Approximate Unitary t-Designs by Short Random Quantum Circuits Using Nearest-Neighbor and Long-Range Gates

Commun. Math. Phys.
Digital Object Identifier (DOI) https://doi.org/10.1007/s00220-023-04675-z

Communications in
Mathematical Physics

Approximate Unitary t-Designs by Short Random Quantum Circuits Using Nearest-Neighbor and Long-Range Gates

Aram W. Harrow[1], **Saeed Mehraban**[2]

[1] MIT Center for Theoretical Physics, Cambridge, Massachusetts, USA. E-mail: aram@mit.edu
[2] Tufts CS, Medford, Massachusetts, USA. E-mail: Saeed.Mehraban@tufts.edu

Received: 27 July 2021 / Accepted: 9 February 2023

Abstract: We prove that $\mathrm{poly}(t) \cdot n^{1/D}$-depth local random quantum circuits with two qudit nearest-neighbor gates on a D-dimensional lattice with n qudits are approximate t-designs in various measures. These include the "monomial" measure, meaning that the monomials of a random circuit from this family have expectation close to the value that would result from the Haar measure. Previously, the best bound was $\mathrm{poly}(t) \cdot n$ due to Brandão–Harrow–Horodecki (Commun Math Phys 346(2):397–434, 2016) for $D = 1$. We also improve the "scrambling" and "decoupling" bounds for spatially local random circuits due to Brown and Fawzi (Scrambling speed of random quantum circuits, 2012). One consequence of our result is that assuming the polynomial hierarchy (PH) is infinite and that certain counting problems are #P-hard "on average", sampling within total variation distance from these circuits is hard for classical computers. Previously, exact sampling from the outputs of even constant-depth quantum circuits was known to be hard for classical computers under these assumptions. However the standard strategy for extending this hardness result to approximate sampling requires the quantum circuits to have a property called "anti-concentration", meaning roughly that the output has near-maximal entropy. Unitary 2-designs have the desired anti-concentration property. Our result improves the required depth for this level of anti-concentration from linear depth to a sub-linear value, depending on the geometry of the interactions. This is relevant to a recent experiment by the Google Quantum AI group to perform such a sampling task with 53 qubits on a two-dimensional lattice (Arute in Nature 574(7779):505–510, 2019; Boixo et al. in Nate Phys 14(6):595–600, 2018) (and related experiments by USTC), and confirms their conjecture that $O(\sqrt{n})$ depth suffices for anti-concentration. The proof is based on a previous construction of t-designs by Brandão et al. (2016), an analysis of how approximate designs behave under composition, and an extension of the quasi-orthogonality of permutation operators developed by Brandão et al. (2016). Different versions of the approximate design condition correspond to different norms, and part of our contribution is to introduce the norm corresponding to anti-concentration and to establish equivalence between these various norms for low-depth circuits. For random

Published online: 04 May 2023

circuits with long-range gates, we use different methods to show that anti-concentration happens at circuit size $O(n \ln^2 n)$ corresponding to depth $O(\ln^3 n)$. We also show a lower bound of $\Omega(n \ln n)$ for the size of such circuit in this case. We also prove that anti-concentration is possible in depth $O(\ln n \ln \ln n)$ (size $O(n \ln n \ln \ln n)$) using a different model.

Contents

1. Introduction

Random unitaries are central resources in quantum information science. They appear in many applications including algorithms, cryptography, and communication. Moreover, they are important toy models for random chaotic systems, capturing phenomena like thermalization or scrambling of quantum information.

An idealized model of a random unitary is the uniform distribution over the unitary group, also known as the Haar measure. However, the Haar measure is an unrealistic model for large systems because the number of random coins and gates needed to generate an element of the Haar distribution scale exponentially with the size of the system (i.e. polynomially with dimension, meaning exponentially in the number of qubits or independent degrees of freedom). To resolve this dilemma, approximate t-designs have been proposed as physically and computationally realistic alternatives to the Haar measure. They approximate the behavior of the Haar measure if one only cares about up to the first t moments.

Several constructions of t-designs have been proposed based on either random or structured circuits. While structured circuits can in some cases be more efficient [22,25,48], random quantum circuits have other advantages. They are plausible models for chaotic random processes in nature, such as scrambling in black holes [17,55], or the spread of entanglement in condensed matter systems [46,47], growth of quantum complexity [12], and decoupling [18]. Moreover, they are practical candidates to benchmark computational advantage for quantum computation over classical models, since they seem to capture the power of a generic polynomial-size unitary circuit. Indeed, the Google quantum AI group has recently run a random unitary circuit on a 53-qubit superconducting device and has argued that this should be hard to simulate classically [5,8] (see Fig. 1 for a demonstration of their proposal). Here the random gates are useful not only for the 2-design property, specifically "anti-concentration", but also for evading the sort of structure which would lend itself to easy simulation, such as being made of Clifford gates.

All previous random circuit based constructions for t-designs required the circuits to have linear depth. In this paper, we show that certain random circuit models with small depth are approximate t-designs. We consider two models of random circuits. The first one is nearest-neighbor local interactions on a D-dimensional lattice. In this model, we apply random $\mathrm{U}(d^2)$ gates on neighboring qudits of a D-dimensional lattice in a certain order.

Depending on the application we want, we can define convergence to the Haar measure in different ways. For example, for scrambling [17] we measure convergence w.r.t. the norm $\mathbb{E}_C \|\rho_S(s) - \frac{1}{2^{|S|}}\|_1^2$, where $\rho_S(s)$ is the density matrix $\rho(s)$ reduced to a subset S of qudits and $\rho(s)$ is the quantum state that results from s steps of the random process. But for anti-concentration, which corresponds loosely to a claim that typical circuit outputs have nearly maximal entropy, we use a norm related to $\mathbb{E}_C \sum_x |\langle x|C|0\rangle|^4$. For other measures of convergence to the Haar measure see [42] or Sect. 2.4. In general, these measures are equivalent but moving between them involves factors that are exponentially large in the number of qudits, i.e., if one norm converges to ϵ the translation implies that another norm converges to $2^{O(n)}\epsilon$. Some of the known size/depth bounds for designs are of the form $O(f(n,t)(n+\ln 1/\epsilon))$ (e.g. [13]) and in 1-D simple arguments yield an $\Omega(n+\ln(1/\epsilon))$ lower bound [17]. In this case, replacing ϵ with $2^{-O(n)}\epsilon$ will not change the asymptotic scaling. [13] defined a strong notion of convergence which implies all the mentioned definitions.

However, in D dimensional lattices the natural lower bound is $\Omega(n^{1/D}+\ln(1/\epsilon))$. Our main challenge in this work is to show that this depth bound is asymptotically achievable, and along the way, we need to deal with the fact that we can no longer freely pay norm-conversion costs of $2^{O(n)}$. We are able to achieve the desired $\mathrm{poly}(t)(n^{1/D}+\ln(1/\epsilon))$ in many operationally relevant norms, but due in part to the difficulty of converting between norms, we do not establish it in all cases. The asymptotic dependency on t in

our result for $D = 2$ is $O(t \ln t)$ times the best asymptotic dependency on t for the $D = 1$ architecture, according to the strong measure defined in [13]. [13] gave a bound of $t^{10.5}$. Recently this bound was improved to $t^{5+o_t(1)}$ by Haferkamp [31]. The dependency on t in our result is hence $t^{6+o_t(1)} \ln t$.

Approximate unitary designs. We will consider several notions of approximate designs in this paper. First, we will introduce some notation. A degree-(t, t) monomial in $C \in \mathrm{U}((\mathbb{C}^d)^{\otimes n})$ is degree t in the entries of C and degree t in the entries of C^*. We can collect all these monomials into a single matrix of dimension d^{2nt} by defining $C^{\otimes t,t} := C^{\otimes t} \otimes C^{*\otimes t}$. We say that μ is an exact [unitary] t-design if expectations of all t, t moments of μ match those of the Haar measure. We can express this succinctly in terms of the operator

$$G_\mu^{(t)} = \mathop{\mathbb{E}}_{C\sim\mu} \left[C^{\otimes t} \otimes C^{*\otimes t}\right]. \tag{1}$$

Then μ is an exact t-design iff $G_\mu^{(t)} = G_{\mathrm{Haar}}^{(t)}$. Since $G_{\mathrm{Haar}}^{(t)}$ is a projector, we sometimes call $G_\mu^{(t)}$ a quasi-projector operator and we will later use the fact that it can sometimes be shown to be very close to a projector.

Most definitions of approximate designs demand that some norm of $G_\mu^{(t)} - G_{\mathrm{Haar}}^{(t)}$ be small. Three norms that we will consider are based on viewing $G_\mu^{(t)}$ as either a vector of length d^{4nt}, a matrix of dimension d^{2nt} or a quantum operation acting on a space of dimension d^{nt}. In each case, one can show that the t-design property implies the t'-design property for $1 \le t' \le t$.

Definition 1 (*Monomial definition of t-designs*). μ is a monomial-based ϵ-approximate t-design if any monomial has expectation within ϵd^{-nt} of that resulting from the Haar measure. In other words,

$$\left\| \mathsf{vec}\left[G_\mu^{(t)}\right] - \mathsf{vec}\left[G_\mu^{(t)}\right] \right\|_\infty \le \frac{\epsilon}{d^{nt}}. \tag{2}$$

$\mathsf{vec}(A)$ is a vector consisting of the elements of matrix A (in the computational basis) and $\|\cdot\|_\infty$ refers to the vector ℓ_∞ norm.

The monomial measure is natural when studying anti-concentration, since a sufficient condition for anti-concentration is that $\mathbb{E}_C |\langle 0|C|0\rangle|^4$ is close to the quantity that arises from the Haar measure, namely $\frac{2}{2^n(2^n+1)}$. This is achieved by [monomial measure] 2-designs.

If the operator-norm distance between $G_\mu^{(t)}$ and $G_{\mathrm{Haar}}^{(t)}$ is small then instead of calling μ an approximate design we call it a t-tensor product expander [36]. This controls the rate at which certain nonlinear (i.e. degree-t polynomial) functions of the state converge to the average value they would have under the Haar measure. We can also measure the distance between $G_\mu^{(t)}$ and $G_{\mathrm{Haar}}^{(t)}$ in the 1-norm (i.e. trace norm) and this notion of approximate designs has been considered before [4,54], although it does not have direct operational meaning. We will show $\mathrm{poly}(t)(n^{1/D} + \ln(1/\epsilon))$-depth convergence in each of these measures.

Finally, we can consider $G_\mu^{(t)}$ to be a superoperator using the following canonical map. Define $\mathrm{Ch}\left[\sum_i X_i \otimes Y_i^T\right]$ by $\mathrm{Ch}\left[\sum_i X_i \otimes Y_i^T\right](Z) := \sum_i X_i Z Y_i$. Thus

$$\mathrm{Ch}\left[G_\mu^{(t)}\right](Z) = \mathop{\mathbb{E}}_{C\sim\mu} \left[C^{\otimes t} Z C^{\dagger\otimes t}\right]. \tag{3}$$

Note that $\mathrm{Ch}\left[G_\mu\right]$ is completely positive and trace preserving, i.e., a quantum channel. For superoperators $\mathcal{M}, \mathcal{N}$ we say that $\mathcal{M} \preceq \mathcal{N}$ if $\mathcal{N} - \mathcal{M}$ is a completely positive (cp) map. Based on this ordering, a strong notion of being an approximate design was proposed by Andreas Winter and first appeared in [13].

Definition 2 (*Strong definition of t-designs*). A distribution μ is a strong ϵ-approximate t-design if

$$(1-\epsilon)\mathrm{Ch}\left[G_{\mathrm{Haar}}^{(t)}\right] \preceq \mathrm{Ch}\left[G_\mu^{(t)}\right] \preceq (1+\epsilon)\mathrm{Ch}\left[G_{\mathrm{Haar}}^{(t)}\right]. \tag{4}$$

Circuit models. The result of [13] constructs t-designs in the strong measure (Definition 2) for $D = 1$ and linear depth, and we generalize this result to construct weak monomial designs for arbitrary D and $O(n^{1/D})$ depth. We also show that the same construction converges to the Haar measure in other norms: diamond, infinity and trace norm. Our proof techniques do not seem to yield t-designs in the strong measure. We do not even know whether the construction of "strong" t-designs in sub-linear depth is possible.

The second model we consider is circuits with long-range two-qubit interactions. In this model, at each step, we pick a pair of qubits uniformly at random and apply a random U(4) gate on them. This model is the standard one when considering bounded-depth circuit classes, such as QNC. Physically, it could model chaotic systems with long-range interactions. Following Oliveira, Dahlsten and Plenio [51] (see also [17,18,34]), we can map the $t = 2$ moments of this process onto a simple random walk on the points $\{1, 2, \ldots, n\}$. We map this random walk to the classical (and exactly solvable) Ehrenfest model, meaning a random walk with a linear bias towards the origin. Further challenges are that this mapping introduces random and heavy-tailed delays and that the norm used for anti-concentration is exponentially sensitive to some of the probabilities. However, we are able to show (in Sect. 4) that after $O(n \ln^2 n)$ rounds of this process the resulting distribution over the unitary group converges to the Haar measure in the mentioned norm.

For a distribution p its collision probability is defined as $\mathrm{Coll}(p) = \sum_x p_x^2$. If $\mathrm{Coll}(p)$ is large ($\Omega(1)$) then the support of p is concentrated around a constant number of outcomes, and if it is small ($\approx 1/2^n$) then it is anti-concentrated. The norm that we consider for anti-concentration is basically the expected collision probability of the output distribution of a random circuit. The expected collision probability for the Haar measure is $\frac{2}{2^n+1}$ and our result shows that a typical circuit of size $O(n \ln^2 n)$ outputs a distribution with expected collision probability $\frac{2}{2^n}\left(1 + \frac{1}{\mathrm{poly}(n)}\right)$. Along with the Paley–Zygmund anti-concentration inequality this result proves that these circuits have the following anti-concentration property:

$$\min_x \Pr_{C\sim\mu}\left[\,|\langle x|C|0\rangle|^2 \geq \frac{1}{2^{n+1}}\right] \geq \text{constant}. \tag{5}$$

Here μ is the distribution of random circuits we consider, and x is any n-bit string. This bound is related to the hardness of classical simulation for random circuits. We furthermore show that sub-logarithmic depth quantum circuits in this model have expected collision probability $\frac{2}{2^n+1}\omega(1)$. The best anti-concentration depth bound we get from this model is $O(\ln^2 n)$. However, we are able to construct a natural family of random circuits with depth $O(\ln n \ln\ln n)$ that are anti-concentrated.

The organization of this paper. The rest of this introductory section states the basic results, ideas and implications related to the main results. In particular, in Sect. 1.1 we explain the connections between our result and the result experiments performed by groups such as the Google AI group aiming towards demonstrating the superiority of quantum computing compared with classical computers on specific tasks. In Sect. 1.2, we describe the models we consider in this paper. In Sect. 1.3, we express the main results of this paper including proof sketches and basic ideas. We then give a brief overview of the previous works related to this paper in Sect. 1.4 and several open questions in Sect. 1.5.

The organization of the rest of this paper is as follows. In Sect. 2 we introduce the preliminary concepts, definitions and tools needed for our proofs. In Sect. 3 we give detailed proofs about how we get approximate t-designs on D-dimensional lattices. In Sect. 4 we give detailed proofs related to anti-concentration bounds from circuits with all-to-all interactions. In Sect. 5 we provide alternative proofs for anti-concentration via low-depth D-dimensional lattices and in Sect. 6 we provide improvements on the existing scrambling and decoupling bounds. Appendix A gives a proof of Theorem 3 about the implications of anti-concentration bound we obtain on computational difficulty of simulating low-depth random quantum circuits. Finally Appendix B gives a background about the basic properties of Krawtchouk polynomials which we use in Sect. 4.

1.1. Connections with quantum computational supremacy experiments. Outperforming classical computers, even for an artificial problem such as sampling from the output of an ideal quantum circuit would be a significant milestone for early quantum computers which has recently been called "quantum computational supremacy" [35,52]. The reason to study quantum computational supremacy in its own right (as opposed to general quantum algorithms) is that it appears to be a distinctly easier task than full-scale quantum computing and even various non-universal forms [2,3,8,14,15,29] of quantum computing can be shown to be hard to simulate classically. For example, the outputs of constant-depth quantum circuits cannot be simulated exactly by classical computers unless the PH collapses [56]. In general, families of quantum circuits have this property if they are universal under postselection, meaning that after measuring all the qubits at the end of the circuit and producing a string of bits, we condition on the values of some of these bits and use the other bits for the output.

However, these hardness results are not robust under noise and error in measurements. A central open question in the theory of quantum computational supremacy is whether simulating these distributions to within constant or $1/\operatorname{poly}(n)$ variational distance would still be hard. It is plausible to conjecture that if such a robust hardness of sampling is true, it would also hold for generic circuits [1,3] (although see [49] for a counterexample). A standard approach to proving such a robust hardness result for generic circuits has been to prove that "anti-concentration" holds, and to use this to relate additive error approximation to average-case relative error approximation; see e.g. [16]. Here "anti-concentration" means having near-maximal entropy in the output of a quantum circuit, which implies that any fixed amplitude of a quantum circuit is likely to be $\geq \frac{\Omega(1)}{2^n}$. This property implies that the complexity of estimating the amplitudes additively (within $\pm\frac{1}{\operatorname{poly}(n)\cdot 2^n}$) is on-average as hard as computing them within inverse polynomial relative error. This lets us turn an assumption about the average-case hardness of relative-error approximation of the amplitudes into a hardness result for the

sampling problems. Approximate t-designs (and even approximate 2-designs) have the desired "anti-concentration" property.

For experimental verification of quantum computational supremacy we can consider the following sampling task: let μ be a distribution over random circuits that satisfies

$$\Pr_{C\sim\mu}\left[\,|\,\langle 0|C|0\rangle\,|^2 \geq \frac{1}{2^{n+1}}\right] \geq 1/8 - 1/\operatorname{poly}(n). \tag{6}$$

(which we call the anti-concentration property). Let $\mathcal{C}_x$ be the family of circuits constructed by first applying a circuit $C \sim \mu$ and then an X gate to each qubit with probability $1/2$ (and identity with probability $1/2$). A similar line of reasoning as in Bremner-Montanaro-Shepherd (see Theorem 6 and 7 of [16]) implies that

Theorem 3. *Fix $\epsilon > 0$ and $0 < \delta < 1/8$. Let μ be a $\frac{1}{\operatorname{poly}(n)}$-approximate 2-design. If there exists a* BPP *machine which takes $C \sim \mu$ as input and for at least $1-\delta$ fraction of such inputs outputs a probability distribution that is within ϵ total variation distance from the probability distribution $p_x = |\,\langle x|C|0\rangle\,|^2$, then there exists an* $\mathsf{FBPP}^{\mathsf{NP}}$ *algorithm that succeeds with probability $1-\delta$ and computes the value $|\,\langle 0|C'|0\rangle\,|^2$ within multiplicative error $\frac{2(\epsilon+1/\operatorname{poly}(n))}{\delta}$ for $1/8 - \frac{1}{\operatorname{poly}(n)}$ fraction of circuits $C' \sim \mathcal{C}_x$.*

This theorem is proved in Appendix A. If we further conjecture the PH is infinite and that amplitudes of the random circuits in Theorem 3 are #P-hard to approximate on average, then this implies that classical computers cannot efficiently sample from any distribution close to the ones generated by these circuits. At the moment, it is only known that nearly exact computation of these amplitudes is hard for #P [10,44,45]. It is an open question whether average-case hardness for the approximation task remains #P-hard.

The linear to sub-linear improvement of the depth required for anti-concentration provided in this paper is likely to be significant for near-term quantum computers that will be constrained both in terms of the number of qubits (n) and noise rate per gate (δ). Due to the constraints in the number of qubits (say 50-100), quantum computational supremacy will only be possible without the overhead of error correction, since even the most efficient known schemes for fault-tolerant quantum computation reduce the number of qubits by more than a factor of two [21]. Thus a quantum circuit with S gates will have an expected $S\delta$ errors. Recent work due to Yung and Gao [57] and the Google group [9] states that noisy random quantum circuits with $O(\ln n)$ random errors output distributions that are nearly uniform, and thus are trivially classically simulable. Thus S can be at most $\ln(n)/\delta$. In proposed near-term quantum devices [6,8,26,50] we can expect $n \sim 10^2$ and $\delta \sim 10^{-2}$. Thus the $S = O(n \ln^2 n)$ for long-range interactions or $S = O(n\sqrt{n})$ bound for 2-D lattices from our work is much closer to being practical than the previous $S = O(n^2)$. (This assumes that the constants are reasonable. We have not made an effort to calculate them rigorously but for the case of long-range interactions we do present a heuristic that suggests that in fact $\approx \frac{5}{6} n \ln n$ gates are necessary and sufficient.)

1.2. Our models. We consider two models of random quantum circuits. The first involves nearest-neighbor local interactions on a D-dimensional lattice and the second involves long-range random two-qubit gates. The order of gates in the first model has some structure but in the second model it is chosen at random. Hence, we can view the second model as the natural dynamics of an n-qubit system, connected as a complete graph.

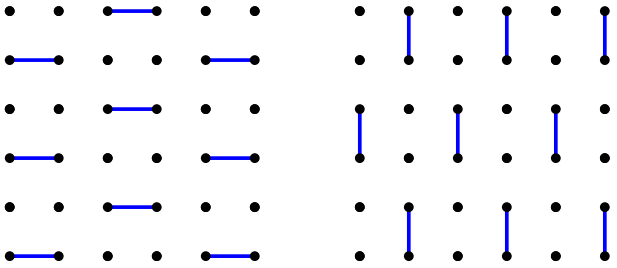

Fig. 1. The architecture proposed by the quantum AI group at Google to demonstrate quantum supremacy consists of a 2D lattice of superconducting qubits. This figure depicts two illustrative timesteps in this proposal. At each timestep, 2-qubit gates (blue) are applied across some pairs of neighboring qubits

We first define the following random circuit model for $D = 1$ which was also considered in [13]:

Definition 4 (*Random circuits on one-dimensional lattices*). $\mu_{1,s}^{\text{lattice},n}$ is the distribution over unitary circuits resulting from the following random process.

For $j = 1 : s$ % for t-designs, view s as poly$(t)n$

- Apply independent random gates from $\mathrm{U}(d^2)$ on qudits $(1, 2), (3, 4), \ldots, (n-1, n)$.
- Apply independent random gates from $\mathrm{U}(d^2)$ on qudits $(2, 3), (4, 5), \ldots, (n-2, n-1)$.

This definition assumes that n is even but we modify it in the obvious way when n is odd. Another modification which would not change our results would be to put the qudits on a ring so that sites n and 1 are connected.

Building on this, we define the following distribution of random circuits on a two-dimensional lattice.

Definition 5 (*Random circuits on two-dimensional lattices*). Consider a two-dimensional lattice with n qudits. Let $r_{\alpha,i}$ be the i$^{\text{th}}$ row of the lattice in direction $\alpha \in \{1, 2\}$, for $1 \le i \le \sqrt{n}$. For each $\alpha \in \{1, 2\}$ let SampleAllRows(α) denote the following procedure (see Fig. 2):

For each $i \in [\sqrt{n}]$, sample a random circuit from $\mu_{1,s}^{\text{lattice},\sqrt{n}}$ and apply it to $r_{\alpha,i}$.

Now define $\mu_{2,c,s}^{\text{lattice},n}$ to be the distribution over unitary circuits resulting from the following random process:

- Repeat these steps c times: apply SampleAllRows(1) and then SampleAllRows(2).
- Apply SampleAllRows(1) a final time.

This distribution has depth $(2c + 1)2s$ and is related but not identical to the Google AI group's experiment [5,8], see Fig. 1. For our results on t-designs, we will take c to be

poly(t) and s to be poly$(t) \cdot \sqrt{n}$. We believe that our result can be extended to any natural family of circuits with nearest-neighbor interactions. We also assume for convenience that $\sqrt{n}$ is an integer, but believe that this assumption is not fundamentally necessary.

Next, we give a recursive definition for our random circuits model on arbitrary D-dimensional lattices. We view a D-dimensional lattice as a collection of $n^{1/D}$ sub-lattices of size $n^{1-1/D}$, labeled as $\xi_1, \ldots, \xi_{n^{1-1/D}}$. We label the rows of the lattice in the D-th direction by $r_1, \ldots, r_{n^{1/D}}$.

Definition 6 (*Random circuits on D-dimensional lattices*). $\mu_{D,c,s}^{\text{lattice},n}$ is the distribution resulting from the following random process.

1. Repeat these steps c times.
   (a) For each $i \in [n^{1/D}]$,
   - Sample a random circuit from $\mu_{D-1,c,s}^{\text{lattice},n^{1-1/D}}$ and apply it to ξ_i.

   (b) For each $j \in [n^{1-1/D}]$
   - Sample a random circuit from $\mu_{1,s}^{\text{lattice},n^{1/D}}$ and apply it to r_j.
2. For each $i \in [n^{1/D}]$,
   (a) Sample a random circuit from $\mu_{D-1,c,s}^{\text{lattice},n^{1-1/D}}$ and apply it to ξ_i.

Next, we define the model with long-range interactions on a complete graph.

Definition 7 (*Random circuit models on complete graphs*). μ_s^{CG} is the distribution over unitary circuits resulting from the following random process.

Repeat this step s times % view s as $O(n \ln^2 n)$.

- Pick a random pair of qudits (i, j) and apply a random $\mathrm{U}(d^2)$ gate between them.

The size of the circuits in this ensemble is s.

1.3. Our results. Our first result is the following.

Theorem 8. *Let $s, c, n > 0$ be positive integers with $\mu_{2,c,s}^{\text{lattice},n}$ defined as in Definition 5.*

1. $s = \text{poly}(t)\left(\sqrt{n} + \ln \frac{1}{\delta}\right)$, $c = O\left(t \ln t + \frac{\ln(1/\delta)}{\sqrt{n}}\right) \implies \left\| \text{vec}\left[G^{(t)}_{\mu^{\text{lattice},n}_{2,c,s}} - G^{(t)}_{Haar} \right] \right\|_\infty \leq \frac{\delta}{d^{nt}}$.
2. $s = \text{poly}(t)\left(\sqrt{n} + \ln \frac{1}{\delta}\right)$, $c = O\left(t \ln t + \frac{\ln(1/\delta)}{\sqrt{n}}\right) \implies \left\| Ch\left[G^{(t)}_{\mu^{\text{lattice},n}_{2,c,s}} - G^{(t)}_{Haar} \right] \right\|_\diamond \leq \delta$.
3. $s = \text{poly}(t)\left(\sqrt{n} + \ln \frac{1}{\delta}\right)$, $c = O\left(t \ln t + \frac{\ln(1/\delta)}{\sqrt{n}}\right) \implies \left\| G^{(t)}_{\mu^{\text{lattice},n}_{2,c,s}} - G^{(t)}_{Haar} \right\|_1 \leq \delta$.
4. $\left\| G^{(t)}_{\mu^{\text{lattice},n}_{2,c,s}} - G^{(t)}_{Haar} \right\|_\infty \leq c \cdot \sqrt{n} \cdot e^{-s/\text{poly}(t)} + \frac{1}{d^{O(c\sqrt{n})}}$.

The three norms in the above theorem refer to the vector ℓ_∞ norm, the superoperator diamond norm $\|\cdot\|_\diamond$ (see Sect. 2.1) and the operator S_∞ norm, also known simply as the operator norm.

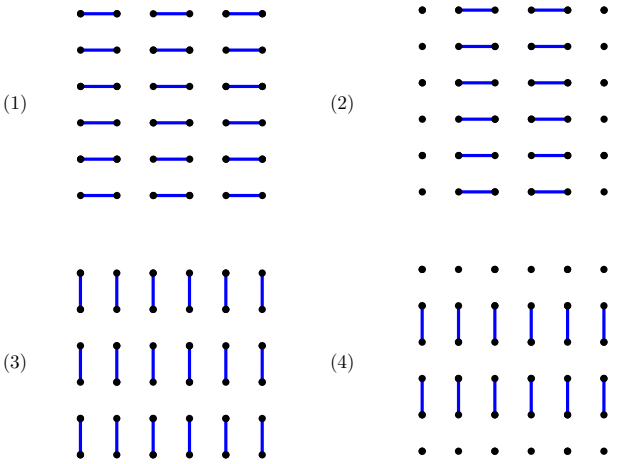

Fig. 2. The random circuit model in definition 5. Each black circle is a qudit and each blue link is a random $\mathrm{SU}(d^2)$ gate. The model does $O(\sqrt{n}\,\mathrm{poly}(t))$ rounds alternating between applying (1) and (2). Then for $O(\sqrt{n}\,\mathrm{poly}(t))$ rounds it alternates between (3) and then (4). This entire loop is then repeated $O(\mathrm{poly}(t))$ times

Proof sketch for part 1. We first give a brief overview of the proof in [13] and explain why their construction requires a circuit to have linear depth. Let $G_{i,i+1}$ be the projector operator for a random two-qudit gate applied to qudits i and $i+1$, and let $G = \frac{1}{n-1}\sum_i G_{i,i+1}$. Therefore $G_s = G^s$ is the quasi-projector corresponding to a 1-D random circuit with size s. [13] observed that $G - G_{\mathrm{Haar}}$ corresponds to a certain local Hamiltonian and $\epsilon = 1 - \|G - G_{\mathrm{Haar}}\|_\infty$ is its spectral gap. The central technical result of [13] is the bound $\epsilon \geq \frac{1}{n\cdot\mathrm{poly}(t)}$. As a result, $\|G_s - G_{\mathrm{Haar}}\|_\infty = (1 - \frac{1}{n\cdot\mathrm{poly}(t)})^s$. In general $G - G_{\mathrm{Haar}}$ has rank $e^{O(n)}$, and in order to construct a strong approximate t-design (Definition 2), one needs to apply a sequence of expensive changes of norm that lose factors polynomial in the overall dimension of G, i.e., $e^{O(nt)}$. Thereby in order to compensate for such exponentially large factors one needs to choose $s = O(n^2\cdot\mathrm{poly}(t))$, meaning depth growing linearly with n. Brown and Fawzi [17] furthermore observed that if G is the projector corresponding to one step of a random circuit on a 2-D lattice, the spectral gap still remains $1 - \|G - G_{\mathrm{Haar}}\|_\infty = O\big(\frac{1}{n\cdot\mathrm{poly}(t)}\big)$, and using the same proof strategy one needs linear depth.

The new ingredient we contribute is to show that if $s = O(\sqrt{n})$ one can replace $G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,1,s}}$ with a certain quasi-projector G', such that

(1) $G' - G_{\mathrm{Haar}}$ has rank $t!^{O(\sqrt{n})}$ and
(2) $\|G' - G_{\mathrm{Haar}}\|_\infty \approx 1/e^{\Omega(\sqrt{n})}$,
(3) $G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,1,s}} \approx G'$ in various norms.

We first use (1) to relate the monomials definition of t-designs to the infinity norm and then use (2) to bound the infinity norm

$$\left\|\mathsf{vec}\left[G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}}\right] - \mathsf{vec}\left[G^{(t)}_{\mathrm{Haar}}\right]\right\|_\infty \approx t!^{O(\sqrt{n})}\left\|G'^c - G^{(t)}_{\mathrm{Haar}}\right\|_\infty \cdot \frac{t!}{d^{nt}} \approx \frac{t!^{O(\sqrt{n})}}{e^{\Omega(c\cdot\sqrt{n})}}\cdot\frac{1}{d^{nt}}. \tag{7}$$

For $c = t \ln t$ the error bound is $1/e^{\Omega(\sqrt{n})} \frac{1}{d^{nt}}$. As a result using (3)

$$\left\| \mathsf{vec}\left[G^{(t)}_{\mu^{\text{lattice},n}_{2,c,s}} \right] - \mathsf{vec}\left[G^{(t)}_{\text{Haar}} \right] \right\|_\infty \approx \left\| \mathsf{vec}\left[G^{\prime c} \right] - \mathsf{vec}\left[G^{(t)}_{\text{Haar}} \right] \right\|_\infty \approx 1/e^{\Omega(\sqrt{n})} \cdot \frac{1}{d^{nt}}. \tag{8}$$

This step requires a certain change of norm for which we only have to pay a factor like $e^{O(\sqrt{n})}$, which we justify by bounding the ranks of the right intermediate operators. The factor of $1/d^{nt}$ comes from the fact that the Haar measure itself has monomial expectation values on this order (in fact as large as $t!/d^{nt}$ but we suppressing the t-dependence in this proof sketch.)

We now briefly describe the construction of G'. Let G_R (and G_C) be the projector operators corresponding to applying a Haar unitary to each row (and column) independently. Then $G' = G_R G_C$. G' has rank $t!^{O(\sqrt{n})}$ because G_R and G_C are each tensor products of $\sqrt{n}$ Haar projectors each with rank $t!$. Let V_R, V_C, and V_{Haar} be respectively the subspaces that G_R, G_C and G_{Haar} project onto. In order to prove (1) in Sect. 3.6.1 we first use the fact that our circuits are computationally universal to argue that $V_C \cap V_R = V_{\text{Haar}}$. We then prove that the angle between $V_R \cap V_{\text{Haar}}^\perp$ and $V_C \cap V_{\text{Haar}}^\perp$ is very close to $\pi/2$, i.e., $\approx \pi/2 \pm \frac{1}{d^{\sqrt{n}}}$. This implies that $G_C G_R = G_{\text{Haar}} + P$, where P is a small matrix in the sense that $\|P\|_\infty \approx 1/d^{\sqrt{n}}$. Choosing $c = \text{poly}(t)$ we obtain (1). To show (2) it is not hard to see that the rank of $G' - G_{\text{Haar}}$ is indeed $e^{O(\sqrt{n})}$. For (3) we use the construction of t-designs from [13]. In particular, our random circuits model first applies an $O(\sqrt{n})$ depth circuit to each row and then an $O(\sqrt{n})$ depth circuit to each column and repeats this for $\text{poly}(t)$ rounds. The result [13] implies that each of these rounds is effectively the same as applying a strong approximate t-design to the rows or columns of the lattice. We then analyze how these designs behave under composition in various norms and prove (3). □

Our second result generalizes Theorem 8 to arbitrary dimensions.

Theorem 9. *There exists a value $\delta = 1/d^{\Omega(n^{1/D})}$ such that for some large enough c depending on D and t:*

$$1.\ s > c \cdot n^{1/D} \implies \left\| \mathsf{vec}\left[G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}} - G^{(t)}_{Haar} \right] \right\|_\infty \leq \tfrac{\delta}{d^{nt}}.$$

$$2.\ s > c \cdot n^{1/D} \implies \left\| Ch\left[G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}} - G^{(t)}_{Haar} \right] \right\|_\diamond \leq \delta.$$

$$3.\ s > c \cdot n^{1/D} \implies \left\| G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}} - G^{(t)}_{Haar} \right\|_\infty \leq \delta.$$

$$4.\ s > c \cdot n^{1/D} \implies \left\| G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}} - G^{(t)}_{Haar} \right\|_1 \leq \delta.$$

In order to understand the implication of this result for anti-concentration, let's first define

Definition 10 (*Anti-concentration*). We say a family of circuits μ satisfy the (α, β) anti-concentration property if for any $x \in \{0,1\}^n$

$$\Pr_{U \sim \mu}\left[|\langle x|U|0\rangle|^2 \geq \frac{\alpha}{2^n} \right] \geq \beta \tag{9}$$

As mentioned before, unitary 2-designs imply strong anti-concentration bound. In particular

Theorem 11. *Let μ be a an ϵ-approximate 2-design in the monomial measure. Then μ satisfies the (α, β) anti-concentration property for $\alpha = \delta(1-\epsilon)$, $\beta = \frac{(1-\delta)^2(1-\epsilon)^2}{2(1+\epsilon)}$ and $0 \leq \delta \leq 1$.*

Proof. (See Appendix A and also Theorem 5 of [32]). The proof is based on the Paley–Zigmond anti-concentration inequality: for a non-negative random variable X and $\delta > 0$ we have

$$\Pr[X \geq \delta \cdot \mathbb{E}X] \geq (1-\delta)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}. \tag{10}$$

□

We remark that based on the result of [24], while sufficient, the 2-design property is not necessary for anti-concentration. In Sect. 5 we give an alternative proof for anti-concentration in $O(D \cdot n^{\frac{1}{D}})$ depth based on different ideas. The method directly implies anti-concentration and not the approximate 2-design property.

For these spatially local circuits we also improve on some bounds in [17] and [18] about scrambling and decoupling, removing polylogarithmic factors. Here we give an informal statement of the result with full details and definitions found in Sect. 6.

Theorem 12 (Informal). *Random quantum circuits acting on D-dimensional lattices composed of n qubits are scramblers and decoupler in the sense of [17] and [18] after $O(D \cdot n^{1/D})$ number of steps.*

Our last result concerns the fully connected model. If $s = O(n \ln^2 n)$ and $d = 2$ then μ_s^{CG} satisfies the anti-concentration criterion according to Definition 10 for constant α and β, i.e., (5). We phrase our result in terms of the expected "collision probability" of the output distribution of $C \sim \mu_s^{\mathrm{CG}}$ from which a bound similar to the one in theorem 11 will follow using the Paley–Zygmond inequality (10). In particular, if C is a quantum circuit on n qubits, starting from $|0^n\rangle$ the collision probability is

$$\mathrm{Coll}(C) := \sum_{x \in \{0,1\}^n} |\langle x|C|0\rangle|^4. \tag{11}$$

For the Haar measure $\mathbb{E}_{C \sim \mathrm{Haar}} \mathrm{Coll}(C) = \frac{2}{2^n+1}$, and for the uniform distribution this value is $1/2^n$. In contrast, a depth-1 random circuit has expected collision probability $(\sqrt{\frac{2}{5}})^n$, which is exponentially larger than what we expect from the Haar measure.

Theorem 13. *There exists a c such that when $s > cn \ln^2 n$,*

$$\mathop{\mathbb{E}}_{C \sim \mu_s^{CG}} Coll(C) \leq \frac{29}{2^n}. \tag{12}$$

Moreover if $t \leq \frac{1}{3c'} n \ln n$ for some large enough c', then

$$\mathop{\mathbb{E}}_{C \sim \mu_s^{CG}} Coll(C) \geq \frac{1.6^{n^{1-1/c'}}}{2^n}. \tag{13}$$

Proof Sketch. For the upper bound, we translate the convergence time of the expected collision probability to the mixing time of a certain classical Markov chain (which we call $X_0, X_1, \ldots$). This Markov chain has also been considered in previous work [18,34,51]. Part of our contribution is to analyze this Markov chain in a new norm. The Markov chain has n sites labeled as $1, \ldots, n$, and at each site x it will move only to $x-1$, x or $x+1$. Such chains are known as "birth and death" chains, and in our case it results from representing the state of the system by a Pauli operator and then taking x to be the Hamming weight of that Pauli operator. It is known [51] that the probability of moving to site $x+1$ is $\approx \frac{6}{5}\frac{x(n-x)}{n^2}$ and the probability of moving to site $x-1$ is $\approx \frac{2}{5}\frac{x(x-1)}{n^2}$. The major difficulty in proving mixing for this Markov chain is that the norm which we have to prove mixing in is exponentially sensitive to small fluctuations (measured in either the 1-norm or the 2-norm). Indeed, given starting condition

$$\Pr[X_0 = k] = \frac{\binom{n}{k}}{2^n - 1}. \tag{14}$$

we would like to show that

$$\mathop{\mathbb{E}}_{C}[\mathrm{Coll}(C)] \approx \sum_{k=1}^{n} \frac{\Pr[X_t = k]}{3^k}, \tag{15}$$

is $\leq O(2^{-n})$. We can think of (15) as a weighted 1-norm on probability distributions.

Our proof will compute the distribution of X_t for $t = O(n \ln^2 n)$ nearly exactly. One distinctive feature of this chain is that when $k/n \ll 1$, the probability of moving is $O(k/n)$ and the chain is strongly biased to move towards the right. When k/n reaches $O(n)$, the chain becomes more like the standard discrete Ehrenfest chain, which is a random walk with a linear bias towards (in this case) $k = \frac{3}{4}n$. Thus the small-k region needs to be handled separately. This is especially true for anti-concentration thanks to the $1/3^k$ term in (15), so that even a small probability of waiting for a long time in this region can have a large effect on the collision probability.

