining possibility lies in if the cluster K defines a logical operator. If so, we must have that the size of $|K| = s \geq d$, the distance of the code. By a similar argument behind Lemma 4.21 and Lemma 4.19, if we condition on K not spanning the boundaries (event CC_X), then K contains at least $\frac{s}{4}$ physical errors.

Via the percolation-based union bound, the probability there exists any cluster K with non-trivial $\mathsf{Syn}_{\bar{X}_1\bar{X}_{2T+1}}$ syndrome is then

$$\begin{aligned} &\mathbb{P}_{P\leftarrow N(p)}\left[\mathsf{Syn}_{\bar{X}_1 \bar{X}_{2T+1}}\left(\mathsf{Rep}_X(P)_1 \otimes \mathsf{Rep}_X(P)_{2T+1} \bigotimes_{t\in[1,T-1]} P_{2t+1} \cdot R_{2t+1}\right) = 1 \text{ and } \mathsf{CC}_X\right] \\ &\leq \sum_{s\geq d}\left(\# \text{ Clusters of Size } s\right) \cdot \left(\# \text{ Error Patterns}\right) \cdot p^{s/4} \\ &\leq \sum_{s\geq d}\left(4^s \cdot z^s\right) \cdot \left(2^s\right) \cdot p^{s/4} \leq \left(8zp^{1/4}\right)^d \cdot \frac{(p/p_2)^{d/4}}{1-(p/p_2)^{1/4}} \end{aligned} \tag{4.135}$$

where $p_2 = (8z)^{-4}$. In the above, we leveraged Fact 4.2.

□

4.3.3 Omitted proofs

We dedicate this section to the proofs that the stabilizers defined in Section 4.3.1.3 factorize into X Pauli operators on the Bulk $\mathcal{B}$. To simplify notation, let $N_u = \{v : (u, v) \in E\}$ denote the neighborhood of a vertex $u \in V$ in the graph G.

Lemma 4.24 (The Meta-Checks $\mathcal{S}^0$)**.** *For each even layer t and $c \in [m_z]$, the associated Z meta-check factorizes into a product of X Pauli operators on $\mathcal{B}$:*

$$G_{(c,t-1)} \cdot G_{(c,t+1)} \cdot \prod_{i\sim c} G_{(i,t)} = X_{(c,t-1)} \otimes X_{(c,t+1)} \bigotimes_{i\sim c} X_{(i,t)} \tag{4.136}$$

And similarly for the X meta-checks.

Proof. For any meta-check centered around the "meta-vertex" (c, t), the product of graph state stabilizers around it can be written in terms of its "neighborhood of neighborhoods". The neighborhood $N_{(c,t)}$ of (c, t) consists of the check qubits $(c, t \pm 1)$, and the code qubits (i, t) s.t. $H^Z_{c,i} = 1$ (i.e. $i \sim c$). The "neighborhood of neighborhoods" is then the multi-set (a set with repetitions) $\cup_{u\in N_{(c,t)}} \cup_{v\in N_u}$, which allows us to write the product of stabilizers above as:

$$G_{(c,t-1)} \cdot G_{(c,t+1)} \cdot \prod_{i\sim c} G_{(i,t)} = X_{(c,t-1)} \otimes X_{(c,t+1)} \bigotimes_{i\sim c} X_{(i,t)} \prod_{u\in N_{(c,t)}} \prod_{v\in N_u} Z_v \tag{4.137}$$

We claim that each qubit in $\cup_{u\in N_{(c,t)}} \cup_{v\in N_u}$ appears an even number of times in this multiset, such that the product of Z Paulis above cancels. In fact, there are only two cases we need to consider: First, consider the code qubits $(i, t \pm 1)$ in layers above and below t. Each such node is in both $N_{(c,t\pm1)}$ and in $N_{(i,t)}$, thus counted twice.

The challenge lies in the X check qubits at layer t. For any $d \in [m_x]$, the check qubit (d, t) lies in the "neighborhood of neighborhoods" of (c, t) through a code qubit (i, t) iff $H^Z_{c,i} \cdot H^X_{d,i} = 1$. Thereby, the parity of the number of appearances is

$$\sum_i H^Z_{c,i} \cdot H^X_{d,i} = \left(H^Z (H^X)^T\right)_{c,d} = 0 \tag{4.138}$$

Since (H^X, H^Z) defines a CSS code. □

The stabilizers in $\mathcal{S}^1$ arise in two types. Recall that $\left|\bar{\Phi}\right\rangle$ consists of an encoded maximally entangled state accross two copies of the LDPC code Q. Then, within $\mathcal{S}^1$ there will be stabilizers of the individual boundary codes, and encoded stabilizers of the Bell state.

Lemma 4.25 (The Stabilizers of the Boundary Codes)**.** *Each X-type or Z-type stabilizer S_∂ of the two boundary LDPC codes Q can be written in the form $X(\alpha) \otimes S_\partial \in \mathcal{S}^1$ via a product of graph state stabilizers. Explicitly,*

1. *For each Z-type stabilizer $c \in [m_z]$ of Q, there exists graph state stabilizers $\in \mathcal{S}^1$, satisfying the decomposition*

$$G_{(c,1)} = X_{(c,1)} \bigotimes_{i\sim c} Z_{(i,1)}, \quad G_{(c,2T+1)} = X_{(c,2T+1)} \bigotimes_{i\sim c} Z_{(i,2T+1)} \tag{4.139}$$

2. *For each X-type stabilizer $c \in [m_x]$ of Q, there exists products of graph state stabilizers $\in \mathcal{S}^1$, satisfying the decomposition*

$$G_{(c,2)} \cdot \prod_{i\sim c} G_{(i,1)} = X_{(c,2)} \bigotimes_{i\sim c} X_{(i,1)}, \quad G_{(c,2T)} \cdot \prod_{i\sim c} G_{(i,2T+1)} = X_{(c,2T)} \bigotimes_{i\sim c} X_{(i,2T+1)}. \tag{4.140}$$

We remark that both of the decompositions above are simply applying an X or Z stabilizer of the LDPC code on either the first or last layer of G.

Proof. Case 1 is rather straightforward, as the graph state stabilizer $S_{(c,1)}$ (resp, $2T+1$) is precisely applying a Z stabilizer of Q on the associated boundary code. Case 2 is more subtle. However, we can follow the reasoning in Lemma 4.24: to show that the Z Pauli's arising from the graph state stabilizers cancel, it suffices to show they are counted an even number of times from the neighborhoods of $(c, 2)$ and $(i, 1)$ where $H^X_{c,i} = 1$. Indeed, the code qubits $(i, 2)$ are counted exactly twice (due to the vertical connections), and the Z-type check qubits $(d, 1)$ are counted an even number of times since Q is a CSS code. □

The encoded Bell pairs are stabilized by products $\bar{X}_1 \otimes \bar{X}_{2T+1}, \bar{Z}_1 \otimes \bar{Z}_{2T+1}$ of logical operators. The Lemma below shows how to construct these operators using graph state stabilizers in G, and only X operators on the Bulk.

Lemma 4.26 (The Encoded Stabilizers of Φ). *Every encoded stabilizer S of Φ can be written as a product of graph state stabilizers $S_\partial \otimes X(\alpha) \in \mathcal{S}^1$, for some subset $\alpha \subset \mathcal{B}$. Explicitly,*

1. *Let $\alpha_x \subset [n]$ denote the support of a logical $\bar{X}$ on Q. Then,*

$$\prod_{\substack{i\in\alpha_x \\ t \text{ odd}}} G_{(i,t)} = \bar{X}_1 \otimes \bar{X}_{2T+1} \bigotimes_{\substack{i\in\alpha_x \\ t \text{ odd} \in\mathcal{B}}} X_{(i,t)} \tag{4.141}$$

2. *Let $\alpha_z \subset [n]$ denote the support of a logical $\bar{Z}$ on Q. Then,*

$$\prod_{\substack{i\in\alpha_z \\ t \text{ even}}} G_{(i,t)} = \bar{Z}_1 \otimes \bar{Z}_{2T+1} \bigotimes_{\substack{i\in\alpha_z \\ t \text{ even}}} X_{(i,t)} \tag{4.142}$$

Proof. The argument follows the strategy of the previous two lemmas. The alternating even/odd layers implies that the Z operations arising due to the vertical connections are each counted twice, thus cancelling. The connections to ancilla qubits on each layer cancel due to the CSS condition, as logical operators by definition have support α_x (resp α_z) with even overlap with the support of Z (resp X) parity checks. □

Bibliography

[1] Michael Sipser. *Introduction to the Theory of Computation*. 3rd ed. Cengage Learning, 2012.

