Anti-Concentration for the Unitary Haar Measure and Applications to Random Quantum Circuits

Bill Fefferman
Department of Computer Science, The University of Chicago, IL, USA

Soumik Ghosh
Department of Computer Science, The University of Chicago, IL, USA

Wei Zhan
Department of Computer Science, Purdue University, West Lafayette, IN, USA

Abstract

We prove a Carbery-Wright style anti-concentration inequality for the unitary Haar measure, by showing that the probability of a polynomial in the entries of a random unitary falling into an ε range is at most a polynomial in ε. Using it, we show that the scrambling speed of a random quantum circuit is lower bounded: Namely, every input qubit has an influence that is at least inverse exponential in depth, on any output qubit touched by its lightcone. Our result on scrambling speed works with high probability over the choice of a circuit from an ensemble, as opposed to just working in expectation.

As an application, we give the first polynomial-time algorithm for learning log-depth random quantum circuits with Haar random gates up to polynomially small diamond distance, given oracle access to the circuit. Other applications of this new scrambling speed lower bound include:

- An optimal $\Omega(\log \varepsilon^{-1})$ depth lower bound for ε-approximate unitary designs on any circuit architecture;
- A polynomial-time quantum algorithm that computes the depth of a bounded-depth circuit, given oracle access to the circuit.

Our learning and depth-testing algorithms apply to architectures defined over any geometric dimension, and can be generalized to a wide class of architectures with good lightcone properties.

2012 ACM Subject Classification Theory of computation → Quantum computation theory

Keywords and phrases Haar measure, anti-concentration, random quanytum circuit, learning

Digital Object Identifier 10.4230/LIPIcs.ITCS.2026.57

Related Version *Full Version*: `https://arxiv.org/abs/2407.19561`

Funding B.F., S.G., and W.Z. acknowledge support from AFOSR (FA9550-21-1-0008). This material is based upon work partially supported by the National Science Foundation under Grant CCF-2044923 (CAREER), by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers (Q-NEXT) and by the DOE QuantISED grant DE-SC0020360.

Acknowledgements The authors thank Anurag Anshu, Adam Bouland, Lijie Chen, Jonas Haferkamp, Robert Huang, Issac Kim, Yunchao Liu, Tony Metger, Marcus Michelen, Tommy Schuster and Kewen Wu for helpful comments and discussions, and thank Jacob Watkins for pointing out a mistake in our previous version of Section 5.3.2.

1 Introduction

Random quantum circuits are one of the most popular paradigms of quantum computation in the near-term era. They are well-studied, both theoretically and experimentally, in the context of quantum advantage demonstrations: e.g., see [6, 8, 3, 7, 43, 44, 21]. They also have numerous applications in areas like benchmarking, like in [19, 37], in cryptography, as in e.g., [4, 2, 50], and in the modeling of physical objects like black holes, as in e.g., [30, 49, 52].

One reason random quantum circuits are extensively studied is because they are rapid "scramblers" of information. Intuitively, this means that the output state it generates has non-trivial correlations across spatially separated qubits. The rate of scrambling depends on the depth – the deeper the circuit is, the better it is at scrambling. However, even though previous works on scrambling have put *upper bounds* on the scrambling speed of random circuits – see, e.g., [28, 11, 46, 27, 33, 13, 14, 23, 42, 36, 16, 51, 17, 53] – *lower bounds* on the scrambling speed are not well studied.

In this work, we give a *lower bound* on the speed with which random quantum circuits can scramble information. Our main theorem is as follows:

▶ **Theorem 1.** *Let $\mathcal{C}$ be a random quantum circuit with a fixed architecture, where each gate is a k-qubit independent Haar random unitary. Let ρ and π be a pair of input and output qubits connected by D layer of gates. Arbitrarily fix the inputs to $\mathcal{C}$, except the qubit ρ, and let $\Phi_{\mathcal{C}}$ be the channel that maps ρ to π.*

Then for every $\gamma > 0$, with probability at least $1-\gamma$ over $\mathcal{C}$ the following holds: For every two single-qubit states ρ and ρ',

$$\left\|\Phi_{\mathcal{C}}(\rho) - \Phi_{\mathcal{C}}(\rho')\right\|_{\mathrm{F}} \geq \left\|\rho - \rho'\right\|_{\mathrm{F}} \cdot (2^{-D}\gamma)^{c_k} \tag{1}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, and $c_k > 0$ is a constant that depends only on k.

In particular, our main result gives a lower bound on the influence that changing an input qubit has on a designated output qubit inside the lightcone of that input qubit and shows that it decays at most exponentially fast with depth. We articulate some points that make Theorem 1 useful for our applications:

- The bound in Theorem 1 is uniform, in the sense that ratio $(2^{-D}\gamma)^{c_k}$ does not depend on how the input qubits are chosen, and is applicable to any circuit architecture. Moreover, the probability quantifier over the circuit $\mathcal{C}$ is stated before the choices of ρ and ρ', which is specifically important for our application on learning random quantum circuits (Theorem 5).
- Although we stated the theorem using Frobenius norm, the norm is only over single-qubit states and thus is equivalent to any other matrix norm.
- To obtain the strongest bound, we can choose D to be the shortest path from ρ to π. When π is outside the lightcone of ρ, we think of $D = \infty$, in which case the statement still holds.
- Our bound in Theorem 1 is tight, in the sense that one could prove a matching upper bound

  $$\left\|\Phi_{\mathcal{C}}(\rho) - \Phi_{\mathcal{C}}(\rho')\right\|_{\mathrm{F}} \leq \left\|\rho - \rho'\right\|_{\mathrm{F}} \cdot (2^{-D}\gamma)^{c'_k} \tag{2}$$

  for various architectures with a different constant $c'_k < c_k$, by calculating the moment $\mathbf{E}\left[\left\|\Phi_{\mathcal{C}}(\rho) - \Phi_{\mathcal{C}}(\rho')\right\|_{\mathrm{F}}^2\right]$. Such calculation can be done by analyzing the Markov chain of Pauli operators, using the results in [28, 11, 45]. However, the moment method could not yield our lower bound, which we will explain in details below.

Implications on proving typicality statements. Observe that Theorem 1 is a statement that works *with high probability* over the choice of circuit from the ensemble $\mathcal{C}$, as opposed to just *in expectation* over the ensemble $\mathcal{C}$. To the best of our knowledge, this is the first typicality result regarding the scrambling behavior of random quantum circuits.

Most of the previous work studying random quantum circuits examines their properties by computing some specific quantities of interest in expectation, such as entanglement entropy, OTOC and collision chance, by relating the quantities with the lower moments (usually the second moment) of the quantum circuit. For non-negative quantities the expectation already provides an upper bound that holds with high-probability (like in (2)) by Markov's inequality. But for proving a strong typicality result as in (1), one also needs a corresponding lower bound, which is commonly proved by calculating the variance of the desired quantity X. In particular, we would want that for some appropriately small ϵ, and for some $t \geq 1$,

$$\mathbf{E}[X^{2t}] \leq (1+\epsilon)\,\mathbf{E}[X^t]^2. \tag{3}$$

However, for most architectures a relation like (3) is largely unknown. One reason is that when X is already a second moment of the quantum circuit, and proving (3) requires computing the fourth or higher moments, which is extremely difficult even for standard brickwork circuits [9]. More importantly, in our case, X can be heuristically approximated by the product $\lambda_1 \cdots \lambda_D$ of i.i.d random variables $\lambda_1, \ldots, \lambda_D$ that follow some distribution Λ with constant variance; because of this, we cannot expect (3) to hold as $\mathbf{E}[X^{2t}]$ will be exponentially larger than $\mathbf{E}[X^t]^2$.

In this work, we overcome this barrier by performing a more fine-grained analysis. We do not analyze the randomness present in the choice of gates all at once; Instead, we break up the randomness of the circuit into D layers and lower bound the random variable X with a product of D random variables that are highly correlated. We then decouple these random variables using a new anti-concentration inequality on the unitary Haar measure, which will be stated in Theorem 6. This allows us to avoid calculating higher moments of the circuit, and still retrieve information about the distribution of X that could be lost in the moments. Our result in Theorem 1 directly implies typicality guarantees on other quantities, such as the purity of a single output qubit, and we believe the method we use can be generalized to prove typicality results for many other useful quantities.

1.1 Applications

Intuitively, Theorem 1 means that the output state of random quantum circuits with logarithmic depth carries a signal pertaining to its input state that can be extracted from the output state using single qubit quantum state tomography. This allows us to show many applications.

1.1.1 Optimal 2-design lower bounds

First, as an immediate corollary of Theorem 1, we show a depth lower bound for random quantum circuits, defined on any architecture, forming approximate unitary designs; see Section 5.1.

▶ **Theorem 2.** *Let $\mathcal{C}$ be a random quantum circuit, defined on an arbitrary fixed architecture of minimum depth D, where each gate is a k-qubit independent Haar random unitary. If $\mathcal{C}$ is an ε-approximate 2-design, then*

$$D = \Omega_k(\log \varepsilon^{-1}).$$

Here the minimum depth means the smallest number of gates that any path from input to output goes through, resulting in a stronger requirement than the previous depth lower bounds that uses the lightcone size property. Theorem 2 shows that previous works on

approximate t-design constructions, e.g. [28, 32, 24, 15, 50], are optimal in ϵ, assuming the gates are Haar random unitaries. And combined with the $\Omega(\log n)$ depth lower bound in [18, 50], it means that the approximate t-design construction of depth $O(\log(n/\varepsilon) \cdot t \operatorname{polylog} t)$ from [50] is optimal in both n and ε.

1.1.2 Testing for depth

Our second application concerns testing the depth of a random circuit when the depth is promised to be at most logarithmic.

▶ **Theorem 3.** *Let $\mathcal{C}$ be a brickwork random quantum circuit on n qubits of an unknown depth D promised to be bounded as $D = O(\log n)$, where each gate is independently Haar random. Given oracle access to $\mathcal{C}$, there is a polynomial time algorithm that outputs D with probability $1 - 1/\mathrm{poly}(n)$.*

Although Theorem 3 is stated with brickwork circuits for simplicity, it is applicable to many other architectures satisfying the following: For a random circuit of depth D, the lightcone size is strictly smaller than that of a random circuit of depth larger than D. Hence, there will be a particular output qubit which will be insensitive to a change in a particular input qubit in the former case and will be sensitive in the latter case. For $D = O(\log n)$ this difference is inverse polynomially large and can be tested by state tomography. The explicit algorithm is given in Section 5.2.

Note that our algorithm outputs the exact depth instead of obtaining an approximation, in contrast with the recently proposed depth test algorithm in [26]. This is partially due to the fact that our algorithm works with high probability with respect to the choice of a circuit from the ensemble, as opposed to working only in expectation, unlike [26].

1.1.3 Learning brickwork random circuits

Third, we show that Theorem 1 allows us to learn brickwork random circuits of logarithmic depth. We start by showing that the first layer of gates can be learned given oracle access to the circuit.