The approach of [18,27,34] has been to relate the original Markov chain to an "accelerated" chain which is conditioned on moving at each step. The status of the original chain can be recovered from the accelerated chain by adding a geometrically distributed "wait time" at each step. Then standard tools from the analysis of Markov chains, such as comparison theorems and log-Sobolev inequalities, can be used to bound the convergence rate of the accelerated chain. Finally, it can be related back to the original chain by arguing that the accelerated chain is unlikely to spend too long on small values of k, allowing us to bound the wait time. For our purposes, this process does not produce sharp enough bounds, due to the heavy-tailed wait times combined with fairly weak bounds on how quickly the accelerated chain converges and leaves the small-k region.

We will sharpen this approach by incompletely accelerating; i.e., we will couple the original chain to a chain that moves with a carefully chosen (but always $\Omega(1)$) probability. In particular, we will introduce a chain where the probabilities of moving from x to $x-1$, x or $x+1$ are each affine functions of x. In fact our new "accelerated" chain is only accelerated for $x < \frac{5}{6}n$ and is actually more likely to stand still for $x \geq \frac{5}{6}n$. This will allow us to exactly solve for the probability distribution of the accelerated chain after any number of steps, using a method of Kac to relate this distribution to the solution of a differential equation. Our solution can be expressed simply in terms of Krawtchouk polynomials, which have appeared in other exact solutions to random processes on the hypercube. We relate this back to the original chain with careful estimates of the mean

and large-deviation properties of the wait time. This ends up showing only that the collision probability is small for t in some interval $[t_1, t_2]$, and to show that it is small for a specific time, we need to prove that the collision probability decreases monotonically when we start in the state $|0^n\rangle$. A further subtlety is that (15) technically only applies when all qubits have been hit by gates and we need to extend this analysis to include the non-negligible probability that some qubits have never been acted on by a gate.

Because previous work achieved quantitatively less sharp bounds, they could omit some of these steps. For example, [27,34] used $O(n^2)$ gates, which meant that the probability of most bad events was exponentially small. By contrast, in depth $O(n \ln^2(n))$, there is probability $n^{-O(\ln n)}$ of missing at least one qubit and so we cannot afford to let this be an additive correction to our target collision probability of constant $\cdot\, 2^{-n}$. Likewise, [18] used only $O(n \ln^2(n))$ gates but achieved a collision probability of $2^{\epsilon n - n}$ for small constant ϵ, which allowed them to use a simpler version of the accelerated chain whose convergence they bounded using generic tools from the theory of Markov chains.

For the lower bound we just consider the event that the initial Hamming weight does not change throughout the process. The initial state with Hamming weight k has probability mass $\Pr[X_0 = k] = \frac{\binom{n}{k}}{2^n - 1}$. Starting with Hamming weight k, the probability of not moving in each step is $e^{-O(k/n)}$, so if $t = cn \ln n$ for $c \ll 1$ then we have $\Pr[X_t = k | X_0 = k] \geq e^{-O(kt/n)}$. Hence

$$\begin{aligned} \mathop{\mathbb{E}}_{C \sim \mu_t} \mathrm{Coll}(C) &\geq \sum_{k=1}^{n} \frac{\binom{n}{k}}{2^n - 1} \frac{\Pr[X_t = k | X_0 = k]}{3^k} \geq \sum_{k=1}^{n} \frac{\binom{n}{k}}{2^n - 1} \frac{e^{-O(kt/n)}}{3^k} \\ &\approx \frac{1}{2^n} (1 + e^{-3t/n})^n \geq \frac{2^{n^{1-O(1)}}}{2^n} \end{aligned} \tag{16}$$

□

A natural question is whether there is a common generalization of our Theorems 9 and 13. In physics, the $D \to \infty$ limit is often considered a good proxy for the fully connected model. This raises the question of whether we needed Theorem 13 to handle the fully connected case, or whether it would be enough to use Theorem 9 in the large D limit. However, Theorem 9 works only for $D = O(\ln n / \ln \ln n)$, and the best depth bound we can get from this theorem is $e^{O(\ln n / \ln \ln n)}$, which is far above the $O(\ln^2(n))$ achievable by Theorem 9. However, in Sect. 5 we give an alternative proof for anti-concentration of outputs via circuits on D-dimensional circuits with $t = 2$ and $D = O(\ln n)$. Using that approach we can make the depth as small as $O(\ln n \ln \ln n)$. We conjecture that $O(\ln n)$ depth should also possible.

In order to establish rigorous bounds, our results involve some inequalities that are not always tight. As a result, the upper bound on collision probability in Theorem 13 has a factor of 29 rather than the $2 + o(1)$ that we would expect and the bound on the number of gates required may be too high by a factor of $\ln(n)$. Since determining the precise number of gates needed for anti-concentration may have utility in near-term quantum hardware, we also undertake a heuristic analysis of what depth seems to be required to achieve anti-concentration. Here we ignore the possibility of large fluctuations in the wait time, for example, and simply set it equal to its expected value. We also freely make the continuum approximation for the biased random walk that ignores wait time, obtaining the Ornstein–Uhlenbeck process. The resulting analysis (found in Sect. 4.6) suggests

that $\frac{5}{6}n \ln n + o(n \ln n)$ gates are needed to achieve anti-concentration comparable to the Haar measure.

This result can also be useful for understanding the near-term power of certain variational quantum algorithms, such as VQE and QAOA. [20,43] show that when a gate sequence is drawn from a 2-design, the gradients used for optimizing VQE and other algorithms become exponentially small. This is called the "barren plateau" phenomenon. Our result would suggest that this occurs in 2-D circuits once the depth is $\gtrsim \sqrt{n}$.

1.4. Previous work. The time evolution of the 2nd moments of random quantum circuits was first studied by Oliveira, Dahlsten and Plenio [51], who investigated their entanglement properties. This was extended by [27,34] to show that after linear depth, arndom circuits on the complete graph yield approximate 2-designs. In [13] Brandão-Harrow-Horodecki (BHH) extended this result and showed that for a $1D$-lattice after depth $t^{10.5} \cdot O(n + \ln \frac{1}{\epsilon})$ these random quantum circuits become ϵ-approximate t-designs. This result was subsequently improved to $t^{5+o_t(1)} O(n + \ln \frac{1}{\epsilon})$ by Haferkamp [31]. All of these results (except [51]) directly imply anti-concentration after the mentioned depths. The construction of t-designs in [13] is in a stronger measure than the one in HL [34]. The gap of the second-moment operator was calculated exactly for $D = 1$ and fully connected circuits by Žnidarič [58] and a heuristic estimate for the t^{th} moment operator was given by Brown and Viola for fully connected circuits [19].

In [17,18] Brown and Fawzi considered "scrambling" and "decoupling" with random quantum circuits. In particular, they showed for a D-dimensional lattice scrambling occurs in depth $O(n^{1/D} \operatorname{polylog}(n))$, and for complete graphs, they showed that after polylogarithmic depth these circuits demonstrate both decoupling and scrambling. For the case of D-dimensional lattices they showed that for the Markov chain K, after depth $n^{1/D} \operatorname{polylog}(n)$, a string of Hamming weight 1 gets mapped to a string with linear Hamming weight with probability $1 - 1/\operatorname{poly}(n)$. While this result is related to ours, it does not seem to yield the results we need e.g. for anti-concentration, due to the powers of Hilbert space dimension that are lost when changing norms.

In [46,47] Nahum, Ruhman, Vijay and Haah considered operator spreading for random quantum circuits on D-dimensional lattices. They considered the case when a single Pauli operator starts from a certain point on the lattice and they analyze the probability that after a certain time a non-identity Pauli operator appears at an arbitrary point on the lattice. For $D = 1$ they showed that this probability function satisfies a biased diffusion equation. Their result in this case is exact. For $D = 2$ they explained, both numerically and theoretically, that this probability function spreads as an almost circular wave whose front satisfies the one dimensional Kardar-Parisi-Zhang equation. They moreover explained: 1) the bulk of the occupied region is in equilibrium, 2) fluctuations appear at the boundary of this region with $\sim t^{1/3}$, and 3) the area of the occupied region grows like t^2, where t is the depth of the circuit. As far as we understand this result does not directly lead to the construction of t-designs and rigorous bounds on the quality of the approximations made in that paper are not known.

If we assume that qudits have infinite local dimension ($d \to \infty$) then the evolution of Pauli strings on a 2-D lattice is closely related to Eden's model [28]. Here, Eden has found certain explicit solutions. However, apart from the $d \to \infty$ limit, his model differs from ours also in that his considers only starting with a single occupied site and running for a time much less than the graph diameter (or equivalently, considering an infinitely large 2-D lattice), while we consider the initial distribution obtained by starting in the $|0^n\rangle$ state.

After the first preprint version of this paper was posted online, [24] improved on our results in several ways. Unlike what we expected, they proved that random quantum circuits acting on linear chains or complete graphs anti-concentrated after depth $\Theta(\ln n)$. It is left as an important open question whether the same bound holds for $D = 2, 3, \ldots$. They also proved one of the conjectures of this paper that the constant factor for the depth bound for the complete graph model is $5/6$. The initial presentation of this result had a mistake in the heuristic reasoning and predicted the constant factor to be $5/3$. This was pointed out and corrected in [24].

1.5. Open questions.

1. Is it possible to construct "strong" t-designs (Definition 2) using sub-linear depth random circuits? If we can show that the off-diagonal moments (see Definition 36) of the distribution, which have expectation zero according to the Haar measure, become smaller than $1/d^{3nt}$ in sub-linear depth, then our construction of monomial designs implies the construction of strong designs. On the other hand, we cannot rule out the possibility that strong designs require linear depth.
2. How large are the constant factors in bounds reported in this paper? Based on a heuristic argument in Sect. 4.6 for the complete graph architecture we conjecture that such random circuits of size $s = \frac{5}{6}n(\ln n + \epsilon)$ are $O(\epsilon)$-approximate 2-designs. See Conjecture 1 for a precise conjectured bound for obtaining 2-designs. In work appearing after the first version of our paper, Ref. [24] proved this conjecture for anti-concentration. Our result had achieved an upper-bound of $O(n \ln^2 n)$.
3. We believe our dependence on n is essentially optimal. But our depth scales with t as t^α for some $\alpha \gtrsim 5$ that is almost certainly not optimal. At the moment the best lower bound is $\Omega(t \ln n)$ depth for any circuit, or $\Omega(n^{1/D})$ in D dimensions. Indeed, very recently [37] provided strong analytical evidence that for the one-dimensional architecture, $\alpha = 1$ for $D = 1$. The argument, however, contains uncontrolled approximations and is not known to extend even to $D = 2$, although such an extension seems plausible. Intriguingly, also for constant n and with a different gate model, some results are known that are completely independent of t [11].
4. If we pick an arbitrary graph and apply random gates on the edges of this graph, after what depth do these circuits become t-designs? We conjecture that if the graph has large expansion and diameter l, then the answer is $O(l)$. However, if the graph has a tight bottleneck (like a binary tree), then even though the graph has small diameter, we suspect that certain measures of t-designs (including the monomial measure) require linear depth. Ideally, the t-design time for any graph could be related to other properties of the graph such as mixing time, cover time, etc.
5. Can we prove a comparison lemma for random circuits, i.e., can we show that if two random circuits are close to each other, then they become t-designs after roughly the same amount of time? Such comparison lemma may imply that other natural families of low-depth circuits are approximate t-designs. A related question is whether deleting random gates from a circuit family can ever speed up convergence to being a t-design. Such a bound has been called a "censoring" inequality in the Markov-chain literature.
6. Our results do not say much about the actual constants that appear in the asymptotic bounds for the required size for anti-concentration. We conjecture the leading term in the anti-concentration time for random circuits on complete graphs is $\frac{5}{6}n \ln n$.

For the D-dimensional case our bounds inherit constant factors from [13]. Simple numerical simulation and also the analysis of [8,46,47] suggest that the constant should be ≈ 1.

7. For the case of D-dimensional circuits, our result does not say much about the dynamics of the distribution when depth is $\ll n^{1/D}$. Such a result may explain the dynamics of entanglement in random circuits. [46,47] consider this problem for the case when a single Pauli operator starts at the middle of the lattice; however, their result does not apply to arbitrary initial operators.
8. The best anti-concentration lower bound we are able to prove is $\Omega(\ln n)$ For D-dimensional lattices one would expect a lower-bound of $\Omega(n^{1/D})$ based on the following intuition for circuits of depth $s < n^{1/D}$: For $s \ll n^{1/D}$, we expect any two non-overlapping clusters of s^D qubits will be close to Haar random. Hence, a crude model for such circuits would be n/s^D copies of Haar-random unitaries each on s^D qubits. In this case we would expect the collision probability to be $\approx \frac{2^{n/s^D}}{2^n}$. Very interestingly, the recent result [24] refutes this intuition for $D = 1$ and showed an upper bound of $O(\ln n)$ for the depth at which anti-concentration is achieved. It seems plausible that at $D = 2, 3, \ldots$ we would also have anti-concentration in depth $O(\ln n)$ since it holds both for $D = 1$ and for fully connected circuits.

2. Preliminaries

2.1. Basic definitions. We need the following norms:

Definition 14. For a superoperator $\mathcal{E}$ the diamond norm [40] is defined as $\|\mathcal{E}\|_\diamond := \sup_d \|\mathcal{E}\otimes \mathrm{id}_d\|_{1\to 1}$, where for a superoperator A the $1 \to 1$ norm is defined as $\|A\|_{1\to 1} := \sup_{X\neq 0} \frac{\|A(X)\|_1}{\|X\|_1}$.

A matrix is called positive semi-definite (psd) if it is Hermitian and has all non-negative eigenvalues. A superoperator $\mathcal{A}$ is called completely positive (cp) if for any $d \geq 0$, $\mathcal{A} \otimes \mathrm{id}_d$ maps psd matrices to psd matrices. A superoperator is called trace-preserving completely positive (tpcp) if it maps if it preserves the trace and is furthermore cp.

Let S be a set of qudits, then

Definition 15. Haar(S) is the Haar measure on $U((\mathbb{C}^d)^{\otimes |S|})$. We refer to Haar($i, j$) as the two qudit Haar measure on qudits indexed by i and j and also if m is an integer, the notation Haar(m) means Haar measure on m qudits.

We now define expected monomials, moment superoperators and quasi-projectors for a distribution μ over the unitary group:

Definition 16. Let $n, t > 0$ be positive integers and μ be any distribution over n-qudit unitary group $\mathrm{U}((\mathbb{C}^d)^{\otimes n})$. Then $G_\mu^{(t)} := \mathbb{E}_{C\sim\mu}\left[C^{\otimes t,t}\right]$ is the quasi-projector of μ. Here $C^{\otimes t,t} = C^{\otimes t} \otimes C^{*\otimes t}$. Also $G_{(i,j)}^{(t)} = G_{\mathrm{Haar}(i,j)}^{(t)}$. Using this Definition we will also use the following quantities:

1. Let $i_1, j_1, \ldots, i_t, j_t, k_1, l_1, \ldots, k_t, l_t \in [d]^n$ be any $2t$-tuple of words $\in [d]^n$. Then the $i_1, \ldots, l_t$ monomial is the expected value of a balanced monomial of μ defined as

$$\mathop{\mathbb{E}}_{C\sim\mu}\left[C_{i_1,j_1}\ldots C_{i_t,j_t}C^*_{k_1,l_1}\ldots C^*_{k_t l_t}\right] = \langle i_1, \ldots, j_t|\, G_\mu^{(t)}\, |k_1, \ldots, l_t\rangle \qquad (17)$$

$C_{a,b}$ is the a, b entry of the unitary matrix C.

2. Let $\mathrm{ad}_X(\cdot) := X(\cdot)X^\dagger$. Then $\mathrm{Ch}\left[G_\mu^{(t)}\right] := \mathbb{E}_{C\sim\mu}\left[\mathrm{ad}_{C^{\otimes t}}\right]$ is the t^{th} moment superoperator of μ.

Next, we define the building blocks of our t-design constructions.

Definition 17 (*Rows of a lattice*). For $1 \le i \le n^{1-1/D}$, $r_{\alpha,i}$ is the i-th row of a D-dimensional lattice in the α-th direction. We will label the qubits in row i by $(\alpha, i, 1), \ldots, (\alpha, i, n^{1/D})$. Assume for convenience that $n^{1/D}$ is an even integer and define the sets of pairs $E_{\alpha,i} := \{((\alpha, i, 1), (\alpha, i, 2)), \ldots, ((\alpha, i, n^{1/D}-1), (\alpha, i, n^{1/D}))\}$ and $O_{\alpha,i} := \{((\alpha, i, 2), (\alpha, i, 3)), \ldots, ((\alpha, i, n^{1/D}-2), (\alpha, i, n^{1/D}-1))\}$.

Definition 18 (*Elementary random circuits*). The elementary quasi-projector in direction α is

$$g_{\mathrm{Rows}(\alpha,n)} := \prod_{1\le l\le n^{1-1/D}} \bigotimes_{(i,j)\in E_{\alpha,l}} G_{(i,j)}^{(t)} \cdot \bigotimes_{(i,j)\in O_{\alpha,l}} G_{(i,j)}^{(t)} =: \prod_{1\le l\le n^{1-1/D}} g_{r_{\alpha,l}}. \quad (18)$$

For the 2-D lattice g_R and g_C for g_1 and g_2, respectively.

The following defines the moment superoperator and quasi-projector of the Haar measure on the rows of a D-dimensional lattice in a specific direction.

Definition 19 (*Idealized model with Haar projectors on rows*). Let $1 \le \alpha \le D$ be one of the directions of a D-dimensional lattice then

$$G_{\mathrm{Rows}(\alpha,n)} := \prod_{1\le i\le n^{1-1/D}} G_{\mathrm{Haar}(r_{\alpha,i})}^{(t)} =: \prod_{1\le i\le n^{1-1/D}} G_{r_{\alpha,i}}. \quad (19)$$

For a 2-D lattice we use G_R and G_C for G_1 and G_2, respectively.

Next, we define moment operators and projectors corresponding to the Haar measure on the sub-lattices of a D-dimensional lattice. We view a D-dimensional lattice as a collection of $n^{1/D}$ smaller lattices each with dimension $D-1$, composed of $n^{1-1/D}$ qudits. We label these sub-lattices with $\mathrm{Planes}(D) := \{p_1, \ldots, p_{n^{1/D}}\}$.

Definition 20 (*Haar measure on sub-lattices*). $G_{\mathrm{Planes}(D)} = \bigotimes_{p\in\mathrm{Planes}(D)} G_{\mathrm{Haar}(p)}^{(t)} \equiv G_{\mathrm{Haar}(n^{1-1/D})}^{(t)\otimes n^{1/D}}$.

Definition 21. For $d = 2$, $t = 2$ and a superoperator $\mathcal{A}$ define

$$\mathrm{Coll}(\mathcal{A}) := \mathrm{Tr}\left(\sum_{x\in\{0,1\}^n} |x\rangle\langle x| \otimes |x\rangle\langle x| \, \mathcal{A}(|0^n\rangle\langle 0^n| \otimes |0^n\rangle\langle 0^n|)\right). \quad (20)$$

In particular, for a distribution μ over circuits of size s the expected collision probability is defined as

$$\mathrm{Coll}_s := \mathrm{Coll}\left(\mathrm{Ch}\left[G_\mu^{(2)}\right]\right). \quad (21)$$

Remark 1. For $d = 2, t = 2$ and when ν is the Haar measure on U(4), $\mathrm{Ch}\left[G^{(2)}_{(i,j)}\right]$ is the following map in the Pauli basis:

$$\mathrm{Ch}\left[G^{(2)}_{(i,j)}\right](\sigma_p \otimes \sigma_q) = \begin{cases} \sigma_0 \otimes \sigma_0 & pq = 00 \\ \frac{1}{15}\sum_{s\in\{0,1,2,3\}^2\setminus 0} \sigma_s \otimes \sigma_s & p = q \neq 00 \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

More generally, if S is a collection of qubits, and $p, q \in \{0, 1, 2, 3\}^S$, then

$$\mathrm{Ch}\left[G^{(2)}_{S}\right](\sigma_p \otimes \sigma_q) = \begin{cases} \sigma_0 \otimes \sigma_0 & pq = 00 \\ \frac{1}{4^{|S|}-1}\sum_{s\in\{0,1,2,3\}^{|S|}\setminus 0} \sigma_s \otimes \sigma_s & p = q \neq 00 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

when $p, q \in \{0, 1, 2, 3\}^S$.

See [34,51] for the proof of these remarks.

2.2. *Operator definitions of the models.*

Definition 22 (*Random circuits on a two-dimensional lattice*). The quasi-projector of $\mu^{\mathrm{lattice},n}_{2,c,s}$ is $G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}} = g^s_R (g^s_C g^s_R)^c$.

The generalization of this definition to arbitrary D dimensions is according to:

Definition 23 (*Recursive definition for random circuits on D-dimensional lattices*). The quasi-projector of $\mu^{\mathrm{lattice},n}_{D,c,s}$ is specified by the recursive formula:

$$G^{(t)}_{\mu^{\mathrm{lattice},n}_{D,c,s}} = G^{(t)\otimes n^{1/D}}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}} \left(g^s_{\mathrm{Rows}(D,n)} G^{(t)\otimes n^{1/D}}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}} \right)^c . \tag{24}$$

It will be useful to our proofs to also define:

1. $\tilde{G}_{n,D,c} = \left(\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c} G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c} \right)^c$
2. $\hat{G}_{n,D,c,s} = G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} G_{\mathrm{Rows}(D,n)}$

In particular, $\tilde{G}_{n,D,c,s}$ is the same as $G_{\mu^{\mathrm{lattice},n}_{D,c,s}}$ except that we have replaced $g^s_{\mathrm{Rows}(D,n)}$ with $G_{\mathrm{Rows}(D,n)}$. Definition 22 is a special case of Definition 23, but we included both of them for convenience.

Definition 24. $G^{(t)}_{\mu^{\mathrm{CG}}_s} = \left(\frac{1}{\binom{n}{2}} \sum_{i\neq j} G^{(t)}_{(i,j)} \right)^s$.

2.2.1. Summary of the definitions. See below for a summary of the definitions:

Notation	Definition	Reference
$\|\cdot\|_\diamond$	superoperator diamond norm	Definition 14
$\|\cdot\|_p$	matrix p-norm for $p \in [0,\infty]$	Definition 14
Haar	the Haar measure	Definition 15
Haar(S)	Haar measure on subset S of qudits	Definition 15
Haar(i,j)	Haar measure on qudits i and j	Definition 15
$U^{\otimes t,t}$	$C^{\otimes t} \otimes C^{*,\otimes t}$	Definition 16
$G_\mu^{(t)}$	average of $C^{\otimes t,t}$ over $C \sim \mu$	Definition 16
$G_{\text{Haar}}^{(t)}$	Projects onto vectors invariant under $C^{\otimes t,t}$	Definition 16
$G_{i,j}^{(t)}$	Haar projector of order t on qudit i and j	Definition 16
$\langle i,j\|G_\mu^{(t)}\|k,l\rangle$	moment of order t: $\mathbb{E}_{C\sim\mu}[C_{i_1,j_1}\dots C_{i_t,j_t}C^*_{i_1,j_1}\dots C^*_{i_t,j_t}]$	Definition 16
$\text{Ch}[G_\mu^{(t)}]$	moment superoperator, equal to $\mathbb{E}_{C\sim\mu}[\text{ad}_{C^{\otimes t}}]$	Definition 16
$r_{\alpha,i}$	i-th row in the α direction with $i \in [n^{1/D}]$, $\alpha \in [D]$	Definition 17
Rows(α,n)	the collection of rows of a lattice (with n points) in the α direction	Definition 17
$g_{\text{Rows}(\alpha,n)}$	two-qudit gates applied to even then odd neighbors in each row in the α direction	Definition 18
$g_{r(\alpha,i)}$	two-qudit gates applied to even then odd neighbors in the i-th row in the α direction	Definition 18
g_R and g_C	$g_{\text{Rows}(1,n)}$ and $g_{\text{Rows}(2,n)}$ when $D=2$.	Definition 18
$G_{\text{Rows}(\alpha,n)}$	Haar projector applied to each row in theα^{th} direction	Definition 19
$G_R(G_C)$	Haar projector applied to each row (column) of a 2D lattice	Definition 19
$G_{\text{Planes}(\alpha)}$	Haar projector applied to each plane perpendicular to the direction α	Definition 20
Coll$(\mathcal{A})$	collision probability from superoperator $\mathcal{A}$	Definition 21
Coll_s	the expected collision probability of a random circuit after s steps	Definition 21
$\mu_{D,c,s}^{\text{lattice},n}$	the distribution over D-dimensional circuits with n qudits	Definition 23
$\tilde{G}_{n,D,c}$	same as $G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}}$ except that we replace $g^s_{\text{Rows}(\alpha,n)}$ with $G_{\text{Rows}(\alpha,n)}$	Definition 23
$\hat{G}_{n,D,c,s}$	one block of $\tilde{G}_{n,D,c}$ defined as $G_{\text{Rows}(D,n)}\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s}G_{\text{Rows}(D,n)}$	Definition 23
μ_s^{CG}	the distribution over circuits with s random two-qudit gates	Definition 24
$\measuredangle(A,B)$	$\cos^{-1}\max_{x\in A,y\in B}\langle x,y\rangle$ is the angle between two vector spaces A and B	Section 3.6.1

2.3. Elementary tools. If A is a matrix and σ_i are the singular values of A, then for $p \in [1,\infty)$ the Schatten p-norm of A is defined as $\|A\|_p := (\sum_i \sigma_i^p)^{1/p}$. The ∞-norm of A is $\|A\|_\infty := \max(i)\sigma_i$. The 1-norm is related to the ∞-norm by $\|A\|_1 \le \text{rank}(A)\cdot\|A\|_\infty$. Moreover, for $p \in [1,\infty]$ and any two matrices A and B, $\|A\otimes B\|_p = \|A\|_p \cdot \|B\|_p$.

If $\mathcal{A}$ and $\mathcal{B}$ are superoperators, then $\|\mathcal{A}\otimes\mathcal{B}\|_\diamond = \|\mathcal{A}\|_\diamond \cdot \|\mathcal{B}\|_\diamond$.

$\text{Ch}[\cdot]$ is the linear map from matrices to superoperators such that for any two equally sized matrices A and B, $\text{Ch}\left[A\otimes B^*\right] = A[\cdot]B^\dagger$. Note that $\text{Ch}[\cdot]$ is associative in the sense that $\text{Ch}\left[A\otimes B^*\right]\circ\text{Ch}\left[C\otimes D^*\right] = \text{Ch}\left[AC\otimes B^*C^*\right]$, for any equally sized matrices A, B, C, D.

Consider the Haar measure over $\text{U}(d)$. $\text{Ch}\left[G^{(t)}_{\text{Haar}}\right]$ (defined in the previous section) is the projector onto the matrix vector space of permutation operators (permuting length t words over the alphabet $[d]$). In particular, for any matrix $X \in \mathbb{C}^{d^t\times d^t}$ we can write

$$\text{Ch}\left[G^{(t)}_{\text{Haar}}\right][X] = \sum_{\pi\in S_t}\text{Tr}(V(\pi)X)Wg(\pi), \tag{25}$$

where $V(\pi)$ is the permutation matrix $\sum_{(i_1,\dots,i_t)\in[d]^t}|i_1,\dots,i_t\rangle\langle i_{\pi(1)},\dots,i_{\pi(t)}|$, and $Wg(\pi)$ is a linear combination of permutations. Specifically

$$Wg(\pi) = \sum_{\sigma \in S_t} \alpha(\pi^{-1}\sigma) V(\sigma). \tag{26}$$

Here the coefficients $\alpha(\cdot)$ are known as Weingarten functions (see [23]). If $\mu, \nu \in S_t$ then let $\mathsf{dist}(\mu, \nu)$ denote the number of transpositions needed to generate $\mu^{-1}\nu$ from the identity permutation. Then we can define $\alpha(\cdot)$ by the following relation.

$$\sum_{\mu,\nu \in S_t} \alpha(\mu^{-1}\nu) |\mu\rangle \langle\nu| = \left(\sum_{\mu,\nu \in S_t} \mathsf{dist}(\mu, \nu) |\mu\rangle \langle\nu| \right)^{-1}. \tag{27}$$

Note that $\alpha(\pi)$ is always real and $|\alpha(\lambda)| = O(1/d^{t+\mathsf{dist}(\lambda)})$. Thus for large d, $Wg(\pi) \approx V(\pi)/d^t$.

Furthermore,

$$\mathrm{Ch}\left[G^{(t)}_{\mathrm{Haar}}\right][X] = \sum_{\pi \in S_t} \operatorname*{Tr}_M \left((V(\pi)_M \otimes I_N) X_{MN}\right) \otimes \mathrm{Wg}(\pi)_M. \tag{28}$$

Let A, B be matrices. For the superoperator $\mathcal{D} \equiv B\mathrm{Tr}[A\cdot]$ we use the notation $\mathcal{D} = BA^*$. We need the following observation:

$$V(\pi)V^*(\sigma) = \mathrm{Ch}\left[|\psi_\pi\rangle \langle\psi_\sigma|\right], \tag{29}$$

where $|\psi_\pi\rangle = (I \otimes V(\pi)) \frac{1}{\sqrt{d^t}} \sum_{i \in [d]^t} |i\rangle |i\rangle$.

We need the following lemma:

Lemma 25. *If A is a (possibly rectangular) matrix, then $AA^\dagger$ and $A^\dagger A$ have the same spectra.*

Lemma 26. *If A and B are matrices and $\|\cdot\|_*$ is a unitarily invariant norm, then $\|AB\|_* \leq \|A\|_* \|B\|_\infty$.*

Proof. This lemma can be viewed as a consequence of Russo-Dye theorem, which states that the extreme points of the unit ball for $\|\cdot\|_\infty$ are the unitary matrices. Thus we can write $B = \|B\|_\infty \sum_i p_i U_i$ for $\{p_i\}$ a probability distribution and $\{U_i\}$ a set of unitary matrices. We use this fact along with the triangle inequality and then unitary invariance to obtain

$$\begin{aligned} \|AB\|_* = \|A \cdot \left(\|B\|_\infty \sum_i p_i U_i \right) \|_* &\leq \|B\|_\infty \sum_i p_i \|AU_i\|_* = \|B\|_\infty \sum_i p_i \|A\|_* \\ &= \|A\|_* \|B\|_\infty. \end{aligned} \tag{30}$$

□

A similar argument applies to superoperators.

Lemma 27. *If $\mathcal{A}$ is a superoperator and $\mathcal{B}$ is a tpcp superoperator then $\|\mathcal{AB}\|_\diamond \leq \|\mathcal{A}\|_\diamond$.*

Proof. Let d be $\geq$ the input dimensions of both $\mathcal{A}$ and $\mathcal{B}$. Then $\|\mathcal{A}\|_\diamond = \max_{\|X\|_1 \leq 1} \|(\mathcal{A} \otimes \mathrm{id}_d)(X)\|_1$ and $\|\mathcal{AB}\|_\diamond = \max_{\|X\|_1 \leq 1} \|(\mathcal{A} \otimes \mathrm{id}_d)(\mathcal{B} \otimes \mathrm{id}_d)(X)\|_1$. Since $\mathcal{B}$ is a tpcp superoperator $\|(\mathcal{B} \otimes \mathrm{id}_d)(X)\|_1 \leq 1$ and so $\|\mathcal{AB}\|_\diamond$ is maximizing over a set which is contained in the set maximized over by $\|\mathcal{A}\|_\diamond$. □

These give rise to the following well-known bound, which often is called "the hybrid argument."

Lemma 28. *Let $\|\cdot\|_*$ be a unitarily invariant norm. If $A_1, \ldots, A_t$ and $B_1, \ldots, B_t$ have ∞-norm ≤ 1. Then*

$$\|A_1 \ldots A_t - B_1 \ldots B_t\|_* \leq \sum_i \|A_i - B_i\|_*. \tag{31}$$

This is also true for superoperators and the diamond norm, if each superoperator is a tpcp map.

We will need a similar bound for tensor products.

Lemma 29. *Suppose $\|A - B\|_* \leq \epsilon$ for some norm $\|\cdot\|_*$ that is multiplicative under tensor product. Then for any integer $M > 0$*

$$\left\|A^{\otimes M} - B^{\otimes M}\right\|_* \leq (\|B\|_* + \epsilon)^M - \|B\|_*. \tag{32}$$

The same holds for superoperators and the diamond norm. In particular $\|A^{\otimes M} - B^{\otimes M}\|_* \leq 2M\|B\|_*^M \epsilon$ for $\epsilon \leq \frac{1}{2M}$.

We need the following definition and lemma:

Definition 30. Let X and Y be two real valued random variables on the same totally ordered sample space Ω. Then we say X is stochastically dominated by Y, if for all $x \leq y \in \Omega$, $\Pr[X \geq x] \leq \Pr[Y \geq y]$. We represent this by $X \preceq Y$.

Lemma 31 (Coupling). *$X \preceq Y$ if and only if there exists a coupling (a joint probability distribution) between X and Y such that the marginals of this coupling are exactly X and Y and that with probability 1, $X \leq Y$.*

2.4. Various measures of convergence to the Haar measure.

Definition 32. Let μ be a distribution over n-qudit gates. Let ϵ be a positive real number.

1. (Strong designs) μ is a strong ϵ-approximate t-design if

$$(1-\epsilon)\cdot \mathrm{Ch}\left[G_{\mathrm{Haar}}^{(t)}\right] \preceq \mathrm{Ch}\left[G_{\mu}^{(t)}\right] \preceq (1+\epsilon)\cdot \mathrm{Ch}\left[G_{\mathrm{Haar}}^{(t)}\right], \tag{33}$$

   or equivalently if

$$\begin{aligned}(1-\epsilon)\cdot \left(\mathrm{Ch}\left[G_{\mathrm{Haar}}^{(t)}\right]\otimes \mathrm{id}\right)\Phi_{d^n}^{\otimes t} &\preceq \left(\mathrm{Ch}\left[G_{\mu}^{(t)}\right]\otimes \mathrm{id}\right)\phi_{d^n}^{\otimes t} \\ &\preceq (1+\epsilon)\cdot \left(\mathrm{Ch}\left[G_{\mathrm{Haar}}^{(t)}\right]\otimes \mathrm{id}\right)\Phi_{d^n}^{\otimes t}.\end{aligned} \tag{34}$$

   The first $\preceq$ is cp ordering and the second $\preceq$ is psd ordering.
2. (Monomial definition) μ is a monomial based ϵ-approximate t-design if for any balanced monomial $m(C)$ of degree at most t

$$\left\|\mathsf{vec}\left[G_{\mu}^{(t)}\right] - \mathsf{vec}\left[G_{\mathrm{Haar}}^{(t)}\right]\right\|_\infty \leq \frac{\epsilon}{d^{nt}}. \tag{35}$$

   Here for a matrix A, $\mathsf{vec}(A)$ is a vector consisting of the entries of A (in the computational basis).

3. (Diamond definition) μ is an ϵ-approximate t-design in the diamond measure if

$$\left\| \mathrm{Ch}\left[G_{\mu}^{(t)}\right] - \mathrm{Ch}\left[G_{\mathrm{Haar}}^{(t)}\right] \right\|_{\diamond} \leq \epsilon. \tag{36}$$

4. (Trace definition) μ is an ϵ-approximate t-design in the trace measure if

$$\left\| G_{\mu}^{(t)} - G_{\mathrm{Haar}}^{(t)} \right\|_{1} \leq \epsilon. \tag{37}$$

5. (TPE) μ is a (d, ϵ, t) t-copy tensor product expander (TPE) if

$$\left\| G_{\mu}^{(t)} - G_{\mathrm{Haar}}^{(t)} \right\|_{\infty} \leq \epsilon. \tag{38}$$

6. (Anti-concentration) μ is an ϵ approximate anti-concentration design if

$$\mathop{\mathbb{E}}_{C\sim\mu} |\langle 0|C|0\rangle|^4 \leq \mathop{\mathbb{E}}_{C\sim\mathrm{Haar}} |\langle 0|C|0\rangle|^4 \cdot (1+\epsilon). \tag{39}$$

7. (Approximate scramblers) μ is an ϵ-approximate scrambler if for any density matrix ρ and subset S of qubits with $|S| \leq n/3$

$$\mathop{\mathbb{E}}_{C\sim\mu} \left\| \rho_S(C) - \frac{I}{2^{|S|}} \right\|_{1}^{2} \leq \epsilon. \tag{40}$$

   where $\rho_S(C) = \mathrm{Tr}_{\setminus S} C\rho C^{\dagger}$ and $\mathrm{Tr}_{\setminus S}$ is trace over the subset of qubits that is complimentary to S.
8. (Weak approximate decouplers) Let M, M', A, A' be systems composed of m, m, $n-m$ and $n-m$, and let $\phi_{MM'}$, $\phi_{AA'}$ and $\psi_{A'}$ be respectively maximally entangled states along M, M', maximally entangled state along AA' and a pure state along A'. μ is an (m, α, ϵ)-approximate weak decoupler if for any subsystem S of $M'A'$ with size $\leq \alpha \cdot n$, when μ applies to $M'A'$,

$$\mathop{\mathbb{E}}_{C\sim\mu} \left\| \rho_{MS}(C) - \frac{I}{2^m} \otimes \frac{I}{2^{|S|}} \right\|_{1} \leq \epsilon. \tag{41}$$

   We consider two definitions. In the first definition the initial state is $\phi_{MM'} \otimes \phi_{AA'}$ and in the second model it is $\phi_{MM'} \otimes \psi_{A'}$. Here $\rho_{MS}(C)$ is the reduced density matrix along MS after the application of $C \sim \mu$.

3. Approximate t-Designs by Random Circuits with Nearest-Neighbor Gates on D-Dimensional Lattices

In this section we prove theorems 8 and 9, which state that our random circuit models defined for D-dimensional lattices (definitions 5) form approximate t-designs in several measures.

We begin in Sect. 3.1 by outlining some basic utility lemmas. The technical core of the proof is contained in the lemmas in Sect. 3.2 in which we bound various norms of products of Haar projectors onto overlapping sets of qubits. These are proved in Sects. 3.5 and 3.6 respectively. We show how to use these lemmas to prove our main theorems in Sect. 3.3 (for a 2-D grid) and in Sect. 3.4 (for a lattice in $D > 2$ dimensions).

3.1. Basic lemmas. In this section we state some utilities lemmas which are largely independent of the details of our circuit models.

3.1.1. Comparison lemma for random quantum circuits.

Definition 33. A superoperator $\mathcal{C}$ is completely positive (cp) if for any psd matrix X, $(\mathcal{C} \otimes \mathrm{id})(X)$ is also psd. For superoperators $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{A} \preceq \mathcal{B}$ if $\mathcal{B} - \mathcal{A}$ is cp.

Our comparison lemma is simply the following:

Lemma 34 (Comparison). *Suppose we have the following cp ordering between superoperators $\mathcal{A}_1 \preceq \mathcal{B}_1, \ldots, \mathcal{A}_t \preceq \mathcal{B}_t$. Then $\mathcal{A}_t \ldots \mathcal{A}_1 \preceq \mathcal{B}_t \ldots \mathcal{B}_1$.*

Corollary 35 (Overlapping designs). *If $K_1, \ldots, K_t$ are respectively the moments superoperators of $\epsilon_1, \ldots, \epsilon_t$-approximate strong k-designs each on a potentially different subset of qudits, then*

$$\begin{aligned} &Ch\left[G^{(t)}_{Haar(S_1)} \ldots G^{(t)}_{Haar(S_t)}\right](1-\epsilon_1)\ldots(1-\epsilon_t) \preceq K_1 \ldots K_t \\ &\preceq Ch\left[G^{(t)}_{Haar(S_1)} \ldots G^{(t)}_{Haar(S_t)}\right](1+\epsilon_1)\ldots(1+\epsilon_t). \end{aligned} \tag{42}$$

3.1.2. Bound on the value of off-diagonal monomials. We first formally define an off-diagonal monomial.

Definition 36 (*Off-diagonal monomials*). A diagonal monomial of balanced degree t of a unitary matrix C is a balanced monomial that can be written as product of absolute square of terms, i.e., $|C_{a_1,b_1}|^2 \ldots |C_{a_t,b_t}|^2$. A monomial is off-diagonal if it is balanced and not diagonal.