[2] Ethan Bernstein and Umesh Vazirani. "Quantum Complexity Theory". In: *SIAM Journal on Computing* 26.5 (1997), pp. 1411–1473. DOI: 10.1137/S0097539796300921.

[3] Peter W. Shor. "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer". In: *SIAM Journal on Computing* 26.5 (1997), pp. 1484–1509. DOI: 10.1137/S0097539795293172.

[4] Adam Bouland, Bill Fefferman, Zeph Landau, and Yunchao Liu. "Noise and the Frontier of Quantum Supremacy". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. 2022, pp. 1308–1317. DOI: 10.1109/FOCS52979.2021.00127.

[5] Dorit Aharonov, Xun Gao, Zeph Landau, Yunchao Liu, and Umesh Vazirani. "A Polynomial-Time Classical Algorithm for Noisy Random Circuit Sampling". In: *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. STOC 2023. Orlando, FL, USA: Association for Computing Machinery, 2023, pp. 945–957. DOI: 10.1145/3564246.3585234.

[6] Yunchao Liu, Matthew Otten, Roozbeh Bassirianjahromi, Liang Jiang, and Bill Fefferman. *Benchmarking near-term quantum computers via random circuit sampling*. 2021. arXiv: 2105.05232 [quant-ph].

[7] Frank Arute, Kunal Arya, Ryan Babbush, et al. "Quantum supremacy using a programmable superconducting processor". In: *Nature* 574.7779 (Oct. 2019), pp. 505–510. DOI: 10.1038/s41586-019-1666-5.

[8] Dominik Hangleiter and Jens Eisert. "Computational advantage of quantum random sampling". In: *Rev. Mod. Phys.* 95 (3 July 2023), p. 035001. DOI: 10.1103/RevModPhys.95.035001.

[9] Scott Aaronson and Sam Gunn. "On the Classical Hardness of Spoofing Linear Cross-Entropy Benchmarking". In: *Theory of Computing* 16.11 (2020), pp. 1–8. DOI: 10.4086/toc.2020.v016a011.

[10] Johnnie Gray and Stefanos Kourtis. "Hyper-optimized tensor network contraction". In: *Quantum* 5 (Mar. 2021), p. 410. DOI: 10.22331/q-2021-03-15-410.

[11] Cupjin Huang, Fang Zhang, Michael Newman, et al. *Classical Simulation of Quantum Supremacy Circuits*. May 2020. arXiv: 2005.06787 [quant-ph].

[12] Feng Pan and Pan Zhang. "Simulation of Quantum Circuits Using the Big-Batch Tensor Network Method". In: *Phys. Rev. Lett.* 128 (3 Jan. 2022), p. 030501. DOI: 10.1103/PhysRevLett.128.030501.

[13] Gleb Kalachev, Pavel Panteleev, and Man-Hong Yung. *Multi-Tensor Contraction for XEB Verification of Quantum Circuits*. Aug. 2021. arXiv: 2108.05665 [quant-ph].

[14] Feng Pan, Keyang Chen, and Pan Zhang. "Solving the Sampling Problem of the Sycamore Quantum Circuits". In: *Phys. Rev. Lett.* 129 (9 Aug. 2022), p. 090502. DOI: 10.1103/PhysRevLett.129.090502.

[15] Gleb Kalachev, Pavel Panteleev, PengFei Zhou, and Man-Hong Yung. *Classical Sampling of Random Quantum Circuits with Bounded Fidelity*. Dec. 2021. arXiv: 2112.15083 [quant-ph].

[16] Yulin Wu, Wan-Su Bao, Sirui Cao, et al. "Strong Quantum Computational Advantage Using a Superconducting Quantum Processor". In: *Phys. Rev. Lett.* 127 (18 Oct. 2021), p. 180501. DOI: 10.1103/PhysRevLett.127.180501.

[17] Qingling Zhu, Sirui Cao, Fusheng Chen, et al. "Quantum computational advantage via 60-qubit 24-cycle random circuit sampling". In: *Science Bulletin* 67.3 (2022), pp. 240–245. DOI: https://doi.org/10.1016/j.scib.2021.10.017.

[18] Xun Gao, Marcin Kalinowski, Chi-Ning Chou, Mikhail D. Lukin, Boaz Barak, and Soonwon Choi. "Limitations of Linear Cross-Entropy as a Measure for Quantum Advantage". In: *PRX Quantum* 5 (1 Feb. 2024), p. 010334. DOI: 10.1103/PRXQuantum.5.010334.

[19] Scott Aaronson and Lijie Chen. "Complexity-Theoretic Foundations of Quantum Supremacy Experiments". In: *Proceedings of the 32nd Computational Complexity Conference*. CCC '17. Riga, Latvia: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.

[20] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J. Bremner, John M. Martinis, and Hartmut Neven. "Characterizing quantum supremacy in near-term devices". In: *Nature Physics* 14.6 (June 2018), pp. 595–600. DOI: 10.1038/s41567-018-0124-x.

[21] A. Morvan, B. Villalonga, X. Mi, et al. *Phase transition in Random Circuit Sampling*. 2023. arXiv: 2304.11119 [quant-ph].

[22] Alexander M. Dalzell, Nicholas Hunter-Jones, and Fernando G. S. L. Brandão. "Random Quantum Circuits Transform Local Noise into Global White Noise". In: *Communications in Mathematical Physics* 405.3 (Mar. 2024), p. 78. DOI: 10.1007/s00220-024-04958-z.

[23] Joonhee Choi, Adam L. Shaw, Ivaylo S. Madjarov, et al. "Preparing random states and benchmarking with many-body quantum chaos". In: *Nature* 613.7944 (Jan. 2023), pp. 468–473. DOI: 10.1038/s41586-022-05442-1.

[24] Scott Aaronson and Alex Arkhipov. "The Computational Complexity of Linear Optics". In: *Theory of Computing* 9.4 (2013), pp. 143–252. DOI: 10.4086/toc.2013.v009a004.

[25] Larry Stockmeyer. "On Approximation Algorithms for #P". In: *SIAM Journal on Computing* 14.4 (1985), pp. 849–861. DOI: 10.1137/0214060.

[26] Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. "On the complexity and verification of quantum random circuit sampling". In: *Nature Physics* 15.2 (Feb. 2019), pp. 159–163. DOI: 10.1038/s41567-018-0318-2.

[27] Ramis Movassagh. *Quantum supremacy and random circuits*. 2020. arXiv: 1909.06210 [quant-ph].

[28] Richard J. Lipton. "New Directions In Testing". In: *Distributed Computing And Cryptography, Proceedings of a DIMACS Workshop, Princeton, New Jersey, USA, October 4-6, 1989*. Ed. by Joan Feigenbaum and Michael Merritt. Vol. 2. DIMACS Series in Discrete Mathematics and Theoretical Computer Science. DIMACS/AMS, 1989, pp. 191–202. DOI: 10.1090/dimacs/002/13.

[29] Peter Gemmell, Richard Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson. "Self-Testing/Correcting for Polynomials and for Approximate Functions". In: *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*. STOC '91. New Orleans, Louisiana, USA: Association for Computing Machinery, 1991, pp. 33–42. DOI: 10.1145/103418.103429.

[30] Scott Aaronson and Alex Arkhipov. "Bosonsampling is Far from Uniform". In: *Quantum Info. Comput.* 14.15–16 (Nov. 2014), pp. 1383–1423.

[31] E. A. Rakhmanov. "Bounds for Polynomials with a Unit Discrete Norm". In: *Annals of Mathematics* 165.1 (2007), pp. 55–88.

[32] Ramamohan Paturi. "On the Degree of Polynomials That Approximate Symmetric Boolean Functions (Preliminary Version)". In: *Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing*. STOC '92. Victoria, British Columbia, Canada: ACM, 1992, pp. 468–474. DOI: 10.1145/129712.129758.

[33] G.G. Lorentz. *Approximation of Functions*. AMS Chelsea Publishing Series. Holt, Rinehart and Winston, 2005.

[34] Alexander A. Sherstov. "Making Polynomials Robust to Noise". In: *Theory of Computing* 9.18 (2013), pp. 593–615. DOI: 10.4086/toc.2013.v009a018.

[35] Alexander M. Dalzell, Nicholas Hunter-Jones, and Fernando G. S. L. Brandão. "Random Quantum Circuits Anticoncentrate in Log Depth". In: *PRX Quantum* 3 (1 Mar. 2022), p. 010333. DOI: 10.1103/PRXQuantum.3.010333.