▶ **Theorem 4.** *Let $\mathcal{C}$ be a brickwork random quantum circuit on n qubits of known depth $D = O(\log n)$, where each gate is independently Haar random. Given oracle access to $\mathcal{C}$, there is a polynomial time algorithm that with high probability outputs each gate in the first layer with polynomially small error.*

Furthermore, in real life scenarios we can assume the distribution over the gates is a discrete approximation of the Haar measure (see Definition 28), and in this case we can actually learn the entire circuit:

▶ **Theorem 5.** *Let $\mathcal{C}$ be a brickwork random quantum circuit on n qubits of known depth $D = O(\log n)$, where each gate is independently drawn from a discretized version of the Haar measure. Given oracle access to $\mathcal{C}$, there is a polynomial time algorithm that with high probability outputs $\mathcal{C}$ with polynomially small error.*

We will prove Theorem 4 and Theorem 5 in Section 5.3. Note that the learning algorithm in Theorem 5 is proper, that the learnt circuit has the exact same depth and architecture as the actual circuit $\mathcal{C}$.

Before this work, the state-of-art learning algorithm for brickwork quantum circuits is due to [31, 35], which runs in polynomial time for circuits of k-dimensional geometry up to $O(\log^{1/(k+1)} n)$ depth, and is not proper (that the outputted circuit has depth larger than the input circuit by at least a constant factor). Our algorithm works for all geometric

dimensions, works up to $O(\log n)$ depth, and still has polynomial runtime. There is evidence that the depth dependence of our learning algorithm is optimal, because random circuits form high enough designs beyond $\log n$ depth [50], and possibly also pseudorandom unitaries, which hints at learning hardness.

Here we provide an informal description of the learner:

- For each gate in the first layer, the learner iteratively searches through the gate set for the correct gate. There's only polynomially many gates in the gate set because we chose an appropriate ε-net over 2-qubit Haar random unitaries. During each round, the learner applies the inverse of a gate chosen from the gate set at the particular location it wishes to learn.
- Select an input qubit to the gate we are trying to learn. The key observation is that, depending on whether we applied the correct inverse, the output lightcone will consist of different qubits. In particular, one qubit will lie outside the lightcone and be unaffected by a change in the selected input qubit if we hit the correct inverse, as opposed to if we didn't. As a corollary of Theorem 1, in the former case, we can detect the influence of the input qubit by performing single qubit state tomography.

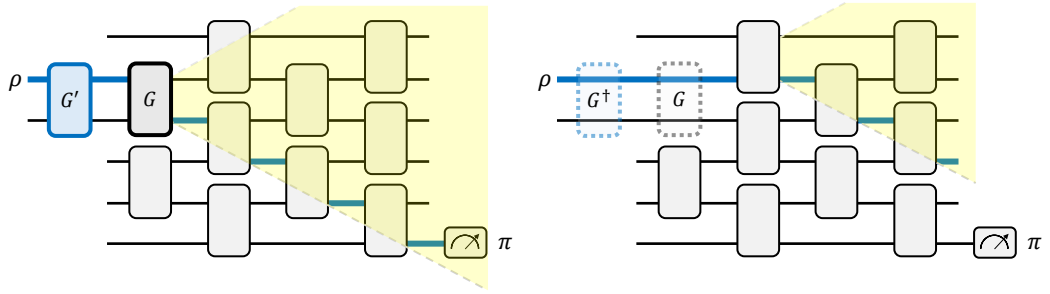

Figure 1 An illustration of the learning algorithm in Theorem 5, where we try to uncompute the gate G by applying some two-qubit unitary G' before it. Left: When G' failed to uncompute G, the lightcone of the input qubit ρ touches an output qubit π at distance D. Right: When $G' = G^\dagger$, the output qubit π falls out of the lightcone and will not be affected by the change in ρ.

Note that the learner of [31, 35] uses a very different strategy: it looks at each output qubit, and then employs a brute force search over all possible lightcones of that circuit to find the correct one. Thereafter, it stitches together all the different local lightcones of every qubit, and then corrects for the errors in the overlapping regions of the different lightcones. The fact that we do not do a brute force search over all possible lightcones makes our algorithm efficient for any dimension, and not just 1-dimensional brickwork circuits.

1.2 Main Technical Tool: Anti-Concentration for Haar Measure

The intuition behind the proof of Theorem 1 is the following: We consider the path of $D+1$ qubits $\rho = \rho_0, \rho_1, \ldots, \rho_D = \Phi_{\mathcal{C}}(\rho)$ in the circuit $\mathcal{C}$, where gate G_i has ρ_i as an input and ρ_{i+1} as an output. Let ρ_i' be the corresponding qubits when the input is ρ', and we want to bound the ratios $\lambda_i = \|\rho_i - \rho_i'\|_{\mathrm{F}} / \|\rho_{i-1} - \rho_{i-1}'\|_{\mathrm{F}}$ and hence their product.

It turns out that we can prove the lower bound $\lambda_i \geq |F(G_i)|$, where F is a polynomial function over the entries of G_i in its matrix representation. Therefore, we only need to show that $|F(G_i)|$ is often not too small, when G_i is a Haar random unitary. In other words, we need to show that the polynomial does not concentrate around zero. Our main technical contribution is the following theorem which proves this anti-concentration phenomenon:

▶ **Theorem 6.** *Let U be a Haar random $n \times n$ unitary matrix, and let $F : \mathbb{C}^{2n^2} \to \mathbb{C}$ be a degree-d polynomial on the entries of U and $U^\dagger$. Then for every $\varepsilon > 0$, it holds that*

$$\Pr\Big[\big|F(U,U^\dagger)\big|^2 \le \varepsilon\, \mathbf{E}\big[|F(U,U^\dagger)|^2\big]\Big] \le C'(n,d)\cdot \varepsilon^{C(n,d)}$$

where $C(n,d) > 0$ and $C'(n,d) > 0$ are constants that depend only on n and d.

The anti-concentration inequality of polynomials over Gaussian random variables was famously proved by Carbery and Wright [12], and their result actually applies to any log-concave distribution over $\mathbb{R}^n$. However, as the Haar measure does not even have a convex support, the proof techniques in [12] does not apply. The alternative inductive proof in [38] also fails to apply, as we are dealing with correlated input variables. We present a very different inductive proof in Section 3.

Note that if we consider $F(U,U^\dagger)$ as a complex random variable, and define the complex variance

$$\mathbf{Var}[F] = \mathbf{E}[|F|^2] - |\mathbf{E}[F]|^2 = \min_{z\in\mathbb{C}} \mathbf{E}[|F-z|^2],$$

then we obtain the form closer to the classical anti-concentration inequalities:

▶ **Corollary 7.** *Let U be a Haar random $n \times n$ unitary matrix, and let F be a degree-d polynomial on the entries of U and $U^\dagger$. Then for every $\varepsilon > 0$ and every $z \in \mathbb{C}$, it holds that*

$$\Pr\Big[|F-z|^2 \le \varepsilon \mathbf{Var}[F]\Big] \le C'(n,d)\cdot \varepsilon^{C(n,d)}.$$

However in this work we will not use the form in Corollary 7, as Theorem 6 suffices for our applications.

▶ Remark 8. Unlike the Carbery-Wright inequality which is dimension-free, meaning the right hand side is $C'(d)\cdot\varepsilon^{C(d)}$ and does not depend on n, we have $C(n,d) = (4n^2d)^{-1}$ and $C'(n,d) = O(n^3d)$ in our proof. In fact, using the concentration bounds it is not hard to show that $C'(n,d)$ must depends polynomially on n. However, we conjecture that $C(n,d)$ could be independent of n, in which case the result would be applicable to random quantum circuits with gates of higher locality.

1.3 Related Works

Concentration phenomenon on unitary Haar measures has been extensively studied, and the readers can refer to [39] for a comprehensive review of the results. In comparison, much less has been shown for the reverse direction, namely the anti-concentration inequalities.

One common way to prove such inequalities is by calculating higher moments and apply the Paley-Zygmund inequality, which was indeed used for showing the anti-concentration property of *output distributions* of Haar random unitaries and random quantum circuits [1, 25, 18]. However, the inequality proved this way is not strong enough for our applications, while calculating moments of a random quantum circuit is also non-trivial and depends highly on the architecture [22, 9]. Instead, we resort to prove a general anti-concentration inequality for polynomials, whose theory has been well developed for Gaussian distributions [12] and product distributions (namely the Littlewood-Offord theory) and has found numerous applications in computational complexity theory [41, 40, 34]. Our inductive proof of Theorem 6 also shares a similar spirit with the elementary proof of Carbery-Wright inequality in [38].

Multiple notions of scrambling property of random quantum circuits has been previously studied. In particular, [11] showed that in a random circuit consists of $O(n \log^2 n)$ sequential applications of Haar random gates on a complete graph of n qubits, every subset of cn qubits is polynomially close to maximally mixed with high probability for some constant $c > 0$. Our Theorem 1 can be viewed as a result in the reverse direction which bounds the scrambling speed of such random circuits. Specifically, at least $\Omega(n \log n \log\log n)$ sequential gates are required, as otherwise with high probability a pair of input and output qubits are $o(\log n)$ depth apart due to a generalization of the coupon-collector problem [48]. It is also reasonable to believe that our method of proving Theorem 1, via the anti-concentration inequality, is applicable to obtain lower bounds for other measures of scrambling such as entanglement and out-of-time-ordered correlation (OTOC) [45, 46, 5, 27]. Upper bounds in the above-mentioned works are obtained by calculating moments and analyzing the averaged Markov chain on Pauli operators, which are not sufficient to prove lower bounds in the typical case.

We also review some previous works related to our applications and clarify the connections. For approximate unitary designs, many previous constructions, for example [28, 10, 24, 29, 15], employed Haar random unitary gates and achieved the optimal $O(\log \varepsilon^{-1})$ dependence on ε. The construction of depth $O(\log(n/\varepsilon) \cdot t \operatorname{polylog} t)$ in the recent work of [50] also falls into this category. Meanwhile, they also proposed a construction of approximate 3-design with only $O(\log\log(n/\varepsilon))$ depth. This does not contradict our lower bound Theorem 2 as the construction uses random Clifford unitaries, which are not independent between layers and also fail the anti-concentration property in Theorem 6. It is intriguing, however, to see if our argument can be extended to show a matching $\Omega(\log\log \varepsilon^{-1})$ depth lower bound.

The depth test algorithm in [26] is based on the entanglement dynamics of random circuits and implemented with their Bell sampling framework. For brickwork circuits, as there is a constant gap between the upper and lower bounds for the entanglement entropy in the typical case, their algorithm gives constant approximation of the depth. For general architectures, the algorithm in [26] also requires knowledge of the entanglement velocity, while our algorithm for Theorem 3 only relies on a specific property of the architecture that the lightcone strictly expands with depth.