We now define the set of diagonal indices as $\mathcal{D} = \{|i, j\rangle \langle i', j'| : i = i', j = j', i, i', j, j' \in [d]^{nt}\}$ and the set of off-diagonal indices as $\mathcal{O} = \{|i, j\rangle \langle i', j'| : i \neq i' \text{ or } j \neq j', i, i', j, j' \in [d]^{nt}\}$. We note that a diagonal monomial can be written as $\mathrm{Tr}(C^{\otimes t,t} x)$ for some $x \in \mathcal{D}$ and similarly, an off-diagonal monomial can be written as $\mathrm{Tr}(C^{\otimes t,t} x)$ for some $x \in \mathcal{O}$.

We relate the strong definition of designs to the monomial definiton via the following lemma.

Lemma 37. *Let $\delta > 0$. Assume that $Ch\left[G^{(t)}_{\mu}\right]$ and $Ch\left[G^{(t)}_{\nu}\right]$ are two moment superoperators that satisfy the following completely positive ordering*

$$(1-\delta) \cdot Ch\left[G^{(t)}_{\nu}\right] \preceq Ch\left[G^{(t)}_{\mu}\right] \preceq (1+\delta) \cdot Ch\left[G^{(t)}_{\nu}\right]. \tag{43}$$

Let $\mathcal{O}$ and $\mathcal{D}$ be respectively the set of off-diagonal and diagonal indices for monomials. Then

$$\max_{x \in \mathcal{O}} |\mathrm{Tr}\left(x G^{(t)}_{\mu}\right)| \leq \max_{x \in \mathcal{O}} |\mathrm{Tr}\left(x G^{(t)}_{\nu}\right)|(1+\delta) + 2\delta \cdot \max_{y \in \mathcal{D}} |\mathrm{Tr}\left(y G^{(t)}_{\nu}\right)|. \tag{44}$$

3.1.3. Bound on the moments of the Haar measure. We need the following bound on the t-th monomial moment of the Haar measure. Assume we have m qudits.

Lemma 38 (Moments of the Haar measure). *Let $G^{(t)}_{Haar(m)}$ be the quasi-projector operator for the Haar measure on m qudits. Then*

$$\max_y \left\| G^{(t)}_{Haar(m)} y G^{(t)}_{Haar(m)} \right\|_1 \leq \frac{t^{O(t)}}{d^{mt}}. \tag{45}$$

Here the maximization is taken over matrix elements in the computational basis like $y = |i_1, \ldots, i_t, i'_1, \ldots, i'_t\rangle \langle j_1, \ldots, j_t, j'_1, \ldots, j'_t|$. Each label (e.g. i_j) is in $[d]^m$.

3.2. Gap bounds for the product of overlapping Haar projectors. We will later need the following results, with proofs deferred until Sect. 3.6.

Lemma 39. $\|G_C G_R - G^{(t)}_{Haar}\|_\infty \leq 1/d^{\Omega(\sqrt{n})}$.

Lemma 40. *Let $D = O(\ln n / \ln \ln n)$ with small enough constant factor, then $\|G_{Planes(D)} G_{Rows(D,n)} - G_{Haar}\|_\infty \leq 1/d^{\Omega(n^{1-1/D})}$.*

Lemma 41. *Let $|x\rangle$ and $|y\rangle$ be two computational basis states. For small enough $D = O(\ln n / \ln \ln n)$ and large enough c, $|\langle x|\tilde{G}_{n,D,c} - G_{Haar}|y\rangle| \leq \frac{\epsilon}{d^{nt}}$ for some $\epsilon = 1/d^{\Omega(n^{1/D})}$.*

Lemma 42. *For large enough c, $\left\| Ch\left[(G_R G_C G_R)^c - G^{(t)}_{Haar} \right] \right\|_\diamond = \frac{t^{O(\sqrt{n}t)}}{d^{\Omega(c\sqrt{n})}}$.*

Lemma 43. *For small enough $D = O(\ln n / \ln \ln n)$ and large enough c,*

$$\left\| Ch\left[(G_{Rows(D,n)} G_{Planes(D)} G_{Rows(D,n)})^c - G^{(t)}_{Haar} \right] \right\|_\diamond = \frac{t^{O(tn^{1-1/D})}}{d^{\Omega(cn^{1-1/D})}}. \tag{46}$$

In these last two lemmas, we see that c will need to grow with t. We believe that a sharper analysis could reduce this dependence, but since we already have a poly(t) dependence in s, improving Lemmas 42 and 43 would not make a big difference. In fact, even in 1-D, [13] found a sharp n dependence but their factor of poly(t) (which we inherit) is probably not optimal.

3.3. Proof of Theorem 8; t-designs on two-dimensional lattices.

Theorem (Restatement of Theorem 8). *Let $s, c, n > 0$ be positive integers with $\mu^{lattice,n}_{2,c,s}$ defined as in Definition 5.*

1. $s = \text{poly}(t)\left(\sqrt{n} + \ln \frac{1}{\delta}\right)$, $c = O\left(t \ln t + \frac{\ln(1/\delta)}{\sqrt{n}}\right) \implies \left\| \text{vec}\left[G^{(t)}_{\mu^{lattice,n}_{2,c,s}} - G^{(t)}_{Haar} \right] \right\|_\infty \leq \frac{\delta}{d^{nt}}$.
2. $s = \text{poly}(t)\left(\sqrt{n} + \ln \frac{1}{\delta}\right)$, $c = O\left(t \ln t + \frac{\ln(1/\delta)}{\sqrt{n}}\right) \implies \left\| Ch\left[G^{(t)}_{\mu^{lattice,n}_{2,c,s}} - G^{(t)}_{Haar} \right] \right\|_\diamond \leq \delta$.
3. $s = \text{poly}(t)\left(\sqrt{n} + \ln \frac{1}{\delta}\right)$, $c = O\left(t \ln t + \frac{\ln(1/\delta)}{\sqrt{n}}\right) \implies \left\| G^{(t)}_{\mu^{lattice,n}_{2,c,s}} - G^{(t)}_{Haar} \right\|_1 \leq \delta$.

4. $\left\| G^{(t)}_{\mu^{lattice,n}_{2,c,s}} - G^{(t)}_{Haar} \right\|_\infty \leq c \cdot \sqrt{n} \cdot e^{-s/\operatorname{poly}(t)} + \frac{1}{d^{O(c\sqrt{n})}}$.

Proof. 1. This item corresponds to convergence of the individual moments of the Haar measure. A balanced moment of a distribution μ can be written as

$$\mathop{\mathbb{E}}_{C\sim\mu} [C_{i_1,j_1} \dots C_{i_t,j_t} C^*_{i'_1,j'_1} \dots C^*_{i'_t,j'_t}] = \langle i, i' | \, G^{(t)}_\mu \, | j, j' \rangle = \operatorname{Tr}[G^{(t)}_\mu \cdot |j, j'\rangle \langle i, i'|] \tag{47}$$

where $|i\rangle := |i_1, \dots, i_t\rangle$ and so on for $|i'\rangle$, $|j\rangle$, $|j'\rangle$. The same moment can also be written as

$$\operatorname{Tr}\left(|j\rangle \langle j'| \operatorname{Ch}\left[G^{(t)}_\mu \right] (|i\rangle \langle i'|) \right) \tag{48}$$

We will see that the strong design condition established by gives us strong bounds first for the "diagonal" case ($i = i'$, $j = j'$) then the off-diagonal case. This is because when we interpret $G^{(t)}_\mu$ as a quantum operation, the diagonal monomials correspond to $\operatorname{Tr} Y G^{(t)}_\mu X$ for psd matrices X, Y, and so the strong design condition applies directly. For off-diagonal moments we need to do a bit more work.

For each the diagonal and off-diagonal monomials, our strategy will be to first compare with the entries of $G_R(G_C G_R)^c G_R$ and then to compare to G_{Haar}.

First observe that since $\operatorname{Ch}\left[G^{(t)}_{\mu^{lattice,n}_{2,c,s}} \right] = (g^s_R g^s_C)^c g^s_R$ and $s = \operatorname{poly}(t)\cdot(\sqrt{n}+\ln(1/\delta))$ then corollary 6 of [13] implies that each g^s_i for $i \in \{R, C\}$ is an δ-approximate t-design. Hence, using corollary 35,

$$\operatorname{Ch}[G_R \, (G_C G_R])^c \, G_R(1 - \frac{\delta}{4t!})] \preceq \operatorname{Ch}\left[G^{(t)}_{\mu^{lattice,n}_{2,c,s}} \right] \preceq \operatorname{Ch}[G_R(G_C G_R])^c G_R(1 + \frac{\delta}{4t!}). \tag{49}$$

Note that we chose $\operatorname{poly}(t)$ large enough so that the error is as small as $\frac{\delta}{4t!}$. This choice will be helpful later.

Focusing first on diagonal monomials $|i\rangle \langle i|$, $|j\rangle \langle j|$ we can bound

$$\begin{aligned}
&(1 + \frac{\delta}{4t!})\operatorname{Tr}\left(|j\rangle \langle j| \operatorname{Ch}\left[G_R(G_C G_R])^c G_R \right] (|i\rangle \langle i|) \right) - \langle i, j | \, G^{(t)}_{\mu^{lattice,n}_{2,c,s}} \, |i, j\rangle \\
&= \operatorname{Tr}\left(|j\rangle \langle j| [\operatorname{Ch}[G_R(G_C G_R])^c \, G_R(1 + \frac{\delta}{4t!}) - G^{(t)}_{\mu^{lattice,n}_{2,c,s}}]](|i\rangle \langle i|) \right) \geq 0.
\end{aligned} \tag{50}$$

In other words, for diagonal monomials

$$\begin{aligned}
&\operatorname{Tr}\left(|j\rangle \langle j| \operatorname{Ch}\left[G^{(t)}_{\mu^{lattice,n^{(t)}}_{2,c,s}} \right] (|i\rangle \langle i|) \right) \\
&\leq (1 + \frac{\delta}{4t!})\operatorname{Tr}\left(|j\rangle \langle j| \operatorname{Ch}[G_R \, (G_C G_R])^c \, G_R](|i\rangle \langle i|) \right) \\
&= (1 + \frac{\delta}{4t!})\operatorname{Tr}\left(G_R(G_C G_R)^c G_R \, |i, j\rangle \langle i, j| \right).
\end{aligned} \tag{51}$$

Similarly, using the first inequality in (49)

$$\mathrm{Tr}\left(|j\rangle\langle j|\,\mathrm{Ch}\left[G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}(t)}\right](|i\rangle\langle i|)\right) \geq (1-\frac{\delta}{4t!})\mathrm{Tr}\left(G_R(G_CG_R)^cG_R\,|i,j\rangle\langle i,j|\right). \tag{52}$$

The next step is to bound $\mathrm{Tr}\,(yG_R(G_CG_R)^cG_Rx)$:

$$\begin{aligned}\left|\mathrm{Tr}\left(G_R(G_CG_R)^cG_Rx\right)-\mathrm{Tr}\left(G^{(t)}_{\mathrm{Haar}}x\right)\right| &= \left|\mathrm{Tr}\left((G_R(G_CG_R)^cG_R-G^{(t)}_{\mathrm{Haar}})x\right)\right| \\ &= \left|\mathrm{Tr}\left(((G_CG_R)^c-G^{(t)}_{\mathrm{Haar}})G_RxG_R\right)\right| \\ &\leq \left\|((G_CG_R)^c-G^{(t)}_{\mathrm{Haar}})\right\|_\infty\cdot\left\|G_RxG_R\right\|_1 \\ &\leq \left\|G_CG_R-G^{(t)}_{\mathrm{Haar}}\right\|^c_\infty\cdot\left(\max_{y\in[d]^{2\sqrt{n}t}}\|G_{r_{1,1}}yG_{r_{1,1}}\|_1\right)^{\sqrt{n}}.\end{aligned} \tag{53}$$

In the third line we have used the Hölder's inequality. In the last inequality we have used the fact that G_1 is a tensor product of $G_{r_{1,i}}$ across each column in the first direction; by symmetry we can just consider $G_{r_{1,1}}$.
Using Lemma 38

$$\max_{y\in[d]^{2\sqrt{n}t}}\left\|G_{r_{1,1}}yG_{r_{1,1}}\right\|_1=\frac{t^{O(t)}}{d^t\sqrt{n}}. \tag{54}$$

Furthermore, using Lemma 39

$$\left\|G_CG_R-G^{(t)}_{\mathrm{Haar}}\right\|_\infty\leq\frac{1}{d^{\Omega(\sqrt{n})}}. \tag{55}$$

therefore

$$\left\|G_CG_R-G^{(t)}_{\mathrm{Haar}}\right\|^c_\infty\cdot(\max_{y\in[d]^{2\sqrt{n}t}}\|G_{r_{1,1}}yG_{r_{1,1}}\|_1)^{\sqrt{n}}\leq\frac{1}{d^{O(c\cdot\sqrt{n})}}\cdot(\frac{t^{O(t)}}{d^t\sqrt{n}})^{\sqrt{n}}. \tag{56}$$

As a result, for some large enough $c=O(t\ln t+\frac{\ln 1/\delta}{\sqrt{n}})$ we conclude

$$\begin{aligned}&|\mathrm{Tr}\left(G_R(G_CG_R)^cG_Rx\right)-M^{(\mathrm{Haar},t)}_x| \\ &\quad\leq\|G_CG_R-G^{(t)}_{\mathrm{Haar}}\|^c_\infty\cdot(\max_{y\in[d]^{2\sqrt{n}t}}\|G_{r_{1,1}}yG_{r_{1,1}}\|_1)^{\sqrt{n}} \\ &\quad\leq\frac{\delta}{4d^{nt}}.\end{aligned} \tag{57}$$

As a result, using Lemma 38 any diagonal monomial satisfies

$$\begin{aligned}|\mathrm{Tr}\left(G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}}\right)-\mathrm{Tr}\left(G^{(t)}_{\mathrm{Haar}}\right)| \leq\ &|\mathrm{Tr}\left(G_R(G_CG_R)^cG_Rx\right)-\mathrm{Tr}\left(G_R(G_CG_R)^cG_Rx\right)| \\ &+\frac{\delta}{4t!}|\mathrm{Tr}\left(G_R(G_CG_R)^cG_Rx\right)|\end{aligned}$$

$$\leq \frac{\delta}{4d^{nt}} + \frac{\delta}{4t!}(M_x^{(\mathrm{Haar},t)} + \frac{\delta}{4d^{nt}})$$
$$\leq \frac{\delta}{4d^{nt}} + \frac{\delta}{4t!}(t!/d^{nt} + \frac{\delta}{4d^{nt}})$$
$$\leq \frac{\delta}{d^{nt}}. \tag{58}$$

Next, we bound the expected off-diagonal monomials of the distribution. The value of the off-diagonal monomials according to the Haar measure is zero. So it is enough to bound $\max_{x\in\mathcal{O}} |\mathrm{Tr}\left(G_\mu^{(t)} x\right)|$, where $\mathcal{O}$ is the set of off-diagonal indices for moments. In order to do this we use Lemma 37 for $\mu = \mu_{2,c,s}^{\mathrm{lattice},n}$ and ν being a distribution with moment superoperator $\mathrm{Ch}[G_R](\mathrm{Ch}[G_R]\mathrm{Ch}[G_C])^c\mathrm{Ch}[G_R]$.

$$\max_{x\in\mathcal{O}} |\mathrm{Tr}(G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}})x| \leq \max_{x\in\mathcal{O}} \mathrm{Tr}(G_R(G_CG_R)^cG_Rx)(1+\frac{\delta}{4t!})$$
$$+\delta/t! \cdot \max_{y\in\mathcal{D}} \mathrm{Tr}(G_R(G_CG_R)^cG_Ry). \tag{59}$$

Here $\mathcal{D}$ is the set of diagonal monomials. Using (57)

$$\max_{y\in\mathcal{D}} \mathrm{Tr}(G_R(G_CG_R)^cG_Ry) \leq \max_{y\in\mathcal{D}} \mathrm{Tr}\left(G^{(t)}_{\mathrm{Haar}}y\right) + \frac{\delta}{4d^{nt}} \leq \frac{t!}{d^{nt}} + \frac{\delta}{4d^{nt}}. \tag{60}$$

In order to bound $\max_{x\in\mathcal{O}} \mathrm{Tr}(G_R(G_CG_R)^cG_Rx)$, we first make the observation that since $x \in \mathcal{O}$, $\mathrm{Tr}(G^{(t)}_{\mathrm{Haar}}x) = 0$. Therefore

$$\max_{x\in\mathcal{O}} |\mathrm{Tr}(G_R(G_CG_R)^cG_Rx)| = \max_{x\in\mathcal{O}} |\mathrm{Tr}((G_R(G_CG_R)^cG_R - G^{(t)}_{\mathrm{Haar}})x)|$$
$$\leq \max_{x\in\mathcal{O}} |\mathrm{Tr}((G_RG_C)^c - G^{(t)}_{\mathrm{Haar}})G_RxG_R)|$$
$$\leq \frac{\delta}{4d^{nt}}. \tag{61}$$

therefore using (57), (59) and (61) we conclude

$$\max_{x\in\mathcal{O}} |\mathrm{Tr}(G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}}x)| \leq \frac{\delta}{4d^{nt}}(1+\delta/(4t!)) + \frac{\delta}{4t!}\cdot\left(\frac{t!}{d^{nt}} + \frac{\delta}{4d^{nt}}\right)$$
$$\leq \frac{\delta}{2d^{nt}} + 2\delta/(4d^{nt}) \leq \frac{\delta}{d^{nt}}. \tag{62}$$

2.

$$\left\|\mathrm{Ch}\left[G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \leq \left\|\mathrm{Ch}\left[g_R^s(g_C^sg_R^s)^c - (G_RG_CG_R]\right)^c\right\|_\diamond$$
$$+ \left\|(\mathrm{Ch}\left[G_RG_CG_R\right])^c - G^{(t)}_{\mathrm{Haar}}]\right\|_\diamond$$
$$\leq 4c\cdot\left\|\mathrm{Ch}\left[g^s_{r_{1,1}}\right]^{\otimes\sqrt{n}} - \left[G_{r_{1,1}}\right]^{\otimes\sqrt{n}}\right\|_\diamond + \left(\frac{t^t}{d^c}\right)^{O(\sqrt{n})}$$
$$\leq 4c\cdot\sqrt{n}\cdot\left\|\mathrm{Ch}\left[g^s_{r_{1,1}} - G_{r_{1,1}}\right]\right\|_\diamond + \left(\frac{t^t}{d^c}\right)^{O(\sqrt{n})}$$

$$\begin{aligned} &\leq \delta/2 + \delta/2 \\ &\leq \delta. \end{aligned} \tag{63}$$

In the first line we have used triangle inequality and the definition $K^{(t)}_{\mu^{\text{lattice},n}_{2,c,s}} = (\prod_\alpha g^s_{\text{Rows}(\alpha,n)})^c$. In the second line, for the first term we have used Lemma 28 and that all operators are compositions of moment superoperators. For the second part we have used Lemma 42. In the third inequality we have used Lemma 29. In fourth inequality, the first term ($\delta/2$) comes from lemma 3 and corollary 6 of [13] for $s = \text{poly}(t) \cdot (\sqrt{n} + \ln \frac{1}{\delta})$, and the second $\delta/2$ is by the choice $c = O(t \ln t + \frac{\ln(1/\delta)}{\sqrt{n}})$.

3. Let $Q_0 := G_{r_{1,1}}$ and $Q_1 := G_{r_{1,1}} - g^s_{r_{1,1}}$, and for $x \in \{0,1\}^{\sqrt{n}}$ let $Q_x = Q_{x_1} \dots Q_{x_{\sqrt{n}}}$. Here, $\|Q_0\|_1 = t!$ and $\|Q_x\|_1 = t!^{\sqrt{n}-|x|} \cdot \|G_{r_{1,1}} - g^s_{r_{1,1}}\|^{|x|}$.

$$\begin{aligned} \|G^{(t)}_{\mu^{\text{lattice},n}_{2,c,s}} - G^{(t)}_{\text{Haar}}\|_1 &\leq \|(g^s_C g^s_R)^c - (G_C G_R)^c\|_1 + \|(G_C G_R)^c - G^{(t)}_{\text{Haar}}\|_1 \\ &\leq 4c \cdot \|(g^s_{r_{1,1}})^{\otimes\sqrt{n}} - G^{\otimes\sqrt{n}}_{r_{1,1}}\|_1 + t^{O(t)\sqrt{n}} \|G_C G_R - G^{(t)}_{\text{Haar}}\|^c_\infty \end{aligned} \tag{64}$$

We bound the two terms separately. First

$$\begin{aligned} 4c\|(g^s_{r_{1,1}})^{\otimes\sqrt{n}} - G^{\otimes\sqrt{n}}_{r_{1,1}}\|_1 &\leq 4c \cdot \sum_{x \in \{0,1\}^{\sqrt{n}} : x \neq 0} \|Q_x\|_1 \\ &\leq 4c \cdot [(t! + \|g^s_{r_{1,1}} - G_{r_{1,1}}\|_1)^{\sqrt{n}} - t!^{\sqrt{n}}] \\ &= 4ct!((1 + \|g^s_{r_{1,1}} - G_{r_{1,1}}\|_1/t!)^{\sqrt{n}} - 1) \\ &\leq 4c \cdot 2\sqrt{n}\|g^s_{r_{1,1}} - G_{r_{1,1}}\|_1 \end{aligned}$$

The last line needs s to be large enough that $\sqrt{n}\|g^s_{r_{1,1}} - G_{r_{1,1}}\|_1 \leq 1/(2\sqrt{n})$.

$$\begin{aligned} &\leq 8c\sqrt{n}(dt!)^{\sqrt{n}} \cdot \|g^s_{r_{1,1}} - G_{r_{1,1}}\|_\infty \\ &\leq 8c\sqrt{n}(dt!)^{\sqrt{n}}(1 - 1/\text{poly}(t))^s \\ &\leq \delta/2 \end{aligned} \tag{65}$$

Now we bound the second term of (64).

$$t^{O(t)\sqrt{n}} \|G_C G_R - G^{(t)}_{\text{Haar}}\|^c_\infty \leq t^{O(t)\sqrt{n}} (d^{-\Omega(\sqrt{n})})^c \quad \text{using Lemma 39} \tag{66}$$

$$\leq t^{C_1 t\sqrt{n}} d^{-cC_2\sqrt{n}} \quad \text{for some universal constants } C_1, C_2 > 0 \tag{67}$$

$$\begin{aligned} &= (t^{C_1 t}/d^{cC_2})^{\sqrt{n}} \\ &\leq \delta/2. \end{aligned} \tag{68}$$

In the last step we need to choose the implicit constant in the definition of c based on C_1, C_2.

4.

$$\begin{aligned}\|G^{(t)}_{\mu^{\text{lattice},n}_{2,c,s}} - G^{(t)}_{\text{Haar}}\|_\infty &\le \|(g^s_C g^s_R)^c - (G_C G_R)^c\|_\infty + \|(G_C G_R)^c - G^{(t)}_{\text{Haar}}\|_\infty \\ &\le 4c \cdot \|(g^s_{r_{1,1}})^{\otimes\sqrt{n}} - G^{\otimes\sqrt{n}}_{r_{1,1}}\|_\infty + \|G_C G_R - G^{(t)}_{\text{Haar}}\|^c_\infty \\ &\le 4c \cdot \sqrt{n} \cdot \|g^s_{r_{1,1}} - G_{r_{1,1}}\|_\infty + \frac{1}{d^{\Omega(c\sqrt{n})}} \\ &\le 4c \cdot \sqrt{n} \cdot e^{-s/\operatorname{poly}(t)} + \frac{1}{d^{\Omega(c\sqrt{n})}}. \end{aligned} \tag{69}$$

These steps follow from the proof of part 1. □

3.4. Proof of Theorem 9; t-designs on D-dimensional lattices. Throughout this section we treat D and t as constants.

Theorem (Restatement of Theorem 9). *There exists a value $\delta = 1/d^{\Omega(n^{1/D})}$ such that for some large enough c depending on D and t:*

1. $s > c \cdot n^{1/D} \implies \left\| \text{vec}\left[G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}} - G^{(t)}_{Haar} \right] \right\|_\infty \le \frac{\delta}{d^{nt}}.$
2. $s > c \cdot n^{1/D} \implies \left\| Ch\left[G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}} - G^{(t)}_{Haar} \right] \right\|_\diamond \le \delta.$
3. $s > c \cdot n^{1/D} \implies \left\| G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}} - G^{(t)}_{Haar} \right\|_\infty \le \delta.$
4. $s > c \cdot n^{1/D} \implies \left\| G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}} - G^{(t)}_{Haar} \right\|_1 \le \delta.$

Proof. 1. Consider the moment superoperator for the D-dimensional random circuit distribution $\text{Ch}\left[G_{\mu^{\text{lattice},n}_{D,c,s}}\right]$, where for $3 \le \alpha \le D$, $\kappa_{\text{Rows}(\alpha,n)}$ is defined according to the recursive formula $\kappa_\alpha = \kappa^{\otimes n^{1/\alpha}}_{\alpha-1}((\text{Ch}[g_i])^s \kappa^{\otimes n^{1/\alpha}}_{\alpha-1})^c$.
Using corollary 6 of [13], if $s = O(n^{1/D})$ then each $g^s_{\text{Rows}(\alpha,n)}$ for $1 \le \alpha \le D$ satisfies a $1/d^{\Omega(n^{1/D})}$-approximate t-design property. Hence, using corollary 35

$$\text{Ch}[\tilde{G}_{n,D,c}](1 - 1/d^{\Omega(n^{1/D})}) \preceq \text{Ch}\left[G^{(t)}_{\mu^{\text{lattice},n}_{2,c,s}}\right] \preceq \text{Ch}\left[\tilde{G}_{n,D,c}\right](1 + 1/d^{\Omega(n^{1/D})}). \tag{70}$$

Therefore,

$$(1 - 1/d^{\Omega(n^{1/D})})\text{Tr}(\tilde{G}_{n,D,c}x) \le \text{Tr}\left(G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}}x\right) \le (1 + 1/d^{\Omega(n^{1/D})})\text{Tr}(\tilde{G}_{n,D,c}x). \tag{71}$$

Where x is a matrix $|i, j\rangle\langle i', j'|$ for $i, j, i', j' \in [d]^{nt}$.
Next, we use Lemma 41. This lemma along with the bound in (71) and Lemma 38 proves the stated bound for diagonal monomials:

$$|\text{Tr}\left(G^{(t)}_{\mu^{\text{lattice},n}_{D,c,s}}x\right) - \text{Tr}(G^{(t)}_{\text{Haar}}x)| \le |\text{Tr}(\tilde{G}_{n,D,c}x) - \text{Tr}(G^{(t)}_{\text{Haar}}x)| + |\text{Tr}(\tilde{G}_{n,D,c}x)|1/d^{\Omega(n^{1/D})}$$

$$\leq \frac{1/d^{\Omega(n^{1/D})}}{d^{nt}} + (|\mathrm{Tr}(G^{(t)}_{\mathrm{Haar}}x)| + \frac{1/d^{\Omega(n^{1/D})}}{d^{nt}})1/d^{\Omega(n^{1/D})}$$
$$\leq \frac{1/d^{\Omega(n^{1/D})}}{d^{nt}} + (t!/d^{nt} + \frac{1/d^{\Omega(n^{1/D})}}{d^{nt}})1/d^{\Omega(n^{1/D})}$$
$$\leq \frac{1/d^{\Omega(n^{1/D})}}{d^{nt}}. \tag{72}$$

Next, we bound off-diagonal monomials $\max_{x\in\mathcal{O}} |\mathrm{Tr}\left(G^{(t)}_{\mu}x\right)|$. Again, we use Lemma 37 for $\mu = \mu^{\mathrm{lattice},n}_{D,c,s}$ and ν being a distribution with moment superoperator $K_{\mu^{\mathrm{lattice},n}_{D,c,s}}$ (or the quasi-projector $\tilde{G}_{n,D,c}$):

$$\max_{x\in\mathcal{O}} |\mathrm{Tr}\left(G^{(t)}_{\mu^{\mathrm{lattice},n}_{D,c,s}}x\right)| \leq \max_{x\in\mathcal{O}} \mathrm{Tr}(\tilde{G}_{n,D,c}x)(1 + 1/d^{\Omega(n^{1/D})}) + 1/d^{\Omega(n^{1/D})} \cdot \max_{y\in\mathcal{D}} \mathrm{Tr}(\tilde{G}_{n,D,c}y). \tag{73}$$

Using Lemma 41

$$\max_{y\in\mathcal{D}} \mathrm{Tr}(\tilde{G}_{n,D,c}y) \leq \max_{y\in\mathcal{D}} \mathrm{Tr}\left(G^{(t)}_{\mathrm{Haar}}y\right) + \frac{1/d^{\Omega(n^{1/D})}}{d^{nt}} \leq \frac{t!}{d^{nt}} + \frac{1/d^{\Omega(n^{1/D})}}{d^{nt}}. \tag{74}$$

Similar to (61) we can show

$$\max_{x\in\mathcal{O}} |\mathrm{Tr}\left(\tilde{G}_{n,D,c}x\right)| = \max_{x\in\mathcal{O}} |\mathrm{Tr}\left((\tilde{G}_{n,D,c} - G^{(t)}_{\mathrm{Haar}})x\right)|$$
$$\leq \max_{x\in\mathcal{O}} |\mathrm{Tr}\left((\hat{G}_{n,D,c})^c - G^{(t)}_{\mathrm{Haar}}\right) \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c} x \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c})|$$
$$\leq \frac{1/d^{\Omega(n^{1/D})}}{d^{nt}}, \tag{75}$$

therefore using (73), (74) and (75) we conclude that any monomial $M_x^{(\mu^{\mathrm{lattice},n}_{D,c,s},t)}$ satisfies

$$\max_{x\in\mathcal{O}} |\mathrm{Tr}\left(G^{(t)}_{\mu^{\mathrm{lattice},n}_{D,c,s}}x\right)| \leq \frac{1/d^{\Omega(n^{1/D})}}{d^{nt}}. \tag{76}$$

2. Let $\epsilon_{D,n} := \left\|\mathrm{Ch}\left[G_{\mu^{\mathrm{lattice},n}_{D,c,s}}\right] - \mathrm{Ch}\left[G^{(t)}_{\mathrm{Haar}}\right]\right\|_{\diamond}$. We use induction to show that $\epsilon_{D,n} = 1/d^{\Omega(n^{1/D})}$ for any integers n and D. This is true for $D = 2$ by Theorem 8. Assuming $\epsilon_{D-1,n} = 1/d^{\Omega(n^{1/(D-1)})}$ for any n, we show that $\epsilon_{D,n} = 1/d^{\Omega(n^{1/D})}$.

$$\epsilon_{D,n} := \left\|\mathrm{Ch}\left[G_{\mu^{\mathrm{lattice},n}_{D,c,s}} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_{\diamond}$$
$$\leq \left\|\mathrm{Ch}\left[G_{\mu^{\mathrm{lattice},n}_{D,c,s}} - \tilde{G}_{n,D,c}\right]\right\|_{\diamond} + \left\|\mathrm{Ch}\left[\tilde{G}_{n,D,c} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_{\diamond}$$
$$\leq \mathrm{poly}(n) \cdot \left\|\mathrm{Ch}\left[(g^s_{r_{1,1}})^{\otimes n^{1-1/D}} - G^{\otimes n^{1-1/D}}_{r_{1,1}}\right]\right\|_{\diamond}$$

$$
\begin{aligned}
&+ \left\| \mathrm{Ch}\left[\tilde{G}_{n,D,c} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \\
&\leq O(n^{1-1/D}) \cdot \| \mathrm{Ch}\left[g^s_{r_{1,1}} - G_{r_{1,1}}\right] \|_\diamond \\
&\quad + \| \mathrm{Ch}\left[(\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s})^c - G^{(t)}_{\mathrm{Haar}}\right] \|_\diamond \\
&\leq O(n) \cdot 1/d^{\Omega(n^{1/D})} \\
&\quad + \left\| \mathrm{Ch}\left[(\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s})^c - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \\
&\leq 1/d^{\Omega(n^{1/D})} + \left\| \mathrm{Ch}\left[(\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s})^c - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond
\end{aligned} \tag{77}
$$

The third line is by triangle inequality. The fourth inequality is by Lemma 28. The fifth line is by Lemma 29 and the definition $\tilde{G}_{n,D,c} = ([\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s}])^c$. The sixth line is by lemma 3 and corollary 6 of [13], which assert that after linear depth in the number of qudits ($n^{1/D}$), the random circuit model we consider is ϵ-approximate t-design in the diamond measure, and that ϵ can be made exponentially small in $n^{1/D}$.

Next, we bound $\left\| \mathrm{Ch}\left[(\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s})^c - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond$. We first relate this expression to the superoperator $\mathrm{Ch}\left[G_{\mathrm{Planes}(D)}\right]$. Using triangle inequality and Lemma 28:

$$
\begin{aligned}
&\left\| \mathrm{Ch}\left[(\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s})^c - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \\
&\leq \left\| \mathrm{Ch}\left[(G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} G_{\mathrm{Rows}(D,n)})^{c-1} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \\
&\leq \left\| \mathrm{Ch}\left[(G_{\mathrm{Rows}(D,n)} (\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} - G_{\mathrm{Planes(D,n)}}) G_{\mathrm{Rows}(D,n)}\right.\right. \\
&\quad \left.\left. + G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)})^{c-1} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \\
&\leq O\left(\| \mathrm{Ch}\left[\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c,s} - G_{\mathrm{Planes}(D)}\right] \|_\diamond \right) \\
&\quad + \left\| \mathrm{Ch}\left[(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)})^{c-1} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \\
&\leq O(n) \left\| \mathrm{Ch}\left[\tilde{G}_{n^{1-1/D},D-1,c,s} - G_{\mathrm{Haar}(p_1)}\right]\right\|_\diamond \\
&\quad + \left\| \mathrm{Ch}\left[(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)})^{c-1} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \\
&\leq O(n)\epsilon_{D-1,n^{1-1/D}} + \left\| \mathrm{Ch}\left[(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)})^{c-1} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \\
&\leq O(n)\frac{1}{d^{n^{1/D}}} + \left\| (\mathrm{Ch}\left[G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)})^{c-1} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond \\
&\leq \frac{1}{d^{n^{1/D}}} + \left\| (\mathrm{Ch}\left[G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)})^{c-1} - G^{(t)}_{\mathrm{Haar}}\right]\right\|_\diamond .
\end{aligned}
$$

The first line is by Lemma 27. The third line is by triangle inequality and Lemma 27. The fourth line is by Lemma 29 and that $\tilde{G}_{\mathrm{Planes}(D)}$ is a tensor product of Haar

moment operators. Note in the sixth line we have used the induction hypothesis: $\epsilon_{D-1,n^{1-1/D}} = 1/d^{O\left(\frac{n^{1-1/D}}{D-1}\right)} = \frac{1}{d^{\Omega(n^{1/D})}}$.

Using Lemma 43 $\left\|(\mathrm{Ch}[G_{\mathrm{Rows}(D,n)}]\tilde{G}_{\mathrm{Planes}(D)}\mathrm{Ch}[G_{\mathrm{Rows}(D,n)}])^{c-1} - \mathrm{Ch}[G^{(t)}_{\mathrm{Haar}}]\right\|_\diamond = \frac{1}{d^{\Omega(n^{1/D})}}$ and this completes the proof.

3. Define $\epsilon_{D,n} := \left\|G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}} - G^{(t)}_{\mathrm{Haar}}\right\|_1$. By induction assume $\epsilon_{D-1,n} = 1/d^{\Omega(n^{1/D-1})}$ for all n. We would like to show that $\epsilon_{D,n} = 1/d^{\Omega(n^{1/D})}$.

$$\begin{aligned}\epsilon_{D,n} &:= \left\|G^{(t)}_{\mu^{\mathrm{lattice},n}_{D,c,s}} - G^{(t)}_{\mathrm{Haar}}\right\|_1 \\ &= \left\|G^{(t)\otimes n^{1/D}}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}}(g^s_{\mathrm{Rows}(D,n)}G^{(t)\otimes n^{1/D}}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}})^c - G^{(t)}_{\mathrm{Haar}}\right\|_1 \end{aligned} \tag{78}$$

Write $G^{(t)\otimes n^{1/D}}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}} = G_{\mathrm{Planes}(D)} + (G^{(t)\otimes n^{1/D}}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}} - G_{\mathrm{Planes}(D)}) =: Z_0 + Z_1$. Our strategy is to expand (78) in terms of $G_{\mathrm{Planes}(D)}$:

$$\begin{aligned}&\left\|(\delta + G_{\mathrm{Planes}(D)})(g^s_{\mathrm{Rows}(D,n)}(\delta + G_{\mathrm{Planes}(D)}))^c - G^{(t)}_{\mathrm{Haar}}\right\|_1 \\ &= \sum_{\phi\in\{0,1\}^{c+1}}\left\|Z_{\phi_0}\prod_{i=1}^{c}(g^s_{\mathrm{Rows}(D,n)}Z_{\phi_i}) - G^{(t)}_{\mathrm{Haar}}\right\|_1 \\ &\le \underbrace{\sum_{\phi\in\{0,1\}^{c+1}\setminus 0^{c+1}}\left\|Z_{\phi_0}\prod_{i=1}^{c}(g^s_{\mathrm{Rows}(D,n)}Z_{\phi_i})\right\|_1}_{(1)} + \underbrace{\left\|Z_0(g^s_{\mathrm{Rows}(D,n)}Z_0)^c - G^{(t)}_{\mathrm{Haar}}\right\|_1}_{(2)}\end{aligned} \tag{79}$$

To bound (1), observe that each term contains at least one Z_1. We would like to bound $\|Z_1\|_1$. Observe that $G_{\mathrm{Planes}} = G^{(t)\otimes n^{1/D}}_{\mathrm{Haar}(n^{1-1/D})}$, so

$$\begin{aligned}\|Z_1\|_1 &= \left\|G^{(t)\otimes n^{1/D}}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}} - G^{(t)\otimes n^{1/D}}_{\mathrm{Haar}(n^{1-1/D})}\right\|_1 \\ &= \left\|\sum_{i=1}^{n^{1/D}} G^{(t)\otimes i-1}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}}(G^{(t)}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}} - G^{(t)}_{\mathrm{Haar}(n^{1-1/D})})G^{(t)\otimes n^{1/D}-i}_{\mathrm{Haar}(n^{1-1/D})}\right\|_1 \\ &\le \sum_{i=1}^{n^{1/D}}\left\|G^{(t)}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}}\right\|_1^{i-1}\left\|G^{(t)}_{\mu^{\mathrm{lattice},n^{1-1/D}}_{D-1,c,s}} - G^{(t)}_{\mathrm{Haar}(n^{1-1/D})}\right\|_1\left\|G^{(t)}_{\mathrm{Haar}(n^{1-1/D})}\right\|_1^{n^{1/D}-i} \end{aligned} \tag{80}$$

$$\le \sum_{i=1}^{n^{1/D}}(t! + \epsilon_{D-1,n})^{i-1}\epsilon_{D-1,n}t!^{n^{1/D}-i} \tag{81}$$

$$\le n^{1/D}(t! + \epsilon_{D-1,n})^{n^{1/D}}d^{-\Omega(n^{1/(D-1)})}. \tag{82}$$

This final expression is $\leq d^{-\Omega(n^{1/D})}$ for n sufficiently large relative to d, t, D. Equation (81) uses the induction hypothesis as well as the fact that $G^{(t)}_{\mathrm{Haar}(m)}$ is a projector of rank $\leq t!$ for any m. (In fact this is an equality when $m \geq \ln(t)$.) This last fact is standard and can be found in Lemma 17 of [13], with the relevant math background in [30,33].