[36] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan. *Limitations of Noisy Reversible Computation*. 1996. arXiv: quant-ph/9611028 [quant-ph].

[37] Xun Gao and Luming Duan. *Efficient classical simulation of noisy quantum computation*. Oct. 2018. arXiv: 1810.03176 [quant-ph].

[38] Abhinav Deshpande, Pradeep Niroula, Oles Shtanko, Alexey V. Gorshkov, Bill Fefferman, and Michael J. Gullans. *Tight bounds on the convergence of noisy random circuits to the uniform distribution*. Dec. 2021. arXiv: 2112.00716 [quant-ph].

[39] Boaz Barak, Chi-Ning Chou, and Xun Gao. "Spoofing Linear Cross-Entropy Benchmarking in Shallow Quantum Circuits". In: *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Ed. by James R. Lee. Vol. 185. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021, 30:1–30:20. DOI: 10.4230/LIPIcs.ITCS.2021.30.

[40] Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd. "Achieving quantum supremacy with sparse and noisy commuting quantum computations". In: *Quantum* 1 (Apr. 2017), p. 8. DOI: 10.22331/q-2017-04-25-8.

[41] Julia Kempe, Oded Regev, Falk Uunger, and Ronald de Wolf. "Upper Bounds on the Noise Threshold for Fault-Tolerant Quantum Computing". In: *Quantum Info. Comput.* 10.5 (May 2010), pp. 361–376.

[42] Y. Kondo, R. Mori, and R. Movassagh. "Quantum supremacy and hardness of estimating output probabilities of quantum circuits". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. Los Alamitos, CA, USA: IEEE Computer Society, Feb. 2022, pp. 1296–1307. DOI: 10.1109/FOCS52979.2021.00126.

[43] Hari Krovi. *Average-case hardness of estimating probabilities of random quantum circuits with a linear scaling in the error exponent*. June 2022. arXiv: 2206.05642 [quant-ph].

[44] John C. Napp, Rolando L. La Placa, Alexander M. Dalzell, Fernando G. S. L. Brandão, and Aram W. Harrow. "Efficient Classical Simulation of Random Shallow 2D Quantum Circuits". In: *Phys. Rev. X* 12 (2 Apr. 2022), p. 021021. DOI: 10.1103/PhysRevX.12.021021.

[45] Aram W. Harrow and Richard A. Low. "Random Quantum Circuits are Approximate 2-designs". In: *Communications in Mathematical Physics* 291.1 (Oct. 2009), pp. 257–302. DOI: 10.1007/s00220-009-0873-6.

[46] John Preskill. "Quantum Computing in the NISQ era and beyond". In: *Quantum* 2 (Aug. 2018), p. 79. DOI: 10.22331/q-2018-08-06-79.

[47] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi. "Quantum certification and benchmarking". In: *Nature Reviews Physics* 2.7 (July 2020), pp. 382–390. DOI: 10.1038/s42254-020-0186-4.

[48] Alexander Erhard, Joel J. Wallman, Lukas Postler, Michael Meth, Roman Stricker, Esteban A. Martinez, Philipp Schindler, Thomas Monz, Joseph Emerson, and Rainer Blatt. "Characterizing large-scale quantum computers via cycle benchmarking". In: *Nature Communications* 10.1 (Nov. 2019), p. 5347. DOI: 10.1038/s41467-019-13068-7.

[49] Robin Harper, Steven T. Flammia, and Joel J. Wallman. "Efficient learning of quantum noise". In: *Nature Physics* 16.12 (Dec. 2020), pp. 1184–1188. DOI: 10.1038/s41567-020-0992-8.

[50] Mohan Sarovar, Timothy Proctor, Kenneth Rudinger, Kevin Young, Erik Nielsen, and Robin Blume-Kohout. "Detecting crosstalk errors in quantum information processors". In: *Quantum* 4 (Sept. 2020), p. 321. DOI: 10.22331/q-2020-09-11-321.

[51] Akel Hashim, Ravi K. Naik, Alexis Morvan, et al. "Randomized Compiling for Scalable Quantum Computing on a Noisy Superconducting Quantum Processor". In: *Phys. Rev. X* 11 (4 Nov. 2021), p. 041039. DOI: 10.1103/PhysRevX.11.041039.

[52] Adam Winick, Joel J. Wallman, and Joseph Emerson. "Simulating and Mitigating Crosstalk". In: *Phys. Rev. Lett.* 126 (23 June 2021), p. 230502. DOI: 10.1103/PhysRevLett.126.230502.

[53] Pavithran Iyer, Aditya Jain, Stephen D. Bartlett, and Joseph Emerson. *Efficient diagnostics for quantum error correction.* 2021. arXiv: 2108.10830 [quant-ph].

[54] Joseph Emerson, Robert Alicki, and Karol Życzkowski. "Scalable noise estimation with random unitary operators". In: *Journal of Optics B: Quantum and Semiclassical Optics* 7.10 (Sept. 2005), S347–S352. DOI: 10.1088/1464-4266/7/10/021.

[55] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. "Randomized benchmarking of quantum gates". In: *Phys. Rev. A* 77 (1 Jan. 2008), p. 012307. DOI: 10.1103/PhysRevA.77.012307.

[56] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. "Exact and approximate unitary 2-designs and their application to fidelity estimation". In: *Phys. Rev. A* 80 (1 July 2009), p. 012304. DOI: 10.1103/PhysRevA.80.012304.

[57] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. "Scalable and Robust Randomized Benchmarking of Quantum Processes". In: *Phys. Rev. Lett.* 106 (18 May 2011), p. 180504. DOI: 10.1103/PhysRevLett.106.180504.

[58] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. "Characterizing quantum gates via randomized benchmarking". In: *Phys. Rev. A* 85 (4 Apr. 2012), p. 042311. DOI: 10.1103/PhysRevA.85.042311.

[59] David C. McKay, Sarah Sheldon, John A. Smolin, Jerry M. Chow, and Jay M. Gambetta. "Three-Qubit Randomized Benchmarking". In: *Phys. Rev. Lett.* 122 (20 May 2019), p. 200502. DOI: 10.1103/PhysRevLett.122.200502.

[60] Steven T. Flammia and Joel J. Wallman. "Efficient Estimation of Pauli Channels". In: *ACM Transactions on Quantum Computing* 1.1 (Dec. 2020). DOI: 10.1145/3408039.

[61] Robin Harper, Wenjun Yu, and Steven T. Flammia. "Fast Estimation of Sparse Quantum Noise". In: *PRX Quantum* 2 (1 Feb. 2021), p. 010322. DOI: 10.1103/PRXQuantum.2.010322.

[62] Steven T Flammia and Ryan O'Donnell. "Pauli error estimation via population recovery". In: *Quantum* 5 (2021), p. 549.

[63] Steven T. Flammia. *Averaged circuit eigenvalue sampling*. 2021. arXiv: 2108.05803 [quant-ph].

[64] John Martinis. "Quantum supremacy using a programmable superconducting processor". In: *Talk at California Institute of Technology* (Nov. 2019).

[65] Scott Aaronson, Dorit Aharonov, Boaz Barak, Sergio Boixo, Adam Bouland, Sandy Irani, Gil Kalai, and Umesh Vazirani. "Quantum supremacy panel discussion". In: *The 4th Winter School in Computer Science and Engineering, IIAS* (Dec. 2019).

[66] Fernando G. S. L. Brandão, Aram W. Harrow, and Michał Horodecki. "Local Random Quantum Circuits are Approximate Polynomial-Designs". In: *Communications in Mathematical Physics* 346.2 (Sept. 2016), pp. 397–434. DOI: 10.1007/s00220-016-2706-8.

[67] Aram W. Harrow and Saeed Mehraban. "Approximate Unitary t-Designs by Short Random Quantum Circuits Using Nearest-Neighbor and Long-Range Gates". In: *Communications in Mathematical Physics* 401.2 (July 2023), pp. 1531–1626. DOI: 10.1007/s00220-023-04675-z.

[68] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth. "Efficient Unitary Designs with a System-Size Independent Number of Non-Clifford Gates". In: *Communications in Mathematical Physics* 397.3 (Feb. 2023), pp. 995–1041. DOI: 10.1007/s00220-022-04507-6.

[69] Jonas Haferkamp and Nicholas Hunter-Jones. *Improved spectral gaps for random quantum circuits: large local dimensions and all-to-all interactions*. 2020. arXiv: 2012.05259 [quant-ph].