The learning algorithm for shallow quantum circuits in [31] is based on the idea of brute-force enumerating all possibilities in a light cone, and stitching the parts together. Therefore, their algorithm works any quantum circuit in worse case, and has complexity exponential in the lightcone size. This means that in order to have polynomial efficiency, the depth has to be $O(\log^{1/k} n)$ for k-dimensional geometrically local circuits and $O(\log\log n)$ for general architectures. However, the learning algorithm is improper and the outputted circuit has a large polynomial blow-up in depth. The blow-up factor could be reduced to constant, at the cost of lowering the depth bound to $O(\log^{1/(k+1)} n)$, and thus could not perform in polynomial time for log-depth circuits even for 1-dimensional geometry. In comparison, our algorithm works for random quantum circuits, on the fixed brickwork or similar geometrically local architecture, where neighboring qubits can be distinguished by their lightcones. The strength of our algorithm Theorem 5 is that it works up to logarithmic depth in polynomial time regardless of the geometric dimension, and it learns properly in that it outputs a circuit with the exact same architecture. This makes our result incomparable with [31].

We also mention that, there is a different learning task where instead given oracle access to the circuit $\mathcal{C}$, the learner is only given copies of the state $\mathcal{C}|0^n\rangle$, and needs to learn a circuit that prepares the same state with error in trace distance. Our algorithm relies on having different inputs and hence does not work in this case, whereas [31] gave a quasi-polynomial efficiency algorithm for 2D and the algorithm was extended to higher dimensions in [35].

2 Preliminaries

We start with some basic notations. We use $\mathbb{U}(d)$ to denote the unitary group of dimension d, and use $\mathcal{U}(d)$ to denote the Haar measure over $\mathbb{U}(d)$. We use Greek letters such as ρ, π, τ to denote density matrices of quantum states. We use $\|\cdot\|_1$ and $\|\cdot\|_{\mathrm{F}}$ for trace norm and Frobenius norm, and $d_\diamond(\cdot,\cdot)$ for diamond distance between unitary channels, defined as

$$d_\diamond(\Phi_1, \Phi_2) = \max_\rho \|(\Phi_1 \otimes \mathbb{I})\rho - (\Phi_2 \otimes \mathbb{I})\rho\|_1,$$

with maximum taken over all density matrices ρ.

A circuit architecture determines the positions of gates in the circuit. We define the depth of an architecture as follows.

▶ **Definition 9.** *In a circuit architecture, a path in space-time between two qubits ρ and ρ' is a sequence of qubits $\rho = \rho_0, \rho_1, \ldots, \rho_D = \rho'$ where for each i, there is a gate in the circuit that has ρ_i as an input and ρ_{i+1} as an output.*

We say the architecture has minimum depth D, if there exists a pair of input and output qubits connected by a path of length D, and every other path from input to output has depth at least D.

A specific architecture of interest is the (1-dimensional) brickwork architecture:

▶ **Definition 10.** *A brickwork quantum circuit on n qubits of depth D consists of D layers of gates, where on layer j, there is a two-qubit gate $G_{i,j} = G_{i+1,j}$ acting on the i-th and $(i+1)$-th qubit if and only if i and j have the same parity.*

The brickwork architecture could be generalized to higher dimensional geometry, and our results still hold for any constant dimension. However, for simplicity we stick with the 1-dimensional architecture in this paper.

We will need the following statements about quantum state tomography on single qubits for our algorithms (see e.g. [47])

▶ **Proposition 11.** *Given access to copies of a single-qubit state ρ, one can output an estimation $\tilde{\rho}$ with $\|\rho - \tilde{\rho}\|_{\mathrm{F}} \le \varepsilon$ in $\mathrm{poly}(1/\varepsilon)$ time.*

The following simple lemma is particularly useful, which bounds the difference between states through a channel:

▶ **Lemma 12.** *Let Φ be a quantum channel that takes k qubits as the input. For every input states ρ and ρ', we have*

$$\left\|\Phi(\rho) - \Phi(\rho')\right\|_{\mathrm{F}} \le 2^{k/2}\left\|\rho - \rho'\right\|_{\mathrm{F}}.$$

As we are frequently dealing with differences between quantum states, here we present some facts about the space of such differences. We start from the Bloch sphere presentation of a single-qubit state:

$$\rho = \frac{1}{2}(I + r_x X + r_y Y + r_z Z), \qquad r_x, r_y, r_z \in \mathbb{R}, \quad r_x^2 + r_y^2 + r_z^2 \le 1 \tag{4}$$

where

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{5}$$

Then the difference $\rho - \rho'$ between two single-qubit states can be written in the Pauli basis

$$\rho - \rho' = r_x X + r_y Y + r_z Z, \quad r_x^2 + r_y^2 + r_z^2 = \frac{1}{2}\|\rho - \rho'\|_{\mathrm{F}}^2 \leq 1. \tag{6}$$

Therefore, if we view $\rho - \rho'$ under the coordinate system $\sqrt{2} \cdot (r_x, r_y, r_z)$, then the set Δ_1 of all possible single-qubit differences can be identified with a ball of radius $\sqrt{2}$ in $\mathbb{R}^3$. The Euclidean space $\mathbb{R}^3$ is equipped with the standard trace inner product $\langle A, B\rangle = \mathrm{Tr}[A^\dagger B]$, so that the norm coincides with the Frobenius norm.

More generally, let Δ_k be the set of all possible differences between two k-qubit states ρ and ρ'. The difference $\rho - \rho'$ can be written as a real linear combination over the Pauli basis

$$\{I, X, Y, Z\}^{\otimes k} \setminus \{I^{\otimes k}\}$$

where the identity is removed as $\mathrm{Tr}[\rho - \rho'] = 0$. Since the Pauli basis are orthonormal, we can think of Δ_k as a subset of the Euclidean space $\mathbb{R}^{4^k-1}$ (although the set is much more complicated than a ball for $k > 1$). The Euclidean space is also equipped with the standard trace inner product and the Frobenius norm. As a result, a quantum channel Φ with k-qubit input and n-qubit output induces a linear map from Δ_k to Δ_n:

$$\rho - \rho' \mapsto \Phi(\rho) - \Phi(\rho'),$$

which is also a real linear map from $\mathbb{R}^{4^k-1}$ to $\mathbb{R}^{4^n-1}$.

3 Anti-concentration Bound

In this section we prove Theorem 6. For simplicity, we introduce the notion of *semi-polynomials*: A function is a degree-d semi-polynomial in complex variables $z_1, \ldots, z_n$, if it is a polynomial in $z_1, \ldots, z_n, \overline{z_1}, \ldots, \overline{z_n}$ of degree at most d. Now Theorem 6 is implied by the following more general form:

▶ **Theorem 13.** *Let $m \leq n$, and $F : \mathbb{C}^{nm} \to \mathbb{C}$ be a degree-d semi-polynomial. Suppose that F takes as inputs the entries of the first m columns of an $n \times n$ unitary matrix, and that the value of F is always a non-negative real number over this domain. Then for $U \sim \mathcal{U}(n)$,*

$$\Pr\left[F(U_{1,1}, \ldots, U_{n,m}) \leq \varepsilon\, \mathbf{E}[F]\right] \leq C'(n, m, d) \cdot \varepsilon^{C(n,m,d)}$$

holds for every $\varepsilon > 0$, where $C(n, m, d) > 0$ and $C'(n, m, d) > 0$ are constants that depend only on n, m and d.

To see that Theorem 13 implies Theorem 6, it suffices to take $m = n$ and notice that $|F|^2 = F\overline{F}$ is a degree-$2d$ semi-polynomial that is always non-negative. We prove Theorem 13 via induction, and the proof is divided into four stages.

3.1 $m = 1, n = 1$

We start with the simplest case when $m = n = 1$. In this case $F : \mathbb{C} \to \mathbb{C}$ is a single-variable degree-d semi-polynomial over the unit circle $\{z : |z| = 1\}$. Since $\overline{z} = 1/z$ over this domain, assuming $F \neq 0$ we can write F as

$$F(z) = G(z)/z^d$$

where $G(z) = \alpha(z - z_1) \cdots (z - z_{2d})$ is a degree $2d$ polynomial in z. Without loss of generality we can assume that $|\alpha| = 1$, and therefore

$$F(z) = |F(z)| = |G(z)| = |z - z_1| \cdots |z - z_{2d}|. \tag{7}$$

Then we can bound the expectation of F as

$$\mathbf{E}[F] \leq \sup_{|z|=1} \prod |z - z_i| \leq \prod (1 + |z_i|). \tag{8}$$

On the other hand, if $|z - z_i| > \delta \geq 0$ for some $|z| \leq 1$, it is easy to show that

$$\frac{|z - z_i|}{1 + |z_i|} > \frac{\delta}{\delta + 2}. \tag{9}$$

Thus if $|z - z_i| > \delta$ holds for all $i = 1, \ldots, 2d$, then

$$F(z) = \prod |z - z_i| > \left(\frac{\delta}{\delta + 2}\right)^{2d} \prod (1 + |z_i|) \geq \left(\frac{\delta}{\delta + 2}\right)^{2d} \mathbf{E}[F]. \tag{10}$$

That means, if we take $\varepsilon = \left(\frac{\delta}{\delta+2}\right)^{2d}$, then $F(z) \leq \varepsilon \, \mathbf{E}[F]$ only happens when z falls into one of the δ-balls around some z_i. Each δ-ball intersect with the unit circle as an arc of angle at most 4δ, and thus has measure at most $2\delta/\pi$ under the Haar measure over the unit circle. Therefore we conclude that

$$\begin{aligned} \Pr_{|z|=1}[F(z) \leq \varepsilon \, \mathbf{E}[F]] &\leq \min\{4d\delta/\pi, 1\} \\ &= \min\left\{\frac{8d}{\pi} \cdot \frac{\varepsilon^{1/(2d)}}{1 - \varepsilon^{1/(2d)}}, 1\right\} \leq \left(\frac{8d}{\pi} + 1\right)\varepsilon^{1/(2d)}, \end{aligned} \tag{11}$$

and we can take $C(1, 1, d) = 1/(2d)$ and $C'(1, 1, d) = 8d/\pi + 1$.

3.2 $m = 1, n = 1$, Alternative Distribution

For the sake of later use, we also need a version where $z = u_1$ follows the distribution of the first coordinate of a Haar-random unit vector $(u_1, \ldots, u_n) \in \mathbb{C}^n$, $n \geq 2$. In this case $\overline{z} = 1/z$ no longer holds, and we need an alternative method.

For every $r \in \mathbb{R}$, $0 \leq r \leq 1$ we define

$$P(r) = \mathop{\mathbf{E}}_{|u_1|=r}[F(u_1)]$$

where the expectation is over a Haar-random $z \in \mathbb{C}$ with $|z| = r$. Notice that a monomial $u_1^k \overline{u_1}^\ell$ in F has expectation 0 unless $k = \ell$, and when $k = \ell$ we have $u_1^k \overline{u_1}^\ell = r^{2k}$. That means $P(r)$ is a degree-d polynomial in r^2.

Notice that r^2 follows the Beta distribution $\mathrm{Beta}(1, n - 1)$, with the density function

$$f(r; 1, n - 1) = (n - 1)(1 - r)^{n-2},$$

and $\mathbf{E}[P(r)]$ under this distribution coincides with $\mathbf{E}[F]$.