For (2), we observe that $(G_{\mathrm{Planes}(D)} g^s_{\mathrm{Rows}(D,n)})^c - G^{(t)}_{\mathrm{Haar}}$ has rank $t!^{O(n^{1/D})}$ so the cost of moving to the infinity norm is moderate:

$$\left\|(G_{\mathrm{Planes}(D)} g^s_{\mathrm{Rows}(D,n)})^c - G^{(t)}_{\mathrm{Haar}}\right\|_1 \leq t!^{O(n^{1/D})} \left\|(G_{\mathrm{Planes}(D)} g^s_{\mathrm{Rows}(D,n)})^c - G^{(t)}_{\mathrm{Haar}}\right\|_\infty \tag{83}$$

$$= t!^{O(n^{1/D})} \left\|G_{\mathrm{Planes}(D)} g^s_{\mathrm{Rows}(D,n)} - G^{(t)}_{\mathrm{Haar}}\right\|^c_\infty \tag{84}$$

We now bound $\left\|G_{\mathrm{Planes}(D)} g^s_{\mathrm{Rows}(D,n)} - G^{(t)}_{\mathrm{Haar}}\right\|_\infty$ using a variant of the proof of part 3 of this theorem.

$$\left\|G_{\mathrm{Planes}(D)} g^s_{\mathrm{Rows}(D,n)} - G^{(t)}_{\mathrm{Haar}}\right\|_\infty \leq \left\|g^s_{\mathrm{Rows}(D,n)} - G^{\otimes n^{1-1/D}}_{r_{1,1}}\right\|_\infty + \left\|G_{\mathrm{Planes}(D)} G^{\otimes n^{1-1/D}}_{r_{1,1}} - G^{(t)}_{\mathrm{Haar}}\right\|^c_\infty \tag{85}$$

Using [13] and Lemma 29

$$\left\|g^s_{\mathrm{Rows}(D,n)} - G^{\otimes n^{1-1/D}}_{r_{1,1}}\right\|_\infty \leq O(n^{1-1/D}) \left\|g^s_{r_{1,1}} - G_{r_{1,1}}\right\|_\infty = \frac{1}{d^{\Omega(n^{1/D})}}.$$

Moreover, using lemma 40 $\left\|G_{\mathrm{Planes}(D)} G^{\otimes n^{1-1/D}}_{r_{1,1}} - G^{(t)}_{\mathrm{Haar}}\right\|^c_\infty = \frac{1}{d^{\Omega(n^{1/D})}}$.

This completes the proof by taking the constant in the $\Omega(n^{1/D})$ in the last exponent sufficiently larger than the constant in the $O(n^{1/D})$ exponent in (85). Here we are ignoring the dependence on d, t, D. Taking this into account properly would yield a depth that scales polynomially with with t, with the degree of the polynomial depending on D.

4. Define $\epsilon_{D,n} := \left\|G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}} - G^{(t)}_{\mathrm{Haar}}\right\|_\infty$. By induction assume $\epsilon_{D,n} = 1/d^{\Omega(n^{1/D})}$ for any integers n and D. Assuming $\epsilon_{D-1,n} = 1/d^{\Omega(n^{1/D-1})}$ for all n, we show that $\epsilon_{D,n} = 1/d^{\Omega(n^{1/D})}$.

$$\begin{aligned}
\epsilon_{D,n} &:= \left\|G^{(t)}_{\mu^{\mathrm{lattice},n}_{2,c,s}} - G^{(t)}_{\mathrm{Haar}}\right\|_\infty \\
&\leq \left\|G_{\mu^{\mathrm{lattice},n}_{D,c,s}} - \tilde{G}_{n,D,c}\right\|_\infty + \left\|\tilde{G}_{n,D,c} - G^{(t)}_{\mathrm{Haar}}\right\|_\infty \\
&\leq \mathrm{poly}(n) \cdot \left\|(g^s_{r_{1,1}})^{\otimes n^{1-1/D}} - G^{\otimes n^{1-1/D}}_{r_{1,1}}\right\|_\infty \\
&\quad + \left\|G_{\mathrm{Rows}(D,n)} \tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c} - G^{(t)}_{\mathrm{Haar}}\right\|^c_\infty \\
&\leq \mathrm{poly}(n) \cdot \left\|(g^s_{r_{1,1}})^{\otimes n^{1-1/D}} - G^{\otimes n^{1-1/D}}_{r_{1,1}}\right\|_\infty
\end{aligned}$$

$$
\begin{aligned}
&+\left\| G_{\mathrm{Rows}(D,n)}(\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c} - F_{\mathrm{Rows}(D,n)}) \right. \\
&\left. +G_{\mathrm{Rows}(D,n)} F_{\mathrm{Rows}(D,n)} - G^{(t)}_{\mathrm{Haar}} \right\|^c_\infty \\
&\leq O(n)1/d^{\Omega(n^{1/D})} + O(n)\epsilon_{D-1,n^{1-1/D}} + 1/d^{\Omega(n^{1-1/D})c} \\
&\leq d^{-\Omega(n^{1/D})} + 1/d^{\Omega(n^{1/D})} + 1/d^{\Omega(n^{1-1/D})c} \\
&\leq d^{-\Omega(n^{1/D})}. \qquad (86)
\end{aligned}
$$

These steps follow from the proof of part 2.

□

3.5. Proofs of the basic lemmas stated in Sect. 3.1.

3.5.1. Comparison lemma for random quantum circuits.

Lemma (Restatement of Lemma 34). *Suppose we have the following cp ordering between superoperators $\mathcal{A}_1 \preceq \mathcal{B}_1, \ldots, \mathcal{A}_t \preceq \mathcal{B}_t$. Then $\mathcal{A}_t \ldots \mathcal{A}_1 \preceq \mathcal{B}_t \ldots \mathcal{B}_1$.*

Proof. We first prove the following claim

Claim. If $\mathcal{A} \preceq \mathcal{B}$ and $\mathcal{C} \preceq D$ are cp maps, then $\mathcal{AC} \preceq \mathcal{BD}$.

Proof. The class of cp maps is closed under composition and addition. Therefore $\mathcal{BD} - \mathcal{AC} = (\mathcal{B} - \mathcal{A})D + \mathcal{A}(\mathcal{D} - \mathcal{C})$ is cp.

The proof (of Lemma 34) is by induction. We show for all $1 \leq i \leq t$

$$\mathcal{A}_i \ldots \mathcal{A}_1 \preceq \mathcal{B}_i \ldots \mathcal{B}_1. \qquad (87)$$

Clearly this is true for $i = 1$. Suppose also this is true for $1 < k < t$. So $\mathcal{A}_i \ldots \mathcal{A}_1 \preceq \mathcal{B}_i \ldots \mathcal{B}_1$ and $\mathcal{A}_{i+1} \preceq \mathcal{B}_{i+1}$, and using the claim $\mathcal{A}_{i+1} \ldots \mathcal{A}_1 \preceq \mathcal{B}_{i+1} \ldots \mathcal{B}_1$. □

Corollary (Restatement of Corollary 35). *If $K_1, \ldots, K_t$ are respectively the moments superoperators of $\epsilon_1, \ldots, \epsilon_t$-approximate strong k-designs each on a potentially different subset of qudits, then*

$$
\begin{aligned}
&Ch\left[G^{(t)}_{Haar(S_1)} \ldots G^{(t)}_{Haar(S_t)}\right](1-\epsilon_1)\ldots(1-\epsilon_t) \preceq K_1 \ldots K_t \\
&\preceq Ch\left[G^{(t)}_{Haar(S_1)} \ldots G^{(t)}_{Haar(S_t)}\right](1+\epsilon_1)\ldots(1+\epsilon_t). \qquad (88)
\end{aligned}
$$

Proof. This is immediate from Lemma 34, Definition 16, and the observation that if $A \preceq B$ then $A \otimes \mathrm{id} \preceq B \otimes \mathrm{id}$. □

3.5.2. Bound on the value of off-diagonal monomials.

Lemma (Restatement of Lemma 37). *Let $\delta > 0$. Assume that $Ch\left[G^{(t)}_\mu\right]$ and $Ch\left[G^{(t)}_\nu\right]$ are two moment superoperators that satisfy the following completely positive ordering*

$$(1-\delta)\cdot Ch\left[G^{(t)}_\nu\right] \preceq Ch\left[G^{(t)}_\mu\right] \preceq (1+\delta)\cdot Ch\left[G^{(t)}_\nu\right]. \qquad (89)$$

Let $\mathcal{O}$ and $\mathcal{D}$ be respectively the set of off-diagonal and diagonal indices for monomials. Then

$$\max_{x\in\mathcal{O}} |\mathrm{Tr}\left(xG^{(t)}_\mu\right)| \leq \max_{x\in\mathcal{O}} |\mathrm{Tr}\left(xG^{(t)}_\nu\right)|(1+\delta) + 2\delta \cdot \max_{y\in\mathcal{D}} |\mathrm{Tr}\left(yG^{(t)}_\nu\right)|. \qquad (90)$$

Proof. Let $\phi_N := |\phi_N\rangle\langle\phi_N|$ for

$$|\phi\rangle := \frac{1}{\sqrt{N}} \sum_{x\in[d]^n} |x\rangle\,|x\rangle \tag{91}$$

be the n-qudit maximally entangled state, and $N = d^n$.

We use the following standard lemma which we leave without proof (see [13] for e.g.)

Lemma 44. *Let μ and ν be two distributions over the n-qudit unitary group then $Ch[G_\mu^{(t)}] \preceq Ch[G_\nu^{(t)}]$ if and only if*

$$\left(Ch\left[G_\nu^{(t)}\right] \otimes \mathrm{id} - Ch\left[G_\mu^{(t)}\right] \otimes \mathrm{id}\right) \phi_N^{\otimes t} \tag{92}$$

is a psd matrix.

We now adapt Lemma 37 to Lemma 44. First,

$$\begin{aligned}\phi_N^{\otimes t} &= \frac{1}{N^t} \sum |i_1,\ldots,i_t\rangle\langle j_1,\ldots,j_t| \otimes |i_1,\ldots,i_t\rangle\langle j_1,\ldots,j_t| \\ &\equiv \frac{1}{N^t} \sum |i\rangle\langle j| \otimes |i\rangle\langle j|\,. \end{aligned} \tag{93}$$

For $i, j, k, l \in [d]^{nt}$, if we define

$$M_{k,i,l,j}^{(\mu,t)} = \langle k|\,\mathrm{Ch}\left[G_\mu^{(t)}\right](|i\rangle\langle j|)\,|l\rangle\,. \tag{94}$$

Therefore

$$(\mathrm{Ch}[G_\mu^{(t)}] \otimes \mathrm{id})\phi_N^{\otimes t} = \frac{1}{N^t} \sum M_{a,b,c,d}^{(\mu,t)} |a\rangle\langle c| \otimes |b\rangle\langle d|\,, \tag{95}$$

and

$$(\mathrm{Ch}[G_{\mathrm{Haar}}^{(t)}] \otimes \mathrm{id})\phi_N^{\otimes t} = \frac{1}{N^t} \sum M_{a,b,c,d}^{(\mathrm{Haar},t)} |a\rangle\langle c| \otimes |b\rangle\langle d|\,. \tag{96}$$

Therefore since $\mathrm{Ch}\left[G_\mu^{(t)}\right] \le (1+\delta)\mathrm{Ch}[G_\nu^{(t)}]$ the following matrix

$$\begin{aligned} A &= ((1+\delta)\mathrm{Ch}[G_{\mathrm{Haar}}^{(t)}] \otimes \mathrm{id} - \mathrm{Ch}\left[G_\mu^{(t)}\right] \otimes \mathrm{id})\phi_N^{\otimes t} \\ &= \frac{1}{N^t} \sum ((1+\delta)M_{a,b,c,d}^{(\mathrm{Haar},t)} - M_{a,b,c,d}^{(\mu,t)})\,|a\rangle\,|b\rangle\langle c|\langle d|\,. \end{aligned} \tag{97}$$

Is psd. We use the following fact about psd matrices which we leave without proof.

Fact— if A is psd then the absolute maximum of off-diagonal terms in A is at most the absolute maximum diagonal term.

Then using the above fact

$$\max_{x\in\mathcal{O}} |(1+\delta)\mathrm{Tr}\left(G_\nu^{(t)}x\right) - \mathrm{Tr}\left(G_\mu^{(t)}x\right)| \le \max_{y\in\mathcal{D}} |(1+\delta)\mathrm{Tr}(G_\nu^{(t)}y) - \mathrm{Tr}\left(G_\mu^{(t)}y\right)|. \tag{98}$$

Hence

$$\max_{x\in\mathcal{O}} |\mathrm{Tr}\left(G_\mu^{(t)} x\right)| \leq \max_{x\in\mathcal{O}} |\mathrm{Tr}\left(G_\nu^{(t)} x\right)|(1+\delta) + \max_{y\in\mathcal{D}} |(1+\delta)\mathrm{Tr}\left(G_\nu^{(t)} y\right) - \mathrm{Tr}\left(G_\mu^{(t)} y\right)|. \tag{99}$$

Now if $y \in \mathcal{D}$

$$\mathrm{Tr}\left(G_\nu^{(t)} x\right)(1-\delta) \leq \mathrm{Tr}\left(G_\mu^{(t)} x\right) \leq \mathrm{Tr}\left(G_\nu^{(t)} x\right)(1+\delta). \tag{100}$$

then using this in (99)

$$\max_{x\in\mathcal{O}} |\mathrm{Tr}\left(G_\mu^{(t)} x\right)| \leq \max_{x\in\mathcal{O}} |\mathrm{Tr}\left(G_\nu^{(t)} x\right)|(1+\delta) + 2\delta \cdot \max_{y\in\mathcal{D}} \mathrm{Tr}\left(G_\nu^{(t)} y\right). \tag{101}$$

□

3.5.3. Bounds on the moments of the Haar measure.

Lemma (Restatement of Lemma 38). *Let $G^{(t)}_{Haar(m)}$ be the quasi-projector operator for the Haar measure on m qudits. Then*

$$\max_y \left\| G^{(t)}_{Haar(m)} y G^{(t)}_{Haar(m)} \right\|_1 \leq \frac{t^{O(t)}}{d^{mt}}. \tag{102}$$

Here the maximization is taken over matrix elements in the computational basis like $y = |i_1, \ldots, i_t, i'_1, \ldots, i'_t\rangle\langle j_1, \ldots, j_t, j'_1, \ldots, j'_t|$. Each label (e.g. i_j) is in $[d]^m$.

Proof. First observe that

$$\max_y \|G^{(t)}_{\mathrm{Haar}} y G^{(t)}_{\mathrm{Haar}}\|_1 = \max_{a,b} \mathrm{Tr}\sqrt{G^{(t)}_{\mathrm{Haar}} |a\rangle\langle b| G^{(t)}_{\mathrm{Haar}} G^{(t)}_{\mathrm{Haar}} |b\rangle\langle a| G^{(t)}_{\mathrm{Haar}}} \tag{103}$$

$$= \max_{a,b} \sqrt{\langle a|G^{(t)}_{\mathrm{Haar}}|a\rangle \cdot \langle b|G^{(t)}_{\mathrm{Haar}}|b\rangle} \tag{104}$$

$$= \max_i \langle i|G^{(t)}_{\mathrm{Haar}}|i\rangle\,. \tag{105}$$

The below lemma concludes the proof.

Lemma 45 (Moments of the Haar measure). *The largest t-th monomial moment of the Haar measure is at most $\frac{t!}{d^{tm}}$.*

Proof. Consider a particular balanced moment of Haar, using Hölder's inequality

$$\mathop{\mathbb{E}}_{C\sim\mathrm{Haar}} |C_{a_1,b_1} \ldots C_{a_t,b_t}|^2 \leq \prod_{i\in[k]} (\mathbb{E}|C_{a_i,b_i}|^{2k})^{1/k} \leq k!/d^{km}. \tag{106}$$

If the moment is not balanced the expectation is zero and hence the bound still works. Here we have used a closed form expression for $\mathbb{E}|C_{a_i,b_i}|^{2k}$, see corollary 2.4 and proposition 2.6 of [23] for a reference. □

3.6. Proofs of the projector overlap lemmas from section 3.2.

3.6.1. Extended quasi-orthogonality of permutation operators with application to random circuits on 2-dimensional lattices. In this section we prove Lemma 39

Lemma (Restatement of Lemma 39). $\|G_C G_R - G^{(t)}_{Haar}\|_\infty \leq 1/d^{\Omega(\sqrt{n})}$.

First, we need a description of the subspaces the projectors G_R, G_C and G_{Haar} project onto. Consider a $\sqrt{n} \times \sqrt{n}$ square lattice with n qudits as the collection of points $A := [\sqrt{n}] \times [\sqrt{n}]$. We use the following interpretation of the Hilbert space a quasi-projector acts on. This interpretation is also used in [13]. Denote $R_j(C_j)$ as the j-th row (column) of A for $j \in [\sqrt{n}]$. Here we assume each point of A consists of t pairs of qudits, each with local dimension d. Thereby, the lattice becomes the Hilbert space $\mathcal{H} := \bigotimes_{(x,y)\in A} \mathbb{C}^{d^{2t}_{(x,y)}}$, and has dimension d^{2tn}.

We are interested in a certain subspace of $\mathcal{H}$, and in order to understand it we need the following notation. For each point $(x, y) \in A$ we assign the quantum state $|\psi_\pi\rangle := (I \otimes V(\pi))\,|\Phi_{d,t}\rangle$, for each permutation $\pi \in S_t$. $|\Phi_{d,t}\rangle$ is the maximally entangled state $\frac{1}{\sqrt{d^t}} \sum_{x\in[d]^t} |x, x\rangle$, $V : S_t \to GL(\mathbb{C}^{d^{2t}})$, is a representation of S_t with the map $V(\pi) : |x(1), x_2, \ldots, x_t\rangle \mapsto |x_{\pi^{-1}(1)}, x_{\pi^{-1}(2)}, \ldots, x_{\pi^{-1}(t)}\rangle$, and S_t is the symmetric group over t elements.

Given these definitions define the following basis states in $\mathcal{H}$:

$$|R_{\pi_1,\pi_2,\ldots,\pi_{\sqrt{n}}}\rangle := \bigotimes_{v_1\in R_1} |\psi_{\pi_1}\rangle_{v_1} \otimes \bigotimes_{v_2\in R_2} |\psi_{\pi_2}\rangle_{v_2} \otimes \ldots \otimes \bigotimes_{v_{\sqrt{n}}\in R_{\sqrt{n}}} |\psi_{\pi_{\sqrt{n}}}\rangle_{v_{\sqrt{n}}}, \quad (107)$$

and,

$$|C_{\pi_1,\pi_2,\ldots,\pi_{\sqrt{n}}}\rangle := \bigotimes_{v_1\in C_1} |\psi_{\pi_1}\rangle_{v_1} \otimes \bigotimes_{v_2\in C_2} |\psi_{\pi_2}\rangle_{v_2} \otimes \ldots \otimes \bigotimes_{v_{\sqrt{n}}\in C_{\sqrt{n}}} |\psi_{\pi_{\sqrt{n}}}\rangle_{v_{\sqrt{n}}}, \quad (108)$$

for each $\sqrt{n}$ tuple of permutations $(\pi_1, \pi_2, \ldots, \pi_{\sqrt{n}}) \in S_t^{\sqrt{n}}$. Here $S_t^{\sqrt{n}}$ is the $\sqrt{n}$-fold Cartesian product $S_t \times \ldots \times S_t$ of S_t with itself. Denote $H_{t,n}$ as the subset consisting of tuples of permutations in which not all of the permutations are equal. For example, elements like $(\pi, \pi, \ldots, \pi)$ are not contained in this set. Notice that these basis are not orthogonal to each other and if $t > d^n$ these are not even linearly independent.

Here we define two vector spaces $V_R, V_C \subseteq \mathcal{H}$, with:

$$V_R := \operatorname*{span}_{\mathbb{C}} \left\{ |R_{\pi_1,\pi_2,\ldots,\pi_{\sqrt{n}}}\rangle : (\pi_1, \pi_2, \ldots, \pi_{\sqrt{n}}) \in S_t^{\sqrt{n}} \right\}, \quad (109)$$

and,

$$V_C := \operatorname*{span}_{\mathbb{C}} \left\{ |C_{\pi_1,\pi_2,\ldots,\pi_{\sqrt{n}}}\rangle : (\pi_1, \pi_2, \ldots, \pi_{\sqrt{n}}) \in S_t^{\sqrt{n}} \right\}, \quad (110)$$

and we call them row and column vector spaces, respectively. Also, denote the intersection between them by $V_{\text{Haar}} := V_R \cap V_C$. Equivalently:

$$V_{\text{Haar}} = \operatorname*{span}_{\mathbb{C}} \left\{ \bigotimes_{v\in A} |\psi_\pi\rangle_v : \pi \in S_t \right\}. \quad (111)$$

Then define $\tilde{V}_R := V_R \cap V_H^{\perp}$ and $\tilde{V}_C := V_C \cap V_H^{\perp}$. Define the angle between two vector spaces A and B as

$$\cos \measuredangle(A, B) := \max_{x \in A, y \in B} \langle x, y \rangle . \tag{112}$$

We need the following definition of a Gram matrix

Definition 46 (*Gram matrix*). Let $v_1, \ldots, v_{\mathrm{Rows}(D,n)}$ be normal vectors that are not necessarily orthogonal to each other. Then the Gram matrix corresponding to this set of vectors is defined as $[J_{ij}] = \langle v_i | v_j \rangle$.

We also need the following lemma

Lemma 47 (Perron-Frobenius [53]). *If A is a (not necessarily symmetric) d-dimensional matrix, then:*

$$||A||_\infty \leq \sqrt{\max_{i \in [d]} \sum_j |A_{i,j}| \cdot \max_{j \in [d]} \sum_i |A_{i,j}|}. \tag{113}$$

Let G_R, G_C and G_{Haar} be the quasi-projectors defined in Sect. 2.1. From [13] we know that G_R, G_C and G_{Haar} are indeed projectors onto V_R, V_C and V_{Haar}, respectively. Define the inner-product matrix between V_R and V_C with matrix Q with entries:

$$[Q]_{g,h} := \langle R_g | C_h \rangle , \qquad g, h \in H_{t,n}. \tag{114}$$

The goal is to prove $\|G_C G_R - G_{\mathrm{Haar}}^{(t)}\|_\infty \leq 1/d^{\Omega(\sqrt{n})}$. This basically means that the composition of G_R and G_C is close to G_{Haar}.

Also let $c_{d,n,t} = \frac{1}{1 - \frac{\sqrt{nt}(t-1)}{2d\sqrt{n}}}$ be a number very close to 1.

The proof is in three main steps. First we relate $\|G_C G_R - G_{\mathrm{Haar}}\|_\infty$ to $\measuredangle(\tilde{V}_R, \tilde{V}_C)$:

Proposition 48. $\|G_C G_R - G_{Haar}\|_\infty \leq \cos^2 \measuredangle(\tilde{V}_R, \tilde{V}_C)$.

Next, we relate $\measuredangle(\tilde{V}_R, \tilde{V}_C)$ to $||Q||_\infty$

Proposition 49. $|\cos \measuredangle(\tilde{V}_R, \tilde{V}_C)| \leq c_{d,n,t} . ||Q||_\infty$

Then we bound $||Q||_\infty$:

Proposition 50. $||Q||_\infty \leq \left(\frac{1}{d} + \frac{1}{d^{\sqrt{n}-1}} + \frac{2t^2}{d^{\sqrt{n}}}\right)^{\sqrt{n}}$.

Propositions 48, 49 and 50 imply the proof of Lemma 39.

Proof of Proposition 48. We use the following result of Jordan

Proposition 51 (Jordan). *if P and Q are two projectors, then the Hilbert space V they act on can be decomposed, as a direct sum, into one-dimensional or two-dimensional subspaces, all of which are invariant under the action of both P and Q at the same time.*

which implies

Corollary 52. *There are orthonormal basis* $e_1, \ldots, e_K$, $f_1, \ldots, f_K$, $q_1, \ldots, q_T$, *and angles* $0 \geq \theta_1 \geq \theta_2 \geq \ldots \geq \theta_K \leq \pi/2$ *such that:*

$$V_R = \operatorname*{span}_{\mathbb{C}} \{e_1, \ldots, e_K, q_1, \ldots, q_T\}, \tag{115}$$

and

$$V_C = \operatorname*{span}_{\mathbb{C}} \{\cos\theta_1 e_1 + \sin\theta_1 f_1, \ldots, \cos\theta_K e_K + \sin\theta_K f_K, q_1, \ldots, q_T\}, \tag{116}$$

and

$$V_{Haar} = \operatorname*{span}_{\mathbb{C}} \{q_1, \ldots, q_T\}. \tag{117}$$

In other words, both G_R and G_C can be decomposed into 2×2 blocks, each corresponding to one of the angles θ_i, such that G_C on this block looks like

$$G_C^{2\times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{118}$$

and G_R

$$G_R^{2\times 2} = \begin{pmatrix} \cos^2\theta_i & \sin\theta_i \cos\theta_i \\ \sin\theta_i \cos\theta_i & \sin^2\theta_i \end{pmatrix}. \tag{119}$$

Hence $G_C G_R$ looks like

$$G_C^{2\times 2} G_R^{2\times 2} = \begin{pmatrix} \cos^2\theta_i & \sin\theta_i \cos\theta_i \\ 0 & 0 \end{pmatrix}, \tag{120}$$

which has largest singular value $|\cos^2\theta_i|$. Propositions 49 and 50 along with this observation imply that the largest singular value of $G_C G_R - G_{\text{Haar}}$ is $1/d^{n^{O(n^{1/D})}}$. □

Proof of Proposition 49. An arbitrary normal vector in $\tilde{V}_R$ can be written as $|\psi_x\rangle = \frac{\sum_{\tilde{\pi}\in H_{t,n}} x_{\tilde{\pi}} |R_{\tilde{\pi}}\rangle}{\sqrt{\sum_{\tilde{\pi},\tilde{\sigma}\in H_{t,n}} x_{\tilde{\pi}} x_{\tilde{\sigma}} \langle R_{\tilde{\pi}}|R_{\tilde{\sigma}}\rangle}}$. Let $|x\rangle$ be a vector with entries corresponding to $x_{\tilde{\pi}_1,\ldots,\tilde{\pi}_{\sqrt{n}}}$. Similarly, a typical vector inside $\tilde{V}_C$ can be represented as $|\psi_y\rangle = \frac{\sum_{\tilde{\pi}\in H_{t,n}} y_{\tilde{\pi}} |R_{\tilde{\pi}}\rangle}{\sqrt{\sum_{\tilde{\pi},\tilde{\sigma}\in H_{t,n}} y_{\tilde{\pi}} y_{\tilde{\sigma}} \langle C_{\tilde{\pi}}|C_{\tilde{\sigma}}\rangle}}$. Also represent the corresponding vector $|y\rangle$ similarly.

Let $\tilde{J}$ and $\tilde{J}'$ be the Gram matrices corresponding to the basis described for $\tilde{V}_R$ and $\tilde{V}_C$, respectively. Then:

$$\begin{aligned} \langle\psi_x|\psi_y\rangle &= \frac{\sum_{\tilde{\pi},\tilde{\sigma}\in H_{t,n}} x_{\tilde{\pi}} \langle R_{\tilde{\pi}}|R_{\tilde{\sigma}}\rangle y_{\tilde{\sigma}}}{\sqrt{\sum_{\tilde{\pi},\tilde{\sigma}\in H_{t,n}} x_{\tilde{\pi}} x_{\tilde{\sigma}} \langle R_{\tilde{\pi}}|R_{\tilde{\sigma}}\rangle} \cdot \sqrt{\sum_{\tilde{\pi},\tilde{\sigma}\in H_{t,n}} y_{\tilde{\pi}} y_{\tilde{\sigma}} \langle C_{\tilde{\pi}}|C_{\tilde{\sigma}}\rangle}} \\ &= \frac{\langle x|\, Q\, |y\rangle}{\sqrt{\langle x|\, \tilde{J}\, |x\rangle} . \sqrt{\langle y|\, \tilde{J}'\, |y\rangle}}. \end{aligned} \tag{121}$$

To see the equality we go through the below calculation.

$$\begin{aligned}\cos\phi &= \sup_{\substack{||x||_2=1\\||y||_2=1}} \frac{\langle x|\, Q\, |y\rangle}{\sqrt{\langle x|\, \tilde{J}\, |x\rangle}.\sqrt{\langle y|\, \tilde{J}'\, |y\rangle}} \\ &\leq c_{d,n,t}\ .\ \sup_{\substack{||x||_2=1\\||y||_2=1}} \langle x|\, Q\, |y\rangle \\ &\leq c_{d,n,t}\ .\ \sup_{\substack{||x||_2=1\\||y||_2=1}} \sqrt{\langle x|\, Q^\dagger Q\, |x\rangle}\ .\ ||y||_2 \\ &\leq c_{d,n,t}\ .\ ||Q||_\infty. \end{aligned} \tag{122}$$

For the second line we used the following proposition

Proposition 53. *If $\tilde{J}$ is the Gram matrix for the basis states, $|R.\rangle$ or $|C.\rangle$ in (107) and (108) for $\tilde{V}_R$ or $\tilde{V}_R$, then for any $|x\rangle$ with $||x||_2 = 1$:*

$$\langle x|\, \tilde{J}\, |x\rangle \geq \left(1 - \frac{\sqrt{n}t\,(t-1)}{2d\sqrt{n}}\right) = \frac{1}{c_{d,n,t}}. \tag{123}$$

For the third line we used Cauchy-Schwartz. □

In order to prove Proposition 50 we need the following tool. If $x\vec{(}1), \vec{x_2}, \ldots, \vec{x_K}$ are d-dimensional vectors the multi-product of them is defined to be:

$$\mathsf{multiprod}(\vec{x_1}, \vec{x_2}, \ldots, \vec{x_K}) := \sum_{i=1}^{d} x_{1i}x_{2i}\ldots x_{Ki}. \tag{124}$$

Proposition 54 (Majorization). *Let $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_K$ be d-dimensional, non-negative and real vectors. If $\vec{x}_i^{\downarrow}$ is $\vec{x}_i$ in descending order, then:*

$$\mathsf{multiprod}\,(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_K) \leq \mathsf{multiprod}(\vec{x}_1^{\downarrow}, \vec{x}_2^{\downarrow}, \ldots, \vec{x}_K^{\downarrow}). \tag{125}$$

Proof. The $K = 2$ version of claim is that $\langle \vec{x}(1), \vec{x}_2\rangle \leq \langle \vec{x}(1)^{\downarrow}, \vec{x}_2^{\downarrow}\rangle$. This is a standard fact. To prove it, observe that WLOG we can assume $\vec{x}(1) = \vec{x}(1)^{\downarrow}$. Then for any out-of-order pair $x_{2i} < x_{2j}$ with $i < j$, we will increase $\langle \vec{x}(1), \vec{x}_2\rangle$ by swapping x_{2i} and x_{2j}. Applying this repeatedly we end with $\langle \vec{x}(1)^{\downarrow}, \vec{x}_2^{\downarrow}\rangle$.

This same argument works if we replace the inner product with a sum over the first $d' \leq d$ terms, i.e. $\sum_{i=1}^{d'} x_{1i}x_{2i}$. Thus the same argument shows that

$$\vec{x}_1 \circ \vec{x}_2 \preceq \vec{x}_1^{\downarrow} \circ \vec{x}_2^{\downarrow}. \tag{126}$$

The proposition now follows by induction on K. □

We also need the following upper bound:

Proposition 55. *Let $e \in S_t$ be the identity permutation. Define $f_t : \mathbb{R}_{>1} \to \mathbb{R}_{>1}$ with the map:*

$$f_t(\alpha) = \sum_{\sigma \in S_t} \frac{1}{\alpha^{\mathsf{dist}(e,\sigma)}}, \tag{127}$$

for $\alpha > 1$. Then as long as $2t^2 \leq \alpha$

$$f_t(\alpha) \leq 1 + \frac{2t^2}{\alpha}. \tag{128}$$

For $\sigma_1, \ldots, \sigma_M \in S_t$ define the function:

$$h(D, t, \sigma_1, \ldots, \sigma_M) := \sum_{\pi \in S_t} \frac{1}{D^{\mathsf{dist}(\pi,\sigma_1)+\ldots+\mathsf{dist}(\pi,\sigma_M)}}. \tag{129}$$

Proposition 56. *Let $(\sigma_1, \ldots, \sigma_M) \in H$ be permutations that not all of them are equal to each other then:*

$$h(D, t, \sigma_1, \ldots, \sigma_M) \leq \frac{1}{D} + \frac{1}{D^{M-1}} + \frac{2t^2}{D^M}. \tag{130}$$

Proof of Proposition 50. In or to prove this, we show that the sum of terms in each row is a small number. Then use Lemma 47 to obtain the result. Consider the particular row $(\sigma_1, \ldots, \sigma_{sqrtn}) \in H$, then the sum of terms in each row is:

$$\sum_{(\pi_1,\ldots,\pi_{\sqrt{n}}) \in H} \langle R_{\pi_1,\ldots,\pi_{\sqrt{n}}} | C_{\sigma_1,\ldots,\sigma_{\sqrt{n}}} \rangle = \sum_{\pi_1,\ldots,\pi_{\sqrt{n}} \in S_t} \langle R_{\pi_1,\ldots,\pi_{\sqrt{n}}} | C_{\sigma_1,\ldots,\sigma_{\sqrt{n}}} \rangle - \sum_{\pi \in S_t} \langle R_{\pi,\ldots,\pi} | C_{\sigma_1,\ldots,\sigma_{\sqrt{n}}} \rangle . \tag{131}$$

The lower bound:

$$\sum_{\pi \in S_t} \langle R_{\pi,\ldots,\pi} | C_{\sigma_1,\ldots,\sigma_{\sqrt{n}}} \rangle \geq 0, \tag{132}$$

is good enough. The goal is to find a good upper bound for $S := \sum_{\pi_1,\ldots,\pi_{\sqrt{n}} \in S_t} \langle R_{\pi_1,\ldots,\pi_{\sqrt{n}}} | C_{\sigma_1,\ldots,\sigma_{\sqrt{n}}} \rangle$. But S simplifies to:

$$S = \left(\sum_{\pi \in S_t} \frac{1}{d^{\mathsf{dist}(\pi,\sigma_1)+\ldots+\mathsf{dist}(\pi,\sigma_{\sqrt{n}})}} \right)^{\sqrt{n}} = h(d, t, \sigma_1, \ldots, \sigma_{\sqrt{n}})^{\sqrt{n}}. \tag{133}$$

Now we use Proposition 56 and find the upper bound:

$$S \leq \left(\frac{1}{d} + \frac{1}{d^{\sqrt{n}-1}} + \frac{2t^2}{d^{\sqrt{n}}} \right)^{\sqrt{n}}. \tag{134}$$

Which is a global maximum and in turn is a bound on the ∞-norm. □

Proof of Proposition 53. We will prove the statement for the row space, and the same thing works for the column space. First, for any normal vector $|x\rangle$, $\langle x|\tilde{J}_R|x\rangle \geq \lambda_{\min}(\tilde{J}_R)$. Let $J(\sqrt{n})$ be the Gram matrix for the Haar subspace on one row of the grid. The entries of $J(\sqrt{n})$ are according to:

$$J(\sqrt{n})_{\pi,\sigma} := \left(\frac{1}{d^{\mathsf{dist}(\pi,\sigma)}}\right)^{\sqrt{n}} = \left(\frac{1}{d^{\sqrt{n}}}\right)^{\mathsf{dist}(\pi,\sigma)}. \tag{135}$$

Let P be the projector that projects out the subspace spanned by $\{|R_{\pi,\ldots,\pi}\rangle : \pi \in S_t\}$. Then $\tilde{J} = PJ(\sqrt{n})^{\otimes\sqrt{n}}P^\dagger$. We first need the following proposition

Proposition 57. *If J is the Gram matrix of the vector space spanned by $\{|\psi_\pi\rangle^{\otimes m} : \pi \in S_t\}$., then:*

$$1 - \frac{t(t-1)}{2d^m} \leq \lambda_{\min}(J) \tag{136}$$

Using this proposition $\lambda_{\min}(J(\sqrt{n})) \geq 1 - \frac{t(t-1)}{2d^{\sqrt{n}}}$, and therefore $\lambda_{\min}(J(\sqrt{n})^{\otimes\sqrt{n}}) \geq (1 - \frac{t(t-1)}{2d^{\sqrt{n}}})^{\sqrt{n}} \geq 1 - \frac{\sqrt{n}t(t-1)}{2d^{\sqrt{n}}}$. This implies that $J(\sqrt{n})^{\otimes\sqrt{n}} \succeq I(1 - \frac{\sqrt{n}t(t-1)}{2d^{\sqrt{n}}})$, and therefore $\tilde{J} \succeq PP^\dagger(1 - \frac{\sqrt{n}t(t-1)}{2d^{\sqrt{n}}})$. This means that restricted to $\tilde{V}_R$ the minimum eigenvalue of $\tilde{J}_R$ is at least $(1 - \frac{\sqrt{n}t(t-1)}{2d^{\sqrt{n}}})$.

□

Proof of Proposition 56. Let $C = \{\sigma_1, \ldots, \sigma_M\}$. Then $h = h_1 + h_2$, where:

$$h_1 = \sum_{\pi\in C} \frac{1}{D^{\mathsf{dist}(\pi,\sigma_1)+\ldots+\mathsf{dist}(\pi,\sigma_M)}}, \tag{137}$$

and,

$$h_2 = \sum_{\pi\in S_t/C} \frac{1}{D^{\mathsf{dist}(\pi,\sigma_1)+\ldots+\mathsf{dist}(\pi,\sigma_M)}}. \tag{138}$$

We then find useful upper bounds for h_1 and h_2 separately. Suppose that C has distinct elements $\{\tau_1, \ldots, \tau_K\}$ with τ_1 appearing μ_1 times, τ_2 appearing μ_2 times, etc. Define

$$S = \left\{(\mu_1, \ldots, \mu_K) \in \mathbb{Z}^K_{\geq 0} : \mu_1 + \ldots + \mu_K = M, \max(i)\mu_i < M\right\} \tag{139}$$

$$P = \left\{(\mu_1, \ldots, \mu_K) \in S : \exists i, j,\ \mu_i = M - 1\ \&\ \mu_j = 1\right\} \tag{140}$$

Now we can bound h_1 by

$$\begin{aligned} h_1 &= \sum_{\pi\in C} \frac{1}{D^{\mu_1\mathsf{dist}(\pi,\tau_1)+\ldots+\mu_K\mathsf{dist}(\pi,\tau_K)}} \\ &\leq \max_{(\mu_1,\ldots,\mu_K)\in S} \frac{D^{\mu_1} + \ldots + D^{\mu_K}}{D^M} \\ &\leq \max_{(\mu_1,\ldots,\mu_K)\in \mathrm{conv}(P)} \frac{D^{\mu_1} + \ldots + D^{\mu_K}}{D^M} \end{aligned}$$

$$\leq \frac{D^{M-1}+D}{D^M}$$
$$= \frac{1}{D}+\frac{1}{D^{M-1}}. \tag{141}$$

Here conv denotes the convex hull and (141) uses the fact that $K \geq 2$ since $\sigma_1, \ldots, \sigma_M$ are assumes to be not all equal. To justify (141), observe that $f(\mu) = D^{\mu_1} + \ldots + D^{\mu_K}$ is a convex function and the maximization is over a convex set whose extreme points are P. Therefore the maximum is achieved at a point in P.