[70] Adam Nahum, Sagar Vijay, and Jeongwan Haah. "Operator Spreading in Random Unitary Circuits". In: *Phys. Rev. X* 8 (2 Apr. 2018), p. 021014. DOI: 10.1103/PhysRevX.8.021014.

[71] Tianci Zhou and Adam Nahum. "Emergent statistical mechanics of entanglement in random unitary circuits". In: *Phys. Rev. B* 99 (17 May 2019), p. 174205. DOI: 10.1103/PhysRevB.99.174205.

[72] Nicholas Hunter-Jones. *Unitary designs from statistical mechanics in random quantum circuits*. 2019. arXiv: 1905.12053 [quant-ph].

[73] Yimu Bao, Soonwon Choi, and Ehud Altman. "Theory of the phase transition in random unitary circuits with measurements". In: *Phys. Rev. B* 101 (10 Mar. 2020), p. 104301. DOI: 10.1103/PhysRevB.101.104301.

[74] Matthew Otten and Stephen K. Gray. "Accounting for errors in quantum algorithms via individual error reduction". In: *npj Quantum Information* 5.1 (Jan. 2019), p. 11. DOI: 10.1038/s41534-019-0125-3.

[75] Matthew Otten, Raman A. Shah, Norbert F. Scherer, Misun Min, Matthew Pelton, and Stephen K. Gray. "Entanglement of two, three, or four plasmonically coupled quantum dots". In: *Phys. Rev. B* 92 (12 Sept. 2015), p. 125432. DOI: 10.1103/PhysRevB.92.125432.

[76] Steven T. Flammia and Yi-Kai Liu. "Direct Fidelity Estimation from Few Pauli Measurements". In: *Phys. Rev. Lett.* 106 (23 June 2011), p. 230501. DOI: 10.1103/PhysRevLett.106.230501.

[77] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin. "Practical Characterization of Quantum Devices without Tomography". In: *Phys. Rev. Lett.* 107 (21 Nov. 2011), p. 210404. DOI: 10.1103/PhysRevLett.107.210404.

[78] Yosef Rinott, Tomer Shoham, and Gil Kalai. *Statistical Aspects of the Quantum Supremacy Demonstration*. 2020. arXiv: 2008.05177 [quant-ph].

[79] *IBM Quantum*. https://quantum-computing.ibm.com/. 2021.

[80] Jay M. Gambetta, A. D. Córcoles, S. T. Merkel, et al. "Characterization of Addressability by Simultaneous Randomized Benchmarking". In: *Phys. Rev. Lett.* 109 (24 Dec. 2012), p. 240504. DOI: 10.1103/PhysRevLett.109.240504.

[81] Andrew W. Cross, Lev S. Bishop, Sarah Sheldon, Paul D. Nation, and Jay M. Gambetta. "Validating quantum computers using randomized model circuits". In: *Phys. Rev. A* 100 (3 Sept. 2019), p. 032328. DOI: 10.1103/PhysRevA.100.032328.

[82] Samuele Ferracin, Theodoros Kapourniotis, and Animesh Datta. "Accrediting outputs of noisy intermediate-scale quantum computing devices". In: *New Journal of Physics* 21.11 (Nov. 2019), p. 113038. DOI: 10.1088/1367-2630/ab4fd6.

[83] Samuele Ferracin, Seth T. Merkel, David McKay, and Animesh Datta. *Experimental accreditation of outputs of noisy quantum computers*. 2021. arXiv: 2103.06603 [quant-ph].

[84] Robin Blume-Kohout and Kevin C. Young. "A volumetric framework for quantum computer benchmarks". In: *Quantum* 4 (Nov. 2020), p. 362. DOI: 10.22331/q-2020-11-15-362.

[85] Daniel Mills, Seyon Sivarajah, Travis L. Scholten, and Ross Duncan. *Application-Motivated, Holistic Benchmarking of a Full Quantum Computing Stack*. 2020. arXiv: 2006.01273 [quant-ph].

[86] Yulong Dong and Lin Lin. *Random circuit block-encoded matrix and a proposal of quantum LINPACK benchmark*. 2020. arXiv: 2006.04010 [quant-ph].

[87] Timothy Proctor, Kenneth Rudinger, Kevin Young, Erik Nielsen, and Robin Blume-Kohout. *Measuring the Capabilities of Quantum Computers*. 2020. arXiv: 2008.11294 [quant-ph].

[88] Matthew Otten and Stephen K. Gray. "Recovering noise-free quantum observables". In: *Phys. Rev. A* 99 (1 Jan. 2019), p. 012338. DOI: 10.1103/PhysRevA.99.012338.

[89] C. Neill, P. Roushan, K. Kechedzhi, et al. "A blueprint for demonstrating quantum supremacy with superconducting qubits". In: *Science* 360.6385 (2018), pp. 195–199. DOI: 10.1126/science.aao4309.

[90] Jonas Helsen, Ingo Roth, Emilio Onorati, Albert H. Werner, and Jens Eisert. *A general framework for randomized benchmarking*. 2020. arXiv: 2010.07974 [quant-ph].

[91] Timothy J. Proctor, Arnaud Carignan-Dugas, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young. "Direct Randomized Benchmarking for Multi-qubit Devices". In: *Phys. Rev. Lett.* 123 (3 July 2019), p. 030503. DOI: 10.1103/PhysRevLett.123.030503.

[92] Winton G. Brown and Bryan Eastin. "Randomized benchmarking with restricted gate sets". In: *Phys. Rev. A* 97 (6 June 2018), p. 062323. DOI: 10.1103/PhysRevA.97.062323.

[93] D S França and A K Hashagen. "Approximate randomized benchmarking for finite groups". In: *Journal of Physics A: Mathematical and Theoretical* 51.39 (Aug. 2018), p. 395302. DOI: 10.1088/1751-8121/aad6fa.

[94] Jonas Helsen, Xiao Xue, Lieven M. K. Vandersypen, and Stephanie Wehner. "A new class of efficient randomized benchmarking protocols". In: *npj Quantum Information* 5.1 (Aug. 2019), p. 71. DOI: 10.1038/s41534-019-0182-7.

[95] Easwar Magesan, Jay M. Gambetta, B. R. Johnson, et al. "Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking". In: *Phys. Rev. Lett.* 109 (8 Aug. 2012), p. 080505. DOI: 10.1103/PhysRevLett.109.080505.

[96] Arnaud Carignan-Dugas, Joel J. Wallman, and Joseph Emerson. "Characterizing universal gate sets via dihedral benchmarking". In: *Phys. Rev. A* 92 (6 Dec. 2015), p. 060302. DOI: 10.1103/PhysRevA.92.060302.

[97] Andrew W. Cross, Easwar Magesan, Lev S. Bishop, John A. Smolin, and Jay M. Gambetta. "Scalable randomised benchmarking of non-Clifford gates". In: *npj Quantum Information* 2.1 (Apr. 2016), p. 16012. DOI: 10.1038/npjqi.2016.12.

[98] Robin Harper and Steven T Flammia. "Estimating the fidelity of T gates using standard interleaved randomized benchmarking". In: *Quantum Science and Technology* 2.1 (Mar. 2017), p. 015008. DOI: 10.1088/2058-9565/aa5f8d.

[99] C. H. Baldwin, B. J. Bjork, J. P. Gaebler, D. Hayes, and D. Stack. “Subspace benchmarking high-fidelity entangling operations with trapped ions”. In: *Phys. Rev. Research* 2 (1 Mar. 2020), p. 013317. DOI: 10.1103/PhysRevResearch.2.013317.

[100] Jonas Helsen, Sepehr Nezami, Matthew Reagor, and Michael Walter. *Matchgate benchmarking: Scalable benchmarking of a continuous family of many-qubit gates.* 2020. arXiv: 2011.13048 [quant-ph].

[101] Jahan Claes, Eleanor Rieffel, and Zhihui Wang. “Character Randomized Benchmarking for Non-Multiplicity-Free Groups With Applications to Subspace, Leakage, and Matchgate Randomized Benchmarking”. In: *PRX Quantum* 2 (1 Mar. 2021), p. 010351. DOI: 10.1103/PRXQuantum.2.010351.

[102] Joel J. Wallman and Joseph Emerson. “Noise tailoring for scalable quantum computation via randomized compiling”. In: *Phys. Rev. A* 94 (5 Nov. 2016), p. 052325. DOI: 10.1103/PhysRevA.94.052325.

[103] Hsin-Yuan Huang, Richard Kueng, and John Preskill. “Predicting many properties of a quantum system from very few measurements”. In: *Nature Physics* 16.10 (Oct. 2020), pp. 1050–1057. DOI: 10.1038/s41567-020-0932-7.