With the analysis in the previous section which also works on $P(r)$, we can show that if we take $\varepsilon = \left(\frac{\delta}{\delta+2}\right)^{d}$, then $P(r) \leq \varepsilon \, \mathbf{E}[P(r)] = \varepsilon \, \mathbf{E}[F]$ only happens when r^2 falls into one of the δ-balls around d complex roots of P, which are intervals of length at most 2δ on the real line. Since the density function of r^2 has a maximum of $n - 1$, we have

$$\Pr[P(r) \leq \varepsilon \, \mathbf{E}[F]] \leq \min\{2d(n - 1)\delta, 1\} \leq 4dn\varepsilon^{1/d}. \tag{12}$$

On the other hand, applying Equation (11) from the previous section on $F(rz)$ for every fixed r directly provides

$$\Pr_{|z|=r}[F(z) \le \varepsilon P(r)] \le \left(\frac{8d}{\pi}+1\right)\varepsilon^{1/(2d)}. \tag{13}$$

Thus by a union bound we have

$$\begin{aligned}\Pr[F(u_1) \le \varepsilon \,\mathbf{E}[F]] &\le \Pr\left[P(r) \le \sqrt{\varepsilon}\,\mathbf{E}[F]\right] + \Pr\left[F(z) \le \sqrt{\varepsilon}P(|z|)\right] \\ &\le 4dn\varepsilon^{1/(2d)} + \left(\frac{8d}{\pi}+1\right)\varepsilon^{1/(4d)} \\ &\le 4d(n+1)\varepsilon^{1/(4d)}. \end{aligned} \tag{14}$$

We note that not only this result will be used in the next stage, the technique itself will also be reapplied multiple times in the later proofs.

3.3 $m = 1, n > 1$

In this stage we consider $m = 1$ with general n, and thus the inputs to F is a Haar-random unit vector $(u_1, \ldots, u_n) \in \mathbb{C}^n$. The strategy is to use induction on n, and show that with high probability over the choice of u_1,

$$P(u_1) = \mathop{\mathbf{E}}_{u_2,\ldots,u_n}[F(u_1, \ldots, u_n)]$$

is not too small conditioned on the fixed u_1. To handle this, we need the following lemma.

▶ **Lemma 14.** *If $F : \mathbb{C}^n \to \mathbb{C}$ is a degree-d semi-polynomial, and $(u_1, \ldots, u_n) \in \mathbb{C}^n$ is a Haar-random unit vector, then $P(u_1) = \mathbf{E}_{u_2,\ldots,u_n}[F(u_1, \ldots, u_n)]$ is a degree-d semi-polynomial on u_1.*

Proof. Omitted in this version. ◀

Since P is an expectation over F, it is also always non-negative and has the same expectation as $\mathbf{E}[F]$. Thus Equation (14) from the previous section gives

$$\Pr[P(u_1) \le \varepsilon \,\mathbf{E}[F]] \le 4d(n+1)\varepsilon^{1/(4d)}. \tag{15}$$

Now we fix some u_1, and consider the degree-d semi-polynomial

$$F_{u_1}(u_2, \ldots, u_n) = F(u_1, u_2, \ldots, u_n)$$

which is always non-negative. Since $(u_2, \ldots, u_n) = ru'$ for some $r \in \mathbb{R}$, and u' follows the Haar measure over the unit sphere in $\mathbb{C}^{n-1}$, we can apply the induction hypothesis for $n-1$ on $F_{u_1}(ru')$ to get

$$\Pr\left[F_{u_1}(ru') \le \varepsilon\, \mathbf{E}_{u'}[F_{u_1}(ru')]\right] \le C'(n-1,1,d)\cdot\varepsilon^{C(n-1,1,d)}. \tag{16}$$

Therefore we conclude that, for every $p \in (0,1)$,

$$\begin{aligned} &\Pr[F(u_1, \ldots, u_n) \le \varepsilon\,\mathbf{E}[F]] \\ \le\ &\Pr[P(u_1) \le \varepsilon^p\,\mathbf{E}[F]] + \Pr\left[F(u_1, \ldots, u_n) \le \varepsilon^{1-p}P(u_1)\right] \\ =\ &\Pr[P(u_1) \le \varepsilon^p\,\mathbf{E}[F]] + \Pr\left[F_{u_1}(u_2, \ldots, u_n) \le \varepsilon^{1-p}\,\mathbf{E}[F_{u_1}]\right] \\ \le\ &4d(n+1)\varepsilon^{p/(4d)} + C'(n-1,1,d)\cdot\varepsilon^{(1-p)C(n-1,1,d)}. \end{aligned} \tag{17}$$

We can take

$$C(n,1,d) = \max_p \min\Big\{\frac{p}{4d}, (1-p)C(n-1,1,d)\Big\} = \frac{1}{4d + C(n-1,1,d)^{-1}} = \frac{1}{4nd},$$

and $C'(n,1,d) = 4d(n+1) + C'(n-1,1,d) = O(n^2 d)$.

3.4 $m > 1, n > 1$

Now we handle the general case when the inputs to the semi-polynomial consist of m columns of a Haar random unitary. The proof is similar to the last stage, using the fact that the input distribution can be viewed as a unitary-invariant distribution over m orthonormal vectors in $\mathbb{C}^n$. In particular, let the vectors be $v_1, \ldots, v_m$, and we consider the function

$$P(v_1, \ldots, v_{m-1}) = \mathop{\mathbf{E}}_{v_m}[F(v_1, \ldots, v_m)].$$

Here v_m is a Haar-random vector over the unit sphere in the $(n-m)$-dimensional orthogonal subspace of $\mathrm{span}(v_1, \ldots, v_m)$. We would like to show that P is a semi-polynomial in order to use induction, and we first show it with v_m replaced with a Gaussian distribution.

▶ **Lemma 15.** *Let $v_1, \ldots, v_m$ be a set of $m < n$ orthonormal vectors in $\mathbb{C}^n$, and let $g = (g_1, \ldots, g_n) \in \mathbb{C}^n$ distributed as a standard $(n-m)$-dimensional complex Gaussian in the orthogonal subspace of $\mathrm{span}(v_1, \ldots, v_m)$.*

Let $G : \mathbb{C}^n \to \mathbb{C}$ be a degree-d semi-polynomial in $(g_1, \ldots, g_n)$. Then

$$Q(v_1, \ldots, v_m) = \mathbf{E}[G(g_1, \ldots, g_n)]$$

is a degree-d semi-polynomial in the entries of $v_1, \ldots, v_m$.

Proof. Omitted in this version. ◀

▶ **Corollary 16.** *$P(v_1, \ldots, v_{m-1}) = \mathbf{E}_{v_m}[F(v_1, \ldots, v_m)]$ is a degree-d semi-polynomial in the entries of $v_1, \ldots, v_{m-1}$.*

Proof. Notice that v_m is equidistributed as $g/\|g\|_2$, where $g = (g_1, \ldots, g_n)$ is the Gaussian in Lemma 15. In other words, $g = r \cdot v_m$ where $r \in \mathbb{R}$ follows a fixed χ distribution independent of v_m.

Now consider each monomial in F, and let G be the part of monomial over entries of v_m, then it suffices to show that $\mathbf{E}[G(v_m)]$ is a degree-d semi-polynomial in $v_1, \ldots, v_{m-1}$. Lemma 15 already showed this for $\mathbf{E}[G(g)]$, and since G is a monomial of degree $\ell \le d$, we have

$$\mathbf{E}[G(g)] = \mathbf{E}[G(r \cdot v_m)] = \mathbf{E}[r^\ell]\,\mathbf{E}[G(v_m)], \tag{18}$$

where $\mathbf{E}[r^\ell]$ is a non-zero constant irrelevant to the choice of $v_1, \ldots, v_{m-1}$. Therefore $\mathbf{E}[G(v_m)]$ is also a degree-d semi-polynomial in $v_1, \ldots, v_m$. ◀

Since P is an expectation over F, it is also always non-negative and has the same expectation as $\mathbf{E}[F]$. Thus the induction hypothesis gives

$$\Pr[P(v_1, \ldots, v_{m-1}) \le \varepsilon\,\mathbf{E}[F]] \le C'(n, m-1, d) \cdot \varepsilon^{C(n,m-1,d)}. \tag{19}$$

Now we fix some $v_1, \ldots, v_{m-1}$ and consider the degree-d semi-polynomial

$$F_{v_1,\ldots,v_{m-1}}(v_m) = F(v_1, \ldots, v_{m-1}, v_m),$$

which is always non-negative. Since there exists a linear map $A : \mathbb{C}^{n-m+1} \to \mathbb{C}^n$ such that $v_m = Au$, where u is a Haar-random unit vector in $\mathbb{C}^{n-m+1}$, we can apply Equation (17) from the previous stage for $m = 1$ on $F_{v_1,\ldots,v_{m-1}}(Au)$ to get

$$\Pr\Big[F_{v_1,\ldots,v_{m-1}}(Au) \le \varepsilon\, \mathbf{E}_u\Big[F_{v_1,\ldots,v_{m-1}}(Au)\Big]\Big] \le C'(n,1,d)\cdot\varepsilon^{C(n,1,d)}. \tag{20}$$

Therefore we conclude that, for every $p \in (0,1)$,

$$\begin{aligned}
&\Pr[F(v_1,\ldots,v_m) \le \varepsilon\, \mathbf{E}[F]]\\
\le\ &\Pr[P(v_1,\ldots,v_{m-1}) \le \varepsilon^p\, \mathbf{E}[F]] + \Pr\Big[F(v_1,\ldots,v_m) \le \varepsilon^{1-p} P(v_1,\ldots,v_{m-1})\Big]\\
=\ &\Pr[P(v_1,\ldots,v_{m-1}) \le \varepsilon^p\, \mathbf{E}[F]] + \Pr\Big[F_{v_1,\ldots,v_{m-1}}(v_m) \le \varepsilon^{1-p}\, \mathbf{E}[F_{v_1,\ldots,v_{m-1}}]\Big]\\
\le\ &C'(n,m-1,d)\cdot\varepsilon^{pC(n,m-1,d)} + C'(n,1,d)\cdot\varepsilon^{(1-p)C(n,1,d)},
\end{aligned} \tag{21}$$

Similar to the previous stage, we can take

$$C(n,m,d) = \frac{1}{C(n,m-1,d)^{-1} + C(n,1,d)^{-1}} = \frac{1}{4nmd},$$

and $C'(n,m,d) = C'(n,m-1,d) + C'(n,1,d) = O(n^2md)$. This completes the proof of Theorem 13.

4 Mixing Bound for Random Circuits

In this section we prove Theorem 1. Since the output qubit $\Phi_{\mathcal{C}}(\rho)$ is in the lightcone of the input qubit ρ, there exists gates $G_1,\ldots,G_D$ in the circuit $\mathcal{C}$ that connect the input and output qubits. That is, there are qubits $\rho_0,\rho_1,\ldots,\rho_D$ with $\rho_0 = \rho$ and $\rho_D = \Phi_{\mathcal{C}}(\rho)$, such that the gate G_i has ρ_{i-1} as an input and ρ_i as an output. Each gate G_i is independently drawn from the Haar measure $\mathcal{U}(2^k)$.