In order to find a bound on h_2, for each $\sigma \in C$ we will define the following vector $\vec{X}_\sigma$ whose entries are labeled by $\pi \in S_t$.

$$\vec{X}_{\sigma,\pi} = \begin{cases} 0 & \text{if } \pi \in C \\ D^{-\mathsf{dist}(\sigma,\pi)} & \text{if } \pi \notin C \end{cases} \tag{142}$$

Then $h_2 = \mathsf{multiprod}(\vec{X}_{\sigma_1}, \ldots, \vec{X}_{\sigma_M})$. We can use Proposition 54 to show that

$$h_2 = \mathsf{multiprod}(\vec{X}_{\sigma_1}, \ldots, \vec{X}_{\sigma_M}) \leq \mathsf{multiprod}(\vec{X}^{\downarrow}_{\sigma_1}, \ldots, \vec{X}^{\downarrow}_{\sigma_M}). \tag{143}$$

We will also define $\vec{X}_e$ (where e denotes the identity element of S_t) by

$$\vec{X}_{e,\pi} = D^{-\mathsf{dist}(e,\pi)}. \tag{144}$$

Observe that $\vec{X}_\sigma$ can be obtained from $\vec{X}_e$ by zeroing out the elements in locations corresponding to C and reordering the remaining elements. Thus for each $\sigma \in C$

$$\vec{X}^{\downarrow}_\sigma \preceq \vec{X}^{\downarrow}_e. \tag{145}$$

We use Proposition 54 again to bound

$$h_2 \leq \mathsf{multiprod}(\underbrace{\vec{X}_e, \ldots \vec{X}_e}_{M \text{ times}}) = f_t(D^{-M}) - 1 \leq \frac{2t^2}{D^M}. \tag{146}$$

□

3.6.2. Extended quasi-orthogonality of permutation operators with application to random circuits on D-dimensional lattices. In this section we prove lemmas 40, 41, 42 and 43. Before getting to the proof we go over some notation and definitions.

Let $\mathrm{Rows}(D,n) := \{r_1, \ldots, r_{n^{1-1/D}}\}$ be the set of rows in the D-th direction and let $V_{\mathrm{Rows}(D,n)}$ be the subspace $G_{\mathrm{Rows}(D,n)}$ projects onto. Then $V_{\mathrm{Rows}(D,n)} = V_{\mathrm{Haar}(r_1)} \otimes \ldots \otimes V_{\mathrm{Haar}(r_{n^{1-1/D}})}$. A spanning set for $V_{\mathrm{Rows}(D,n)}$ is $H_{\mathrm{Rows}(D,n)} := \{|D_{\sigma_1,\ldots,\sigma_{n^{1-1/D}}}\rangle : \sigma_1, \ldots, \sigma_{n^{1-1/D}} \in S_t\}$. Here $V_{\mathrm{Haar}(S)}$ is the Haar subspace (like V_{Haar}) on a subset of qudits S. $|D_{\sigma_1,\ldots,\sigma_{n^{1-1/D}}}\rangle$ is the basis state representing maximally entangled states for each qudit such that the qudits in the first row are permuted by π_1, the qudits in the second row are permuted by π_2, and so on. In other words:

$$|D_{\sigma_1,\ldots,\sigma_{n^{1-1/D}}}\rangle = \bigotimes_{r_i \in \mathrm{Rows}(D,n)} \bigotimes_{v \in r_i} |\psi_{\sigma_i}\rangle_v. \tag{147}$$

We view the D dimensional lattice as $n^{1/D}$ $D-1$-dimensional sub-lattices, each composed of $n^{1-1/D}$ qudits. More concretely, the full lattice is the set $A = [n^{1/D}]^D$. For $1 \le \beta \le n^{1/D}$, denote $p_\beta = \{(x(1), \ldots, x_{\mathrm{Rows}(D,n)}) \in A : x_{\mathrm{Rows}(D,n)} = \beta\}$. We denote the set of these lattices by $\mathrm{Planes}(D) := \{p_1, \ldots, p_{n^{1-1/D}}\}$. (This terminology is chosen to match the $D = 3$ case but the arguments here apply to any $D > 2$.) These lattices are connected to each other by the rows in $\mathrm{Rows}(D, n)$. $V_{\mathrm{Planes}(D)} = V_{\mathrm{Haar}(p_1)} \otimes \ldots \otimes V_{\mathrm{Haar}(p_{n^{1-1/D}})}$ is the span of $H_{\mathrm{Planes}(D)} := \{|F_{\pi_1,\ldots,\pi_{n^{1-1/D}}}\rangle : \pi_1, \ldots, \pi_{n^{1-1/D}} \in S_t\}$. Here $|F_{\pi_1,\ldots,\pi_{n^{1-1/D}}}\rangle$ is the basis state of maximally entangled states for each qudit, such that the qudits in p_1 are permuted by π_1, qudits in p_2 are permuted by π_2 and so on. In other words:

$$|F_{\pi_1,\ldots,\pi_{n^{1/D}}}\rangle = \bigotimes_{p_i \in \mathrm{Planes}(D)} \bigotimes_{v \in p_i} |\psi_{\pi_i}\rangle_v . \tag{148}$$

Then $G_{\mathrm{Planes}(D)}$ is the projector onto $V_{\mathrm{Planes}(D)}$.

Let $\tilde{V}_{\mathrm{Planes}(D)} := V_{\mathrm{Planes}(D)} \cap V_{\mathrm{Haar}}^{\perp}$ and $\tilde{V}_{\mathrm{Rows}(D,n)} =: V_{\mathrm{Rows}(D,n)} \cap V_{\mathrm{Haar}}^{\perp}$ be respectively the orthogonal complements of $V_{\mathrm{Planes}(D)}$ and $V_{\mathrm{Rows}(D,n)}$ with respect to V_{Haar}. Also define $\tilde{H}_{\mathrm{Rows}(D,n)}$ and $\tilde{H}_{\mathrm{Planes}(D)}$ the same as $H_{\mathrm{Rows}(D,n)}$ and $H_{\mathrm{Planes}(D)}$, excluding basis marked with permutations that are all equal to each other. For example, $F_{\pi,\ldots,\pi} \notin \tilde{H}_{\mathrm{Planes}(D)}$. Define the overlap matrix $[Q]_{gh} := \langle g|h\rangle$, for $g \in H_{\mathrm{Planes}(D)}$ and $h \in H_{\mathrm{Rows}(D,n)}$. Let $\tilde{J}_{\mathrm{Planes}(D)}$ and $\tilde{J}_{\mathrm{Rows}(D,n)}$ be the Gram matrices corresponding to $\tilde{H}_{\mathrm{Planes}(D)}$ and $\tilde{H}_{\mathrm{Rows}(D,n)}$, respectively. In other words, $[\tilde{J}_D]_{g,h} = \langle g|h\rangle$ for $g, h \in \tilde{H}_{\mathrm{Rows}(D,n)}$ and $[\tilde{J}_{\mathrm{Planes}(D)}]_{g,h} = \langle g|h\rangle$ for $g, h \in \tilde{H}_{\mathrm{Planes}(D)}$.

We first prove Lemma 40, which basically states that the composition of $G_{\mathrm{Rows}(D,n)}$ and $F_{\mathrm{Rows}(D,n)}$ is very close to $G_{\mathrm{Haar}}^{(t)}$, or equivalently, $\tilde{V}_{\mathrm{Rows}(D,n)}$ and $\tilde{V}_{\mathrm{Planes}(D)}$ are almost orthogonal:

Lemma (Restatement of Lemma 40). *Let $D = O(\ln n / \ln\ln n)$ with small enough constant factor, then $\|G_{Planes(D)} G_{Rows(D,n)} - G_{Haar}\|_\infty \le 1/d^{\Omega(n^{1-1/D})}$.*

Proof. The proof is very similar to the proof of Lemma 39. In particular, we need generalized versions of propositions 48, 49 and 50. The generalization of proposition 48 states that $\cos^2(\measuredangle(\tilde{V}_{\mathrm{Planes}(D)}, \tilde{V}_{\mathrm{Rows}(D,n)}))$ equals the largest singular value of $F_D G_{\mathrm{Rows}(D,n)} - G_{\mathrm{Haar}}$. Proposition 49 generalizes to the statement that the cosine of this angle is equal to

$$\frac{1}{\sqrt{\lambda_{\min}(\tilde{J}_{\mathrm{Planes}(D)})\lambda_{\min}(\tilde{J}_{\mathrm{Rows}(D,n)})}} \|Q\|_\infty \le c_{D,d,n,t} \|Q\|_\infty . \tag{149}$$

Where $1/c_{D,d,n,t}$ is a lower bound on $\sqrt{\lambda_{\min}(\tilde{J}_{\mathrm{Planes}(D)})\lambda_{\min}(\tilde{J}_{\mathrm{Rows}(D,n)})}$.

We first bound $\|Q\|_\infty$. Using Lemma 47

$$\|Q\|_\infty \le \sqrt{\max_h \sum_g Q_{gh} \max_g \sum_h Q_{gh}} =: \omega. \tag{150}$$

Similar to the calculations in Sect. 3.6.1

$$Q_{F_{\pi_1,\ldots,\pi_{n^{1-1/D}};D_{\sigma_1,\ldots,\sigma_{n^{1/D}}}}} = \frac{1}{d^{\sum_{i=1}^{n^{1-1/D}} \sum_{j=1}^{n^{1/D}} \mathrm{dist}(\pi_i,\sigma_j)}} . \tag{151}$$

Let α (β) be respectively the set of permutations $\sigma_1, \ldots, \sigma_{n^{1/D}}$ ($\pi_1, \ldots, \pi_{n^{1-1/D}}$) that are not all equal. We compute

$$\max_{\pi_1,\ldots,\pi_{n^{1-1/D}}\in\beta} \sum_{\sigma_1,\ldots,\sigma_{n^{1/D}}} \frac{1}{d^{\sum_{i=1}^{n^{1-1/D}}\sum_{j=1}^{n^{1/D}}\mathsf{dist}(\pi_i,\sigma_j)}}$$

$$= \max_{\pi_1,\ldots,\pi_{n^{1-1/D}}\in\beta} \Big(\sum_{\sigma} \frac{1}{d^{\sum_{i=1}^{n^{1-1/D}}\mathsf{dist}(\pi_i,\sigma)}}\Big)^{n^{1/D}} \tag{152}$$

$$= \max_{\pi_1,\ldots,\pi_{n^{1-1/D}}\in\beta} h(d^{n^{1/D}}, t, \pi_1, \ldots, \pi_{n^{1-1/D}}) \tag{153}$$

$$\leq \left(\frac{1}{d} + \frac{1}{d^{n^{1-1/D}-1}} + \frac{2t^2}{d^{n^{1-1/D}}}\right)^{n^{1/D}} \tag{154}$$

$$= \frac{1}{d^{\Omega(n^{1-1/D})}}. \tag{155}$$

and

$$\max_{\sigma_1,\ldots,\sigma_{n^{1/D}}\in\alpha} \sum_{\pi_1,\ldots,\pi_{n^{1-1/D}}} \frac{1}{d^{\sum_{i=1}^{n^{1-1/D}}\sum_{j=1}^{n^{1/D}}\mathsf{dist}(\pi_i,\sigma_j)}}$$

$$= \max_{\sigma_1,\ldots,\sigma_{n^{1/D}}\in\alpha} \Big(\sum_{\pi} \frac{1}{d^{\sum_{j=1}^{n^{1/D}}\mathsf{dist}(\pi,\sigma_j)}}\Big)^{n^{1/D}} \tag{156}$$

$$= \max_{\sigma_1,\ldots,\sigma_{n^{1/D}}\in\alpha} h(d^{n^{1-1/D}}, t, \sigma_1, \ldots, \sigma_{n^{1/D}}) \tag{157}$$

$$\leq \Big(\frac{1}{d} + \frac{1}{d^{n^{1/D}-1}} + \frac{2t^2}{d^{n^{1/D}}}\Big)^{n^{1-1/D}} \tag{158}$$

$$= \frac{1}{d^{\Omega(n^{1/D})}}. \tag{159}$$

Hence

$$\omega = \frac{1}{d^{\Omega(n^{1-1/D})}}. \tag{160}$$

Next, we have to show that $c_{D,d,n,t}$ is not too large. Using exactly the same steps in the proof of Proposition 53 we can show that

$$\lambda_{\min}(\tilde{J}_{\mathrm{Planes}(D)}) \geq 1 - \frac{n^{1/D}t(t-1)}{2d^{n^{1-1/D}}}, \tag{161}$$

and

$$\lambda_{\min}(\tilde{J}_{\mathrm{Rows}(D,n)}) \geq 1 - \frac{n^{1-1/D}t(t-1)}{2d^{n^{1/D}}}. \tag{162}$$

□

Next, we use this result to prove Lemma 41. Recall the expression $\tilde{G}_{n,D,c}$ from Definition 23

$$\tilde{G}_{n,D,c} = (\tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}} G_{\mathrm{Rows}(D,n)} \tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}})^c, \tag{163}$$

where c is a constant depending on D and t, but independent of n. Note that $\tilde{G}_{n,D,c} = \tilde{G}_{n,D,c}^\dagger$ if $\tilde{G}_{n^{1-1/D},D-1,c} = \tilde{G}_{n^{1-1/D},D-1,c}^\dagger$. Also let $\hat{G}_{n,D,c} := G_{\mathrm{Rows}(D,n)} \tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}} G_{\mathrm{Rows}(D,n)}$. Hence $\tilde{G}_{n,D,c} = \tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}} (\hat{G}_{n,D,c})^{c-1} \tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}}$.

Lemma (Restatement of Lemma 41). *Let $|x\rangle$ and $|y\rangle$ be two computational basis states. For small enough $D = O(\ln n / \ln \ln n)$ and large enough c, $|\langle x|\tilde{G}_{n,D,c} - G_{Haar}|y\rangle| \leq \frac{\epsilon}{d^{nt}}$ for some $\epsilon = 1/d^{\Omega(n^{1/D})}$.*

Proof. The proof is by induction. The base case $D = 2$ is by Lemma 39. We assume that for any large enough m, $\|\hat{G}_{m,D-1,c} - G_{\mathrm{Haar}}\|_\infty \leq \frac{1}{d^{O(m^{1/D-1})}}$, and we show that $\|\hat{G}_{n,D,c} - G_{\mathrm{Haar}}\|_\infty \leq \frac{1}{d^{\Omega(n^{1/D})}}$.

$$\begin{aligned}
\left\|\hat{G}_{n,D,c} - G_{\mathrm{Haar}}\right\|_\infty &\leq \left\| G_{\mathrm{Rows}(D,n)} \tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}} G_{\mathrm{Rows}(D,n)} - G_{\mathrm{Haar}} \right\|_\infty && (164)\\
&\leq \left\| \tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}} G_{\mathrm{Rows}(D,n)} - G_{\mathrm{Haar}} \right\|_\infty && (165)\\
&= \left\| (\tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}} - G_{\mathrm{Planes}(D)}) G_{\mathrm{Rows}(D,n)} \right. && (166)\\
&\quad \left. + G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)} - G_{\mathrm{Haar}} \right\|_\infty && (167)\\
&\leq \left\| (\tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}} - G_{\mathrm{Planes}(D)}) G_{\mathrm{Rows}(D,n)} \right\|_\infty && (168)\\
&\quad + \left\| G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)} - G_{\mathrm{Haar}} \right\|_\infty && (169)\\
&\leq \left\| \tilde{G}_{n^{1-1/D},D-1,c}^{\otimes n^{1/D}} - G_{\mathrm{Planes}(D)} \right\|_\infty \\
&\quad + \left\| G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)} - G_{\mathrm{Haar}} \right\|_\infty && (170)\\
&\leq n^{1/D} \left\| \tilde{G}_{n^{1-1/D},D-1,c} - G_{\mathrm{Haar}(p_1)} \right\|_\infty \\
&\quad + \left\| G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)} - G_{\mathrm{Haar}} \right\|_\infty && (171)\\
&\leq n^{1/D} \frac{1}{d^{O(n^{(1-1/D)\cdot 1/(D-1)})}} + 1/d^{\Omega(n^{1-1/D})} && (172)\\
&\leq \frac{n^{1/D}}{d^{\Omega(n^{1/D})}} + 1/d^{\Omega(n^{1-1/D})} && (173)\\
&\leq \frac{1}{d^{\Omega(n^{1/D})}}. && (174)
\end{aligned}$$

□

Lemma (Restatement of Lemma 41). *Let $|x\rangle$ and $|y\rangle$ be two computational basis states. For small enough $D = O(\ln n / \ln \ln n)$ and large enough c, $|\langle x|\tilde{G}_{n,D,c} - G_{Haar}|y\rangle| \leq \frac{\epsilon}{d^{nt}}$ for some $\epsilon = 1/d^{\Omega(n^{1/D})}$.*

Proof. The proof is by induction. Our induction hypothesis is $\max_x |\langle x|(\tilde{G}_{n,D,c} - G_{\text{Haar}})^2|x\rangle| \leq \frac{\epsilon}{d^{nt}}$. First, we show this bound (for sub-lattices of dimension $D-1$) implies the statement of this theorem:

$$\begin{aligned}
|\langle x|\tilde{G}_{n,D,c} - G_{\text{Haar}}|y\rangle| &= |\langle x|\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c}(\tilde{G}'^{c-1}_{n,D,c} - G_{\text{Haar}})\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c}|y\rangle| \\
&\leq \left\|\hat{G}_{n,D,c} - G_{\text{Haar}}\right\|_\infty^{c-1} \left\|\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c}|y\rangle\langle x|\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c}\right\|_1 \\
&\leq \left\|\hat{G}_{n,D,c} - G_{\text{Haar}}\right\|_\infty^{c-1} \\
&\quad \times \max_{x,y\in[d]^{2tn^{1-1/D}}} \left\|\tilde{G}_{n^{1-1/D},D-1,c}|y\rangle\langle x|\tilde{G}_{n^{1-1/D},D-1,c}\right\|_1^{n^{1/D}} \\
&\leq \frac{1}{d^{O(c\cdot n^{1/D})}} \max_{x,y\in[d]^{2tn^{1-1/D}}} \left\|\tilde{G}_{n^{1-1/D},D-1,c}|y\rangle\langle x|\tilde{G}_{n^{1-1/D},D-1,c}\right\|_1^{n^{1/D}} \\
&\leq \frac{1}{d^{O(c\cdot n^{1/D})}} \max_{x\in[d]^{2tn^{1-1/D}}} |\langle x|\tilde{G}^2_{n^{1-1/D},D-1,c}|x\rangle|^{n^{1/D}} \\
&\leq \frac{1}{d^{O(c\cdot n^{1/D})}} \\
&\quad \times \max_{x\in[d]^{2tn^{1-1/D}}} (\langle x|G_{\text{Haar}}|x\rangle + |\langle x|(\tilde{G}_{n^{1-1/D},D-1,c} - G_H)^2|x\rangle|)^{n^{1/D}} \\
&\leq \frac{1}{d^{O(c\cdot n^{1/D})}} \left(\max_{x\in[d]^{2tn^{1-1/D}}} \frac{t! + 1/d^{n^{(1-1/D)\cdot\frac{1}{D-1}}}}{d^{n^{1-1/D}t}}\right)^{n^{1/D}} \\
&\leq \frac{\epsilon}{d^{nt}}. \qquad (175)
\end{aligned}$$

Next, assumming $\max_x |\langle x|(\tilde{G}_{n^{1-1/D},D-1,c} - G_{\text{Haar}})^2|x\rangle| \leq \frac{\epsilon}{d^{n^{1-1/D}t}}$, we show $\max_x |\langle x|(\tilde{G}_{n,D,c} - G_{\text{Haar}})^2|x\rangle| \leq \frac{\epsilon}{d^{nt}}$. The proof is very similar to the above calculation:

$$\begin{aligned}
|\langle x|(\tilde{G}_{n,D,c} - G_{\text{Haar}})^2|y\rangle| &= \langle x|\,|\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c}(\hat{G}^{c-1}_{n,D,c} - G_{\text{Haar}})\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c} \\
&\quad \times(\hat{G}^{c-1}_{n,D,c} - G_{\text{Haar}})\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c}|\,|y\rangle \\
&\leq \left\|(\hat{G}^{c-1}_{n,D,c} - G_{\text{Haar}})\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c}(\hat{G}^{c-1}_{n,D,c} - G_{\text{Haar}})\right\|_\infty \\
&\quad \times \left\|\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c}|y\rangle\langle x|\tilde{G}^{\otimes n^{1/D}}_{n^{1-1/D},D-1,c}\right\|_1 \\
&\leq \left\|\hat{G}^{c-1}_{n,D,c} - G_{\text{Haar}}\right\|_\infty \\
&\quad \times \max_{x,y\in[d]^{2tn^{1-1/D}}} \left\|\tilde{G}_{n^{1-1/D},D-1,c}|y\rangle\langle x|\tilde{G}_{n^{1-1/D},D-1,c}\right\|_1^{n^{1/D}} \\
&\leq \frac{\epsilon}{d^{nt}}. \qquad (176)
\end{aligned}$$

In the third line we have used Lemma 26. We skip the calculations after the third line because it is similar to the calculations of (175). □

Next, we prove Lemma 43. Lemma 42 is a special case of this lemma and we skip its proof.

Lemma (Restatement of Lemma 43). *For small enough $D = O(\ln n / \ln\ln n)$ and large enough c,*

$$\left\| Ch\left[(G_{Rows(D,n)} G_{Planes(D)} G_{Rows(D,n)})^c - G^{(t)}_{Haar} \right] \right\|_\diamond = \frac{t^{O(tn^{1-1/D})}}{d^{\Omega(cn^{1-1/D})}}. \quad (177)$$

Proof. As discussed in Sect. 2.3, the superoperator $\mathrm{Ch}[G^{(t)}_{\mathrm{Haar}}]$ can be written in the following canonical form

$$\mathrm{Ch}[G^{(t)}_{\mathrm{Haar}}][X] = \sum_{\pi \in S_t} \mathrm{Tr}(V(\pi)X)\,\mathrm{Wg}(\pi). \quad (178)$$

Using the notation defined in Sect. 2.3, $\mathcal{X}[\mathrm{Ch}[G^{(t)}_{\mathrm{Haar}}]] = G^{(t)}_{\mathrm{Haar}}$ and

$$\left(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)}\right)^c - G^{(t)}_{\mathrm{Haar}} =: \sum_{a,b \in S_t^{\times n^{1-1/D}}} |D_a\rangle\, \Lambda_{a,b}\, \langle D_b| \,. \quad (179)$$

Using the definition of Λ we can write

$$\begin{aligned} &\mathrm{Ch}\left[(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)})^c - G^{(t)}_{\mathrm{Haar}} \right] \\ &= \mathrm{Ch}\left[\sum_{a,b \in S_t^{\times n^{1-1/D}}} |D_a\rangle\, \Lambda_{a,b}\, \langle D_b| \right] = \sum_{a,b \in S_t^{\times n^{1-1/D}}} \frac{1}{d^{nt}} V(a) \Lambda_{a,b} V^*(b). \end{aligned} \quad (180)$$

Therefore

$$\begin{aligned} &\|\mathrm{Ch}[(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)})^c - G^{(t)}_{\mathrm{Haar}}]\|_\diamond \\ &\le \sum_{a,b \in S_t^{\times n^{1-1/D}}} |\Lambda_{a,b}| \left\| \frac{1}{d^{nt}} V(a) V^*(b) \right\|_\diamond \\ &\le \sum_{a,b \in S_t^{\times n^{1-1/D}}} |\Lambda_{a,b}| \le t^{O(n^{1-1/D})} \|\Lambda\|_\infty^c. \end{aligned} \quad (181)$$

Here we have used $\| \frac{1}{d^{nt}} V(a) V^*(b) \|_\diamond \le 1$. This is because $V(a)V^*(b)$ is a tensor product of $n^{1-1/D}$ superoperators, i.e., $\otimes_i V(a_i) V^*(b_i)$, and hence $\|V(a)V^*(b)\|_\diamond = \prod_i \|V(a_i)V^*(b_i)\|_\diamond$. It is enough to show that each of $\|V(a_i)V^*(b_i)\|_\diamond$ is bounded by 1.

$$\begin{aligned} \frac{1}{d^{nt}} \left\| V(a_1) V(b_1)^* \right\|_\diamond &= \frac{1}{d^{nt}} \sup_{X : \|X\|_1 = 1} \left\| \mathrm{Tr}_A (V(a_1)_A \otimes \mathrm{id}_B\, X_{AB}) \otimes V_A(b_1) \right\|_1 \\ &= \sup_{X : \|X\|_1 = 1} \left\| \mathrm{Tr}_A (V(a_1)_A \otimes \mathrm{id}_B\, X_{AB}) \right\|_1 \cdot \frac{1}{d^{nt}} \left\| V_A(b_1) \right\|_1 \\ &\le \sup_{X : \|X\|_1 = 1} \left\| V_{(a_1)} \otimes \mathrm{id}\, X_{AB} \right\|_1 \cdot 1 \\ &= \sup_{X : \|X\|_1 = 1} \|X_{AB}\|_1 \end{aligned}$$

$$\leq 1. \tag{182}$$

It is enough to compute $\|\Lambda\|_\infty$. Let $|a\rangle$ be an orthonormal basis labeled according to the indices of Λ. Define

$$T := \sum_{a,b} \sqrt{\Lambda}_{ab} \, |D_a\rangle \langle b| \, . \tag{183}$$

$TT^\dagger = \sum_{a,b} |D_a\rangle \Lambda_{a,b} \langle D_b|$ and $T^\dagger T = \sum_{a,b} |a\rangle (\sqrt{\Lambda} J \sqrt{\Lambda})_{ab} \langle b|$, where $[J]_{a,b} = \langle D_a | D_b \rangle$. First of all, using Lemma 25 $TT^\dagger$ and $T^\dagger T$ have the same spectra. Hence

$$\| \sum_{a,b} |D_a\rangle \Lambda_{a,b} \langle D_b| \|_\infty = \|TT^\dagger\|_\infty = \|T^\dagger T\|_\infty = \|\sqrt{\Lambda} J \sqrt{\Lambda}\|_\infty \tag{184}$$

Therefore

$$\begin{aligned} \|\Lambda\|_\infty &\leq \Big\| \sum_{a,b} |D_a\rangle \Lambda_{a,b} \langle D_b| \Big\|_\infty + \|\sqrt{\Lambda}(J - \mathrm{id})\sqrt{\Lambda}\|_\infty \\ &\leq \Big\| \left(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)}\right)^c - G^{(t)}_{\mathrm{Haar}} \Big\|_\infty + \|\sqrt{\Lambda}\|_\infty \|J - \mathrm{id}\,\|_\infty \|\sqrt{\Lambda}\|_\infty \\ &= \Big\| \left(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)}\right)^c - G^{(t)}_{\mathrm{Haar}} \Big\|_\infty + \|\Lambda\|_\infty \|J - \mathrm{id}\,\|_\infty \end{aligned} \tag{185}$$

As a result

$$\|\Lambda\|_\infty \leq \frac{\Big\| \left(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)}\right)^c - G^{(t)}_{\mathrm{Haar}} \Big\|_\infty}{1 - \|J - \mathrm{id}\,\|_\infty}. \tag{186}$$

In Lemma 40 we showed that $\|(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Columns}(D)} G_{\mathrm{Rows}(D,n)})^c - G^{(t)}_{\mathrm{Haar}}\|_\infty \leq \|G_{\mathrm{Columns}(D)} G_{\mathrm{Rows}(D,n)} - G^{(t)}_{\mathrm{Haar}}\|^c_\infty = 1/d^{O(cn^{1-1/D})}$. It is enough to show that $\|J - \mathrm{id}\,\|_\infty$ is small. But J is tensor product of $n^{1-1/D}$ Gram matrices J_1 such that $\|J_1 - \mathrm{id}\,\|_\infty = \frac{O(t^2)}{d^{n^{1/D}}}$ (see Lemma 57), hence $\|J - \mathrm{id}\,\|_\infty = n^{1-1/D} \frac{O(t^2)}{d^{n^{1/D}}}$ which is bounded by $1/2$ for large enough n and constant t and D. As a result, $\|\Lambda\|_\infty = 1/d^{O(cn^{1-1/D})}$. Combining this with (181) we find that

$$\Big\| \mathrm{Ch} \Big[\left(G_{\mathrm{Rows}(D,n)} G_{\mathrm{Planes}(D)} G_{\mathrm{Rows}(D,n)}\right)^c - G^{(t)}_{\mathrm{Haar}} \Big] \Big\|_\diamond \leq t^{O(tn^{1-1/D})} 1/d^{O(cn^{1-1/D})}. \tag{187}$$

□

4. $O(n \ln^2 n)$-Size Random Circuits with Long-Range Gates Output Anti-concentrated Distributions

Recall that for a circuit C, $\mathrm{Coll}(C)$ is the collision probability,

$$\sum_{x \in \{0,1\}^n} |\langle x|C|0\rangle|^4, \tag{188}$$

of C in the computational basis. Also recall that $\mu_t^{(\mathrm{CG})}$ is the distribution over random circuits obtained from application of t random long-range gates. Unlike the previous section where we used t to denote the degree of a monomial, here we use t for time, i.e. the number of time-steps in a random circuit.

The goal of this section is to prove the following theorem:

Theorem (Restatement of Theorem 13). *There exists a c such that when $s > cn \ln^2 n$,*

$$\mathop{\mathbb{E}}_{C\sim\mu_s^{CG}} Coll(C) \leq \frac{29}{2^n}. \tag{189}$$

Moreover if $t \leq \frac{1}{3c'} n \ln n$ for some large enough c', then

$$\mathop{\mathbb{E}}_{C\sim\mu_s^{CG}} Coll(C) \geq \frac{1.6^{n^{1-1/c'}}}{2^n}. \tag{190}$$

Our strategy is to relate the convergence of the expected collision probability to a classical Markov chain mixing problem. In Sect. 4.1 we go over the notation and definitions we use in the proof of this theorem. In Sect. 4.2 we prove the theorem. This proof is based on several lemmas which we will prove in sections 4.3 and 4.5.

4.1. Background: random circuits with long-range gates and Markov chains. Previous work [17,18,34,51] demonstrates that if we only care about the second moment of $\mu_t^{(\mathrm{CG})}$, then the corresponding moment superoperator is related to a certain classical Markov chain. In particular the application of the moment superoperator on the basis $\mathrm{P}_n^2 := \left\{\sigma_p \otimes \sigma_p : p \in \{0,1,2,3\}^n\right\}$ is a classical Markov chain. We now describe this connection.

We first start with some basic properties of moment superoperators.

Fact 58. *Let μ and $\mu_1, \ldots, \mu_K$ be distributions over circuits.*

1. *If μ is a convex combination of $\mu_1, \ldots, \mu_K$ then $Ch\left[G_\mu^{(2)}\right]$ is the same convex combination of $Ch\left[G_{\mu_1}^{(2)}\right], \ldots, Ch\left[G_{\mu_K}^{(2)}\right]$.*
2. *If μ is the composition of a circuit from μ_1 with a circuit with μ_2, then $Ch\left[G_\mu^{(2)}\right] = Ch\left[G_{\mu_2}^{(2)}\right] \circ Ch\left[G_{\mu_1}^{(2)}\right]$.*

Recall that $\mathrm{Ch}\left[G_{i,j}^{(2)}\right]$ denotes $\mathrm{Ch}\left[G_{\mathrm{U}(4)}^{(2)}\right]$ applied to qubits i and j. Since μ_1^{CG} is a convex combination of two-qubit random $\mathrm{U}(4)$ gates, the first point above implies that

$$\mathrm{Ch}\left[G_{\mu_1^{\mathrm{CG}}}^{(2)}\right] = \frac{2}{n(n-1)} \sum_{i<j} \mathrm{Ch}\left[G_{i,j}^{(2)}\right] \tag{191}$$

and since μ_t^{CG} is t times compositions of μ_1^{CG} with itself, the second item implies that

$$\mathrm{Ch}\left[G_{\mu_t^{\mathrm{CG}}}^{(2)}\right] = \left(\frac{2}{n(n-1)} \sum_{i<j} \mathrm{Ch}\left[G_{i,j}^{(2)}\right]\right)^t. \tag{192}$$

The moment superoperator $\mathrm{Ch}[G_{\mathrm{U}(4)}^{(2)}]$ has the following simple action on the Pauli basis:

$$\mathrm{Ch}[G_{\mathrm{U}(4)}^{(2)}](\sigma_p \otimes \sigma_q \otimes \sigma_a \otimes \sigma_b)$$

$$= \begin{cases} \sigma_0 \otimes \sigma_0 \otimes \sigma_0 \otimes \sigma_0 & pq = ab = 00 \\ \frac{1}{15} \sum_{\substack{c,d \in \{0,1,2,3\}^2 \\ cd \neq 00}} \sigma_c \otimes \sigma_d \otimes \sigma_c \otimes \sigma_d & pq = ab \neq 00 \\ 0 & \text{otherwise} \end{cases} \tag{193}$$

In particular the action of $\mathrm{Ch}[G^{(2)}_{\mathrm{U}(4)}]$ on the Pauli basis P_2^2 is a stochastic matrix, and for any pair $i \neq j$ the action of $\mathrm{Ch}[G^{(2)}_{\mathrm{U}(4)}]$ on qubits i, j can be represented by a stochastic matrix acting on P_n^2. Using (192) $\mathrm{Ch}\left[G^{(2)}_{\mu_s^{\mathrm{CG}}}\right]$ on P_n^2 is also a stochastic matrix. We can describe this stochastic matrix as a Markov chain on state space $\mathcal{S} = \{0, 1, 2, 3\}^n$, with $S_t \in \mathcal{S}$ describing the string at time t.

It turns out that the expected collision probability depends on the subset of qubits that have been hit by the random circuit. In case a subset of size m of qubits (out of n qubits) never have a gate applied to them, then the expected collision probability converges to a value like $\approx \frac{2^m}{2^n}$ and not $\frac{1}{2^n+1}$. So we need to separately track which qubits have ever been hit by a gate throughout this process. Let $H_t \in 2^{[n]}$ denote the set of qubits that have been hit by at least one gate by time t, where $2^{[n]}$ denotes the power set of $[n]$.

Together (S_t, H_t) can be modeled as the following Markov chain.

Definition 59. Let $(S_0, H_0), (S_1, H_1), (S_2, H_2), \ldots \in \mathcal{S} \times 2^{[n]}$ be the following classical Markov chain. Initially $H_0 = \emptyset$ and S_0 is a random element of $\{0, 3\}^n \backslash 0^n$. At each time step t we choose a random pair $i, j \in [n]$ with $i \neq j$. We let $H_{t+1} = H_t \cup \{i, j\}$ so that H_t represents the set of all indices chosen up to time t. We determine S_{t+1} from S_t using the transition rule of (193). Specifically if the i, j positions of S_t are both 0, then we leave them equal to 00, and otherwise we replace them with a uniformly random element of $\{01, 02, \ldots, 33\}$.

Suppose we condition on $H_t \supseteq H$ for some set H with $|H| = n - m$. Let

$$P_t^{(n-m)}(k) := \Pr\left[|S_t(H)| = k | H_t \subseteq H\right]. \tag{194}$$

We can use this notation since the RHS of (194) depends only on $|H|, t, n, k$ and not on H.

For a function $f : [n] \to \mathbb{R}$ we define $\|\cdot\|_*$ to be the following norm

$$\|f\|_* := \sum_{k \in [n]} \frac{|f(k)|}{3^k}. \tag{195}$$

4.1.1. Summary of the definitions. See below for a summary of the definitions:

Notation	Definition	Reference
$\mathrm{Coll}(C)$	The collision probability of circuit C	Equation (188)
$G_{\mu}^{(t)}$	Average of $C^{\otimes t,t}$ over $C \sim \mu$	Definition 16
$G_{i,j}^{(t)}$	Haar projector of order t on qudits i and j	Definition 16
μ_t^{CG}	The distribution over circuits with t random two-qubit gates	Definition 24
P_n^2	$\{\sigma_p \otimes \sigma_p : p \in \{0,1,2,3\}^n\}$	Section 4.1
$S_0, S_1, \ldots$	Markov chain of Pauli strings	Definition 59
H_t	Subset of $[n]$ that is covered according to the Markov chain of Pauli strings	Definition 59
$S'_0, S'_1, \ldots$	Accelerated Markov chain of binary strings with decoupled coordinates	Definition 82
X_t	$\lvert S_t \rvert$	Section 4.5
Y_t	Steps of the accelerated Markov chain Q	Section 4.5
$P_t^{(n-m)}(k)$	$\Pr[\lvert S_t(H)\rvert = k \lvert H_t \subseteq H]$ for any fixed H with $\lvert H\rvert = n-m$	Equation (194)
$P_t(k)$	$\Pr[\lvert S_t(H)\rvert = k]$. Also equal to $P_t^{(n)}(k)$	Equation (194)
$\lVert f\rVert_*$	$\sum_{x=1}^n \frac{\lvert f(x)\rvert}{3^x}$	Equation (195)
$\lVert f\rVert_\square$	$\sum_{x=1}^n \lvert f(x)\rvert \frac{3n}{x3^x}$	Proposition 74
P	Transition matrix of the birth and death Markov chain	Equation (229)
Q	Transition matrix of the partially accelerated Markov chain	Equation (234)
$T_{\text{left (right)}}(Y^s)$	Wait time for the steps $Y_0, \ldots, Y_s$ on the left (right) hand side of site $\frac{5}{6}n$	Section 4.5.2
ν	$3/4n$	Section 4.5.2
ν_τ	$Y_0 \exp(-\frac{\tau}{\nu}) + \nu(1-\exp(-\frac{\tau}{\nu}))$	Section 4.5.2
β	$8(4+c)\ln n$ for constant c fixed in advance	Section 4.5.2
$x(0)$	ν/β	Section 4.5.2
ρ_x	$\sum_{j=1}^s I\{Y_j = x\}$	Section 4.5.2
A	$\cap_{1\le x\le x(0)}\{N_x \le \beta x\}$	Section 4.5.2
M_s	$\min_{1\le j\le s}\{Y_j\}$	Section 4.5.2
y^s	Short hand for $(y_0, \ldots, y_s)$	Section 4.5.6
$\mathsf{Bin}(n,p)$	Binomial distribution on n elements each occurring with probability p	
$\mathsf{Geo}(\alpha)$	Geometric distribution with mean $\frac{1}{\alpha}$	
$\mathsf{Pois}(\tau)$	Poisson distribution with mean τ	
$\mathsf{Unif}[a,b]$	Continuous uniform distribution on the interval $[a,b]$	

4.2. Proof of Theorem 13: bound on the collision probability. Before giving the proof we state the following three main theorems. The first one relates the expected collision probability to the $\lVert\cdot\rVert_*$ norm of the probability vector on the state space of the Markov chain of weights. More concretely

Theorem 60.

$$\mathop{\mathbb{E}}_{C\sim\mu_t^{CG}} Coll(C) \le \frac{1}{2^n} + \sum_{m=0}^{n} \binom{n}{m} e^{-tm/n} \lVert P_t^{(n-m)}\rVert_* \tag{196}$$

This result is proved in Sect. 4.3.

The second theorem shows that for $t \approx n\ln^2 n$, $\lVert P_t^{(n)}\rVert_* \approx \text{Constant} \times \frac{1}{2^n+1}$, where $\frac{1}{2^n+1}$ is the value of this norm at the stationary state.

Theorem 61. *There exists a constant c such that if $t = cn\ln^2 n$ then $\lVert P_t^{(m)}\rVert_* \le \frac{28}{2^m+1}$.*

This result is proved in Sect. 4.5.

The third theorem gives an exact expression for the collision probability in terms of the Markov chain $S_0, S_1, \ldots$. We use this to compute the lower bound.

Theorem 62. $Coll_{\mu_t^{CG}} = \frac{1}{2^n}\left(1 + \sum_{p,q\in\{0,3\}^n\setminus 0^n} \Pr[S_t = p | S_0 = q]\right)$

The proof of this expression is the same as equation (215) and is derived in section 4.3.

Proof of Theorem 13. We first prove the upper bound. There are two major steps.

Combining Theorems 60 and 61 and choosing $t = cn\ln^2(n)$ we obtain

$$\mathop{\mathbb{E}}_{C\sim\mu_t^{\mathrm{CG}}} \mathrm{Coll}(C) \leq \frac{1}{2^n} + \sum_{m=0}^{n} \binom{n}{m} e^{-tm/n} \frac{28}{2^{n-m}} \tag{197}$$

$$\leq \frac{1}{2^n}(1 + 28(1 + 2e^{-t/n})^n) \tag{198}$$

$$\leq \frac{1}{2^n}(1 + 28(1 + \frac{2}{n^{c\ln n}})^n) \tag{199}$$

$$\leq \frac{29}{2^n + 1}. \tag{200}$$

Here we need to assume n is larger than some universal constant. This can be done by adjusting c to cover the finite set of cases where n is too small.