[104] M. Otten. *QuaC: Open Quantum Systems in C, a time-dependent open quantum systems solver.* https://github.com/0tt3r/QuaC. 2017.

[105] M. B. Plenio and P. L. Knight. “The quantum-jump approach to dissipative dynamics in quantum optics”. In: *Rev. Mod. Phys.* 70 (1 Jan. 1998), pp. 101–144. DOI: 10.1103/RevModPhys.70.101.

[106] M. H. Devoret and R. J. Schoelkopf. “Superconducting Circuits for Quantum Information: An Outlook”. In: *Science* 339.6124 (2013), pp. 1169–1174. DOI: 10.1126/science.1231930.

[107] Norman F Ramsey. “A molecular beam resonance method with separated oscillating fields”. In: *Physical Review* 78.6 (1950), p. 695.

[108] Héctor Abraham et al. *Qiskit: An Open-source Framework for Quantum Computing.* 2019. DOI: 10.5281/zenodo.2562110.

[109] Alexander Zlokapa, Sergio Boixo, and Daniel Lidar. *Boundaries of quantum supremacy via random circuit sampling.* 2020. arXiv: 2005.02464 [quant-ph].

[110] Jin-Sung Kim, Lev S. Bishop, Antonio D. Córcoles, Seth Merkel, John A. Smolin, and Sarah Sheldon. “Hardware-efficient random circuits to classify noise in a multiqubit system”. In: *Phys. Rev. A* 104 (2 Aug. 2021), p. 022609. DOI: 10.1103/PhysRevA.104.022609.

[111] Hsin-Yuan Huang, Yunchao Liu, Michael Broughton, Isaac Kim, Anurag Anshu, Zeph Landau, and Jarrod R. McClean. "Learning Shallow Quantum Circuits". In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. STOC 2024. Vancouver, BC, Canada: Association for Computing Machinery, 2024, pp. 1343–1351. DOI: 10.1145/3618260.3649722.

[112] Zeph Landau and Yunchao Liu. *Learning quantum states prepared by shallow circuits in polynomial time*. 2024.

[113] Sergey Bravyi, David Gosset, and Robert Koenig. "Quantum advantage with shallow circuits". In: *Science* 362.6412 (2018), pp. 308–311.

[114] Sergey Bravyi, David Gosset, Robert König, and Marco Tomamichel. "Quantum advantage with noisy shallow circuits". In: *Nature Physics* 16.10 (Oct. 2020), pp. 1040–1045. DOI: 10.1038/s41567-020-0948-z.

[115] Adam Bene Watts and Natalie Parham. *Unconditional Quantum Advantage for Sampling with Shallow Circuits*. 2023. arXiv: 2301.00995 [quant-ph].

[116] Adam Bene Watts, Robin Kothari, Luke Schaeffer, and Avishay Tal. "Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2019. Phoenix, AZ, USA: Association for Computing Machinery, 2019, pp. 515–526. DOI: 10.1145/3313276.3316404.

[117] Barbara M. Terhal and David P. DiVincenzo. *Adaptive Quantum Computation, Constant Depth Quantum Circuits and Arthur-Merlin Games*. 2004. arXiv: quant-ph/0205133 [quant-ph].

[118] Xun Gao, Sheng-Tao Wang, and L.-M. Duan. "Quantum Supremacy for Simulating a Translation-Invariant Ising Spin Model". In: *Phys. Rev. Lett.* 118 (4 Jan. 2017), p. 040502. DOI: 10.1103/PhysRevLett.118.040502.

[119] Juan Bermejo-Vega, Dominik Hangleiter, Martin Schwarz, Robert Raussendorf, and Jens Eisert. "Architectures for Quantum Simulation Showing a Quantum Speedup". In: *Phys. Rev. X* 8 (2 Apr. 2018), p. 021010. DOI: 10.1103/PhysRevX.8.021010.

[120] Jonas Haferkamp, Dominik Hangleiter, Adam Bouland, Bill Fefferman, Jens Eisert, and Juani Bermejo-Vega. "Closing gaps of a quantum advantage with short-time hamiltonian dynamics". In: *Physical Review Letters* 125.25 (2020), p. 250501.

[121] Edward Farhi and Hartmut Neven. *Classification with Quantum Neural Networks on Near Term Processors*. 2018. arXiv: 1802.06002 [quant-ph].

[122] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. "Parameterized quantum circuits as machine learning models". In: *Quantum Science and Technology* 4.4 (2019), p. 043001.

[123] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf. "Training deep quantum neural networks". In: *Nature communications* 11.1 (2020), p. 808.

[124] Johannes Bausch. "Recurrent quantum neural networks". In: *Advances in neural information processing systems* 33 (2020), pp. 1368–1379.

[125] Andrea Skolik, Jarrod R McClean, Masoud Mohseni, Patrick van der Smagt, and Martin Leib. "Layerwise learning for quantum neural networks". In: *Quantum Machine Intelligence* 3 (2021), pp. 1–11.

[126] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. "The power of quantum neural networks". In: *Nature Computational Science* 1.6 (2021), pp. 403–409.

[127] Matthias C Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J Coles. "Generalization in quantum machine learning from few training data". In: *Nature communications* 13.1 (2022), p. 4919.

[128] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. "Cost function dependent barren plateaus in shallow parametrized quantum circuits". In: *Nature communications* 12.1 (2021), p. 1791.

[129] Mateusz Ostaszewski, Edward Grant, and Marcello Benedetti. "Structure optimization for parameterized quantum circuits". In: *Quantum* 5 (2021), p. 391.

[130] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T. Sornborger, and Patrick J. Coles. "Absence of Barren Plateaus in Quantum Convolutional Neural Networks". In: *Phys. Rev. X* 11 (4 Oct. 2021), p. 041011. DOI: 10.1103/PhysRevX.11.041011.

[131] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, Shan You, and Dacheng Tao. "Learnability of quantum neural networks". In: *PRX Quantum* 2.4 (2021), p. 040337.

[132] Zoë Holmes, Kunal Sharma, Marco Cerezo, and Patrick J Coles. "Connecting ansatz expressibility to gradient magnitudes and barren plateaus". In: *PRX Quantum* 3.1 (2022), p. 010313.

[133] Kunal Sharma, Marco Cerezo, Lukasz Cincio, and Patrick J Coles. "Trainability of dissipative perceptron-based quantum neural networks". In: *Physical Review Letters* 128.18 (2022), p. 180505.

[134] Eric R Anschuetz and Bobak T Kiani. "Quantum variational algorithms are swamped with traps". In: *Nature Communications* 13.1 (2022), p. 7760.

[135] M Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J Coles. "Challenges and opportunities in quantum machine learning". In: *Nature Computational Science* 2.9 (2022), pp. 567–576.

[136] Nathan Linial, Yishay Mansour, and Noam Nisan. “Constant depth circuits, Fourier transform, and learnability”. In: *Journal of the ACM (JACM)* 40.3 (1993), pp. 607–620.

[137] Elchanan Mossel, Ryan O’Donnell, and Rocco P. Servedio. “Learning Juntas”. In: *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing.* STOC ’03. San Diego, CA, USA: Association for Computing Machinery, 2003, pp. 206–212. DOI: 10.1145/780542.780574.

[138] Marco L. Carmosino, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. “Learning Algorithms from Natural Proofs”. In: *31st Conference on Computational Complexity (CCC 2016).* Ed. by Ran Raz. Vol. 50. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, 10:1–10:24. DOI: 10.4230/LIPIcs.CCC.2016.10.

[139] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. “Barren plateaus in quantum neural network training landscapes”. In: *Nature communications* 9.1 (2018), p. 4812.

[140] Zoë Holmes, Andrew Arrasmith, Bin Yan, Patrick J Coles, Andreas Albrecht, and Andrew T Sornborger. “Barren plateaus preclude learning scramblers”. In: *Physical Review Letters* 126.19 (2021), p. 190501.

[141] Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. “Noise-induced barren plateaus in variational quantum algorithms”. In: *Nature communications* 12.1 (2021), p. 6961.

[142] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. “The complexity of NISQ”. In: *Nature Communications* 14.1 (Sept. 2023), p. 6001. DOI: 10.1038/s41467-023-41217-6.

[143] Lukasz Cincio, Yiğit Subaşı, Andrew T Sornborger, and Patrick J Coles. “Learning the quantum algorithm for state overlap”. In: *New Journal of Physics* 20.11 (2018), p. 113022.

[144] Sumeet Khatri, Ryan LaRose, Alexander Poremba, Lukasz Cincio, Andrew T Sornborger, and Patrick J Coles. “Quantum-assisted quantum compiling”. In: *Quantum* 3 (2019), p. 140.