Note that even when gate G_i is given, we cannot directly claim any relationship between ρ_{i-1} and ρ_i, as the other input qubits to G_i are correlated and possibly entangled with ρ_{i-1}. To handle this, we let τ_{i-1} be the composite state of the k input qubits to G_i, and by Lemma 12 we have

$$\big\|\tau_{i-1} - \tau'_{i-1}\big\|_{\mathrm{F}} \ge 2^{-k/2}\big\|\rho_{i-1} - \rho'_{i-1}\big\|_{\mathrm{F}}, \tag{22}$$

where ρ'_i and τ'_i are the corresponding states when the input state is $\rho'_0 = \rho'$. Therefore we only need to bridge the remaining gaps by proving

$$\big\|\rho_i - \rho'_i\big\|_{\mathrm{F}} \ge \lambda_i\big\|\tau_{i-1} - \tau'_{i-1}\big\|_{\mathrm{F}} \tag{23}$$

where λ_i is some function of G_i, and then bound the distribution of λ_i by applying the anti-concentration bound from Theorem 6.

At a first glance, this looks impossible as τ_{i-1} is a k-qubit state while ρ_i is single-qubit, which means that as long as $k > 1$, whatever G_i is, there will be input states $\tau_{i-1} \neq \tau'_{i-1}$ to G_i with the output qubits $\rho_i = \rho'_i$. The key observation is that the states τ_i cannot be arbitrary k-qubit states: Since all the input qubits to the circuit $\mathcal{C}$ are fixed except ρ, when the circuit $\mathcal{C}$ is given, each τ_i is the result of ρ through a fixed quantum channel. As the difference $\rho - \rho'$ ranges within the 3-dimensional Euclidean space Δ_1, the difference $\tau_{i-1} - \tau'_{i-1}$ could also only range within a 3-dimensional subspace of Δ_k.

This allows us to define and bound the distribution of λ_i that uniformly holds for every such subspace as follows:

▶ **Lemma 17.** *For every $k \in \mathbb{N}$, there exists a distribution Λ_k over $[0,\infty)$ that satisfies:*

- *There exists $t > 0$ such that $\mathbf{E}[\Lambda_k^{-t}] < \infty$, where we use $\mathbf{E}[\Lambda_k^{-t}]$ as shorthand for $\mathbf{E}_{x\sim\Lambda_k}[x^{-t}]$.*
- *For every fixing of the input qubits other than ρ, and layers $1,\dots,i-1$ of the circuit $\mathcal{C}$, there exists a function $\lambda_i\colon \mathbb{U}(2^k) \to [0,\infty)$ at layer i such that*

$$\left\|\rho_i - \rho_i'\right\|_{\mathrm{F}} \geq \lambda_i(G_i)\left\|\tau_{i-1} - \tau_{i-1}'\right\|_{\mathrm{F}}$$

holds for all input states ρ and ρ', and $\lambda_i(G_i) \sim \Lambda_k$ when $G_i \sim \mathcal{U}(2^k)$.

We will prove Lemma 17 in Section 4.1. For now let us show how Lemma 17 would imply Theorem 1.

Proof of Theorem 1. From Lemma 17 and Equation (22) we get

$$\left\|\rho_i - \rho_i'\right\|_{\mathrm{F}} \geq \lambda_i(G_i)\left\|\tau_{i-1} - \tau_{i-1}'\right\|_{\mathrm{F}} \geq 2^{-k/2}\lambda_i(G_i)\left\|\rho_{i-1} - \rho_{i-1}'\right\|_{\mathrm{F}} \tag{24}$$

for every $i = 1,\dots,D$. Here each function $\lambda_i(G_i)$ depends on the previous layers, but as random variables $\lambda_i = \lambda_i(G_i)$ they are independent, since λ_i follows the same distribution Λ_k no matter how $\lambda_1,\dots,\lambda_{i-1}$ are fixed.

Since $\left\|\rho_0 - \rho_0'\right\|_{\mathrm{F}} = \left\|\rho - \rho'\right\|_{\mathrm{F}}$ and $\left\|\rho_D - \rho_D'\right\|_{\mathrm{F}} = \left\|\Phi_{\mathcal{C}}(\rho) - \Phi_{\mathcal{C}}(\rho')\right\|_{\mathrm{F}}$, we have

$$\left\|\Phi_{\mathcal{C}}(\rho) - \Phi_{\mathcal{C}}(\rho')\right\|_{\mathrm{F}} \geq 2^{-kD/2}\lambda_1\cdots\lambda_D\left\|\rho - \rho'\right\|_{\mathrm{F}}. \tag{25}$$

To bound the product of λ_i we use Markov's inequality, which states that for every $\alpha > 0$,

$$\begin{aligned}\Pr[\lambda_1\cdots\lambda_D \leq \alpha] &= \Pr[(\lambda_1\cdots\lambda_D)^{-t} \geq \alpha^{-t}] \\ &\leq \alpha^t\,\mathbf{E}[\lambda_1^{-t}\cdots\lambda_D^{-t}] = \alpha^t\,\mathbf{E}[\Lambda_k^{-t}]^D\end{aligned} \tag{26}$$

where we take $t > 0$ to be the constant in Lemma 17. Take α such that $\gamma = \alpha^t\,\mathbf{E}[\Lambda_k^{-t}]^D$, we conclude that with probability at least $1-\gamma$,

$$\begin{aligned}\left\|\Phi_{\mathcal{C}}(\rho) - \Phi_{\mathcal{C}}(\rho')\right\|_{\mathrm{F}} &\geq 2^{-kD/2}\alpha\left\|\rho - \rho'\right\|_{\mathrm{F}} \\ &= 2^{-kD/2}\gamma^{1/t}\,\mathbf{E}[\Lambda_k^{-t}]^{-D/t}\left\|\rho - \rho'\right\|_{\mathrm{F}} \\ &= (2^{-D}\gamma)^{O_k(1)}\left\|\rho - \rho'\right\|_{\mathrm{F}}.\end{aligned}$$

◀

4.1 Proof of Lemma 17

The basic idea of the proof is to lower bound the ratio $\left\|\rho_i - \rho_i'\right\|_{\mathrm{F}} / \left\|\tau_{i-1} - \tau_{i-1}'\right\|_{\mathrm{F}}$ with a polynomial function on the entries of G_i and apply Theorem 6.

The quantum channel defined by G_i that maps τ_{i-1} to ρ_i induces the linear map

$$M_i : \tau_{i-1} - \tau_{i-1}' \mapsto \rho_i - \rho_i'.$$

The domain of M_i is a 3-dimensional subspace of Δ_k which we denote as $\mathcal{S}_i$, while the range of M_i is Δ_1. Notice that the map M_i is completely determined by the domain $\mathcal{S}_i$ and the gate G_i, while $\mathcal{S}_i$ depends only on the fixed inputs to the circuit $\mathcal{C}$ and the gates of $\mathcal{C}$ in layers $1,\dots,i-1$.

Since both the domain and the range are Euclidean spaces, the absolute determinant $|\det M_i|$ is independent of the choices of bases when writing M_i as matrix in $\mathbb{R}^{3\times 3}$. We will show that $|\det M_i|$ is basically the lower bound that we seek for, via the following propositions whose proof can be found in the full version.

▶ **Proposition 18.** *It always holds that*

$$\left\|\rho_i - \rho_i'\right\|_{\mathrm{F}} \geq 2^{-k}|\det M_i| \cdot \left\|\tau_{i-1} - \tau_{i-1}'\right\|_{\mathrm{F}}.$$

▶ **Proposition 19.** *For each fixed domain $\mathcal{S}_i$, $\det M_i$ is a degree-6 semi-polynomial in the entries of G_i.*

▶ **Proposition 20.** *There exists a constant $\mu_k > 0$ such that for every possible domain $\mathcal{S}_i$,*

$$\mathop{\mathbf{E}}_{G_i \sim \mathcal{U}(2^k)}\left[|\det M_i|^2\right] \geq \mu_k.$$

Now we are ready to prove Lemma 17.

Proof of Lemma 17. Applying Theorem 6 with Propositions 19 and 20 gives

$$\begin{aligned}\Pr\left[2^{-k}|\det M_i| \leq \varepsilon\right] &\leq \Pr\left[|\det M_i|^2 \leq 2^{2k}\mu_k^{-1}\varepsilon^2\,\mathbf{E}\left[|\det M_i|^2\right]\right] \\ &\leq C'(2^k,6)\cdot(2^{2k}\mu_k^{-1}\varepsilon^2)^{C(2^k,6)} \end{aligned} \tag{27}$$

which holds for every domain $\mathcal{S}_i$ and every $\varepsilon \geq 0$. We define Λ_k as the distribution over $[0,\infty)$ with the following cumulative function:

$$\Pr[\Lambda_k \leq x] = \min\left\{C'(2^k,6)\cdot(2^{2k}\mu_k^{-1}x^2)^{C(2^k,6)}, 1\right\}.$$

Take $t = C(2^k,6) > 0$ and let $C = C'(2^k,6)\cdot(2^{2k}\mu_k^{-1})^{C(2^k,6)} > 0$, then we have

$$\begin{aligned}\mathbf{E}[\Lambda_k^{-t}] &= \int_0^\infty x^{-t}\mathrm{d}\Pr[\Lambda_k \leq x] \\ &= \int_0^{C^{-1/(2t)}} x^{-t}\mathrm{d}(Cx^{2t}) \\ &= \int_0^{C^{-1/(2t)}} 2tC\cdot x^{t-1}\mathrm{d}x = 2tC^{1/2} < \infty. \end{aligned} \tag{28}$$

Now suppose the input qubits other than ρ are fixed, and the layers $1,\ldots,i-1$ of the circuit $\mathcal{C}$ is given. This fixes the domain $\mathcal{S}_i$ for M_i, and provides the cumulative function

$$P(x) = \Pr\left[2^{-k}|\det M_i| \leq x\right].$$

Notice that P is continuous and non-decreasing, and by defining its inverse $P^{-1}(y) = \sup_x P(x) \leq y$ on $[0,1]$, we can show that for every $y \in [0,1]$,

$$\Pr\left[P\big(2^{-k}|\det M_i|\big) \leq y\right] = \Pr\left[2^{-k}|\det M_i| \leq P^{-1}(y)\right] = P(P^{-1}(y)) = y.$$

We then define the function $\lambda_i : \mathbb{U}(2^k) \to [0,\infty)$ as follows: For each $G_i \in \mathbb{U}(2^k)$, let $\lambda_i(G_i)$ be the smallest $\lambda \geq 0$ such that

$$P\big(2^{-k}|\det M_i|\big) = \Pr[\Lambda_k \leq \lambda].$$

When $G_i \sim \mathcal{U}(2^k)$, we have $\lambda_i(G_i) \sim \Lambda_k$ since

$$\Pr[\lambda_i(G_i) \leq x] = \Pr\left[P\big(2^{-k}|\det M_i|\big) \leq \Pr[\Lambda_k \leq x]\right] = \Pr[\Lambda_k \leq x]. \tag{29}$$

We also have $2^{-k}|\det M_i| \geq \lambda_i(G_i)$ since $P(x) \leq \Pr[\Lambda_k \leq x]$ holds for all $x \geq 0$. Combined with Proposition 18 we get

$$\left\|\rho_i - \rho_i'\right\|_{\mathrm{F}} \geq 2^{-k}|\det M_i| \cdot \left\|\tau_{i-1} - \tau_{i-1}'\right\|_{\mathrm{F}} \geq \lambda_i(G_i)\left\|\tau_{i-1} - \tau_{i-1}'\right\|_{\mathrm{F}}. \qquad \blacktriangleleft$$

5 Applications

5.1 Depth Lower Bound for Approximate Designs

In this section we prove Theorem 2. We first recall the definition of an approximate unitary design.