For the lower bound we use the expression in Theorem 62 and bound it according to

$$\mathrm{Coll}_{\mu_t^{\mathrm{CG}}} \geq \frac{1}{2^n} \sum_{p\in\{0,3\}^n} \Pr[S_t = p | S_0 = p], \tag{201}$$

$$= \frac{1}{2^n} \sum_{k=0}^{n} \sum_{p\in\{0,3\}^n : |p|_H = k} \Pr[S_t = p | S_0 = p], \tag{202}$$

$$\geq \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} r_k^t, \tag{203}$$

where

$$r_k = \frac{14}{15}(1 - \frac{k}{n})(1 - \frac{k}{n-1}) + \frac{1}{15} \geq e^{-3\frac{k}{n}}, \tag{204}$$

is the probability that a string of Hamming weight k does not change after one step of the Markov chain. Assume $t \leq \frac{1}{3c'} n\ln n$ then

$$\mathrm{Coll}_{\mu_t^{\mathrm{CG}}} \geq \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} e^{-3\frac{kt}{n}}, \tag{205}$$

$$= \frac{1}{2^n}(1 + e^{-3t/n})^n, \tag{206}$$

$$\geq \frac{1}{2^n} \exp\left(\frac{n^{1-1/c'}}{1 + n^{-1/c'}}\right) \tag{207}$$

$$\geq \frac{1}{2^n} \cdot 1.6^{n^{1-1/c'}} \tag{208}$$

□

4.3. Proof of Theorem 60: relating collision probability to a Markov chain. In this section we relate the expected collision probability of a random circuit with long-range gates to the $\|\cdot\|_*$ norm of the probability vector $P_t^{(m)}$ defined in Sect. 4.1. We will prove several intermediate results along the way to Theorem 60.

Theorem 63 (Section 3 of [34]). *We can write*

$$Ch\left[G^{(2)}_{\mu_t^{CG}}\right](\sigma_q \otimes \sigma_q) = \sum_{p\in\{0,1,2,3\}^n} \Pr[S_t = p|S_0 = q]\sigma_p \otimes \sigma_p. \tag{209}$$

Proof of Theorem 62. We can write the expected collision probability in terms of the moment superoperator $\mathrm{Ch}\left[G^{(2)}_{\mu_t^{\mathrm{CG}}}\right]$. We use the notation $\mathrm{Coll}_{\mu_t^{\mathrm{CG}}} = \mathbb{E}_{C\sim\mu_t^{\mathrm{CG}}}\mathrm{Coll}(C)$:

$$\begin{aligned}
\mathrm{Coll}_{\mu_t^{\mathrm{CG}}} &= \sum_{z\in\{0,1\}^n} \mathop{\mathbb{E}}_{C\sim\mu_t^{\mathrm{CG}}} |\langle z|C|0\rangle|^4 \\
&= \sum_{z\in\{0,1\}^n} \langle z|\otimes\langle z| \mathop{\mathbb{E}}_{C\sim\mu_t^{\mathrm{CG}}} \left(C\,|0^n\rangle\langle 0^n|\,C^\dagger \otimes C\,|0^n\rangle\langle 0^n|\,C^\dagger\right)|z\rangle\otimes|z\rangle \\
&= \mathrm{Tr}\sum_{z\in\{0,1\}^n} |z\rangle\langle z|\otimes|z\rangle\langle z|\,\mathrm{Ch}\left[G^{(2)}_{\mu_t^{\mathrm{CG}}}\right]\left(|0^n\rangle\langle 0^n|\otimes|0^n\rangle\langle 0^n|\right)
\end{aligned} \tag{210}$$

It is useful to write $|0^n\rangle\langle 0^n|\otimes|0^n\rangle\langle 0^n|$ and $\sum_{z\in\{0,1\}^n}|z\rangle\langle z|\otimes|z\rangle\langle z|$ in the Pauli basis:

$$|0^n\rangle\langle 0^n|\otimes|0^n\rangle\langle 0^n| = \frac{1}{4^n}\sum_{p,q\in\{0,3\}^n}\sigma_p\otimes\sigma_q. \tag{211}$$

$$\sum_{z\in\{0,1\}^n}|z\rangle\langle z|\otimes|z\rangle\langle z| = \frac{1}{2^n}\sum_{p\in\{0,3\}^n}\sigma_p\otimes\sigma_p. \tag{212}$$

Then the collision probability becomes:

$$\begin{aligned}
\mathrm{Coll}_\nu = \frac{1}{2^n} + (1-\frac{1}{2^n})\frac{1}{2^n}\mathrm{Tr}&\left(\sum_{z\in\{0,1\}^n}|z\rangle\langle z|\otimes|z\rangle\langle z|\right)\mathrm{Ch}\left[G^{(2)}_{\mu_t^{\mathrm{CG}}}\right] \\
&\times\left(\frac{1}{2^n-1}\sum_{q\in\{0,3\}^n\setminus 0^n}\sigma_q\otimes\sigma_q\right)
\end{aligned} \tag{213}$$

Using Theorem 63

$$\begin{aligned}
&\mathrm{Ch}\left[G^{(2)}_{\mu_t^{\mathrm{CG}}}\right]\left(\frac{1}{2^n-1}\sum_{q\in\{0,3\}^n\setminus 0^n}\sigma_q\otimes\sigma_q\right) \\
&\quad= \frac{1}{2^n-1}\sum_{\substack{p\in\{0,1,2,3\}^n\setminus 0^n \\ q\in\{0,3\}^n\setminus 0^n}}\Pr[S_t = p|S_0 = q]\sigma_p\otimes\sigma_p.
\end{aligned} \tag{214}$$

As a result

$$\mathrm{Coll}_{\mu_t^{\mathrm{CG}}} = \frac{1}{2^n}\left(1 + \sum_{p,q\in\{0,3\}^n\setminus 0^n} \Pr[S_t = p|S_0 = q]\right) \tag{215}$$

□

For a string $a \in \{0,1,2,3\}^n$ and a subset $A \in 2^{[n]}$ we let $a(A)$ denote the substring of a restricted to A.

Lemma 64. *For $H \subseteq [n]$ and $p, q \in \{0,1,2,3\}^n$*

$$\Pr[S_t = p|S_0 = q, H_t = H] = \frac{1}{\binom{|H|}{|p(H)|}3^{|p(H)|}} \Pr[|S_t(H)| = |p(H)|\big|S_0 = q, H_t = H] \tag{216}$$

if $q([n]\setminus H) = p([n]\setminus H)$ and 0 otherwise.

In other words, once we condition on $H_t = H$, the probability distribution of $S_t(H)$ depends only on its Hamming weight.

Proof. Conditioned on $H_t = H$ the sites of $[n]\setminus H$ are not hit, so the event that $q([n]\setminus H) \neq p([n]\setminus H)$ has zero probability. Now since the set H is covered, 1, 2 or 3 have equal probabilities of appearing at any position of the string $S_t(H)$. As a result for each non-zero bit of $S_t(H)$ we get a factor of $1/3$. □

Using Theorem 62 and Lemma 64 we obtain

Corollary 65.

$$Coll_{\mu_t^{CG}} = \frac{1}{2^n} + (1 - 1/2^n) \sum_{H\subseteq[n]} \Pr[H_t = H] \sum_{1\le k\le |H|} \frac{\Pr[|S_t(H)| = k\big|H_t = H]}{3^k}. \tag{217}$$

Proof. Expanding Theorem 62 we have

$$\mathrm{Coll}_{\mu_t^{\mathrm{CG}}} = \frac{1}{2^n}\left(1 + \sum_{p,q\in\{0,3\}^n\setminus 0^n} \Pr[S_t = p|S_0 = q]\right) \tag{218}$$

$$= \frac{1}{2^n}\left(1 + \sum_{H\subseteq[n]} \sum_{p,q\in\{0,3\}^n\setminus 0^n} \Pr[H_t = H|S_0 = q] \Pr[S_t = p|S_0 = q, H_t = H]\right) \tag{219}$$

$$= \frac{1}{2^n}\left(1 + \sum_{H\subseteq[n]} \sum_{p,q\in\{0,3\}^n\setminus 0^n} \Pr[H_t = H] \Pr[S_t = p|S_0 = q, H_t = H]\right). \tag{220}$$

Using Lemma 64 in the above we have

$$= \frac{1}{2^n} + \frac{1}{2^n} \sum_{H\subseteq[n]} \Pr[H_t = H] \sum_{p,q\in\{0,3\}^n\setminus 0^n} \Pr[S_t = p|S_0 = q, H_t = H], \tag{221}$$

$$= \frac{1}{2^n} + \frac{1}{2^n} \sum_{H\subseteq[n]} \Pr[H_t = H] \sum_{\substack{p,q\in\{0,3\}^n\setminus 0^n \\ p([n]\setminus H)=q([n]\setminus H)}} \times \frac{1}{\binom{|H|}{|p(H)|}3^{|p(H)|}} \Pr[|S_t(H)| = |p(H)| \big| S_0 = q, H_t = H], \tag{222}$$

$$= \frac{1}{2^n} + \frac{1}{2^n} \sum_{H\subseteq[n]} \Pr[H_t = H] \sum_{q\in\{0,3\}^n} \sum_{1\le k\le |H|} \sum_{\substack{p\in\{0,3\}^n\setminus 0^n \\ p([n]\setminus H)=q([n]\setminus H) \\ |p(H)|=k}} \times \frac{1}{\binom{|H|}{|p(H)|}3^{|p(H)|}} \Pr[|S_t(H)| = |p(H)| \big| S_0 = q, H_t = H], \tag{223}$$

$$= \frac{1}{2^n} + \frac{1}{2^n} \sum_{H\subseteq[n]} \Pr[H_t = H] \sum_{1\le k\le |H|} \sum_{q\in\{0,3\}^n} \sum_{\substack{p\in\{0,3\}^n\setminus 0^n \\ p([n]\setminus H)=q([n]\setminus H) \\ |p(H)|=k}} \times \frac{1}{\binom{|H|}{k}3^k} \Pr[|S_t(H)| = k \big| S_0 = q, H_t = H], \tag{224}$$

$$= \frac{1}{2^n} + \frac{1}{2^n} \sum_{H\subseteq[n]} \Pr[H_t = H] \sum_{1\le k\le |H|} \sum_{q\in\{0,3\}^n} \times \frac{1}{3^k} \Pr[|S_t(H)| = k \big| S_0 = q, H_t = H], \tag{225}$$

$$= \frac{1}{2^n} + (1 - 1/2^n) \sum_{H\subseteq[n]} \Pr[H_t = H] \sum_{1\le k\le |H|} \frac{\Pr[|S_t(H)| = k \big| H_t = H]}{3^k}. \tag{226}$$

□

The standard coupon-collector bound is

Lemma 66 (coupon collector). *Let* $H \subseteq [n]$. *Then* $\Pr[H_t \subseteq H] \le e^{-(n-|H|)t/n}$.

Proof. Let $E_H^{(i)}$ be the event that at step i of the circuit a random gate lands completely inside the set H. Then $\Pr[E_H^{(i)}] = \frac{|H|(|H|-1)}{n(n-1)}$. Now $\Pr[H_t \subseteq H] = \Pr[E_H^{(i)}]^t = \left(\frac{|H|(|H|-1)}{n(n-1)}\right)^t \le \left(\frac{|H|}{n}\right)^t \le e^{-(n-|H|)t/n}$. □

We now have all the pieces to prove Theorem 60.

Proof of Theorem 60. Using corollary 65 the total collision probability is

$$\begin{aligned}
\mathrm{Coll}_{\mu_t^{\mathrm{CG}}} &= \frac{1}{2^n} + (1 - 1/2^n) \sum_{H\subseteq[n]} \Pr[H_t = H] \sum_{k=1}^{n} \frac{\Pr[|S_t(H)| = k | H_t = H]}{3^k} \\
&= \frac{1}{2^n} + (1 - 1/2^n) \sum_{H\subseteq[n]} \sum_{k=1}^{n} \frac{\Pr[|S_t(H)| = k, H_t = H]}{3^k} \\
&\le \frac{1}{2^n} + (1 - 1/2^n) \sum_{H\subseteq[n]} \sum_{k=1}^{n} \frac{\Pr[|S_t(H)| = k, H_t \subseteq H]}{3^k}
\end{aligned}$$

$$\leq \frac{1}{2^n} + \sum_{H\subseteq[n]} \Pr[H_t \subseteq H] \sum_{k=1}^{n} \frac{\Pr[|S_t(H)| = k | H_t \subseteq H]}{3^k}$$

$$\leq \frac{1}{2^n} + \sum_{H\subseteq[n]} \Pr[H_t \subseteq H] \sum_{k=1}^{n} \frac{P_t^{(|H|)}(k)}{3^k}$$

$$\leq \frac{1}{2^n} + \sum_{H\subseteq[n]} e^{-(n-|H|)t/n} \sum_{k} P_t^{(|H|)}(k)/3^k \quad \text{Lemma 66}$$

$$= \frac{1}{2^n} + \frac{1}{(1-\epsilon)} \sum_{m=0}^{n} \binom{n}{m} e^{-mt/n} \| P_t^{(n-m)} \|_*. \quad \text{setting } m = n - |H| \tag{227}$$

□

4.4. Proof of Proposition 67: collision probability is non-increasing in time. When we try to recover the original chain from the accelerated chain we find that s steps of the accelerated chain typically correspond to $t = O(s)$ steps of the original chain, but with a significant variance. This means that our bounds on the collision probability of the accelerated chain translate only into bounds for a distribution of times of running the original chain. This issue can be addressed using the following fact.

Proposition 67. $\mathbb{E}_{C\sim\mu_t^{CG}} Coll(C)$ *is a non-increasing function of* t.

Proof. $\mathrm{Ch}[G_{\mu_1^{(\mathrm{CG})}}]$ corresponds to an average of $n(n-1)/2$ projectors (using the Hilbert-Schmidt inner product). Hence it is a psd matrix with maximum eigenvalue ≤ 1. Let $\alpha = \sum_{p\in\{0,3\}^n} \sigma_p \otimes \sigma_p$. (210) may be written as

$$\sum_{z\in\{0,1\}^n} \mathrm{Tr}\left(|z\rangle\langle z| \otimes |z\rangle\langle z| \, \mathrm{Ch}\left[G^{(2)}_{\mu_t^{\mathrm{CG}}}\right] \left(|0^n\rangle\langle 0^n| \otimes |0^n\rangle\langle 0^n|\right)\right)$$
$$= \mathrm{Tr}\left(\frac{\alpha}{2^n} \mathrm{Ch}\left[G^{(2)}_{\mu_t^{\mathrm{CG}}}\right] \left(|0^n\rangle\langle 0^n| \otimes |0^n\rangle\langle 0^n|\right)\right) \tag{228}$$

Using (212), terms of the form $\sigma_p \otimes \sigma_q$ for $p \neq q$ in the decomposition of $|0^n\rangle\langle 0^n| \otimes |0^n\rangle\langle 0^n|$ do not contribute to the collision probability. Therefore using this observation and (228), the collision probability after t steps is proportional to $\mathrm{Tr}(\alpha \mathrm{Ch}[G_{\mu_1^{(\mathrm{CG})}}]^t \alpha)$. Since $\mathrm{Ch}[G_{\mu_1^{(\mathrm{CG})}}]$ has all eigenvalues between 0 and 1, we conclude that the collision probability after t steps cannot increase in t. □

This argument relied on the starting state being $|0^n\rangle$. There exist starting states, such as $|+\rangle^{\otimes n}$, for which the collision probability increases when random gates are applied.

4.5. Proof of Theorem 61: the Markov chain analysis. Consider the following birth-and-death Markov chain on the state space $\{0, 1, 2, \ldots, n\}$.

$$P(k,l) := \begin{cases} 1 - \dfrac{2k(3n-2k-1)}{5n(n-1)} & l = k \\ \dfrac{2k(k-1)}{5n(n-1)} & l = k-1 \\ \dfrac{6k(n-k)}{5n(n-1)} & l = k+1 \\ 0 & \text{otherwise} \end{cases} \tag{229}$$

This Markov chain is reducible in general, however if we restrict the state space to $\{0\}$ or $\{1, 2, \ldots, n\}$ it is irreducible. Consider the following initial distribution over the state space $\{1, 2, \ldots, n\}$:

$$P_0^{(n)}(k) = \frac{\binom{n}{k}}{2^n - 1} \qquad k \in \{1, 2, \ldots, n\} \tag{230}$$

We claim that

Lemma 68.

$$P_t^{(n)} = P^t P_0^{(n)} \tag{231}$$

Proof. The proof follows from the fact that $\Pr[|S_t| = l \big| |S_0| = k] = P^t(k,l)$ which was shown in Lemma 5.2 of [34]. □

We now prove Theorem 61 which gives a sharp upper bound on $\| P_t^{(n)} \|_*$. Throughout this section we drop the superscript (n). Moreover we use the notation $X_t := |S_t|$.

Proof overview: The philosophy of our analysis is to consider an acceleration of the chain P: a chain with transition matrix Q which is the same as P but moves faster. As mentioned in the introduction, previous work [17,34] considered a "fully accelerated" chain, but we will instead carefully choose the amount of acceleration so that the transition probabilities are affine functions of x. This will allow an exact solution of the dynamics of this partially accelerated chain using a method of Kac [39], as we describe in Sect. 4.5.4. We then analyze how much time should P wait in each step of its walk in order to simulate steps of Q. In order to do this we need to prove bounds on how many times each site of the Markov chain has been visited during the accelerated walk and based on that we count how many steps the original chain should wait. This analysis is demonstrated in Sect. 4.5.2. Along the way during the wait-time analysis we will further modify the partially accelerated chain to run in continuous time, so that in time t we sample t' from $\mathrm{Pois}(t)$ (the Poisson distribution) and move t' steps. This resulting chain is also exactly solvable, and the solution turns out to be extremely simple, and exemplifies the connection of the accelerated walk with the well-known Ornstein–Uhlenbeck process (see Proposition 76). We need to analyze the error from moving to continuous time, which turns out to be a straightforward analysis of the Poisson distribution.

Now suppose that the accelerated chain goes through a sequence of transitions $Y_0, Y_1, \ldots, Y_s$.

Let $p(x) = P(x, x+1)$ and $q(x) = P(x, x-1)$. We first consider the chain P conditioned on moving at every single step. This chain at site x has probability of moving forward and backwards $\frac{p(x)}{p(x)+q(x)}$ and $\frac{q(x)}{p(x)+q(x)}$, respectively. We can compute these probabilities

$$\begin{aligned} Q^a(x,x) &= 0, \\ Q^a(x,x+1) &= \frac{3(n-x)}{3n-2x-1}, \\ Q^a(x,x-1) &= \frac{x-1}{3n-2x-1}. \end{aligned} \tag{232}$$

Such a chain is called accelerated. The chain Q^a was used in [18,27,34] but we will not use it in this paper.

Instead of an accelerated chain we now define a partially accelerated chain as:

$$\begin{aligned} Q^w(x,x) &= w(x), \\ Q^w(x,x+1) &= (1-w(x))\frac{3(n-x)}{3n-2x-1}, \\ Q^w(x,x-1) &= (1-w(x))\frac{x-1}{3n-2x-1}. \end{aligned} \tag{233}$$

for arbitrary probability value $w(x)$. Setting $w(x) = \frac{2x}{3n-1}$ the partially accelerated chain becomes affine:

$$\begin{aligned} Q(x,x) &= \frac{2x}{3n-1}, \\ Q(x,x+1) &= \frac{3(n-x)}{3n-1}, \\ Q(x,x-1) &= \frac{x-1}{3n-1}. \end{aligned} \tag{234}$$

By "affine" we mean that the transition probabilities are degree-1 polynomials in x. Let $X_0, X_1, \ldots$ be the steps of the Markov chain evolving according to the transition matrix P and $Y_0, Y_1, \ldots$ be the Markov chain according to Q. We now describe a coupling between these two.

4.5.1. Coupling between X and Y chains. For $x < \frac{5}{6}n$ let $\alpha(x) = 1 - \frac{p(x)+q(x)}{1-w(x)} = 1 - \frac{2x(3n-1)}{5n(n-1)}$. If $0 < x < \frac{5}{6}n$, $0 < \alpha(x) < 1$. So for this range we can view $\alpha(x)$ as a probability.

For $x \geq \frac{5}{6}n$, let $\beta(x)$ be the solution to the following equation

$$p(x) + q(x) = 1 - w(x) + \beta(x)w(x)(p(x) + q(x)). \tag{235}$$

We can solve for $\beta(x)$ to find

$$\beta(x) = \frac{1}{w(x)}\frac{-\alpha(x)}{1-\alpha(x)} = \frac{2x(3n-1) - 5n(n-1)}{4x^2}. \tag{236}$$

For $x \geq \frac{5}{6}n$ we have $\alpha(x) < 0$, so from the first expression for $\beta(x)$ we see that $\beta(x) > 0$. From the second expression for $\beta(x)$ we can calculate the upper bound $\beta(x) \leq 1/4 + \frac{6}{5n}$.

Coupling 69. *The following describes a coupling between $X_0, X_1, \ldots$ and $Y_0, Y_1, \ldots$. It takes as input an arbitrary $x \in [n]$. We write $A \leftarrow a$ to mean that we assign value a to variable A.*

- *Set $X_0 \leftarrow x$ and $Y_0 \leftarrow x$.*
- *Sample $Y_1, Y_2, \ldots,$ according to the Markov chain Q.*
- *Set $s \leftarrow 0$ and $t \leftarrow 0$.*
- *Repeat the following steps.*

– If $\alpha(X_t) > 0$ then

In this case, the X chain may move more slowly than the Y chain, so one step of the Y chain corresponds to one or more steps of the X chain.

1. *With probability $1 - \alpha(X_t)$, set $s \leftarrow s + 1$.*
2. *Set $t \leftarrow t + 1$.*
3. *Set $X_t \leftarrow Y_s$.*

– Else

Otherwise, there is the possibility of advancing the X chain while the Y chain waits. This is only possible if $Y_s = Y_{s+1}$.

1. *If $Y_s \neq Y_{s+1}$ then*
   a. *Set $s \leftarrow s + 1$.*
   b. *Set $t \leftarrow t + 1$.*
   c. *Set $X_t \leftarrow Y_s$.*
2. *Else*
   a. *With probability $\beta(X_t)$, set $s \leftarrow s + 1$*
   b. *Otherwise (with probability $1 - \beta(X_t)$)*
      i. *$t \leftarrow t + 1$*
      ii. *$X_t \leftarrow Y_t$*

Definition 70. For a tuple L and a number x let $L_{\mathrm{left}(x)}$ be the same as L except that we remove elements which are $> x$. Similarly define $L_{\mathrm{right}(x)}$ to be the tuple resulted from removing the elements that are $< x$.

Theorem 71. *Assume $X_0 = Y_0$ and fix $s > 0$, and let $\mathcal{Y} := (Y_0, Y_1, \ldots, Y_s)$. Define*

$$S := \{i : \mathcal{Y}_{right\left(\frac{5}{6}n\right)}[i] = \mathcal{Y}_{right\left(\frac{5}{6}n\right)}[i+1]\}. \tag{237}$$

Let

$$T_{left}(\mathcal{Y}) = \sum_{y \in \mathcal{Y}_{left\left(\frac{5}{6}n\right)}} \mathsf{Geo}(\alpha_y) \tag{238}$$

and

$$T_{right}(\mathcal{Y}) = \sum_{y \in S} \mathsf{Bern}(\beta(y)) \tag{239}$$

then the process in Coupling 69 satisfies

$$Y_s = X_{s + T_{left}(\mathcal{Y}) - T_{right}(\mathcal{Y})} \tag{240}$$

Proof. We prove this by induction on Coupling 69. For the base case we have $X_0 = Y_0$. Now suppose for $s > 0$, $Y_s = X_{s+T_{\text{left}}-T_{\text{right}}}$. Let Y_{s+1} the $s+1$-th step. There are two possibilities: if $Y_s < \frac{5}{6}n$, then $\alpha(Y_s) > 0$. In this case, s will be incremented once while t may be incremented many times. The number of times t will advance is distributed according to $\mathsf{Geo}(\alpha(Y_s))$. Let $X' = Y_{s+1}$, i.e. the location on the chain after one step of Q. We show that X' is distributed according to $X_{s+T_{\text{left}}-T_{\text{right}}+\mathsf{Geo}(\alpha(Y_s))}$. To see this note:

$$\begin{aligned}
\Pr[X' = x | X_{s+T_{\text{left}}-T_{\text{right}}} = x] &= \alpha(x) + (1-\alpha(x))w(x) \\
&= 1 - p(x) - q(x) = P(x,x) \\
\Pr[X' = x+1 | X_{s+T_{\text{left}}-T_{\text{right}}} = x] &= (1-\alpha(x))(1-w(x))\frac{p(x)}{p(x)+q(x)} = p(x) \\
&= P(x, x+1) \\
\Pr[X' = x-1 | X_{s+T_{\text{left}}-T_{\text{right}}} = x] &= (1-\alpha(x))(1-w(x))\frac{q(x)}{p(x)+q(x)} \\
&= q(x) = P(x, x-1). \qquad (241)
\end{aligned}$$

Now if $Y_s \geq \frac{5}{6}n$ then if $Y_{s+1} \neq Y_s$, $X_{s+T_{\text{left}}-T_{\text{right}}+1} = Y_{s+1}$. But if $Y_{s+1} = Y_s$, then with probability $\beta(Y_s)$ the X process skips this, ie, $Y_s = X_{s+1+T_{\text{left}}-T_{\text{right}}-1}$. Let $x \geq \frac{5}{6}n$. Let E_+ be the event that $X_{s+T_{\text{left}}-T_{\text{right}}+1} = x+1$ conditioned on $X_{s+T_{\text{left}}-T_{\text{right}}} = x$. Then

$$\Pr[E_+] = (1-w(x))\frac{p(x)}{p(x)+q(x)} + \beta(x)w(x)\Pr[E_+] \qquad (242)$$

which implies that $\Pr[E_+] = P(x, x+1)$. Similarly if we define E_- to be the event that $X_{s+T_{\text{left}}-T_{\text{right}}+1} = x-1$ conditioned on $X_{s+T_{\text{left}}-T_{\text{right}}} = x$, then

$$\Pr[E_-] = (1-w(x))\frac{q(x)}{p(x)+q(x)} + \beta(x)w(x)\Pr[E_-] \qquad (243)$$

which implies that $\Pr[E_-] = P(x, x-1)$. Using this, if E_0 is the event that $X_{s+T_{\text{left}}-T_{\text{right}}+1} = x$ conditioned on $X_{s+T_{\text{left}}-T_{\text{right}}} = x$ then $\Pr[E_0] = \Pr[(E_+ \cup E_-)^{\mathsf{c}}] = P(x,x)$. □

We need the following two theorems which basically assert that 1) the wait time during the accelerated process is not too long and 2) the accelerated chain mixes after $O(n \ln^2 n)$ steps in the $\|\cdot\|_*$ norm.

Theorem 72 (Wait-time bound). *Let $Y_0, Y_1, \ldots, Y_s$ be s steps of the accelerated Markov chain defined in (234) for $Y_0 \sim \text{Bin}(n, 1/2)$, and W_s be the number of steps Markov chain $X_0, X_1, \ldots$ has waited after s steps of the accelerated chain. Then for $s = O(n \ln n)$, and for any constant $\alpha > 0$ there exists a constant c such that*

$$\Pr[T_{left}(s) \geq cn \ln^2 n] \leq 2^{-n} \cdot n^{-\alpha}. \qquad (244)$$

Theorem 73 (Accelerated-chain mixing). *If $s \geq 3n \ln n$ then*

$$\|Q_s\|_* \leq \frac{27}{2^n + 1}\left(1 + \frac{1}{\text{poly}(n)}\right). \qquad (245)$$

Also, the following theorem combines Theorems 72 and 73 to argue that the original Markov chain mixes rapidly in the $\|\cdot\|_*$ norm.

Proposition 74. *Let*

$$\|f\|_{\square} := \sum_{k=1}^{n} |f(k)| \frac{3n}{k3^k} \tag{246}$$

For any $t_0 \leq t_1 \leq t_2$*:*

$$\mathop{\mathbb{E}}_{\tau} \|P_\tau\|_* \leq \frac{t_0}{2} \cdot \Pr[T_{left}(t_0) \geq t_1 - t_0] + \frac{1}{T} \sum_{t_0 \leq s \leq 4t_2} \|Q_s\|_{\square} + 6t_2 \frac{1}{1.4^{t_2}} \tag{247}$$

where, $T = t_2 - t_1 + 1$.

Proof of Theorem 61. We need to find suitable values for t_0, t_1, t_2. Let $t_0 = 3n \ln n$ so that $\max_{s \in [t_0, t_2]} \|Q_s\|_{\square} \leq \frac{27}{2^n+1}(1 + \frac{1}{\text{poly}(n)})$ in Proposition 74. Next, choose c to be large enough so that (using Theorem 72) if $t_1 = cn \ln^2 n$

$$\Pr[T_{\text{left}}(t_0) \geq t_1 - t_0] \leq \frac{1}{2^n + 1} \frac{1}{n^3}. \tag{248}$$

Finally, let $c' > c$ be any constant and choose $t_2 = c' t_1$. Using Theorem 73 we conclude that:

$$\frac{1}{T} \sum_{\tau=t_1}^{t_2} \|P_\tau\|_* \leq \frac{28}{2^n + 1}. \tag{249}$$

This implies that there exists a value $t_1 \leq t^* \leq t_2$ such that

$$\|P_{t^*}\|_* \leq \frac{1}{T} \sum_{\tau=t_1}^{t_2} \|P_\tau\|_* \leq \frac{28}{2^n + 1}. \tag{250}$$

Since t^* is related to $n \ln^2 n$ by a constant, this implies the proof.

□

It remains to prove Theorems 72 and 73 and Proposition 74. We prove Theorem 72 in Sects. 4.5.2 and 4.5.3, Theorem 73 in Sects. 4.5.4 and 4.5.5, and Proposition 74 in Sect. 4.5.6.

4.5.2. Wait-time analysis. In this section we prove Theorem 72. Before getting to the proof we need some preliminaries. Sites with low Hamming weight have the largest wait times. Hence, intuitively, we want to say that during the accelerated walk, these sites are not hit so often. More formally, let $N_x = \sum_{\tau=1}^{s} I\{Y_\tau \leq x\}$ and let $\beta > 1$.

Proposition 75. *Let* $\nu = 3/4n$. *For* $x \leq \nu/\beta$, $\Pr[N_x \geq \beta x] \leq s^{3/2} e \cdot e^{-\frac{\beta}{8} x}$.

If we set $\beta = 8(4+c)\ln n$ then Proposition 75 implies that

$$\Pr[N_x \geq \beta x] \leq \frac{1}{\binom{n}{x} n^c}. \tag{251}$$

Let $x(0)$ denote the corresponding ν/β, i.e.

$$x(0) := \frac{\nu}{8(4+c)\ln n}. \tag{252}$$

Proof. We observe that N_x conditioned on $Y_0 = z \geq 1$ is stochastically dominated by the same variable conditioned on $Y_0 = 1$. The proof is by just taking the natural coupling that makes sure the latter walk is always $\leq$ the former. Hence we can assume that the walk starts out from $Y_0 = 1$ and we will obtain a valid upper bound.

In [34] (see the proof of lemma A.5) the authors show that

$$\Pr[N_x \geq \beta x] \leq \sum_{\tau=\beta x}^{s} \Pr[Y_\tau \leq x]. \tag{253}$$

To understand these probabilities we will develop an exactly solvable analogue for Y_τ. Although Y_τ is a random walk in discrete time and space, we can approximate it by a process that takes place in continuous time and space. If Y_τ were an unbiased random walk then we could approximate it with Brownian motion. However, it is biased to always drift towards the point $\frac{3}{4}n$. The continuous-time-and-space random process which diffuses like Brownian motion but is biased to drift towards a fixed point is called the Ornstein–Uhlenbeck process. We will not prove a formal connection between Y_τ and the Ornstein–Uhlenbeck process, but instead will prove bounds on Y_τ that are inspired by the analogous facts about Ornstein–Uhlenbeck.

Proposition 76 (Connection with the Ornstein–Uhlenbeck process). *Define*

$$\nu_\tau := ze^{-\frac{4\tau}{3n}} + \frac{3}{4}n\left(1 - e^{-\frac{4\tau}{3n}}\right). \tag{254}$$

Then we can bound

$$\Pr[Y_\tau \leq x] \leq \sqrt{\tau}e \cdot e^{-\frac{(\nu_\tau - x)^2}{2\nu_\tau}} \tag{255}$$

The proof is in Sect. 4.5.3.

This proposition is inspired by the fact that the exact solution to the Ornstein–Uhlenbeck process is a Gaussian with mean and variance both equal to ν_τ. We can see that once $\tau \gtrsim n\ln n$, this is close to a Gaussian centered at $\frac{3}{4}n$, i.e. the stationary distribution.

Note that ν_τ is an increasing function of τ, and furthermore, for $\nu_\tau \geq x$, $e^{-\frac{(\nu_\tau - x)^2}{2\nu_\tau}}$ is decreasing in ν_τ, and therefore τ. Hence the sum in (253) can be bounded by

$$\Pr[N_x \geq \beta x] \leq s^2 e \cdot \exp\left(-\frac{(\nu(1 - e^{-\frac{\beta x}{\nu}}) - x)^2}{2\nu(1 - e^{-\frac{\beta x}{\nu}})}\right). \tag{256}$$

Using the following inequalities

$$\frac{u}{1+u} \leq 1 - e^{-u} \leq u, \text{ for } u \leq 1. \tag{257}$$

we find that

$$\Pr[N_x \geq \beta x] \leq s^{3/2} e \cdot \exp\left(- \frac{\left(\frac{\beta x}{1+\frac{\beta x}{\nu}} - x \right)^2}{2\beta x} \right). \tag{258}$$

Since $\frac{\beta x}{\nu} < 1$ then

$$\Pr[N_x \geq \beta x] \leq s^{3/2} e \cdot e^{-\frac{\beta}{8} x}. \tag{259}$$

□

Now following [18,27,34], define the good event $A := \cap_{1 \leq x \leq x(0)} \{N_x \leq \beta \cdot x\}$. Recall that $\beta = 8(4+c) \ln n$ and $x(0) = \nu/\beta$.

Proposition 77. $\Pr[A^{\mathsf{C}} | Y_0] \leq \frac{2}{\binom{n}{Y_0} n^{c-1}}$.

To prove Proposition 77, we will need a bound on the minimum site visited during the accelerated walk. Let $M_s := \min_{1 \leq i \leq s} \{Y_i\}$. Then

Proposition 78. $\Pr[M_s \leq a | Y_0 = z] \leq s \frac{\binom{n}{a} 3^a}{\binom{n}{z} 3^z}$

We need the following lemma which is a standard fact about Markov chains.

Lemma 79. *Let $Y_0, \ldots$ be a Markov chain with stationary distribution π then for any x, y in the state space and integer $s > 0$*

$$\Pr[Y_s = y | Y_0 = x] \leq \frac{\pi_y}{\pi_x}. \tag{260}$$

Proof.

$$\Pr[Y_s = y | Y_0 = x] = \frac{1}{\pi_x} \pi_x \Pr[Y_s = y | Y_0 = x] \tag{261}$$

$$\leq \frac{1}{\pi_x} \sum_z \pi_z \Pr[Y_s = y | Y_0 = z] \tag{262}$$

$$\leq \frac{\pi_y}{\pi_x} \tag{263}$$

□

Proof of Proposition 78.

$$\begin{aligned}\Pr[M_s \le a | Y_0 = z] &\le \Pr[\cup_{1\le i\le s}\{Y_i = a\} | Y_0 = z] \\ &\le \sum_{j=1}^{s} \Pr[Y_j = a | Y_0 = z] \\ &\le s \cdot \frac{\pi_a}{\pi_z} && \text{using Lemma 79} \\ &= s \cdot \frac{\binom{n}{a}3^a}{\binom{n}{z}3^z}. && (264)\end{aligned}$$

□

Now we show that the event $A = \cap_{1\le x\le x(0)}\{N_x \le \beta \cdot x\}$ is very likely.

Proof of Proposition 77. The proof is very similar to the proof of lemma 4.5 in Brown and Fawzi [18].

$$\begin{aligned}\Pr[A^{\mathsf{c}}] &= \Pr[\cup_x N_x > \beta \cdot x] \\ &\le \sum_x \Pr[N_x > \beta \cdot x] \\ &\le \sum_{x<M_s} \Pr[N_x > \beta \cdot x] + \sum_{M_s\le x<Y_0} \Pr[N_x > \beta \cdot x] \\ &\quad + \sum_{x(0)\ge x\ge Y_0} \Pr[N_x > \beta \cdot x] && (265)\end{aligned}$$

In the last line we have used the fact that $M_s \le Y_0$. Now we will handle each term in (265) separately. When $x < M_s$, $N_x = 0$, so $\sum_{x<M_s} \Pr[N_x > \beta \cdot x] = 0$. Next when $x \ge Y_0$, we can use Proposition 75 to bound $\Pr[N_x > \beta x] \le \binom{n}{Y_0} n^{-c}$. Finally, when $M_s \le x < Y_0$, we have

$$\begin{aligned}\Pr[N_x > \beta \cdot x] &= \Pr[N_x > \beta \cdot x | M_s \le x] \Pr[M_s \le x] \\ &\le \Pr[N_x > \beta \cdot x | Y_0 = 1] \Pr[M_s \le x] \\ &\le \Pr[M_s \le x] \cdot \frac{1}{\binom{n}{x} n^c} && \text{using Proposition 75} \quad (266) \\ &\le \frac{\binom{n}{x}}{\binom{n}{Y_0} 3^{Y_0 - x}} \cdot \frac{1}{\binom{n}{x} n^c} && \text{using Proposition 78} \quad (267)\end{aligned}$$

We now combine these contributions and sum over x to obtain

$$\begin{aligned}\Pr[A^{\mathsf{c}}] &\le s \frac{1}{\binom{n}{Y_0} n^c} \sum_{x<Y_0} 3^{Y_0 - x} + \sum_{x(0)\le x\le Y_0} \frac{1}{\binom{n}{Y_0} n^c} && (268) \\ &\le s \frac{1}{2\binom{n}{Y_0} n^c} + \frac{1}{\binom{n}{Y_0} n^{c-1}} && (269) \\ &\le \frac{2}{\binom{n}{Y_0} n^{c-2}}. && (270)\end{aligned}$$

□

Proof of Theorem 72. Recall that the initial position on the chain Y_0 is distributed according to a binomial around $n/2$. Hence it is enough to show that starting from position Y_0 on the chain the probability that the wait time is larger than the bound stated in the theorem is bounded by

$$\frac{1}{\binom{n}{Y_0}\,\mathrm{poly}(n)}. \tag{271}$$

If such bound holds than the probability of waiting too long is bounded by

$$\frac{1}{2^n-1}\sum_{Y_0=1}^{n}\frac{\binom{n}{Y_0}}{\binom{n}{Y_0}\,\mathrm{poly}(n)} = \frac{1}{2^n+1}\cdot\frac{1}{\mathrm{poly}(n)}. \tag{272}$$

We achieve this in the following.

Let a be a constant. Consider the following bound on the wait-time random variable $T_{\mathrm{left}}(s) = T_{\mathrm{left}}(Y_0) + \ldots + T_{\mathrm{left}}(Y_s)$:

$$\begin{aligned}\Pr[T_{\mathrm{left}}(s)\geq a] &\leq \Pr[T_{\mathrm{left}}(s)\geq a|A]+\Pr[A^{\mathsf{c}}]\\ &\leq \sum_{m=1}^{Y_0}\Pr[T_{\mathrm{left}}(s)\geq a|A,M_s=m]\Pr[M_s=m]+\Pr[A^{\mathsf{c}}]\\ &\leq \frac{1}{\binom{n}{Y_0}3^{Y_0}}\sum_{m=1}^{Y_0}\Pr[T_{\mathrm{left}}(s)\geq a|A,M_s=m]\binom{n}{m}3^m+\frac{2}{\binom{n}{Y_0}n^{c-2}},\end{aligned} \tag{273}$$

using Propositions 78 and 77.