[145] Kunal Sharma, Sumeet Khatri, Marco Cerezo, and Patrick J Coles. “Noise resilience of variational quantum compiling”. In: *New Journal of Physics* 22.4 (2020), p. 043006.

[146] Tyson Jones and Simon C Benjamin. “Robust quantum compilation and circuit optimisation via energy minimisation”. In: *Quantum* 6 (2022), p. 628.

[147] Cristina Cirstoiu, Zoe Holmes, Joseph Iosue, Lukasz Cincio, Patrick J Coles, and Andrew Sornborger. “Variational fast forwarding for quantum simulation beyond the coherence time”. In: *npj Quantum Information* 6.1 (2020), p. 82.

[148] Yong-Xin Yao, Niladri Gomes, Feng Zhang, Cai-Zhuang Wang, Kai-Ming Ho, Thomas Iadecola, and Peter P Orth. "Adaptive variational quantum dynamics simulations". In: *PRX Quantum* 2.3 (2021), p. 030307.

[149] Joe Gibbs, Zoë Holmes, Matthias C. Caro, Nicholas Ezzell, Hsin-Yuan Huang, Lukasz Cincio, Andrew T. Sornborger, and Patrick J. Coles. "Dynamical simulation via quantum machine learning with provable generalization". In: *Phys. Rev. Res.* 6 (1 Mar. 2024), p. 013241. DOI: 10.1103/PhysRevResearch.6.013241.

[150] Matthias C Caro, Hsin-Yuan Huang, Nicholas Ezzell, Joe Gibbs, Andrew T Sornborger, Lukasz Cincio, Patrick J Coles, and Zoë Holmes. "Out-of-distribution generalization for learning quantum dynamics". In: *Nature Communications* 14.1 (2023), p. 3751.

[151] Sofiene Jerbi, Joe Gibbs, Manuel S. Rudolph, Matthias C. Caro, Patrick J. Coles, Hsin-Yuan Huang, and Zoë Holmes. *The power and limitations of learning quantum dynamics incoherently*. 2023. arXiv: 2303.12834 [quant-ph].

[152] Seth Lloyd and Christian Weedbrook. "Quantum generative adversarial learning". In: *Physical review letters* 121.4 (2018), p. 040502.

[153] Marcello Benedetti, Delfina Garcia-Pintos, Oscar Perdomo, Vicente Leyton-Ortega, Yunseong Nam, and Alejandro Perdomo-Ortiz. "A generative modeling approach for benchmarking and training shallow quantum circuits". In: *npj Quantum Information* 5.1 (2019), p. 45.

[154] Brian Coyle, Daniel Mills, Vincent Danos, and Elham Kashefi. "The Born supremacy: quantum advantage and training of an Ising Born machine". In: *npj Quantum Information* 6.1 (2020), p. 60.

[155] Xun Gao, Eric R Anschuetz, Sheng-Tao Wang, J Ignacio Cirac, and Mikhail D Lukin. "Enhancing generative models via quantum correlations". In: *Physical Review X* 12.2 (2022), p. 021037.

[156] Manuel S Rudolph, Ntwali Bashige Toussaint, Amara Katabarwa, Sonika Johri, Borja Peropadre, and Alejandro Perdomo-Ortiz. "Generation of high-resolution handwritten digits with an ion-trap quantum computer". In: *Physical Review X* 12.3 (2022), p. 031010.

[157] Elton Yechao Zhu, Sonika Johri, Dave Bacon, Mert Esencan, Jungsang Kim, Mark Muir, Nikhil Murgai, Jason Nguyen, Neal Pisenti, Adam Schouela, et al. "Generative quantum learning of joint probability distribution functions". In: *Physical Review Research* 4.4 (2022), p. 043092.

[158] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. "Efficient quantum state tomography". In: *Nature communications* 1.1 (2010), p. 149.

[159] B. P. Lanyon, C. Maier, M. Holzäpfel, et al. “Efficient tomography of a quantum many-body system”. In: *Nature Physics* 13.12 (Dec. 2017), pp. 1158–1162. DOI: 10.1038/nphys4244.

[160] Valentin Gebhart, Raffaele Santagati, Antonio Andrea Gentile, Erik M Gauger, David Craig, Natalia Ares, Leonardo Banchi, Florian Marquardt, Luca Pezzè, and Cristian Bonato. “Learning quantum systems”. In: *Nature Reviews Physics* 5.3 (2023), pp. 141–156.

[161] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. “Sample-efficient learning of interacting quantum systems”. In: *Nature Physics* 17.8 (Aug. 2021), pp. 931–935. DOI: 10.1038/s41567-021-01232-0.

[162] Cambyse Rouzé and Daniel Stilck França. *Learning quantum many-body systems from a few copies*. 2021. arXiv: 2107.03333 [quant-ph].

[163] Jeongwan Haah, Robin Kothari, and Ewin Tang. “Optimal learning of quantum Hamiltonians from high-temperature Gibbs states”. In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. 2022, pp. 135–146. DOI: 10.1109/FOCS54457.2022.00020.

[164] Ashley Montanaro. *Learning stabilizer states by Bell sampling*. 2017. arXiv: 1707.04012 [quant-ph].

[165] David Gross, Sepehr Nezami, and Michael Walter. “Schur–Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations”. In: *Communications in Mathematical Physics* 385.3 (2021), pp. 1325–1393.

[166] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. “Low-Stabilizer-Complexity Quantum States Are Not Pseudorandom”. In: *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*. Ed. by Yael Tauman Kalai. Vol. 251. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 64:1–64:20. DOI: 10.4230/LIPIcs.ITCS.2023.64.

[167] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. *Improved Stabilizer Estimation via Bell Difference Sampling*. 2023. arXiv: 2304.13915 [quant-ph].

[168] Srinivasan Arunachalam, Sergey Bravyi, Arkopal Dutt, and Theodore J. Yoder. “Optimal Algorithms for Learning Quantum Phase States”. In: *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*. Ed. by Omar Fawzi and Michael Walter. Vol. 266. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 3:1–3:24. DOI: 10.4230/LIPIcs.TQC.2023.3.

[169] Scott Aaronson and Sabee Grewal. "Efficient Tomography of Non-Interacting-Fermion States". In: *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*. Ed. by Omar Fawzi and Michael Walter. Vol. 266. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 12:1–12:18. DOI: 10.4230/LIPIcs.TQC.2023.12.

[170] Ching-Yi Lai and Hao-Chung Cheng. "Learning quantum circuits of some T gates". In: *IEEE Transactions on Information Theory* 68.6 (2022), pp. 3951–3964.

[171] Senrui Chen, Sisi Zhou, Alireza Seif, and Liang Jiang. "Quantum advantages for Pauli channel estimation". In: *Physical Review A* 105.3 (2022), p. 032435.

[172] Ewout van den Berg, Zlatko K. Minev, Abhinav Kandala, and Kristan Temme. "Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors". In: *Nature Physics* 19.8 (Aug. 2023), pp. 1116–1121. DOI: 10.1038/s41567-023-02042-2.

[173] Zhi Li, Liujun Zou, and Timothy H Hsieh. "Hamiltonian tomography via quantum quench". In: *Physical review letters* 124.16 (2020), p. 160502.

[174] Liangyu Che, Chao Wei, Yulei Huang, Dafa Zhao, Shunzhong Xue, Xinfang Nie, Jun Li, Dawei Lu, and Tao Xin. "Learning quantum Hamiltonians from single-qubit measurements". In: *Physical Review Research* 3.2 (2021), p. 023246.

[175] Wenjun Yu, Jinzhao Sun, Zeyao Han, and Xiao Yuan. *Practical and Efficient Hamiltonian Learning*. 2022. arXiv: 2201.00190 [quant-ph].

[176] Dominik Hangleiter, Ingo Roth, Jens Eisert, and Pedram Roushan. *Precise Hamiltonian identification of a superconducting quantum processor*. 2021. arXiv: 2108.08319 [quant-ph].

[177] Daniel Stilck França, Liubov A. Markovich, V. V. Dobrovitski, Albert H. Werner, and Johannes Borregaard. "Efficient and robust estimation of many-qubit Hamiltonians". In: *Nature Communications* 15.1 (Jan. 2024), p. 311. DOI: 10.1038/s41467-023-44012-5.

[178] Assaf Zubida, Elad Yitzhaki, Netanel H. Lindner, and Eyal Bairey. *Optimal short-time measurements for Hamiltonian learning*. 2021. arXiv: 2108.08824 [quant-ph].