▶ **Definition 21.** *For a distribution $\mathcal{D}$ over $\mathbb{U}(n)$ and $t \in \mathbb{N}_+$, define the moment superoperator as the following channel:*

$$\Phi_{\mathcal{D}}^{(t)} : \rho \mapsto \int_{\mathbb{U}(n)} U^{\otimes t} \rho (U^\dagger)^{\otimes t} \mathrm{d}\mathcal{D}(U).$$

The distribution $\mathcal{D}$ is an ε-approximate unitary t-design if

$$\left\| \Phi_{\mathcal{D}}^{(t)} - \Phi_{\mathcal{U}(n)}^{(t)} \right\|_\diamond \leq \varepsilon.$$

Note that there are several other definitions of the approximate design (see e.g. [10]), and we choose the weaker one so that our Theorem 2 is still compatible with the stronger definitions of designs.

Proof of Theorem 2. Without loss of generality, let us assume that the first input qubit ρ and the first output qubit π are depth D apart. We fix all other input qubits to be maximally mixed, so that when ρ is also maximally mixed, the entire output is maximally mixed regardless of the circuit $\mathcal{C}$, and thus $\pi = I/2$. On the other hand, when $\rho = |0\rangle\langle 0|$, we know the following about the output qubit π via Theorem 1 that with probability $1 - 2^{-D}$ over the circuit $\mathcal{C}$,

$$\|\pi - I/2\|_{\mathrm{F}} \geq \||0\rangle\langle 0| - I/2\|_{\mathrm{F}} \cdot (2^{-2D})^{c_k} = \frac{1}{\sqrt{2}} \cdot 2^{-2Dc_k}. \tag{30}$$

We can expand the left hand side of Equation (30) as

$$\|\pi - I/2\|_{\mathrm{F}}^2 = \mathrm{Tr}[(\pi - I/2)^2] = \mathrm{Tr}[\pi^2] - 1/2. \tag{31}$$

Since $\mathrm{Tr}[\pi^2] \leq 1$ always holds, we get

$$\mathbf{E}_{\mathcal{C}}[\mathrm{Tr}[\pi^2]] \geq \frac{1}{2} + \frac{1}{2} \cdot 2^{-4Dc_k} + \frac{1}{2} \cdot 2^{-D}. \tag{32}$$

Now imagine feeding two copies of the input state $|0\rangle\langle 0| \otimes (I/2)^{\otimes(n-1)}$ to the superoperators $\Phi_{\mathcal{C}}^{(2)}$ and $\Phi_{\mathcal{U}(n)}^{(2)}$, and apply a swap test on the first output qubits in the two copies. The output probability is determined by $\mathrm{Tr}[\pi^2]$. We know already from [20] that when going through an n-qubit Haar random unitary, we have $\mathbf{E}_{\mathcal{U}(n)}[\mathrm{Tr}[\pi^2]] = 1/2$. Therefore, we conclude that the difference $2^{-4Dc_k} + 2^{-D} \leq O(\varepsilon)$ which means that $D \geq \Omega_k(\log \varepsilon^{-1})$. ◀

5.2 Depth Test

In this section we prove Theorem 3, where we learn the exact depth of a brickwork random circuit $\mathcal{C}$.

The processed is described in Algorithm 1. Here c_2 is the constant c_k in Theorem 1 with $k = 2$ for brickwork circuits, and $\gamma > 0$ is the target error probability. Notice that in a brickwork circuit of depth D, the $(D+2)$-th output qubit lies outside the lightcone of the first input qubit. Therefore, when the algorithm iterates to the correct depth D, we have $\pi = \pi'$ and D must be returned even when both π and π' are estimated with ε error.

Algorithm 1 Algorithm for depth testing.

Arbitrarily fix all input qubits except the first one ρ.
for $D = 0, 1, \ldots$ **do**
  Let $\varepsilon = (2^{-2D}\gamma)^{c_2}/4$;
  Apply $\mathcal{C}$ with $\rho = |0\rangle\langle 0|$ and let π be the $(D+2)$-th output qubit;
  Apply $\mathcal{C}$ with $\rho = |1\rangle\langle 1|$ and let π' be the $(D+2)$-th output qubit;
  Estimate π and π' up to ε error by state tomography;
  if $\left\|\pi - \pi'\right\|_{\mathrm{F}} \leq 2\varepsilon$ **then return** D.

On the other hand, when D is smaller than the actual depth, the $(D+2)$-th output qubit lies inside the lightcone of the first input qubit. By Theorem 1, with probability $1 - 2^{-D}\gamma$ over $\mathcal{C}$ we have

$$\left\|\pi - \pi'\right\|_{\mathrm{F}} \geq \left\||0\rangle\langle 0| - |1\rangle\langle 1|\right\|_{\mathrm{F}} \cdot (2^{-2D}\gamma)^{c_2} > 4\varepsilon. \tag{33}$$

This means when both π and π' are estimated with ε error, we still have $\left\|\pi - \pi'\right\|_{\mathrm{F}} > 2\varepsilon$. By a union bound over D, with probability $1 - \gamma$ all those D smaller than the actual depth will be skipped, and thus the outputted depth is correct.

Note that the efficiency of the algorithm depends on the single-qubit tomography process, which by Proposition 11 is $\mathrm{poly}(1/\varepsilon)$. As a conclusion, we obtain the following more general statement which implies Theorem 3:

▶ **Theorem 22.** *Let $\mathcal{C}$ be a brickwork random quantum circuit of an unknown depth D , where each gate is independently Haar random.. Given oracle access to $\mathcal{C}$, for any $\gamma \in (0, 1)$, Algorithm 1 outputs D with probability at least $1 - \gamma$ in time* $\mathrm{poly}(2^D, \gamma^{-1})$.

▶ Remark 23. The only property of the brickwork architecture we used here is that the set of qubits within the lightcone of an input qubit is strictly expanding when the depth grows, which allows us to distinguish between different depths. Therefore the algorithm can be easily modified, with the same efficiency, to work with higher dimensional brickwork circuits and other architectures.

5.3 Learning Brickwork Random Circuits

5.3.1 Learning the First Gate

Here we prove Theorem 4, where we learn the gate $G_{1,1}$, namely the gate in the first layer acting on the first and second qubit, in a brickwork circuit of depth $D = O(\log n)$ with Haar random gates. The same arguments also work for other gates in the first layer.

To learn $G_{1,1}$, we try to uncompute $G_{1,1}$ by first apply some two-qubit unitary $G^\dagger \in \mathbb{U}(4)$ and then apply the circuit $\mathbb{C}$. We distinguish whether G is close to $G_{1,1}$ or not using the similar idea as in Section 5.2. Specifically, if $G = G_{1,1}$ so that $G^\dagger$ perfectly uncomputes the gate $G_{1,1}$, then the $(D+1)$-th output qubit will lie outside the lightcone of the first input qubit. Note that this is also true when $G^\dagger$ cancels $G_{1,1}$ into unentangled single-qubit gates, that is when there exists $U_1, U_2 \in \mathbb{U}(2)$ that $G_{1,1}G^\dagger = U_1 \otimes U_2$. In contrast, when $G^\dagger$ does not cancel $G_{1,1}$ into unentangled single-qubit gates, the first two qubits will be entangled after the first layer of gates, and thus the $(D+1)$-th output qubit will be affected when the first input qubit changes.

To put the above intuition more formally, we first define the distance between two-qubit gates when taking the quotient over unentangled single-qubit gates:

▶ **Definition 24.** *For $G, G' \in \mathbb{U}(4)$, we define the distance*

$$d_{\otimes}(G, G') = \min_{U_1, U_2 \in \mathbb{U}(2)} d_{\diamond}(G, (U_1 \otimes U_2) \cdot G').$$

We note that $d_{\otimes}$, like the diamond distance $d_{\diamond}$, is a pseudometric on $\mathbb{U}(4)$. That is, a metric except that two distinct unitaries could have distance zero. However, $d_{\otimes}(G, G') = 0$ if and only if $G' \cdot G^{-1} = U_1 \otimes U_2$ for some $U_1, U_2 \in \mathbb{U}(2)$.

The learning algorithm is described in Algorithm 2, with $\delta, \gamma > 0$ being the target error rate. Notice that within each loop of σ_2, the second input qubit ρ_2 is fixed, while the first one changes with a difference σ_1 that goes through the Pauli basis. The input qubits ρ_1 and ρ_2 are Pauli eigenstates except when $\sigma_2 = \rho_2 = I/2$, which can be obtained by randomly choosing $|0\rangle$ or $|1\rangle$.

Algorithm 2 Algorithm for learning the gate $G_{1,1}$.

```
Arbitrarily fix all input qubits except the first and second ones ρ_1, ρ_2.
Let ε = 10^{-4} δ^2 (2^{-D+1} γ)^{c_2};
foreach G from an ε-net of 𝕌(4) under distance d_⊗ do
    for σ_1 ∈ {X, Y, Z} and σ_2 ∈ {I, X, Y, Z} do
        Let ρ_1 = (I + σ_1)/2, ρ_2 = (I + σ_2)/2 and apply G^† on the first two qubits,
            then apply 𝒞 and let π be the (D + 1)-th output qubit;
        Let ρ_1 = (I − σ_1)/2, ρ_2 = (I + σ_2)/2 and apply G^† on the first two qubits,
            then apply 𝒞 and let π' be the (D + 1)-th output qubit;
        Estimate π and π' up to ε error by state tomography;
        if ‖π − π'‖_F ≥ 5ε then reject G.
    return G if not rejected.
```

To prove the correctness of the algorithm, we need the following lemmas that connects the distance $d_{\otimes}(U, I \otimes I)$ with the behavior of U over the Pauli basis.

▶ **Lemma 25.** *For $U \in \mathbb{U}(4)$, if $d_{\otimes}(U, I \otimes I) \leq \delta$ then for every $\sigma_1 \in \{X, Y, Z\}$ and $\sigma_2 \in \{I, X, Y, Z\}$,*

$$\left\| \mathrm{Tr}_A \left[U(\sigma_1 \otimes \sigma_2) U^{\dagger} \right] \right\|_{\mathrm{F}} \leq 2\delta,$$

where Tr_A is the partial trace that traces out the first qubit.

▶ **Lemma 26.** *For $U \in \mathbb{U}(4)$, if for every $\sigma_1 \in \{X, Y, Z\}$ and $\sigma_2 \in \{I, X, Y, Z\}$ we have*

$$\left\| \mathrm{Tr}_A \left[U(\sigma_1 \otimes \sigma_2) U^{\dagger} \right] \right\|_{\mathrm{F}} \leq \delta,$$

then $d_{\otimes}(U, I \otimes I) \leq 20\sqrt{\delta}$.