Let $\rho_x = N_x - N_{x-1}$ be the number of times site x has been visited during s rounds of the accelerated walk. Recall from Sect. 4.5 that

$$T_{\mathrm{left}}(s) \preceq \sum_{x=1}^{5n/6}\rho_x\cdot\mathsf{Geo}(\frac{6x}{5n}). \tag{274}$$

Hence we need a concentration bound for sums of geometric random variables. Fortunately we know the following Chernoff-type tail bounds on the sum of geometric random variables.

Theorem 80 (Janson [38]). *Let $G=\sum_{i=1}^{n}\mathsf{Geo}(p_i)$ be the sum of independent geometric random variables with parameters $p_1,\ldots,p_n$, and let $p^*=\min_i p_i$ and $\phi := \sum_{i=1}^{n}\frac{1}{p_i}=\mathbb{E}G$, then for any $\lambda\geq 1$*

$$\Pr[G\geq\lambda\phi]\leq\frac{1}{\lambda}(1-p^*)^{(\lambda-1-\ln\lambda)\phi}, \tag{275}$$

The bound we need for our results is:

Corollary 81. *Let G be sum of s geometric random variables with parameters $p^* = p_1 \leq \ldots \leq p_s$, and $\mathbb{E}G = \phi$. If $T > 3c\ln(c)\phi$, then*

$$\Pr[G > T] \leq \frac{1}{3c\ln c}(1-p^*)^{T(1-1/c)}. \tag{276}$$

In particular, we can say that if $T > \mathbb{E}W$, then for any constant c there exists a constant c' such that

$$\Pr[G > c'T] \leq (1-p^*)^{cT}. \tag{277}$$

Proof. It is enough to show that if $\lambda > 3c\ln c$ then $\lambda - 1 - \ln\lambda > \lambda(1-1/c)$. Let $f(\lambda) := \frac{\lambda}{c} - \ln(e\lambda)$ for $c > 1$, we observe that f is an increasing function for $\lambda > c$. We need to find a point λ^* such that $f(\lambda^*) > 0$, and one can check that $\lambda^* = 3c\ln c$ works. □

In order to employ Corollary 81 in the context of wait time (specifically (273)) we just need to find an upper bound on the expected wait time. Now we condition on A. Hence for $x \leq x(0)$, $N_x \leq \beta x$. Among all possibilities given by event A, the wait time is maximized when the minimum visited site (M_s) is visited as often as possible (see Brown-Fawzi [18] for a discussion). So it will suffice to bound the wait time for the situation when $x \leq x(0)$, $\rho_x = \beta$ and for $x = x(0)$, $\rho_x = s - \beta x(0)$. In this case, the expected wait time (conditioned on any starting point) is bounded by

$$\mathbb{E}[T_{\text{left}}(s)|A] \leq \beta \sum_{1 \leq x \leq x(0)} \frac{5n}{2x} + (s - \beta x(0))\frac{5n}{2x(0)} \tag{278}$$

Assuming the parameters in Proposition 75 we find that $\mathbb{E}[T_{\text{left}}(s)|A] = O(n\ln^2 n + s\ln n)$, and in particular if $s = O(n\ln n)$ then $\mathbb{E}[T_{\text{left}}(s)|A] = O(n\ln^2 n)$.

Therefore using Lemma 80 for any $C > 0$ there exists a large enough constant C' such that

$$\Pr[T_{\text{left}}(s) \geq C'n\ln^2 n | H, M_s = m] \leq e^{-C\cdot\frac{m}{n}\cdot n\ln^2 n}. \tag{279}$$

Combining this with (273) and choosing C large enough yields

$$\begin{aligned}\Pr[T_{\text{left}}(s) \geq C'n\ln^2 n] &\leq \frac{1}{\binom{n}{Y_0}3^{Y_0}} \sum_{m=1}^{Y_0} e^{-C\cdot\frac{m}{n}\cdot n\ln^2 n}\binom{n}{m}3^m + \frac{2}{\binom{n}{Y_0}n^{c-2}} \\ &\leq \frac{3}{\binom{n}{Y_0}\cdot n^{c-2}}.\end{aligned} \tag{280}$$

and this completes the proof. □

4.5.3. Proof of Proposition 76: Connection with the Ornstein–Uhlenbeck process. We first define a new Markov chain $S'_0, S'_1, S'_2, \ldots$ which is easier to analyze and gives us useful bounds for the Markov chain $S_0, S_1, S_2, \ldots$.

Definition 82. $S'_0, S'_1, S'_2, \ldots$ is the following Markov chain. The state space is $\{0,1\}^n$. The initial string S'_0 is sampled uniformly at random from $\{0,1\}^n \backslash 0^n$. At each step t, S'_{t+1} results from S'_t by picking a random position of S'_t. If it was a zero we flip it, otherwise if it was a 1 with probability $1/3$ we flip it and with probability $2/3$ it doesn't change.

The Hamming weight of these strings corresponds to the position on a birth-and-death chain on the state space $\{0, 1, 2, \ldots, n\}$. Given a string $S' \in \{0,1\}^n$ the probability that the Hamming weight of S' increases by 1 is $1 - x/n$ and the probability that it decreases is $\frac{x}{3n}$. Let Q' be the transition matrix describing the Hamming weight.

We now claim that:

Proposition 83. *Starting from a string of Hamming weight* ≥ 1*, at any time* t*,* Y_t *stochastically dominates* Y'_t*, meaning that*

$$\Pr[Y'_t \geq k] \leq \Pr[Y_t \geq k] \tag{281}$$

Proof. It is enough to observe that for $0 \leq x \leq n$, the probability of moving forward for Q is larger than the probability of moving forward for Q', and also the probability of moving backwards for Q is smaller than the probability of moving backwards for Q'. □

Now suppose that we simulate Q' for T steps. First, instead of considering T steps we consider this number to be a Poisson random variable $T \sim \mathsf{Pois}(\tau)$, where τ is some positive real number. Let f_l be the number of times that site l is hit after T steps. Then $(f_1, \ldots, f_n) \sim \mathsf{Multi}(T, \frac{1}{n}, \ldots, \frac{1}{n})$ is the number of times each position in $[n]$ is hit after T steps. Here $\mathsf{Multi}(T, \frac{1}{n}, \ldots, \frac{1}{n})$ is the multinomial distribution over n items summing up to T, each happening with probability $1/n$.

We can then consider T in turn to be a random variable distributed according to $T \sim \mathsf{Pois}(\tau)$. It turns out that defining T in this way will make $f_1, \ldots, f_n$ independent. Moreover, for any $l \in \{1, \ldots, n\}$,

$$f_l \sim \mathsf{Pois}(\tau/n). \tag{282}$$

In other words, the number of times each site is hit is independently distributed according to a Poisson distribution. This technique is sometimes called Poissonization.

Now suppose the l'th bit of S'_0 starts out from 0 and that $f_l = k$. We find that the probability of ending up with a 1 in this case is

$$p_k = \frac{3}{4}\left(1 - \left(\frac{-1}{3}\right)^k\right), \tag{283}$$

and the probability of reaching a 0 is

$$1 - p_k = \frac{1}{4} + \frac{3}{4}\left(\frac{-1}{3}\right)^k. \tag{284}$$

Using these two probabilities and taking the expectation over the Poisson measure we can compute

$$\begin{aligned}\Pr[S'_T[l] = 1 | S'_0[l] = 0] &= \sum_{k=0}^{\infty} \frac{e^{-\tau/n}}{k!} (\tau/n)^k \left(\frac{3}{4} - \frac{3}{4}(-1/3)^k\right) \\ &= \frac{3}{4}\left(1 - e^{-\frac{4\tau}{3n}}\right) \\ &=: \alpha_\tau. \end{aligned} \tag{285}$$

Note that the T on the LHS is still a random variable distributed according to $\mathsf{Pois}(\tau)$.

For the case when the l'th bit starts out equal to 1 and $f_l = k$, we find the probabilities in a similar way. The probability of ending up in bit 1 is

$$p_k = \frac{3}{4} + \frac{1}{4}\left(\frac{-1}{3}\right)^k, \tag{286}$$

and the probability of ending up in 0 is

$$1 - p_k = \frac{1}{4} - \frac{1}{4}\left(\frac{-1}{3}\right)^k. \tag{287}$$

We then compute

$$\begin{aligned}\Pr[S'_T[l] = 1 | S'_0[l] \neq 0] &= \sum_{k=0}^{\infty} \frac{e^{-\tau/n}}{k!} (\tau/n)^k \left(\frac{3}{4} + \frac{1}{4}(-1/3)^k\right) \\ &= \frac{3}{4} + \frac{1}{4} e^{-\frac{4\tau}{3n}} \\ &=: \beta_\tau. \end{aligned} \tag{288}$$

As a result conditioned on $|S'_0| = z$,

$$Y'_T \sim Y'_{\mathsf{Pois}(\tau)} \sim \mathsf{Bin}(n - z, \alpha_\tau) + \mathrm{Bin}(z, \beta_\tau). \tag{289}$$

This has expectation equal to

$$\mathbb{E}\left[Y'_T | Y'_0 = z\right] = z e^{-\frac{4\tau}{3n}} + \frac{3}{4} n \left(1 - e^{-\frac{4\tau}{3n}}\right), \tag{290}$$

which is simply equal to ν_τ, which was first introduced in (254). Next, using a simple Chernoff bound for sum of binomial random variables we can show that for all $x < \nu_j$

$$\Pr[Y'_{\mathsf{Pois}(\tau)} \leq x] \leq e^{-\nu_\tau \frac{(1 - x/\nu_\tau)^2}{2}} = e^{-\frac{(\nu_\tau - x)^2}{2\nu_\tau}}. \tag{291}$$

This bound is exactly the one that we expect from an Ornstein–Uhlenbeck process.

Fix a number $x \in [n]$. Let B (the bad event) be $\{|S'_T| \leq k\}$. Then

$$\Pr[B] = \sum_{s=0}^{\infty} \Pr[T = s] \Pr[B | T = s] \tag{292}$$

$$\geq \Pr[T = \tau] \Pr[B | T = \tau] \tag{293}$$

$$\geq \Pr[T = \tau] \Pr[|S'_\tau| \leq x] \tag{294}$$

We can evaluate $\Pr[T = \tau] = \frac{\tau^\tau}{\tau!} e^{-\tau} \geq \frac{1}{\sqrt{\tau} e}$, where we have use Stirling's formula (from wikipedia) which states that $\frac{\tau!}{(\tau/e)^\tau} \leq e\sqrt{\tau}$. Together with the bound in (294) we find that

$$\Pr[Y'_\tau \leq x] \leq \sqrt{\tau} e \cdot \Pr[B] \tag{295}$$

Combining this inequality with (291) we conclude that

$$\Pr[Y'_\tau \leq x] \leq \sqrt{\tau} e \cdot e^{-\frac{(\nu_\tau - x)^2}{2\nu_\tau}} \tag{296}$$

Using Proposition 83

$$\Pr[Y_\tau \leq x] \leq \sqrt{\tau} e \cdot e^{-\frac{(\nu_\tau - x)^2}{2\nu_\tau}} \tag{297}$$

If $\tau \geq \frac{3}{4} n \ln n$ then $\frac{3}{4} n \geq \nu_\tau \geq \frac{3}{4} n - 1$. Therefore

$$\Pr[Y_\tau \leq x] \leq \sqrt{\frac{3}{4} n \ln n} e \cdot e^{-\frac{2\left(\frac{3}{4} n - x - 1\right)^2}{3n}} \tag{298}$$

4.5.4. Proof of Theorem 73: exact solution to the Markov chain Q. In this section we give an exact solution to the Markov chain Q defined in Sect. 4.5. Here, by giving an exact solution we mean we can find the eigenvalues and eigenvectors of the transition matrix explicitly and evaluate the norm $\|Q_t\|_*$. The construction follows nearly directly from a result of Kac [39].

Recall the transition probabilities of Markov chain Q according to Equation (234). In (234), Q is defined over the state space $[n]$. Without loss of generality and for convenience we can relabel the state space to $\{0, 1, 2, \ldots, n-1\}$ and redefine the transition matrix according to:

$$\begin{aligned} p_i &:= Q(i, i+1) = \frac{3(n-i-1)}{3n-1}, \\ q_i &:= Q(i, i-1) = \frac{i}{3n-1}, \\ r_i &:= Q(i, i) = \frac{2(i+1)}{3n-1}. \end{aligned} \tag{299}$$

for $i \in \{0, 1, 2, 3, \ldots, n-1\}$.

Now we consider the eigenvalue problem

$$x^{(\lambda)} Q = \lambda x^{(\lambda)}, \tag{300}$$

where $x^{(\lambda)}$ is a row vector with entries $x^{(\lambda)}(i)$, is the left eigenvector corresponding to the eigenvalue λ. For now we drop the superscript λ in $x^{(\lambda)}$. Expanding this equation we have

$$p_{i-1} x(i-1) + r_i x(i) + q_{i+1} x(i+1) = \lambda x(i). \tag{301}$$

Notice that $q_0 = p_{n-1} = 0$. Define the generating function

$$g_\lambda(z) = \sum_{i=0}^{\infty} x(i) z^i, \tag{302}$$

where for $i \geq n$, we set $x(i) = 0$. It suffices to solve (301) subject to the boundary conditions $x_{-1} = x_n = 0$. For $i > 0$ we can write

$$p_{i-1} x(i-1) z^i + r_i x(i) z^i + q_{i+1} x(i+1) z^i = \lambda x(i) z^i, \tag{303}$$

assuming $x_{-1} = 0$. Using the coefficients of (299) we get

$$\frac{3(n-i)}{3n-1} x(i-1) z^i + \frac{2(i+1)}{3n-1} x(i) z^i + \frac{i+1}{3n-1} x(i+1) z^i = \lambda x(i) z^i. \tag{304}$$

For $i = 0$ the equation is

$$x_1 = ((3n-1)\lambda - 2) x(0). \tag{305}$$

Summing $(\sum_{i=0}^{\infty})$ over the first term in the left-hand side of (304) we obtain

$$\frac{3(n-1)}{3n-1} z \cdot g_\lambda(z) - \Big(\frac{3}{3n-1}\Big) z^2 \frac{d}{dz} g_\lambda(z). \tag{306}$$

Similarly for the second term we get

$$\frac{2}{3n-1} g_\lambda(z) + \Big(\frac{2}{3n-1}\Big) z \frac{d}{dz} g_\lambda(z), \tag{307}$$

and for the third term

$$\Big(\frac{1}{3n-1}\Big) \frac{d}{dz} g_\lambda(z), \tag{308}$$

and for the term on the right-hand side

$$\lambda g_\lambda(z) \tag{309}$$

Let $\lambda' = \lambda \frac{3n-1}{3(n-1)} - \frac{2}{3(n-1)}$. Putting all of these together we obtain the following first order differential equation

$$\frac{1}{g_\lambda(z)} \frac{d}{dz} g_\lambda(z) = (n-1) \frac{3\lambda' - 3z}{-3z^2 + 2z + 1}, \tag{310}$$

with the boundary conditions

$$g_\lambda(0) = x(0), \tag{311}$$

$$\frac{d^n}{dz^n} g(0) = 0. \tag{312}$$

Assume $n-1$ is divisible by 4. Solving this differential equation and applying the first boundary condition ($g_\lambda(0) = x(0)$) we get

$$g_\lambda(z) = x^{(\lambda)}(0)(1+3z)^{\frac{n-1}{4}(1+3\lambda')}(1-z)^{\frac{n-1}{4}(3-3\lambda')}. \tag{313}$$

The second boundary condition basically says that $g_\lambda(z)$ should be a polynomial of degree at most $n-1$. This implies that $3\lambda'(n-1)/4$ should be an integer. Since the exponents of both the $(1+3z)$ and the $(1-z)$ terms should be nonnegative, we can further constrain $3\lambda'(n-1)/4$ to lie in the interval $[-\frac{n-1}{4}, 3\frac{n-1}{4}]$. These constraints are enough to determine the n eigenvalues $\lambda_0, \ldots, \lambda_{n-1}$. They must (up to an irrelevant choice of ordering) satisfy

$$3\lambda'_m \frac{n-1}{4} = 3\frac{n-1}{4} - m. \tag{314}$$

Rearranging and solving for λ_m we have

$$\lambda_m = 1 - \frac{4m}{3n-1}. \tag{315}$$

The eigenvalue gap is exactly $\dfrac{4}{3n-1}$. Note for $m = 0$ we get $\lambda_0 = 1$ and

$$g_1(z) = x^{(1)}(0)(1+3z)^{n-1} = x^{(1)}(0)\sum_{i=0}^{n-1}\binom{n-1}{i}3^i z^i = \sum_i \pi(i) z^i. \tag{316}$$

In the last equation we have introduced $\pi(i)$, which is the stationary distribution. This is a binomial centered around $\frac{3}{4}(n-1)$ and shifted by 1. Its mean $\frac{3}{4}n + \frac{1}{4}$ differs from that of the non-accelerated chain by an offset of $\approx \frac{1}{4}$. We might expect a shift like this because the accelerated chain spends less time on lower values of x.

Since the stationary distribution has unit 1-norm we can evaluate

$$x^{(1)} = \frac{1}{4^{n-1}} \tag{317}$$

The eigenvectors for each eigenvalue λ can be indirectly read from the generating function $g_\lambda(z)$. We use the notation $x^{(\lambda)}$ for the eigenvector corresponding to eigenvalue λ. Also we denote the i-th component of these vectors by $x^{(\lambda)}(i)$, for $i \in \{0, 1, 2, 3, \ldots, n-1\}$.

4.5.5. Exact solution to the Markov chain Q implies a good upper bound on $\|Q_t\|_\square$. We want to use the above exact solution to derive a bound on $\|Q_t\|_\square$. We begin by stating some facts.

1. $\lambda_m = 1 - \frac{4m}{3n-1} \le e^{-\frac{4m}{3n-1}}$ for $m \in [0, n-1]$.
2. $g_m(z) = x^{(m)}(0)(1+3z)^{n-m-1}(1-z)^m = \sum_{i=0}^{n-1} x^{(m)}(i) z^i$ for $m \in [0, n-1]$.
3. $x^{(m)}Q = \lambda_m x^{(m)} = (1 - \dfrac{4m}{3n-1})x^{(m)}$ for $m \in [0, n-1]$.
4. Q is a reversible Markov chain on $\{0, \ldots, n-1\}$ with stationary distribution $\pi(i) = \binom{n-1}{i}3^i/4^{n-1}$.

Since $x^{(m)}$'s are the left eigenvectors of Q, they can be used to find the right eigenvectors $y^{(n)}$:

$$y^{(m)}(i) = \frac{x^{(m)}(i)}{\pi(i)}. \tag{318}$$

Left and right eigenvectors are orthonormal with respect to each other, i.e., for any $l, m \in [n-1]$

$$\sum_{i=0}^{n-1} x^{(m)}(i) y^{(l)}(i) = \sum_{i=0}^{n-1} \frac{x^{(m)}(i) x^{(l)}(i)}{\pi(i)} = \delta_{m,l}. \tag{319}$$

We define the following inner product between functions

$$(f, g) := \sum_i \frac{1}{\pi(i)} f(i) g(i), \tag{320}$$

according to which $\{x^{(m)} : m \in [n-1]\}$ forms an orthonormal basis, i.e.,

$$(x^{(i)}, x^{(j)}) := \delta_{i,j}. \tag{321}$$

We denote the initial distribution by $Q_0(i) = \frac{1}{2^n - 1}\binom{n}{i+1}$. Also we denote the eigenvector corresponding to eigenvalue 1 with $x^{(1)} = \pi$, which is the same as the stationary distribution. We write this initial vector as a combination of eigenvectors of the chain

$$Q_0 = \sum_{i=0}^{n-1} \alpha_i x^{(i)} \qquad \text{with} \qquad \alpha_i = (x^{(i)}, Q_0). \tag{322}$$

Therefore after t steps

$$\begin{aligned} Q_t &= \sum_{m=0}^{n-1} \alpha_m \lambda_m^t x^{(m)} = \sum_{m=0}^{n-1} (x^{(m)}, Q_0) \lambda_m^t x^{(m)}, \\ &= \sum_{m=0}^{n-1} (x^{(m)}, Q_0) \lambda_m^t x^{(m)}. \end{aligned} \tag{323}$$

We are interested in

$$\|Q_t\|_\square := (1 - 1/2^n) \frac{12}{2^n} (Q_0, Q_t) \leq \frac{12}{2^n} (Q_0, Q_t) \tag{324}$$

Using Equation (323) this can be evaluated as

$$\|Q_t\|_\square \leq \frac{12}{2^n} \Big(\sum_{m=0}^{n-1} (x^{(m)}, Q_0) \lambda_m^t x^{(m)}, Q_0 \Big) \tag{325}$$

$$= \frac{12}{2^n} \sum_{m=0}^{n-1} (x^{(m)}, Q_0)^2 \lambda_m^t \tag{326}$$

$$= \frac{12}{2^n} \sum_{m=0}^{n-1} \alpha_m^2 \lambda_m^t \tag{327}$$

As a result the problem reduces to evaluating the overlaps $\alpha_m = (x^{(m)}, Q_0)$.

$$\begin{aligned}
\alpha_m &= (x^{(m)}, Q_0) \\
&= \sum_{i=0}^{n-1} x^{(m)}(i) \frac{\frac{\binom{n}{i+1}}{2^n-1}}{\frac{\binom{n-1}{i}}{4^{n-1}} 3^i} && (328) \\
&= 3 \cdot \frac{4^{n-1}}{2^n-1} \sum_{i=0}^{n-1} x^{(m)}(i) \frac{n}{(i+1)\cdot 3^{i+1}} && (329) \\
&= 3n \cdot \frac{4^{n-1}}{2^n-1} \int_{z=0}^{1/3} g_m(z) dz && (330)
\end{aligned}$$

Now we evaluate the integral $\int_{z=0}^{1/3} g_m(z)dz$. We consider two cases, one for $m = 0$ and one for $m > 0$:

1. $m = 0$:
   In this case $g_0(z) = (1+3z)^{n-1}$. Therefore

$$\begin{aligned}
\int_{z=0}^{1/3} g_0(z)dz &= x^{(0)}(0) \int_{z=0}^{1/3} (1+3z)^{n-1} dz && (331) \\
&= x^{(0)}(0) \frac{2^n}{3\cdot n} && (332) \\
&= \frac{4}{2^n \cdot 3n} \text{using Equation (317)} && (333)
\end{aligned}$$

2. $m > 0$:
   In this case we give an upper bound on the integral

$$\begin{aligned}
\int_{z=0}^{1/3} g_m(z)dz &= x^{(m)}(0) \int_{z=0}^{1/3} (1+3z)^{n-m-1}(1-z)^m dz && (334) \\
&\le x^{(m)}(0) 2^{n-1} \int_{z=0}^{1/3} \left(\frac{1-z}{1+3z}\right)^m dz && (335) \\
&\le x^{(m)}(0) 2^{n-1} \int_{z=0}^{1/3} (1-z)^m dz && (336) \\
&\le x^{(m)}(0) \frac{2^{n-1}}{m+1} && (337)
\end{aligned}$$

   As a result we conclude that

$$\alpha_m \le \begin{cases} 1 + \frac{1}{2^{n-1}} & m = 0 \\ x^{(m)}(0) 4^n \frac{3n}{4(m+1)} & m > 0 \end{cases} \quad (338)$$

The last step is to evaluate $x^{(m)}(0)$. In order to do this we need some insight from a well studied class of polynomials known as the Krawtchouk polynomials. It turns out the Krawtchouk polynomial naturally appears in the expansion of $(1+3z)^{n-m-1}(1-z)^m$ as the coefficients of z monomials. The degree-t Krawtchouk polynomial is defined as:

$$K^{(t)}(x) := \sum_{i=0}^{t} \binom{x}{i} \binom{n-x-1}{t-i} 3^{t-i} (-1)^i. \quad (339)$$

(Elsewhere in the literature the Krawtchouk polynomials have been defined with the 3 above replaced by either 1 or some other number.) Now we evaluate the coordinates in each $x^{(m)}$ vector.

$$\begin{aligned}(1+3z)^{n-m-1}(1-z)^m &= \sum_{i=0}^{n-m-1}\binom{n-m-1}{i}3^i z^i \sum_{j=0}^{m}\binom{m}{j}(-1)^j z^j,\\ &= \sum_{i=0}^{n-m-1}\sum_{j=0}^{m}\binom{n-m-1}{i}\binom{m}{j}3^i(-1)^j z^{i+j}\\ &= \sum_{t=0}^{n-1} z^t \sum_{i=0}^{t}\binom{m}{i}\binom{n-m-1}{t-i}3^{t-i}(-1)^i,\\ &=: \sum_{t=0}^{n-1} z^t K^{(t)}(m). \end{aligned} \tag{340}$$

Hence these Krawtchouk polynomials define the eigenstates, up to overall normalization, according to

$$x^{(m)}(i) = x^{(m)}(0)K^{(i)}(m). \tag{341}$$

Moreover using the orthogonality of the $x^{(m)}$'s, we have

$$(4^n-1)x^{(m)}(0)^2\sum_{t=0}^{n-1}\frac{K^{(t)}(m)^2}{\binom{n}{t}3^t} = 1. \tag{342}$$

In order to compute $x^{(m)}(0)$ we prove the following proposition.

Proposition 84. $\sum_{t=0}^{n-1}\dfrac{K^{(t)}(m)^2}{\binom{n-1}{t}3^t} = \dfrac{4^n}{\binom{n-1}{m}3^m}$.

Proving this will require two lemmas that establish symmetry and orthogonality properties of Krawtchouk polynomials.

Lemma 85 (Orthogonality). *If we define*

$$k^{(t)}(x) := \sum_{i=0}^{t}\binom{x}{i}\binom{N-x}{t-i}p^{t-i}(-q)^i, \tag{343}$$

for $p, q \in [0,1]$ *and* $p+q = 1$. *Then these Krawtchouk polynomials satisfy the following orthogonality relationship*

$$\sum_{x=0}^{n}\binom{N}{x}p^x q^{N-x}k^{(t)}(x)k^{(s)}(x) = \binom{N}{t}(pq)^t\delta_{t,s}. \tag{344}$$

Lemma 86 (Symmetry). *The Krawtchouk polynomials obey the following symmetry relation.*

$$\frac{\binom{n-1}{x}}{3^t} K^{(t)}(x) = \frac{\binom{n-1}{t}}{3^x} K^{(x)}(t). \tag{345}$$

These two lemma are proved in appendix B.

Proof of Proposition 84. Using Lemma 85, setting $p = 3/4$ and $q = 1/4$, and $N = n-1$ and $t = s$, we have

$$4^t k^{(t)}(x) = \sum_{i=0}^{t} \binom{x}{i} \binom{n-x-1}{t-i} 3^{t-i} (-1)^i = K^{(t)}(x), \tag{346}$$

Therefore we obtain the relation

$$\sum_{x=0}^{n} \binom{n-1}{x} 3^x K^{(t)}(x)^2 = \binom{n-1}{t} 3^t 4^{n-1}. \tag{347}$$

We now use the symmetry from Lemma 86 to obtain

$$K^{(t)}(x) = \frac{3^t}{3^x} \frac{\binom{n-1}{t}}{\binom{n-1}{x}} K^{(x)}(t). \tag{348}$$

As a result

$$\sum_{x=0}^{n} \frac{K^{(x)}(t)^2}{3^x \binom{n-1}{x}} = \frac{4^{n-1}}{\binom{n-1}{t} 3^t}. \tag{349}$$

This concludes the proof. □

A corollary of Proposition 84 is that

$$x^{(m)}(0) = \frac{1}{(4^n - 1)} \sqrt{\binom{n-1}{m} 3^m}. \tag{350}$$

Plugging this into Equation (338) we get

$$\alpha_m \leq \begin{cases} 2 & m = 0 \\ \sqrt{\binom{n-1}{m} 3^m} \frac{3n}{2(m+1)} & m > 0 \end{cases} \tag{351}$$

Now we are ready to prove Theorem 73.

Proof of Theorem 73. Using Equations (327) and (351)

$$\| Q_t \|_\diamond \leq \frac{12}{2^n} \sum_{m=0}^{n-1} \alpha_m^2 \lambda_m^t \tag{352}$$

$$\leq \frac{24}{2^n} + \frac{12}{2^n} \sum_{m=1}^{n-1} \alpha_m^2 \lambda_m^t \tag{353}$$

$$\leq \frac{24}{2^n} + \frac{12}{2^n}\sum_{m=1}^{n-1}\left(\sqrt{\binom{n-1}{m}3^m}\frac{3n}{2(m+1)}\right)^2\lambda_m^t \tag{354}$$

$$\leq \frac{24}{2^n} + \frac{27n^2}{2^n}\sum_{m=1}^{n-1}\binom{n-1}{m}3^m\lambda_m^t \tag{355}$$

$$\leq \frac{24}{2^n} + \frac{27n^2}{2^n}\sum_{m=1}^{n-1}\binom{n-1}{m}\left(3e^{-\frac{4t}{3n-1}}\right)^m \tag{356}$$

$$\leq \frac{24}{2^n} + \frac{27n^2}{2^n}\sum_{m=1}^{n-1}\binom{n-1}{m}\left(3e^{-4n\ln n}\right)^m \tag{357}$$

$$\leq \frac{24}{2^n}\left(1+O\left(\frac{1}{n}\right)\right) \tag{358}$$

□

4.5.6. Proof of Proposition 74: Combining wait-time analysis with the analysis of the accelerated chain.

Proposition (Restatement of Proposition 74). *Let*

$$\|f\|_\square := \sum_{k=1}^{n}|f(k)|\frac{3n}{k3^k} \tag{359}$$

For any $t_0 \leq t_1 \leq t_2$:

$$\mathop{\mathbb{E}}_{\tau}\|P_\tau\|_* \leq \frac{t_0}{2}\cdot\Pr[T_{left}(t_0)\geq t_1-t_0] + \frac{1}{T}\sum_{t_0\leq s\leq 4t_2}\|Q_s\|_\square + 6t_2\frac{1}{1.4^{t_2}} \tag{360}$$

where, $T = t_2 - t_1 + 1$.

Proof of Proposition 74. Let $\tau \sim \mathsf{Unif}(t_1, t_2)$. Then

$$\frac{1}{T}\sum_{s=t_1}^{t_2}\|P_s\|_* = \mathop{\mathbb{E}}_{\tau}\|P_\tau\|_* \tag{361}$$

We use the notation $y^s = (y_1, \ldots, y_s)$, for y_j running over $[n]$. Consider the event $\{X_\tau = k\}$. This event is equivalent to the disjoint union $\cup_{s\geq 0}\cup_{y^s\in[n]^s:y_s=k}\{Y^s = y^s\}\cap\{W_{s-1} < \tau \leq W_s\}$. Here $y_0 \sim \mathrm{Bin}(n, 1/2)$, conditioned on $y_0 \neq 0$. Therefore

$$\begin{aligned}\Pr[X_\tau = k] &= \sum_{s\geq 0}\sum_{y^s:y_s=k}\Pr[Y^s = y^s]\Pr[W_{s-1} < \tau \leq W_s] \\ &= \sum_{0\leq s<t_0}\sum_{y^s:y_s=k}\Pr[Y^s = y^s]\Pr[W_{s-1} < \tau \leq W_s] \\ &\quad + \sum_{t_0\leq s}\sum_{y^s:y_s=k}\Pr[Y^s = y^s]\Pr[W_{s-1} < \tau \leq W_s].\end{aligned} \tag{362}$$

We first argue about the time average of the first term.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\tau} \sum_{0\le s<t_0} \sum_{y^s:y_s=k} \Pr[Y^s = y^s] \Pr[W_{s-1} < \tau \le W_s] \\
&\le \mathop{\mathbb{E}}_{\tau} \sum_{0\le s<t_0} \sum_{y^s:y_s=k} \Pr[Y^s = y^s] \Pr[W_s \ge \tau] \\
&= \mathop{\mathbb{E}}_{\tau} \sum_{0\le s<t_0} \sum_{y^s:y_s=k} \Pr[Y^s = y^s] \Pr[s + T_{\mathrm{left}(y^s)} - T_{\mathrm{right}(y^s)} \ge \tau] \\
&\le \mathop{\mathbb{E}}_{\tau} \sum_{0\le s<t_0} \sum_{y^s:y_s=k} \Pr[Y^s = y^s] \Pr[T_{\mathrm{left}(y^s)} \ge \tau - s] \\
&\le \sum_{0\le s<t_0} \sum_{y^s:y_s=k} \Pr[Y^s = y^s] \Pr[T_{\mathrm{left}(y^s)} \ge t_1 - t_0] \\
&\le t_0 \cdot \Pr[T_{\mathrm{left}}(t_0) \ge t_1 - t_0]. \qquad (363)
\end{aligned}$$

In the last step we have used the fact that $T_{\mathrm{left}(y^s)}$ is a nondecreasing function of s. To bound the contribution to the $\|\cdot\|_*$ norm, observe that $\|(1, 1, \ldots, 1)\|_* = 1/3 + 1/3^2 + \ldots \le 1/2$. Thus the contribution from the first term is $\le \frac{t_0}{2} \cdot \Pr[T_{\mathrm{left}(y^{t_0})} \ge t_1 - t_0]$.

Next we argue about the time average of the second term in (362).

$$\begin{aligned}
&\sum_{\substack{s\ge t_0 \\ y^s:y_s=k}} \Pr[Y^s = y^s] \mathop{\mathbb{E}}_{\tau} \Pr[W_{s-1} < \tau \le W_s] \\
&\le \sum_{\substack{0\le s\le 4t_2 \\ y^s:y_s=k}} \Pr[Y^s = y^s] \mathop{\mathbb{E}}_{\tau} \Pr[W_{s-1} < \tau \le W_s]. \qquad \text{(part i)} \qquad (364) \\
&+ \sum_{s>4t_2} \max_{y^s:y_s=k} \mathop{\mathbb{E}}_{\tau} \Pr[W_{s-1} < \tau \le W_s]. \qquad \text{(part ii)} \qquad (365)
\end{aligned}$$

We now analyze each part independently

(part i) Write

$$\mathop{\mathbb{E}}_{\tau} \Pr[W_{s-1} < \tau \le W_s] = \mathop{\mathbb{E}}_{W} \mathop{\mathbb{E}}_{\tau} I[W_{s-1} < \tau \le W_s]. \qquad (366)$$

Here $\mathbb{E}_W$ is the expectation value over wait times $W_{y_1}, \ldots, W_{y_s}$, and $I[W_{s-1} < \tau \le W_s]$ is the indicator of the event $W_{s-1} < \tau \le W_s$.

We first bound $\mathbb{E}_W \mathbb{E}_\tau I[W_{s-1} < \tau \le W_s]$. Fix y^s such that $y_s = k$, and for $a \le b$ integers, let $[a, b]$ denote the set $\{a, a+1, \ldots, b\}$. Then

$$\begin{aligned}
\mathop{\mathbb{E}}_{W} \mathop{\mathbb{E}}_{\tau} I[W_{s-1} < \tau \le W_s], &= \mathop{\mathbb{E}}_{W} \frac{|[t_1, t_2] \cap [W_{s-1}, W_s]|}{T}, \\
&\le \mathop{\mathbb{E}}_{W} \frac{|[W_{s-1}, W_s]|}{T}. \qquad (367)
\end{aligned}$$

There are two possibilities for the random variable $|[W_{s-1}, W_s]| = W_s - W_{s-1}$; one for $k < \frac{5}{6}n$ and one for $k \ge \frac{5}{6}n$:

$$W_s - W_{s-1} \sim \begin{cases} \mathsf{Geo}(1-\alpha(k)) & k < \frac{5}{6}n \\ \mathsf{Bern}(\beta_k) & k \ge \frac{5}{6}n \end{cases} \qquad (368)$$

Therefore

$$\begin{aligned}\mathop{\mathbb{E}}_{W}[W_s - W_{s-1}] &\le \mathsf{Geo}(1-\alpha(k)) + \mathsf{Bern}(\beta_k)\\ &\le \frac{5n(n-1)}{2k(3n-1)} + 1/2\\ &\le \frac{3n}{k}. \end{aligned} \tag{369}$$

Using this in (367) and (364) we find the bound

$$\begin{aligned}&\sum_{\substack{0\le s\le 4t_2\\ y^s: y_s = k}} \Pr[Y^s = y^s] \mathop{\mathbb{E}}_{\tau} \Pr[W_{s-1} < \tau \le W_s]\\ &\le \sum_{\substack{0\le s\le 4t_2\\ y^s: y_s = k}} \Pr[Y^s = y^s] \frac{3n}{kT} \le \sum_{t_0 \le s \le 4t_2} \Pr[Y_s = k] \frac{3n}{kT} \end{aligned} \tag{370}$$

(part ii) For the second part we use

$$\begin{aligned}&\sum_{s>4t_2} \max_{y^s: y_s = k} \mathop{\mathbb{E}}_{\tau} \Pr[W_{s-1} < \tau \le W_s]\\ &\le \sum_{s>4t_2} \max_{y^s: y_s = k} \mathop{\mathbb{E}}_{\tau} \Pr[W_{s-1} < \tau]\\ &\le \sum_{s>4t_2} \max_{y^s: y_s = k} \max_{t_1 \le \tau \le t_2} \Pr[W_{s-1} < \tau]\\ &\le \sum_{s>4t_2} \max_{y^s: y_s = k} \max_{t_1 \le \tau \le t_2} \Pr[s - 1 + T_{\mathrm{left}(y^s)} < \tau + T_{\mathrm{right}(y^s)}]\\ &\le \sum_{s>4t_2} \max_{y^s: y_s = k} \max_{t_1 \le \tau \le t_2} \Pr[s - 1 - \tau < T_{\mathrm{right}(y^s)}]\\ &\le \sum_{s>4t_2} \max_{y^s: y_s = k} \Pr[s - 1 - t_2 < T_{\mathrm{right}(y^s)}] \end{aligned} \tag{371}$$

Now recall from Equation (239) we know that $T_{\mathrm{right}(y^s)}$ is statistically dominated by $\mathsf{Bin}(s, 1/2)$. So the RHS of (371) gets bounded by:

$$\begin{aligned}&\le \sum_{s>4t_2} \Pr[s - t_2 \le \mathsf{Bin}(s, 1/2)]\\ &\le \sum_{s>4t_2} \sum_{k=s-t_2}^{s} \frac{\binom{s}{k}}{2^s}\\ &\le \sum_{s>4t_2} t_2 \frac{\binom{s}{t_2}}{2^s} && \text{(using } s > 4t_2\text{)}\\ &\le \sum_{s>4t_2} t_2 \frac{\binom{s}{s/4}}{2^s} && \text{(using } s > 4t_2\text{)} \end{aligned}$$

$$\leq \sum_{s>4t_2} t_2 \frac{(4e)^{s/4}}{2^s}$$
$$\leq t_2 \cdot \sum_{s>4t_2} \frac{1}{1.09^s}$$
$$\leq 12t_2 \frac{1}{1.4^{t_2}} \tag{372}$$

Using (370), (372), (363) and (362)

$$\mathop{\mathbb{E}}_{\tau} \Pr[X_\tau = k] \leq t_0 \cdot \Pr[T_{\text{left}}(t_0) \geq t_1 - t_0] + \sum_{t_0 \leq s \leq 4t_2} \Pr[Y_s = k] \frac{3n}{kT} + 12t_2 \frac{1}{1.4^{t_2}} \tag{373}$$

Therefore

$$\mathop{\mathbb{E}}_{\tau} \| P_\tau \|_* \leq \frac{t_0}{2} \cdot \Pr[T_{\text{left}}(t_0) \geq t_1 - t_0] + \frac{1}{T} \sum_{t_0 \leq s \leq 4t_2} \| Q_s \|_{\square} + 6t_2 \frac{1}{1.4^{t_2}}. \tag{374}$$

□

4.6. Towards exact constants. Here we discuss what constant factors we may expect from the bound in Theorem 13. We do not consider the case of D-dimensional graphs here.