[179] Eyal Bairey, Itai Arad, and Netanel H Lindner. "Learning a local Hamiltonian from local measurements". In: *Physical review letters* 122.2 (2019), p. 020504.

[180] Christopher E Granade, Christopher Ferrie, Nathan Wiebe, and David G Cory. "Robust online Hamiltonian learning". In: *New Journal of Physics* 14.10 (2012), p. 103013.

[181] Andi Gu, Lukasz Cincio, and Patrick J. Coles. "Practical Hamiltonian learning with unitary dynamics and Gibbs states". In: *Nature Communications* 15.1 (Jan. 2024), p. 312. DOI: 10.1038/s41467-023-44008-1.

[182] Frederik Wilde, Augustine Kshetrimayum, Ingo Roth, Dominik Hangleiter, Ryan Sweke, and Jens Eisert. *Scalably learning quantum many-body Hamiltonians from dynamical data*. 2022. arXiv: 2209.14328 [quant-ph].

[183] Hsin-Yuan Huang, Yu Tong, Di Fang, and Yuan Su. "Learning many-body Hamiltonians with Heisenberg-limited scaling". In: *Physical Review Letters* 130.20 (2023), p. 200403.

[184] Anurag Anshu and Srinivasan Arunachalam. "A survey on the complexity of learning quantum states". In: *Nature Reviews Physics* 6.1 (Jan. 2024), pp. 59–69. DOI: 10.1038/s42254-023-00662-4.

[185] Barbara M Terhal and David P DiVincenzo. "Classical simulation of noninteracting-fermion quantum circuits". In: *Physical Review A* 65.3 (2002), p. 032325.

[186] Scott Aaronson and Daniel Gottesman. "Improved simulation of stabilizer circuits". In: *Physical Review A* 70.5 (2004), p. 052328.

[187] J Ignacio Cirac, David Perez-Garcia, Norbert Schuch, and Frank Verstraete. "Matrix product states and projected entangled pair states: Concepts, symmetries, theorems". In: *Reviews of Modern Physics* 93.4 (2021), p. 045003.

[188] Dominik S Wild and Álvaro M Alhambra. "Classical simulation of short-time quantum dynamics". In: *PRX Quantum* 4.2 (2023), p. 020340.

[189] Chao Yin and Andrew Lucas. *Polynomial-time classical sampling of high-temperature quantum Gibbs states*. 2023. arXiv: 2305.18514 [quant-ph].

[190] Scott Aaronson. "Shadow Tomography of Quantum States". In: *SIAM Journal on Computing* 49.5 (2020), STOC18-368-STOC18–394. DOI: 10.1137/18M120275X.

[191] Costin Bădescu and Ryan O'Donnell. "Improved Quantum data analysis". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2021. Virtual, Italy: Association for Computing Machinery, 2021, pp. 1398–1411. DOI: 10.1145/3406325.3451109.

[192] Ryan Levy, Di Luo, and Bryan K. Clark. "Classical shadows for quantum process tomography on near-term quantum computers". In: *Phys. Rev. Res.* 6 (1 Jan. 2024), p. 013029. DOI: 10.1103/PhysRevResearch.6.013029.

[193] Hsin-Yuan Huang, Sitan Chen, and John Preskill. "Learning to Predict Arbitrary Quantum Processes". In: *PRX Quantum* 4 (4 Dec. 2023), p. 040337. DOI: 10.1103/PRXQuantum.4.040337.

[194] Jonathan Kunjummen, Minh C Tran, Daniel Carney, and Jacob M Taylor. "Shadow process tomography of quantum channels". In: *Physical Review A* 107.4 (2023), p. 042403.

[195] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. "The randomized measurement toolbox". In: *Nature Reviews Physics* 5.1 (Jan. 2023), pp. 9–24. DOI: 10.1038/s42254-022-00535-2.

[196] B. Schumacher and R. F. Werner. *Reversible quantum cellular automata.* 2004. arXiv: quant-ph/0405174 [quant-ph].

[197] D. Gross, V. Nesme, H. Vogts, and R. F. Werner. "Index Theory of One Dimensional Quantum Walks and Cellular Automata". In: *Communications in Mathematical Physics* 310.2 (Jan. 2012), pp. 419–454. DOI: 10.1007/s00220-012-1423-1.

[198] Jeongwan Haah, Lukasz Fidkowski, and Matthew B. Hastings. "Nontrivial Quantum Cellular Automata in Higher Dimensions". In: *Communications in Mathematical Physics* 398.1 (Nov. 2022), pp. 469–540. DOI: 10.1007/s00220-022-04528-1.

[199] Wilbur Shirley, Yu-An Chen, Arpit Dua, Tyler D. Ellison, Nathanan Tantivasadakarn, and Dominic J. Williamson. "Three-Dimensional Quantum Cellular Automata from Chiral Semion Surface Topological Order and beyond". In: *PRX Quantum* 3 (3 Aug. 2022), p. 030326. DOI: 10.1103/PRXQuantum.3.030326.

[200] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. "Quantum state tomography via compressed sensing". In: *Physical review letters* 105.15 (2010), p. 150401.

[201] Nengkun Yu and Tzu-Chieh Wei. *Learning marginals suffices!* 2023. arXiv: 2303.08938 [quant-ph].

[202] Christof Zalka. "Grover's quantum searching algorithm is optimal". In: *Physical Review A* 60.4 (1999), p. 2746.

[203] Michael A Nielsen. "A simple formula for the average gate fidelity of a quantum dynamical operation". In: *Physics Letters A* 303.4 (2002), pp. 249–252.

[204] Ashley Montanaro and Ronald de Wolf. *A Survey of Quantum Property Testing.* Graduate Surveys 7. Theory of Computing Library, 2016, pp. 1–81. DOI: 10.4086/toc.gs.2016.007.

[205] Runyao Duan, Yuan Feng, and Mingsheng Ying. "Perfect distinguishability of quantum operations". In: *Physical Review Letters* 103.21 (2009), p. 210501.

[206] Ingemar Bengtsson and Karol Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement.* Cambridge university press, 2017.

[207] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. "Query-optimal estimation of unitary channels in diamond distance". In: *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS).* 2023, pp. 363–390. DOI: 10.1109/FOCS57990.2023.00028.

[208] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. "Elementary gates for quantum computation". In: *Phys. Rev. A* 52 (5 Nov. 1995), pp. 3457–3467. DOI: 10.1103/PhysRevA.52.3457.

[209] Vivek V Shende, Stephen S Bullock, and Igor L Markov. "Synthesis of quantum logic circuits". In: *Proceedings of the 2005 Asia and South Pacific Design Automation Conference*. 2005, pp. 272–275.

[210] John Watrous. *The theory of quantum information*. Cambridge university press, 2018.

[211] Thomas Chen, Shivam Nadimpalli, and Henry Yuen. "Testing and learning quantum juntas nearly optimally". In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2023, pp. 1163–1185.

[212] Fernando G. S. L. Brandão and Michael J. Kastoryano. "Finite Correlation Length Implies Efficient Preparation of Quantum Thermal States". In: *Communications in Mathematical Physics* 365.1 (Jan. 2019), pp. 1–16. DOI: 10.1007/s00220-018-3150-8.

[213] Charles H Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. "Strengths and weaknesses of quantum computing". In: *SIAM journal on Computing* 26.5 (1997), pp. 1510–1523.

[214] Xie Chen, Zheng-Cheng Gu, and Xiao-Gang Wen. "Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order". In: *Phys. Rev. B* 82 (15 Oct. 2010), p. 155138. DOI: 10.1103/PhysRevB.82.155138.

[215] Dorit Aharonov and Yonathan Touati. *Quantum Circuit Depth Lower Bounds For Homological Codes*. 2018. arXiv: 1810.03912 [quant-ph].

[216] Lorenzo Piroli, Georgios Styliaris, and J. Ignacio Cirac. "Quantum Circuits Assisted by Local Operations and Classical Communication: Transformations and Phases of Matter". In: *Phys. Rev. Lett.* 127 (22 Nov. 2021), p. 220503. DOI: 10.1103/PhysRevLett.127.220503.

[217] S. Bravyi, M. B. Hastings, and F. Verstraete. "Lieb-Robinson Bounds and the Generation of Correlations and Topological Quantum Order". In: *Phys. Rev. Lett.* 97 (5 July 2006), p. 050401. DOI: 10.1103/PhysRevLett.97.050401.