The proof of Lemma 26 is rather technical and thus is omitted in this version. For now, let us show how Lemmas 25 and 26 would imply the completeness and the soundness of Algorithm 2.

We apply Theorem 1 to the circuit $\mathcal{C}$ minus its first layer, which is a circuit of depth $D - 1$. Notice that among the input qubits to the second layer of $\mathcal{C}$, the third to n-th qubits only depends on the gates $G_{1,3}, \ldots, G_{1,n}$ and thus can be viewed as fixed. The first qubit does not affect the $(D + 1)$-th output qubit and thus its entire lightcone can be removed from the picture. That leaves us the second qubit, which is

$$\mathrm{Tr}_A \left[G_{1,1} G^{\dagger} (\rho_1 \otimes \rho_2) G G_{1,1}^{\dagger} \right].$$

Therefore, with ρ_1 changed with a difference σ_1 and $\rho_2 = (I + \sigma_2)/2$ unchanged, Theorem 1 implies that with probability $1 - \gamma$, the following holds for every G.

$$\left\|\pi - \pi'\right\|_{\mathrm{F}} \geq \left\|\mathrm{Tr}_A\left[G_{1,1}G^\dagger(\sigma_1 \otimes \rho_2)GG_{1,1}^\dagger\right]\right\|_{\mathrm{F}} \cdot (2^{-D+1}\gamma)^{c_2}. \tag{34}$$

Meanwhile, Lemma 12 implies that

$$\left\|\pi - \pi'\right\|_{\mathrm{F}} \leq \left\|\mathrm{Tr}_A\left[G_{1,1}G^\dagger(\sigma_1 \otimes \rho_2)GG_{1,1}^\dagger\right]\right\|_{\mathrm{F}} \cdot \sqrt{2}. \tag{35}$$

For completeness, since we take G from an ε-net, the algorithm must have tested some G with $d_{\otimes}(G_{1,1}, G) = d_{\otimes}(G_{1,1}G^\dagger, I \otimes I) \leq \varepsilon \leq \delta$. Hence by Lemma 25 and Equation (35), for such G it always holds that $\left\|\pi - \pi'\right\|_{\mathrm{F}} \leq 2\sqrt{2}\varepsilon$, and thus G will not be rejected even with ε tomography errors in π and π'.

For soundness, assume that $d_{\otimes}(G_{1,1}, G) > \delta$, then by Lemma 26 we know that for some $\sigma_1 \in \{X, Y, Z\}$ and $\sigma_2 \in \{I, X, Y, Z\}$,

$$\left\|\mathrm{Tr}_A\left[G_{1,1}G^\dagger(\sigma_1 \otimes \sigma_2)GG_{1,1}^\dagger\right]\right\|_{\mathrm{F}} > \frac{1}{400}\delta^2. \tag{36}$$

As $\sigma_2 = 2\rho_2 - I$, that means there exists some ρ_2 such that

$$\left\|\mathrm{Tr}_A\left[G_{1,1}G^\dagger(\sigma_1 \otimes \rho_2)GG_{1,1}^\dagger\right]\right\|_{\mathrm{F}} > \frac{1}{1200}\delta^2. \tag{37}$$

Thus by Equation (34) we have $\left\|\pi - \pi'\right\|_{\mathrm{F}} > \delta^2(2^{-D+1}\gamma)^{c_2}/1200 > 7\varepsilon$, and G will be rejected when π and π' are estimated with ε error.

The efficiency of the algorithm depends on the size of the ε-net and the state tomography process, which are both $\mathrm{poly}(1/\varepsilon) = \mathrm{poly}(1/\delta, 1/\gamma)$. Notice that the algorithm similarly works for every gate in the first layer, and as a result, we obtained the following formal statement of Theorem 4:

▶ **Theorem 27.** *Let $\mathcal{C}$ be a brickwork random quantum circuit of depth D, where each gate is independently Haar random. Let $G_{1,1}$ be the gate in the first layer of $\mathcal{C}$ that acts on the first and second qubit. Given oracle access to $\mathcal{C}$, for any $\delta, \gamma \in (0, 1)$, with probability at least $1 - \gamma$ over $\mathcal{C}$, Algorithm 2 outputs some $G \in \mathbb{U}(4)$ such that $d_{\otimes}(G, G_{1,1}) \leq \delta$ in time* $\mathrm{poly}(2^D, 1/\delta, 1/\gamma)$.

5.3.2 Learning the Circuit with Discretized Distribution

It is tempting to use Theorem 27 to learn the entire circuit $\mathcal{C}$. Indeed, if the statement is errorless that $d_{\otimes}(G, G_{1,1}) = 0$, we could view the single-qubit gates $G_{1,1}G^\dagger = U_1 \otimes U_2$ as part of the second layer. That means after learning the first layer, we can perfectly uncompute it and hence use Algorithm 2 to learn the second layer, and proceed until we are only left with single qubit gates, which are easily learnable.

However, when there are learning errors, which are inevitable for a continuous gate distribution like $\mathcal{U}(4)$, the above framework runs into a problem. Since we cannot perfectly uncompute the first layer, the inputs to the rest of the circuit are not clean enough: In particular, the error in $\sigma_1 \otimes \rho_2$ could have much larger influence on the output difference $\pi - \pi'$ than σ_1 itself. To control the influence of the errors, we need to reduce the learn error in the previous layer to be polynomially smaller than the target error in the current layer, and therefore a circuit of depth $D = \Theta(\log n)$ would require error as small as $2^{-\Omega(\log^2 n)}$ and incur quasi-polynomial running time.

Here we present a bypass to the problem which allows us to prove an errorless version of Theorem 27 and thus make the proposed framework work. The idea is to change the distribution from $\mathbb{U}(4)$ into a discrete one that approximates $\mathbb{U}(4)$. Intuitively, any ε-net where the elements are distributed according to $\mathcal{U}(4)$ would be a good approximation. We formalize this intuition as the following:

▶ **Definition 28.** *An ε-net of a distribution $\mathcal{D}$ over a pseudometric space (S, d) is a distribution $\mathcal{D}_\varepsilon$ over S with a finite support, such that $\mathcal{D}_\varepsilon = f(\mathcal{D})$ for some (possibly randomized) map $f : S \to S$, with the following properties:*

- *For every $x \in S$, $d(x, f(x)) \leq \varepsilon$.*
- *For every $x_1, x_2 \in S$, either $f(x_1) = f(x_2)$ or $d(f(x_1), f(x_2)) \geq \varepsilon$.*

Notice that under Definition 28, the set $\operatorname{supp} \mathcal{D}_\varepsilon$ is indeed an ε-net in the normal sense. Actually, the definition is general enough so that we can first choose any ε-net as the support, remove the redundant elements with zero distances, and then take f to be an arbitrary rounding scheme into the support. We show that $\mathcal{D}_\varepsilon$ approximates $\mathcal{D}$ via the following lemma.

▶ **Lemma 29.** *If $F : S \to \mathbb{R}$ is L-Lipschitz, that is for all $x_1, x_2 \in S$,*

$$|F(x_1) - F(x_2)| \leq L \cdot d(x_1, x_2),$$

then for every $\delta \in \mathbb{R}$ we have

$$\Pr_{x \sim \mathcal{D}_\varepsilon}[F(x) \leq \delta] \leq \Pr_{x \sim \mathcal{D}}[F(x) \leq \delta + \varepsilon L].$$

From now on, for each $\varepsilon > 0$ we fix some ε-net of the Haar measure $\mathcal{U}(4)$ under the $d_\otimes$ distance, and denote it by $\mathcal{U}_\varepsilon(4)$. We will show that when the gates in the brickwork circuit are drawn the net, we can actually use the framework at the start of this section to learn the circuit. To do so, we first prove an errorless version of Theorem 27 as follows.

▶ **Theorem 30.** *Let $\mathcal{C}$ be a brickwork random quantum circuit of depth D, where each gate is independently drawn from $\mathcal{U}_\varepsilon(4)$. Let $G_{1,1}$ be the gate in the first layer of $\mathcal{C}$ that acts on the first and second qubit. Given oracle access to $\mathcal{C}$, for every $\gamma \in (0, 1)$ there is an algorithm that with probability at least $1 - \gamma$ over $\mathcal{C}$ outputs $G_{1,1}$ in time $\operatorname{poly}(2^D, 1/\gamma)$.*

Proof. The algorithm is basically the same as Algorithm 2, except that we now iterate G through the support of $\mathcal{U}_\varepsilon(4)$. The proof is also mostly the same: When $G = G_{1,1}$, we have $\pi = \pi'$ and thus G will not be rejected; Otherwise $d_\otimes(G, G_{1,1}) \geq \varepsilon$, and the proof goes through as long as we have the corresponding version of Theorem 1.

Since Theorem 1 is proved via Lemma 17, it suffices to prove Lemma 17 where $\mathcal{U}(4)$ is replaced with $\mathcal{U}_\varepsilon(4)$. The crux is to prove the inequality Equation (27), that is for some $C, C' > 0$, it holds for all $x \geq 0$ that

$$\Pr[|\det M| \leq x] \leq C' x^C, \tag{38}$$

where M is the matrix form of the linear map $M : \tau - \tau' \mapsto \operatorname{Tr}_A[G(\tau - \tau')G^\dagger]$, for $\tau - \tau'$ in a certain fixed 3-dimensional subspace of Δ_2.

Note that by Proposition 18, each entry of M is in $[-2, 2]$, while by Proposition 19, each entry of M is a Lipschitz function of G under distance $d_\diamond$. This is because when $G, G' \in \mathbb{U}(4)$ that $d_\diamond(G, G') \leq \delta$ corresponds to matrices M and M', for any $\tau - \tau' \in \Delta_2$ and $\sigma \in \Delta_1$ with $\left\|\tau - \tau'\right\|_{\mathrm{F}} = \|\sigma\|_{\mathrm{F}} = 1$ we have

$$\left|\operatorname{Tr}[\sigma M(\tau - \tau')] - \operatorname{Tr}[\sigma M'(\tau - \tau')]\right| \leq 2\delta. \tag{39}$$

As $\det M$ consists of 6 monomials of degree 3 in the entries of M, we conclude that $|\det M|$ is $2^3 \cdot 3 \cdot 6 = 144$-Lipschitz in G under $d_\diamond$. But when $G' = (U_1 \otimes U_2)G$ we have $M' = U_2 M U_2^\dagger$ which means $|\det M| = \left|\det M'\right|$, and thus $|\det M|$ is also 144-Lipschitz in G under $d_\otimes$.

As a result, since we already know that Equation (38) holds when $G \sim \mathcal{U}(4)$, by Lemma 29 we have

$$\Pr_{G\sim\mathcal{U}_\varepsilon(4)}[|\det M| \le x] \le \Pr_{G\sim\mathcal{U}(4)}[|\det M| \le x + 144\varepsilon] \le C'(x + 144\varepsilon)^C. \tag{40}$$

However, this will not yield Lemma 17 as the bound is even non-zero when $x = 0$. Fortunately, we can simply use the union bound to ignore the cases when $|\det M| \le \varepsilon$ for any gate G on the path, which will only add $\mathrm{poly}(\varepsilon)D$ to the error probability γ. And conditioned on $|\det M| > \varepsilon$, for every $x \ge 0$ we have

$$\Pr_{G\sim\mathcal{U}_\varepsilon(4)}[|\det M| \le x] \le C'(145x)^C, \tag{41}$$

which allows us to prove Lemma 17 on $\mathcal{U}_\varepsilon(4)$, albeit with different constants. ◀

As a result, we can use Theorem 30 to exactly learn the circuit $\mathcal{C}$ layer by layer, and obtain Theorem 5 formally as follows.

▶ **Theorem 31.** *Let $\mathcal{C}$ be a brickwork random quantum circuit on n qubits of depth D, where each gate is independently drawn from $\mathcal{U}_\varepsilon(4)$. Given oracle access to $\mathcal{C}$, for every $\gamma \in (0,1)$ there is an algorithm that with probability at least $1-\gamma$ outputs $\mathcal{C}$ in time* $\mathrm{poly}(n, 2^D, 1/\gamma)$.

References

1. Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, STOC '11, pages 333–342, New York, NY, USA, 2011. Association for Computing Machinery. doi:10.1145/1993636.1993682.
2. Scott Aaronson and Shih-Han Hung. Certified randomness from quantum supremacy. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, pages 933–944, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3564246.3585145.
3. Frank Arute et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, October 2019. doi:10.1038/s41586-019-1666-5.
4. Roozbeh Bassirian, Adam Bouland, Bill Fefferman, Sam Gunn, and Avishay Tal. On certified randomness from fourier sampling or random circuit sampling, 2024. arXiv:2111.14846.
5. Bruno Bertini and Lorenzo Piroli. Scrambling in random unitary circuits: Exact results. *Phys. Rev. B*, 102:064305, August 2020. doi:10.1103/PhysRevB.102.064305.
6. 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. *Nature Physics*, 14(6):595–600, April 2018. doi:10.1038/s41567-018-0124-x.
7. 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)*. IEEE, February 2022. doi:10.1109/focs52979.2021.00127.
8. Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. On the complexity and verification of quantum random circuit sampling. *Nature Physics*, 15(2):159–163, October 2018. doi:10.1038/s41567-018-0318-2.
9. Paolo Braccia, Pablo Bermejo, Lukasz Cincio, and M. Cerezo. Computing exact moments of local random quantum circuits via tensor networks, 2024. arXiv:2403.01706.

10 Fernando G.S.L. Brandão, Aram W. Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Commun. Math. Phys.*, 346(2):397–434, 2016. doi:10.1007/s00220-016-2706-8.

11 Winton Brown and Omar Fawzi. Scrambling speed of random quantum circuits, 2013. arXiv:1210.6644.

12 Anthony Carbery and James Wright. Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Mathematical Research Letters*, 8(3):233–248, 2001. doi:10.4310/mrl.2001.v8.n3.a1.

13 Chi-Fang Chen, Adam Bouland, Fernando G. S. L. Brandão, Jordan Docter, Patrick Hayden, and Michelle Xu. Efficient unitary designs and pseudorandom unitaries from permutations, 2024. doi:10.48550/arXiv.2404.16751.

14 Chi-Fang Chen, Jordan Docter, Michelle Xu, Adam Bouland, and Patrick Hayden. Efficient unitary t-designs from random sums, 2024. doi:10.48550/arXiv.2402.09335.

15 Chi-Fang Chen, Jeongwan Haah, Jonas Haferkamp, Yunchao Liu, Tony Metger, and Xinyu Tan. Incompressibility and spectral gaps of random circuits, 2024. doi:10.48550/arXiv.2406.07478.

16 Chi-Fang Chen and Andrew Lucas. Operator growth bounds from graph theory. *Communications in Mathematical Physics*, 385(3):1273–1323, July 2021. doi:10.1007/s00220-021-04151-6.

17 Chi-Fang Chen, Andrew Lucas, and Chao Yin. Speed limits and locality in many-body quantum dynamics. *Reports on Progress in Physics*, 86(11):116001, September 2023. doi:10.1088/1361-6633/acfaae.

18 Alexander M. Dalzell, Nicholas Hunter-Jones, and Fernando G. S. L. Brandão. Random quantum circuits anticoncentrate in log depth. *PRX Quantum*, 3(1), March 2022. doi:10.1103/prxquantum.3.010333.

19 Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80:012304, July 2009. doi:10.1103/PhysRevA.80.012304.

20 Joseph Emerson, Robert Alicki, and Karol Zyczkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7, March 2005. doi:10.1088/1464-4266/7/10/021.

21 Bill Fefferman, Soumik Ghosh, Michael Gullans, Kohdai Kuroiwa, and Kunal Sharma. Effect of non-unital noise on random circuit sampling, 2023. doi:10.48550/arXiv.2306.16659.

22 Matthew P.A. Fisher, Vedika Khemani, Adam Nahum, and Sagar Vijay. Random quantum circuits. *Annual Review of Condensed Matter Physics*, 14(1):335–379, 2023. doi:10.1146/annurev-conmatphys-031720-030658.

23 Jeongwan Haah, Yunchao Liu, and Xinyu Tan. Efficient approximate unitary designs from random pauli rotations, 2024. doi:10.48550/arXiv.2402.05239.

24 Jonas Haferkamp. Random quantum circuits are approximate unitary t-designs in depth $O\left(nt^{5+o(1)}\right)$. *Quantum*, 6:795, September 2022. doi:10.22331/q-2022-09-08-795.

25 Dominik Hangleiter, Juan Bermejo-Vega, Martin Schwarz, and Jens Eisert. Anticoncentration theorems for schemes showing a quantum speedup. *Quantum*, 2:65, May 2018. doi:10.22331/q-2018-05-22-65.

26 Dominik Hangleiter and Michael J. Gullans. Bell sampling from quantum circuits. *Phys. Rev. Lett.*, 133:020601, July 2024. doi:10.1103/PhysRevLett.133.020601.

27 Aram W. Harrow, Linghang Kong, Zi-Wen Liu, Saeed Mehraban, and Peter W. Shor. Separation of out-of-time-ordered correlation and entanglement. *PRX Quantum*, 2:020339, June 2021. doi:10.1103/PRXQuantum.2.020339.

28 Aram W. Harrow and Richard A. Low. Random quantum circuits are approximate 2-designs. *Communications in Mathematical Physics*, 291(1):257–302, July 2009. doi:10.1007/s00220-009-0873-6.

29 Aram W. Harrow and Saeed Mehraban. Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates. *Commun. Math. Phys.*, 401(2):1531–1626, 2023. doi:10.1007/s00220-023-04675-z.

30 Patrick Hayden and John Preskill. Black holes as mirrors: quantum information in random subsystems. *Journal of High Energy Physics*, 2007(09):120–120, September 2007. doi:10.1088/1126-6708/2007/09/120.

31 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, pages 1343–1351, New York, NY, USA, 2024. Association for Computing Machinery. doi:10.1145/3618260.3649722.

32 Nicholas Hunter-Jones. Unitary designs from statistical mechanics in random quantum circuits, 2019. arXiv:1905.12053.

33 Shao-Kai Jian, Gregory Bentsen, and Brian Swingle. Linear growth of circuit complexity from brownian dynamics, 2022. arXiv:2206.14205.

34 Daniel Kane. A structure theorem for poorly anticoncentrated polynomials of gaussians and applications to the study of polynomial threshold functions. *The Annals of Probability*, 45(3):1612–1679, 2017. doi:10.1214/16-AOP1097.

35 Zeph Landau and Yunchao Liu. Learning quantum states prepared by shallow circuits in polynomial time, 2024. doi:10.48550/arXiv.2410.23618.

36 Elliott H. Lieb and Derek W. Robinson. The finite group velocity of quantum spin systems. *Communications In Mathematical Physics*, 28(3):251–257, September 1972. doi:10.1007/BF01645779.

37 Yunchao Liu, Matthew Otten, Roozbeh Bassirianjahromi, Liang Jiang, and Bill Fefferman. Benchmarking near-term quantum computers via random circuit sampling, 2022. arXiv:2105.05232.

38 Shachar Lovett. An elementary proof of anti-concentration of polynomials in gaussian variables. Electronic Colloquium on Computational Complexity (ECCC), 2010. URL: https://eccc.weizmann.ac.il/report/2010/182.

39 Elizabeth S Meckes. *The random matrix theory of the classical compact groups*, volume 218. Cambridge University Press, 2019. doi:10.1017/9781108303453.

40 Raghu Meka, Oanh Nguyen, and Van Vu. Anti-concentration for polynomials of independent random variables. *Theory of Computing*, 12(11):1–17, 2016. doi:10.4086/toc.2016.v012a011.

41 Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *SIAM Journal on Computing*, 42(3):1275–1301, 2013. doi:10.1137/100811623.

42 Tony Metger, Alexander Poremba, Makrand Sinha, and Henry Yuen. Simple constructions of linear-depth t-designs and pseudorandom unitaries, 2024. doi:10.48550/arXiv.2404.12647.

43 A. Morvan et al. Phase transition in random circuit sampling, 2023. arXiv:2304.11119.

44 Ramis Movassagh. The hardness of random quantum circuits. *Nature Physics*, 19(11):1719–1724, November 2023. doi:10.1038/s41567-023-02131-2.

45 Adam Nahum, Jonathan Ruhman, Sagar Vijay, and Jeongwan Haah. Quantum entanglement growth under random unitary dynamics. *Physical Review X*, 7(3), July 2017. doi:10.1103/physrevx.7.031016.

46 Adam Nahum, Sagar Vijay, and Jeongwan Haah. Operator spreading in random unitary circuits. *Phys. Rev. X*, 8:021014, April 2018. doi:10.1103/PhysRevX.8.021014.

47 Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

48 Erdős Paul and Alfréd Rényi. On a classical problem of probability theory. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 6(1):215–220, 1961.

49 Lorenzo Piroli, Christoph Sünderhauf, and Xiao-Liang Qi. A random unitary circuit model for black hole evaporation. *Journal of High Energy Physics*, 2020(4), April 2020. doi:10.1007/jhep04(2020)063.

50 Thomas Schuster, Jonas Haferkamp, and Hsin-Yuan Huang. Random unitaries in extremely low depth, 2024. doi:10.48550/arXiv.2407.07754.

51 Henrik Wilming and Albert H. Werner. Lieb-robinson bounds imply locality of interactions. *Phys. Rev. B*, 105:125101, March 2022. doi:10.1103/PhysRevB.105.125101.

52 Lisa Yang and Netta Engelhardt. The complexity of learning (pseudo)random dynamics of black holes and other chaotic systems, 2023. arXiv:2302.11013.

53 Tianci Zhou, Shenglong Xu, Xiao Chen, Andrew Guo, and Brian Swingle. Operator lévy flight: Light cones in chaotic long-range interacting systems. *Phys. Rev. Lett.*, 124:180601, May 2020. doi:10.1103/PhysRevLett.124.180601.