[218] Elliott H. Lieb and Mary Beth Ruskai. "Proof of the strong subadditivity of quantum-mechanical entropy". In: *Journal of Mathematical Physics* 14.12 (Nov. 2003), pp. 1938–1941. DOI: 10.1063/1.1666274.

[219] Thiago Bergamaschi, Chi-Fang Chen, and Yunchao Liu. *Quantum computational advantage with constant-temperature Gibbs sampling*. FOCS 2024, to appear. 2024. arXiv: 2404.14639 [quant-ph].

[220] Thiago Bergamaschi and Yunchao Liu. *Single-shot logical state preparation for arbitrary quantum LDPC codes*. 2024.

[221] Dolev Bluvstein, Simon J. Evered, Alexandra A. Geim, et al. "Logical quantum processor based on reconfigurable atom arrays". In: *Nature* 626.7997 (Feb. 2024), pp. 58–65. DOI: 10.1038/s41586-023-06927-3.

[222] Dorit Aharonov and Michael Ben-Or. “Fault-Tolerant Quantum Computation with Constant Error Rate”. In: *SIAM Journal on Computing* 38.4 (2008), pp. 1207–1282. DOI: 10.1137/S0097539799359385.

[223] Héctor Bombín. “Single-Shot Fault-Tolerant Quantum Error Correction”. In: *Phys. Rev. X* 5 (3 Sept. 2015), p. 031043. DOI: 10.1103/PhysRevX.5.031043.

[224] E. B. Davies. “Markovian master equations”. In: *Communications in Mathematical Physics* 39.2 (June 1974), pp. 91–110. DOI: 10.1007/BF01608389.

[225] Evgeny Mozgunov and Daniel Lidar. “Completely positive master equation for arbitrary driving and small level spacing”. In: *Quantum* 4 (2020), p. 227.

[226] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. *High-Temperature Gibbs States are Unentangled and Efficiently Preparable*. 2024. arXiv: 2403.16850 [quant-ph].

[227] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. “Proof Verification and the Hardness of Approximation Problems”. In: *J. ACM* 45.3 (1998), pp. 501–555. DOI: 10.1145/278298.278306.

[228] Dorit Aharonov, Wim van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. “Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation”. In: *SIAM Journal on Computing* 37.1 (2007), pp. 166–194. DOI: 10.1137/S0097539705447323.

[229] Chi-Fang Chen, Hsin-Yuan Huang, John Preskill, and Leo Zhou. *Local minima in quantum systems*. 2023. arXiv: 2309.16596 [quant-ph].

[230] Alexei Y. Kitaev, A. H. Shen, and Mikhail N. Vyalyi. *Classical and Quantum Computation*. Vol. 47. Graduate studies in mathematics. American Mathematical Society, 2002.

[231] Patrick Rall, Chunhao Wang, and Pawel Wocjan. “Thermal State Preparation via Rounding Promises”. In: *Quantum* 7 (Oct. 2023), p. 1132. DOI: 10.22331/q-2023-10-10-1132.

[232] Chi-Fang Chen, Michael J. Kastoryano, Fernando G. S. L. Brandão, and András Gilyén. *Quantum Thermal State Preparation*. 2023. arXiv: 2303.18224 [quant-ph].

[233] Chi-Fang Chen, Michael J. Kastoryano, and András Gilyén. *An efficient and exact noncommutative quantum Gibbs sampler*. 2023. arXiv: 2311.09207 [quant-ph].

[234] Michael J. Kastoryano and Kristan Temme. “Quantum logarithmic Sobolev inequalities and rapid mixing”. In: *Journal of Mathematical Physics* 54.5 (May 2013), p. 052202. DOI: 10.1063/1.4804995.

[235] Ivan Bardet, Ángela Capel, Li Gao, Angelo Lucia, David Pérez-García, and Cambyse Rouzé. “Rapid Thermalization of Spin Chain Commuting Hamiltonians”. In: *Phys. Rev. Lett.* 130 (6 Feb. 2023), p. 060401. DOI: 10.1103/PhysRevLett.130.060401.

[236] Ivan Bardet, Ángela Capel, Li Gao, Angelo Lucia, David Pérez-García, and Cambyse Rouzé. "Entropy Decay for Davies Semigroups of a One Dimensional Quantum Lattice". In: *Communications in Mathematical Physics* 405.2 (Feb. 2024), p. 42. DOI: 10.1007/s00220-023-04869-5.

[237] Ángela Capel, Cambyse Rouzé, and Daniel Stilck França. *The modified logarithmic Sobolev inequality for quantum spin systems: classical and commuting nearest neighbour interactions*. 2021. arXiv: 2009.11817 [quant-ph].

[238] Joel Rajakumar, James D. Watson, and Yi-Kai Liu. *Polynomial-Time Classical Simulation of Noisy IQP Circuits with Constant Depth*. 2024. arXiv: 2403.14607 [quant-ph].

[239] Rawad Mezher, Joe Ghalbouni, Joseph Dgheim, and Damian Markham. "Fault-tolerant quantum speedup from constant depth quantum circuits". In: *Phys. Rev. Res.* 2 (3 Sept. 2020), p. 033444. DOI: 10.1103/PhysRevResearch.2.033444.

[240] Louis Paletta, Anthony Leverrier, Alain Sarlette, Mazyar Mirrahimi, and Christophe Vuillot. *Robust sparse IQP sampling in constant depth*. 2023. arXiv: 2307.10729 [quant-ph].

[241] Sergey Bravyi and Alexei Kitaev. "Universal quantum computation with ideal Clifford gates and noisy ancillas". In: *Phys. Rev. A* 71 (2 Feb. 2005), p. 022316. DOI: 10.1103/PhysRevA.71.022316.

[242] Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. "Universal Blind Quantum Computation". In: *2009 50th Annual IEEE Symposium on Foundations of Computer Science* (2008), pp. 517–526.

[243] Pawel Wocjan and Kristan Temme. "Szegedy Walk Unitaries for Quantum Maps". In: *Communications in Mathematical Physics* 402 (2021), pp. 3201–3231.

[244] E Onorati, Oliver Buerschaper, Martin Kliesch, Martin Kliesch, William David Brown, Albert H. Werner, Albert H. Werner, and Jens Eisert. "Mixing Properties of Stochastic Quantum Hamiltonians". In: *Communications in Mathematical Physics* 355 (2016), pp. 905–947.

[245] M. Kliesch, T. Barthel, C. Gogolin, M. Kastoryano, and J. Eisert. "Dissipative Quantum Church-Turing Theorem". In: *Physical Review Letters* 107.12 (Sept. 2011). DOI: 10.1103/physrevlett.107.120501.

[246] Michael J. Bremner, Richard Jozsa, and Dan J. Shepherd. "Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 467 (2010), pp. 459–472.

[247] Dominik Hangleiter, Juan Bermejo-Vega, Martin Schwarz, and Jens Eisert. "Anticoncentration theorems for schemes showing a quantum speedup". In: *Quantum* 2 (May 2018), p. 65. DOI: 10.22331/q-2018-05-22-65.

[248] Leonardo Novo, Juan Bermejo-Vega, and Ra'ul Garc'ia-Patr'on. "Quantum advantage from energy measurements of many-body quantum systems". In: *Quantum* 5 (2019), p. 465.

[249] Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd. "Average-case complexity versus approximate simulation of commuting quantum computations". In: *Physical review letters* 117 8 (2015), p. 080501.

[250] Li Gao and Cambyse Rouzé. "Complete entropic inequalities for quantum Markov chains". In: *Archive for Rational Mechanics and Analysis* 245.1 (2022), pp. 183–238.

[251] Dénes Petz. "On certain properties of the relative entropy of states of operator algebras". In: *Mathematische Zeitschrift* 206 (1991), pp. 351–361.

[252] Scott Aaronson. "Quantum computing, postselection, and probabilistic polynomial-time". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 461 (2004), pp. 3473–3482.

[253] Robert Raussendorf, Sergey Bravyi, and Jim Harrington. "Long-range quantum entanglement in noisy cluster states". In: *Phys. Rev. A* 71 (6 June 2005), p. 062313. DOI: 10.1103/PhysRevA.71.062313.

[254] Daniel Gottesman. "Fault-tolerant quantum computation with constant overhead". In: *Quantum Info. Comput.* 14.15–16 (Nov. 2014), pp. 1338–1372. DOI: 10.26421/QIC14.15-16-5.

[255] Alexey A. Kovalev and Leonid P. Pryadko. "Fault tolerance of quantum low-density parity check codes with sublinear distance scaling". In: *Phys. Rev. A* 87 (2 Feb. 2013), p. 020304. DOI: 10.1103/PhysRevA.87.020304.