What is the right time scale in order to get anti-concentration? Since Pauli strings of weight k have contribution $1/3^k$ as well as expected wait-time of $\approx n/k$, it seems reasonable to guess that lower values of k contribute more to the anti-concentration probability. On the other hand, the initial distribution of k is centered around $n/2$. Still, enough probability mass survives at low values of k to yield a non-trivial lower bound in Theorem 13.

Thus, let us focus initially on walks starting with weight $k = 1$. Here the expected "escape time" from the low-k sector (say to $k = n/2$) is $\approx \frac{5}{6} n \ln n$, and, simultaneously, it takes $\approx \frac{5}{6} n \ln n$ time to hit $\frac{3}{4}n - o(n)$. This is the basis for the following conjecture. A special case of this conjecture for anti-concentration was recently resolved in [24].

Conjecture 1. *If* $t = \frac{5}{6} n \ln n + o(n \ln n)$ *then* $\Pr_{C \sim \mu_t^{(CG)}} [|\langle x| C |0\rangle|^2 \geq \frac{\alpha}{2^n}] = \Omega(1)$.

Here is the reasoning behind this conjecture. Recall that the transition matrix P is a birth-death chain, with probability of moving forward, backwards, and staying put being p_l, q_l and r_l, respectively. Let π be the stationary distribution. Let $T_l = \min\{t : X_t \geq l\}$ be the time of hitting the chain site l starting from the first site. For any birth-death chain, starting at site $l - 1$ [41], the expected time of moving one step forward is

$$\mathop{\mathbb{E}}_{l-1} [T_l] = \frac{1}{q_l} \sum_{i=1}^{l-1} \frac{\pi(i)}{\pi(l)}. \tag{375}$$

In our chain

$$\mathop{\mathbb{E}}_{l-1}[T_l] = \frac{5}{2}\sum_{i=1}^{l-1}\frac{\binom{n}{i}}{\binom{n-2}{l-2}3^{l-i}}. \tag{376}$$

In order to bound this we use the inequalities (proven in [7])

$$\binom{n-2}{l-2} \le \binom{n-2}{i-1}\left(\frac{n-i-1}{i}\right)^{l-i-1}, \tag{377}$$

and

$$\binom{n}{i} \le \binom{n}{l-1}\left(\frac{l-1}{n-l+2}\right)^{l-i-1}. \tag{378}$$

Therefore

$$\mathop{\mathbb{E}}_{l-1}[T_l] \le \frac{5}{6}\sum_{i=1}^{l-1}\frac{\binom{n}{l-1}}{\binom{n-2}{l-2}}\left(\frac{l-1}{3(n-l+2)}\right)^{l-i-1}, \tag{379}$$

$$\le \frac{5}{6}n\left(\frac{1}{l-1}+\frac{1}{3n/4-l+7/4}\right)(1+O(1/n)). \tag{380}$$

The last line holds for $l < \frac{3}{4}n$. The transition from (379) to (380) is directly inspired by Equation (2) of [the arXiv version of] [7].

To bound the time of reaching $\frac{3}{4}n-\delta$ for some $\delta \ge 0$ we sum (380) over $1 \le l \le \frac{3}{4}n-\delta$ and neglect the $1+O(1/n)$ corrections.

$$\mathop{\mathbb{E}}_{1}[T_{\frac{3}{4}n-\delta}] \le \frac{5}{6}n\left(\sum_{l=1}^{\frac{3}{4}n-\delta}\frac{1}{\ell}+\frac{1}{3n/4-l+11/4}\right) \tag{381}$$

$$\approx \frac{5}{6}n\left(\ln\left(\frac{\frac{3}{4}n-\delta}{1}\right)+\ln\left(\frac{3n/4+7/4}{\delta+11/4}\right)\right) \tag{382}$$

$$= \frac{5}{6}n\left(\ln\frac{n^2}{\delta+1}+O(1)\right). \tag{383}$$

Using this bound, for $a < b$ we can also compute $\mathbb{E}_a[T_b]$ as $\mathbb{E}_1[T_b] - \mathbb{E}_1[T_a]$. We wish to estimate $\mathbb{E}_a[T_b]$ in two main regimes. Recall that our starting distribution is $\mathrm{Bin}(n, 1/2)$ and the stationary distribution is $\mathrm{Bin}(n, 3/4)$. Thus we need to know the time for most of the probability mass to reach $\approx 3/4n$, and for the left tail of the initial distribution to reach the left tail of the final distribution. (The right tail is less demanding and less important, because it does not have the long wait times and it is suppressed by the $1/3^k$ factors.) For the bulk of the probability distribution we use the estimate $\mathbb{E}_{n/2}[T_{3/4n-O(1)}] \lesssim \frac{5}{6}n\ln n$. For the left tail, we use the bound $\mathbb{E}_1[T_{0.74n}] \lesssim \frac{5}{6}n\ln n$. In each case the time required is $\frac{5}{6}n\log n + O(n)$.

5. Alternative Proof for Anti-concentration of the Outputs of Random Circuits with Nearest-Neighbor Gates on D-Dimensional Lattices

5.1. The $D = 2$ case. In this section we consider a simplified version of $\mu_{2,c,s}^{\text{lattice},n}$, where $c = 1$ and that $K^{(t)}_{\mu^{\text{lattice},n}_{2,1,s}} = k_R^s k_C^s$. We prove the following:

Theorem 87. *If $s = O(\sqrt{n} + \ln(1/\epsilon))$ then $\mu_{2,1,s}^{lattice,n}$ satisfies*

$$\mathop{\mathbb{E}}_{C\sim\mu_{2,s}^{lattice,n}} Coll(C) \leq \frac{2}{2^n+1}(1+\epsilon). \tag{384}$$

This result is already established in Theorem 8, we give an alternative proof based on a reduction to a classical probabilistic process. This alternative approach may help with the analysis of random circuits on arbitrary graphs.

We use the following two statements

Lemma 88 (Brandão-Harrow-Horodecki'13 [13]). *Let $t = O(\sqrt{n} + \ln \frac{1}{\epsilon})$ then*

$$Ch[g^t_{Rows}] \preceq \bigotimes_{i\in Rows} Ch[G_i]\cdot(1+\epsilon). \tag{385}$$

the same holds for $Ch[g^t_{Columns}]$.

Proof. This result is proved by Brandão-Harrow-Horodecki in [13]. □

Proposition 89. *Let K_i an ϵ approximate 2-designs on row or column $i \in \{R, C\}$, in the sense that*

$$K_i \preceq (1+\epsilon)Ch[G_i] \tag{386}$$

then for any sequence of rows or columns $i_1, \ldots, i_t$

$$Coll(K_{i_t}\ldots K_{i_1}) \leq (1+\epsilon)^t Coll(Ch[G_{i_t}]\ldots Ch[G_{i_1}]). \tag{387}$$

Proof. This proposition is proved in Sect. 3.5.1. □

Putting these together

$$\text{Coll}(\mu^{\text{lattice},n}_{2,1,s}) \leq (1+\epsilon)^2\text{Coll}(\text{Ch}[G_R G_C]). \tag{388}$$

Therefore our objective is to show that

Proposition 90.

$$Coll(Ch[G_R G_C]) \leq \frac{2}{2^n+1}\left(1+\frac{1}{\text{poly}(n)}\right). \tag{389}$$

Proof. Using the Markov chain interpretation discussed in Sect. 4, the initial distribution on the chain is

$$V_0 := \frac{1}{2^n}(\sigma_0 \otimes \sigma_0 + \sum_{\substack{p\in\{0,3\}^n \\ p\neq 0}} \sigma_p \otimes \sigma_p), \tag{390}$$

and after the application a large enough random quantum circuit the distribution converges to

$$V^* := \frac{1}{2^n}\sigma_0 \otimes \sigma_0 + (1 - \frac{1}{2^n}) \cdot \frac{1}{4^n - 1} \sum_{\substack{p\in\{0,1,2,3\}^n \\ p\neq 0}} \sigma_p \otimes \sigma_p, \tag{391}$$

and we want to see how fast this convergence happens.

For clarity, throughout this proof we represent distributions along the full lattice by capital letters (such as V) and for individual rows or columns with small letters (such as v^i for distribution v on row or column i). Also, for simplicity we write 0 instead of $\sigma_0 \otimes \sigma_0$, and σ_0^i for all zeros across row or column i.

V_0 is separable across any subset of nodes. So the initial distribution along each row or column is exactly

$$\frac{1}{2^{\sqrt{n}}}(\sigma_0 + \sum_{\substack{p\in\{0,3\}^{\sqrt{n}} \\ p\neq 0}} \sigma_p \otimes \sigma_p) =: v_0. \tag{392}$$

After one application of $\mathrm{Ch}[G_R]$ each such distributions become

$$\begin{aligned} v^* &:= \frac{1}{2^{\sqrt{n}}}\sigma_0 \otimes \sigma_0 + (1 - \frac{1}{2^{\sqrt{n}}})\frac{1}{4^{\sqrt{n}} - 1} \sum_{\substack{p\in\{0,1,2,3\}^{\sqrt{n}} \\ p\neq 0}} \sigma_p \otimes \sigma_p \\ &=: \frac{1}{2^{\sqrt{n}}}\sigma_0 + (1 - \frac{1}{2^{\sqrt{n}}})v. \end{aligned} \tag{393}$$

Here we have defined

$$v := \frac{1}{4^{\sqrt{n}} - 1} \sum_{\substack{p\in\{0,1,2,3\}^{\sqrt{n}} \\ p\neq 0}} \sigma_p \otimes \sigma_p. \tag{394}$$

therefore the distribution along the full chain is $V_1 := (\frac{1}{2^{\sqrt{n}}}\sigma_0 + (1 - \frac{1}{2^{\sqrt{n}}})v)^{\otimes\sqrt{n}}$. We also use the notation $v^y\sigma_0^{\setminus y} := \otimes_{i:y_i=1} v \otimes \bigotimes_{i:y_i=0} \sigma_0$, for $y \in \{0, 1\}^{\sqrt{n}}$.

Before getting to the analysis, we should first understand the main reason why $\mathrm{Coll}(\mathrm{Ch}[G_R])$ is large.

After we apply $\mathrm{Ch}[G_R]$ the collision probability across each row is exactly $\frac{2}{2^{\sqrt{n}}+1}$. So the collision probability across the whole lattice is $\approx \frac{2^{\sqrt{n}}}{2^n}$; which is much larger (by a factor of $2^{\sqrt{n}}$) than what we want. The crucial observation here is that if in (393) we project out all the σ_0 terms across each row, then the bound becomes $\approx \frac{1}{2^n}$. So what really slows this process are the zero σ_0 terms. The issue is that, after an application of

$\mathrm{Ch}[G_R]$, all zeros states get projected to themselves. However, if one applies $\mathrm{Ch}[G_C]$ they get partially mix with other rows. So the objective is to show that after application of $\mathrm{Ch}[G_C]\mathrm{Ch}[G_R]$ for *constant* number of times, these zeros disappear with large enough probability.

Let V_s be the distribution along the full chain after we apply $(\mathrm{Ch}[G_C]\mathrm{Ch}[G_R])^s$. Eventually we want to compute

$$\mathrm{Coll}(\mathrm{Ch}[G_s]) = \frac{1}{2^n}\mathrm{Tr}\left(v_0^{\otimes\sqrt{n}} V_s\right) =: \kappa(V_s). \tag{395}$$

Here we have defined the map

$$\kappa : A \mapsto \frac{1}{2^n}\mathrm{Tr}(V_0 A). \tag{396}$$

As a result

$$V_1 = \otimes_{r\in\mathrm{Rows}} \frac{1}{2\sqrt{n}}\sigma_0^r + (1-\frac{1}{2\sqrt{n}})v^r = \sum_{y\in\{0,1\}^{\sqrt{n}}} \frac{1}{2\sqrt{n}(\sqrt{n}-|y|)}(1-\frac{1}{2\sqrt{n}})^{|y|} v^y \sigma_0^{\setminus y}. \tag{397}$$

An important observation here is that

$$\kappa_i\left(\frac{1}{2\sqrt{n}}\sigma_0^i\right) = \frac{1}{2\sqrt{n}}, \quad \kappa\left((1-\frac{1}{2\sqrt{n}})v^i\right) = \frac{(1-\frac{1}{2\sqrt{n}})}{2\sqrt{n}+1} < \frac{1}{2\sqrt{n}}. \tag{398}$$

the relevant information here is that when κ is applied to the summation in (397), it amounts to

$$\kappa(V_1) < \frac{1}{2^n}\sum_{y\in\{0,1\}^{\sqrt{n}}} 1 = \frac{2\sqrt{n}}{2^n}. \tag{399}$$

In other words, each σ_0 term contributes to the number 1 in the above summation. That means if we had started with the distribution

$$V' = \bigotimes_{r\in\mathrm{Rows}} o(1/\sqrt{n})\frac{1}{2\sqrt{n}}\sigma_0^r + \left(1 - o(1/\sqrt{n})\frac{1}{2\sqrt{n}}\right)v^r, \tag{400}$$

then we would have obtained

$$\kappa(V') = \frac{2}{2^n+1}\left(1+\frac{1}{\mathrm{poly}(n)}\right), \tag{401}$$

which is exactly what we want. The last relevant piece of information is that if v_j'' is a distribution over row j that with probability 1 contains a nonzero item, then when $\mathrm{Ch}[G_j]$ is applied to it, it will instantly get mapped to v_j. This phenomenon is related to strong stationarity in Markov chain theory.

We claim that after the first application of $\mathrm{Ch}[G_C]$, the expected collision probability is according to the bound claimed in this theorem. In order to see this, we consider the distribution V_1 ((397)), this time along each column. Note that the distribution along columns. For any set of columns $j_1, \ldots, j_k$ let $E_{j_1,\ldots,j_k}$ be the event that these columns are all zeros, and the rest of the columns have at least one non-zero element in them. Here

we use the notation $E_{j_1,\ldots,j_k} \equiv E_y$ for $y \in \{0,1\}^{\sqrt{n}}$ such that the $j_1, \ldots, j_k$ locations of y are ones and the rest of its bits are zeros.

Therefore

$$\begin{aligned}\mathrm{Coll}(\mathrm{Ch}[G_C]V_1) &= \sum_{y\in\{0,1\}^{\sqrt{n}}} \Pr[E_y]\kappa\left(\sigma_0^y V^{\setminus y}\right)\\ &= \frac{1}{2^n} + \sum_{y\in\{0,1\}^{\sqrt{n}}\setminus} \Pr[E_y]\left(\frac{1}{2\sqrt{n}+1}\right)^{\sqrt{n}-|y|}.\end{aligned}$$

Let $p_0 := \frac{1}{2\sqrt{n}} + \frac{1}{4}(1 - \frac{1}{2\sqrt{n}})$. The main observation is that for each such y,

$$\Pr[E_y] \le p_0^{\sqrt{n}|y|}\left(1 - p_0^{\sqrt{n}}\right)^{\sqrt{n}-|y|}. \tag{402}$$

Therefore

$$\begin{aligned}\mathrm{Coll}(\mathrm{Ch}[G_C]V_1) &\le \frac{1}{2^n} + \sum_{y\in\{0,1\}^{\sqrt{n}}\setminus 0} p_0^{\sqrt{n}|y|}\left(1 - p_0^{\sqrt{n}}\right)^{\sqrt{n}-|y|}\left(\frac{1}{2\sqrt{n}+1}\right)^{\sqrt{n}-|y|}\\ &= \frac{1}{2^n} + \left(p_0^{\sqrt{n}} + \left(1 - p_0^{\sqrt{n}}\right)\frac{1}{2\sqrt{n}+1}\right)^{\sqrt{n}}\\ &= \frac{1}{2^n} + \frac{1-\frac{1}{2^n}}{2^n+1}\left(1 + \frac{1}{\mathrm{poly}(n)}\right)\\ &= \frac{2}{2^n+1}\left(1 + \frac{1}{\mathrm{poly}(n)}\right),\end{aligned} \tag{403}$$

and this completes the proof. □

5.2. Generalization to arbitrary D-dimensional case. See Sect. 2.1 for definitions in this section. In particular, we need definitions for $\mathrm{Ch}[g_i]$, K_i and $\mathrm{Ch}[G_i]$ for each coordinate i of the lattice, and $K_t = (\prod_i k_i)^t$.

In this section we prove that

Theorem 91. *D-dimensional $O\left(Dn^{1/D} + D\ln\left(\frac{D}{\epsilon}\right)\right)$-depth random circuits on n qubits have expected collision probability $\frac{2}{2^n+1}\left(1 + \frac{1}{\mathrm{poly}(n)}\right)$.*

Proof. The proof is basically a generalization of the proof for Theorem 87. Here we sketch an outline and avoid repeating details. In particular, we need generalizations of Lemma 88 and Proposition 89

The generalization of Lemma 88 is simply that k_i^t for $t = O(n^{1/D} + \ln\frac{D}{\epsilon})$ is an $\frac{\epsilon}{d}$-approximate 2-design. Proposition 89 naturally generalizes to: if for each coordinate K_i is an $\frac{\epsilon}{D}$-approximate 2-design then

$$\mathrm{Coll}\left(\prod_i K_i\right) \le \left(1 + \frac{\epsilon}{D}\right)^D \cdot \mathrm{Coll}\left(\prod_i \mathrm{Ch}[G_i]\right). \tag{404}$$

Our objective is then to show

$$\mathrm{Coll}\left(\prod_i \mathrm{Ch}[G_i]\right) = \frac{2}{2^n+1}\left(1+\frac{1}{\mathrm{poly}(n)}\right). \tag{405}$$

This last step may be the most non-trivial part in this proof.

Here we just outline the proof. For detailed discussions see the proof of Proposition 90. We first separate the all zeros state of the chain which contributes as $1/2^n$ to the expected collision probability. After the application of G_1 on the first coordinate, each row in this coordinate, will be all zeros vector with probability $1/2^{n^{1/D}}$ and V with probability $1-1/2^{n^{1/D}}$. After the application of G_2 each plane in the direction 1, 2 will be all zeros with probability $\approx 1/2^{2n^{1/D}}$ and V with probability $\approx 1-1/2^{2n^{1/D}}$. After the application of G_3 each plane in 1, 2, 3 direction is all zeros with probability $\approx 1/2^{3n^{1/D}}$ and V otherwise, and so on. Eventually after the application of G_d the distribution along the chain is all zeros with probability $\approx 1/2^{Dn^{1/D}}$ and V otherwise. At this point the distribution along each individual row in each coordinate is $\approx 1/2^{Dn^{1/D}}0 + (1 - 1/2^{Dn^{1/D}})V$. So the collision probability across each such row is

$$\approx \frac{1}{2^{Dn^{1/D}}} + \frac{1}{2^{n^{1/D}}}. \tag{406}$$

Therefore the collision probability across the full chain is

$$\approx \frac{1}{2^n} + \left(\frac{1}{2^{Dn^{1/D}}} + \frac{1}{2^{n^{1/D}}}\right)^{n^{1-1/D}} \approx \frac{1}{2^n} + \frac{1}{2^n}\exp\left(\frac{1}{2^{dn^{1/D}}}n^{1-1/D}\right). \tag{407}$$

□

Corollary 92. *$O(\ln n \ln\ln n)$-depth random circuits with long-range gates have expected collision probability $\frac{2}{2^n+1}\left(1+\frac{1}{\mathrm{poly}(n)}\right)$.*

Proof. Set $D = \ln n$ in Theorem 91. □

6. Scrambling and Decoupling with Random Quantum Circuits

In this section we reconstruct some of the results of Brown and Fawzi [17,18]. The paper [17] proves random circuit depth bounds required for scrambling and some weak notions of decoupling. We are able to use our proof technique to reconstruct and improve on the results of this paper. [18] on the other hand introduces a stronger notion of decoupling with random circuits. Unfortunately our method does not seem to yield any results about this model.

We first define an approximate scrambler based on [17].

Definition 93 (*Scramblers*). μ is an ϵ-approximate scrambler if for any density matrix ρ and subset S of qubits with $|S| \le n/3$

$$\mathop{\mathbb{E}}_{C\sim\mu} \|\rho_S(C) - \frac{I}{2^{|S|}}\|_1^2 \le \epsilon. \tag{408}$$

where $\rho_S(C) = \mathrm{Tr}_{\setminus S} C\rho C^\dagger$ and $\mathrm{Tr}_{\setminus S}$ is trace over the subset of qubits that is compliimentary to S.

We show that small depth circuits from $\mu_{D,c,s}^{\text{lattice},n}$ are good scramblers.

Theorem 94. *If $s = O(D \cdot n^{1/D} + \ln D)$ and $c = 1$ then $\mu_{D,c,s}^{lattice,n}$ is a $\frac{1}{\text{poly}(n)}$-approximate scrambler. In particular, for $D = O(\ln n)$ this corresponds to an ensemble of $O(\ln n \ln\ln n)$ depth circuits that are $\frac{1}{\text{poly}(n)}$-approximate scramblers.*

Brown and Fawzi show a circuit depth bound of $O(\ln^2 n)$ for random circuits with long-range interactions. Our result improves this to $O(\ln n \ln\ln n)$ depth. We believe that the right bound should be $O(\ln n)$. Moreover, no bound for the case of D-dimensional lattices was mentioned in their result.

Proof. We first rewrite $\mathbb{E}_{C\sim\mu}\|\rho_S(C) - \frac{I}{2^{|S|}}\|_1^2 \le 2^{|S|}\mathbb{E}_{C\sim\mu}\text{Tr}(\rho_S^2(C)) - 1$ (to see why this is true see [17]). Next, consider an arbitrary density matrix

$$\rho = \sum_{i,j} \rho_{i,j} |i\rangle \langle j| . \tag{409}$$

We first find an expression for $\text{Tr}_{\setminus S}(C\rho C^\dagger)$

$$\begin{aligned}
\text{Tr}_{\setminus S}(C\rho C^\dagger) &= \sum_{i,j} \rho_{i,j} \text{Tr}_{\setminus S}(C |i\rangle \langle j| C^\dagger) \\
&= \sum_{i,j} \rho_{i,j} \text{Tr}_{\setminus S} \sum_{g,h} C_{ig} C^*_{jh} |g\rangle \langle h| \\
&= \sum_{i,j} \rho_{i,j} \sum_{\tilde{g},\tilde{h}} \sum_{p} C_{i,\tilde{g};p} C^*_{j,\tilde{h};p} |\tilde{g}\rangle \langle \tilde{h}| .
\end{aligned} \tag{410}$$

Therefore

$$\begin{aligned}
\mathop{\mathbb{E}}_{C\sim\mu_{D,c,s}^{\text{lattice},n}} \text{Tr}_S \left(\text{Tr}_{\setminus S}\left(C\rho C^\dagger\right)\right)^2 &= \mathop{\mathbb{E}}_{C\sim\mu_{D,c,s}^{\text{lattice},n}} \text{Tr}_S \left(\sum_{i,j} \rho_{i,j} \sum_{\tilde{g},\tilde{h}} \sum_{p} C_{i,\tilde{g};p} C^*_{j,\tilde{h};p} |\tilde{g}\rangle \langle \tilde{h}| \right)^2 \\
&= \mathop{\mathbb{E}}_{C\sim\mu_{D,c,s}^{\text{lattice},n}} \sum_{i,j} \sum_{k,l} \sum_{\tilde{g}_1,\tilde{h}_1} \sum_{\tilde{g}_2,\tilde{h}_2} \sum_{p,q} \\
&\quad \rho_{i,j}\rho_{kl} C_{i,\tilde{g}_1;p} C^*_{j,\tilde{h}_1;p} C_{i,\tilde{g}_2;q} C^*_{j,\tilde{h}_2;q} \delta_{\tilde{h}_1=\tilde{g}_2} \delta_{\tilde{h}_2=\tilde{g}_1} \\
&= \mathop{\mathbb{E}}_{C\sim\mu_{D,c,s}^{\text{lattice},n}} \sum_{ij,k,l} \sum_{a,b,c,d} \rho_{i,j}\rho_{kl} C_{i,a;b} C^*_{j,c;b} C_{i,c;d} C^*_{j,a;d} \\
&= \text{Tr}\left(\rho \otimes \rho \text{Ch}\left[G^{(2)}_{\mu_{D,c,s}^{\text{lattice},n}}\right] \left(\sum_{a,b,c,d} |ab\rangle \langle cb| \otimes |cd\rangle \langle ad| \right)\right) \\
&= \text{Tr}\left(\rho \otimes \rho \text{Ch}\left[G^{(2)}_{\mu_{D,c,s}^{\text{lattice},n}}\right](A)\right).
\end{aligned} \tag{411}$$

both $\rho\otimes\rho$ and A are psd therefore using Lemma 34

$$\text{Tr}\left(\rho \otimes \rho \text{Ch}[G^{(2)}_{\mu}](A)\right) \le (1+\epsilon)^D \cdot \text{Tr}\left(\rho \otimes \rho \prod_{1\le i\le D} \text{Ch}[G_i](A)\right). \tag{412}$$

Next, using Equation 3 of [17] we reduce computation of $\mathrm{Tr}(\rho\otimes\rho\prod_{1\le i\le D}\mathrm{Ch}[G_i](A))$ to the following probabilistic process: starting from a uniform distribution over $\{0,3\}^n\backslash I^n$ show that the probability that after the application $\prod_{1\le i\le D}\mathrm{Ch}[G_i]$ the string on Markov chain K defined in Sect. 4 has weight $\le n/3$ is $\mathrm{poly}(n)/2^n$ and this reconstructs theorem A.1 of [17].

The initial state on the chain is $\frac{1}{2^n}\sum_{p\in\{0,3\}^n\backslash 00}\sigma_p\otimes\sigma_p$ we add the term $\frac{1}{2^n}\sigma_0\otimes\sigma_0$ this can only slower the process. With this modification each site is initially independently $Z\otimes Z$ or $I\otimes I$, each with probability $1/2$.

From using the proof of Theorem 91 after the application $\prod_i\mathrm{Ch}[G_i]$ the distribution along the each row is $\approx 1/2^{Dn^{1/D}}\sigma_0\otimes\sigma_0+(1-1/2^{Dn^{1/D}})V$. Therefore the probability that each site is zero is at most $1/4+1/2^{Dn^{1/D}} =: 1/4+\delta =: p_0$. Hence the probability of having at most $n/3$ is at most

$$\sum_{k=1}^{n/3}\frac{\binom{n}{k}}{4^n-1}p_0^{n-k}(1-p_0)^k=\sum_{k=1}^{n/3}\frac{\binom{n}{k}}{4^n-1}(1/4+\delta)^{n-k}(3/4-\delta)^k$$
$$\le e^{4\cdot 2/3n\cdot\delta}\sum_{k=1}^{n/3}\frac{\binom{n}{k}}{4^n-1}1/4^{n-k}(3/4)^k \tag{413}$$

which is within $1+O\left(n/2^{Dn^{1/D}}\right)$ of what we would expect from the Haar measure. Also when $D=O(\ln n)$ with a proper constant, this value is $1+1/\mathrm{poly}(n)$. □

Next, we consider the following notion of decoupling defined in [17]. Consider a maximally entangled state $\Phi_{MM'}$ along equally sized systems M and M' each with m qubits, and a pair of equally sized systems A and A'. Similar to [17] we consider two models for AA': 1) a pure state $|0\rangle_A\langle 0|$ along system A with $n-m$ qubits and 2) a maximally entangled state $\phi_{AA'}$. We then apply a random circuit to systems $M'A$ and we want that for a small subsystem S of M the final state $\rho_{MS}(t)$ be decoupled in the sense that $\rho_{MS}(t)\approx I/2^{m+s}$.

Definition 95 (*Weak decouplers*). A distribution μ over $\mathrm{U}(2^n)$ is an ϵ-approximate weak decoupler if $\|\rho_{MS}(t)-\frac{I_M}{2^{|M|}}\otimes\frac{I_S}{2^{|S|}}\|_1\le\epsilon$.

Theorem 96. *Let D be a constant integer. If $s=O(D\cdot n^{1/D})$ and $c=1$ then there exists a constant $c'<1$ such that if $m<c'n^{1/D}$ then $\mu_{D,c,s}^{lattice,n}$ is a $\frac{1}{\mathrm{poly}(n)}$-approximate weak decoupler.*

The depth bound Brown and Fawzi find in [17] for this problem is $n^{1/D}\cdot O(\ln n)$ depth for $m=\mathrm{poly}(n)$.

Proof. We first show that the bound we want to calculate for the 1-norm in this theorem can be written as $\mathrm{Tr}\left(E\mathrm{Ch}\left[G_{\mu_{D,c,s}^{\mathrm{lattice},n}}\right]F\right)$ where E and F are psd matrices. Hence using Lemma 34 we can use the overlapping projectors $\prod_i\mathrm{Ch}[G_i]$ instead $\mathrm{Ch}[G_{\mu_{D,c,s}^{\mathrm{lattice},n}}]$ as the second-moment operator.

We first start with the case when ψ_A is the pure state $|0\rangle_A\langle 0|$. The initial state is the (pure) density matrix

$$\rho_{\mathrm{init}}=\frac{1}{2^m}\sum_{i,j}|i\rangle\langle j|\otimes|i0\rangle\langle j0| \tag{414}$$

where $|i\rangle$ runs through the computational basis of M and 0 is the initial state of A. After the application of a circuit C

$$\rho_{\text{init}} \mapsto \rho_C = \frac{1}{2^m} \sum_{i,j,k,l} |i\rangle \langle j| \otimes |k\rangle \langle l| \, C_{i0;k} C^*_{j0;l}. \tag{415}$$

where $C_{a;b}$ is the ab entry of C. The density matrix corresponding to subsystem MS becomes

$$\frac{1}{2^m} \sum_{i,j,k',l',q'} |i\rangle \langle j| \otimes |k'\rangle \langle l'| \, C_{i0;k'q'} C^*_{j0;l'q'}. \tag{416}$$

We use the bound (also used in [17])

$$\|\rho_{MS}(C) - \frac{I_M}{2^{|M|}} \otimes \frac{I_S}{2^{|S|}}\|_1 \leq 2^{m+s} \text{Tr}(\rho^2_{MS}(C)) - 1. \tag{417}$$

Next, using the proof of Theorem 94 $\mathbb{E}_{C\sim\mu} \text{Tr}(\rho^2_{MS}(C))$ can be written as $\text{Tr}(C\text{Ch}[G^{(2)}_\mu]D)$ where C and D are psd, hence $\text{Tr}(C\text{Ch}[G^{(2)}_{\mu^{\text{lattice},n}_{D,c,s}}]D) \leq \text{Tr}(C \prod_i \text{Ch}[G_i]D)(1+\epsilon)$. Hence we can just use $\prod_i \text{Ch}[G_i]$ to bound the expectation $\mathbb{E}_{C\sim\mu^{\text{lattice},n}_{D,c,s}} \|\rho_{MS}(C) - \frac{I_M}{2^{|M|}} \otimes \frac{I_S}{2^{|S|}}\|_1$.

Next, we do the same calculation for the case when $\psi_{AA'}$ is the maximally entangled state $\frac{1}{2^{n-m}} \sum_{i,j} |i\rangle \langle j| \otimes |i\rangle \langle j|$. Therefore the initial density matrix is

$$\begin{aligned}\rho_{\text{init}} &= \frac{1}{2^n} \sum_{i,j,k,l} |i\rangle_M \langle j| \otimes |i\rangle_{M'} \langle j| \otimes |k\rangle_A \langle l| \otimes |k\rangle_{A'} \langle l| \\ &= \frac{1}{2^n} \sum_{i,j,k,l} |i\rangle_M \langle j| \otimes |ik\rangle_{M'A'} \langle jl| \otimes |k\rangle_{A'} . \langle l| \end{aligned} \tag{418}$$

After the application of the random circuit this gets mapped to

$$\rho_{\text{init}} \mapsto \rho(C) = \frac{1}{2^n} \sum_{i,j,k,l} |i\rangle_M \langle j| \otimes |z\rangle_{M'A'} \langle w| \otimes |k\rangle_{A'} \langle l| \, C_{ik,z} C^*_{jl,w}. \tag{419}$$

Again we can use a bound similar to (417) and similar to the proof of Theorem 94 we can show that tracing out a subsystem, the trace of the resulting density matrix squared can be written as $\text{Tr}\left(C\text{Ch}[G^{(2)}_\mu]D\right)$ for C and D psd.

As proved in theorem 3.5 of [17], the task is to show that starting with uniform distribution over all strings with weight $\leq m = O(n^{1/D})$, prove that the probability that after the application of the random circuit the weight of the string on the chain is $\geq n/2$ is at least $1 - 1/4^m$. It is enough to show that this is true for the initial state with Hamming weight 1. Without loss of generality assume the nonzero digit in this string is in the first row of the first direction. After the application of G_1 the first row in this direction becomes V. Using Chernoff bound for independent Bernoulli trials, with probability at least $1 - e^{-\Omega(n^{1/D})}$ there are at most $1/4 \cdot n^{1/D} \cdot 2^{1/D}$ zeros on this row. After the application of G_2 with probability at least $1 - e^{-\Omega(n^{1/D})}$ there are $1/4 \cdot n^{2/D} \cdot 2^{2/D}$, and so on. Hence after the completion of $\prod_i G_i$ with probability at least $1 - e^{-\Omega(n^{1/D})}$ there are at most $1/4 \cdot n^{D/D} \cdot 2^{D/D} = n/2$ zeros on the chain. For constant D the failure probability is at most $e^{-\Omega(n^{1/D})}$ and we can choose the constant c' small enough so that if $m < c' n^{1/D}$ the probability of failure is at most $1/4^m$. □

Acknowledgements. We are grateful to Scott Aaronson and Yury Polyanskiy for detailed comments and to Yuval Peres for telling us about [7]. We thank Sami Boulebnane for pointing us to some minor errors in the first version on arXiv. We also thank Matthew Khoury for numerically studying the collision probability and validating several theoretical time-scales proved in this paper. The first draft of this work was completed when SM was affiliated with CSAIL MIT.

Author Contributions The authors contributed equally to this work.

Funding 'Open Access funding provided by the MIT Libraries' AWH was funded by NSF Grants CCF-1452616, CCF-1729369, PHY-1818914, the NSF QLCI program through Grant Number OMA-2016245 and ARO contract W911NF-17-1-0433. SM was funded by NSF Grant CCF-1729369.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

A. Proof of Theorem 3

In this section we prove Theorem 3. The proof is directly inspired by the work of Bremner, Montanaro and Shepherd (see Theorem 6 and 7 of [16]). A similar theorem was also proved in [32].

Definition 97. Let μ be a $\frac{1}{\text{poly}(n)}$-approximate 2-design over the n-qubit unitary group. $\mathcal{C}_x$ is the family of unitaries constructed by first applying a circuit $C \sim \mu$ and then sampling an n-bit string x uniformly at random, and then applying an X gate to qubit j whenever $x_j = 1$.

Proof of Theorem 3. Let C be a random quantum circuit $\sim \mu$, and define $p_x = |\langle x|C|0\rangle|^2$. Denote this output distribution with p_C. Suppose there exists a BPP algorithm that samples from a distribution q_x that is within total variation distance ϵ of p_x. Therefore

$$\frac{1}{2}\sum_x |p_x - q_x| \leq \epsilon \tag{420}$$

Stockmeyer showed that given a BPP machine, there exists an FBPP^{NP} algorithm that computes its output probabilities within (inverse polynomial) $\frac{1}{\text{poly}(n)}$ multiplicative error. As a result, there is an FBPP^{NP} algorithm that for each string x computes a number $\hat{q}_x$ that satisfies

$$|q_x - \hat{q}_x| = q_x \cdot \frac{1}{\text{poly}(n)}. \tag{421}$$

Therefore using triangle inequality

$$\sum_x |p_x - \hat{q}_x| \leq \sum_x |p_x - q_x| + \sum_x |q_x - \hat{q}_x| \leq 2\epsilon + 1/\operatorname{poly}(n). \tag{422}$$

Let $0 < \delta < 1$. Using Markov's inequality, for at least $1 - \delta$ fraction of n-bit strings (such as y),

$$|p_y - \hat{q}_y| \leq \frac{2\epsilon + 1/\operatorname{poly}(n)}{2^n \delta}. \tag{423}$$

Using the definition of $\mathcal{C}_x$ with probability at least $1-\delta$ over a circuit C' from the family $\mathcal{C}_x$, $p'_0 = |\langle 0|C'|0\rangle|^2$ satisfies (423). Furthermore, we will show below that given a $\frac{1}{\operatorname{poly}(n)}$-approximate 2-design μ, for any output string $y \in \{0,1\}^n$, there exist a constant fraction $\geq 1/8 - \frac{1}{\operatorname{poly}(n)}$ of unitaries $C \sim \mu$, such that $p_y \geq 1/2^{n+1}$. Therefore w.p. at least $1-\delta$ the $\mathsf{FBPP}^{\mathsf{NP}}$ algorithm computes $\hat{q'}_0$ that satisfies

$$|\hat{q'}_0 - p'_0| \leq \frac{2\epsilon + \frac{1}{\operatorname{poly}(n)}}{2^n \delta} \leq \frac{2(\epsilon + \frac{1}{\operatorname{poly}(n)})}{\delta} p'_0. \tag{424}$$

for $1/8 - \frac{1}{\operatorname{poly}(n)}$ fraction of random unitaries C' from the ensemble $\mathcal{C}_x$. In the last line, we have used (427) (which we are going to prove next).
Now we show that for any output string $y \in \{0,1\}^n$, there exist a constant fraction $\geq 1/8 - \frac{1}{\operatorname{poly}(n)}$ of unitaries $C \sim \mu$, such that $p_y \geq 1/2^{n+1}$. To see this first recall the following known moments of the Haar measure

$$\mathop{\mathbb{E}}_{C\sim \text{Haar}} |\langle x|C|0\rangle|^2 = \frac{1}{2^n}, \qquad \mathop{\mathbb{E}}_{C\sim \text{Haar}} |\langle x|C|0\rangle|^4 = \frac{2}{2^n(2^n+1)}. \tag{425}$$

Since μ is a $\frac{1}{\operatorname{poly}(n)}$-approximate 2-design

$$\mathop{\mathbb{E}}_{C\sim \mu} |\langle x|C|0\rangle|^2 = \frac{1 + \frac{1}{\operatorname{poly}(n)}}{2^n}, \qquad \mathop{\mathbb{E}}_{C\sim \mu} |\langle x|C|0\rangle|^4 = \frac{2}{2^n(2^n+1)}\left(1 + \frac{1}{\operatorname{poly}(n)}\right). \tag{426}$$

Using the Paley–Zygmund inequality and the moments of a 2-design above

$$\begin{aligned}\Pr_{C\sim\mu}\left[|\langle x|C|0\rangle|^2 \geq \frac{1}{2^{n+1}}\right] &\geq 1/4 \frac{\left(\mathop{\mathbb{E}}_{C\sim\mu} |\langle x|C|0\rangle|^2\right)^2}{(\mathop{\mathbb{E}}_{C\sim\mu} |\langle x|C|0\rangle|^4)} = 1/4 \frac{\frac{1+\frac{1}{\operatorname{poly}(n)}}{4^n}}{\frac{2\left(1+\frac{1}{\operatorname{poly}(n)}\right)}{2^n\cdot(2^n+1)}} \\ &= 1/8 - \frac{1}{\operatorname{poly}(n)}.\end{aligned} \tag{427}$$

□

B. Basic Properties of the Krawtchouk Polynomials

Theorem. (Restatement of Lemma 86) *The Krawtchouk polynomials obey the following symmetry relation.*

$$\frac{\binom{n-1}{x}}{3^t} K^{(t)}(x) = \frac{\binom{n-1}{t}}{3^x} K^{(x)}(t). \tag{428}$$

Proof. This is implied by the observation that for all $i \in [t]$

$$\binom{n-1}{x}\binom{x}{i}\binom{n-x-1}{t-i} = \frac{(n-1)!}{(x-i)!(t-i)!i!(n-x-t+i-1)!} \tag{429}$$

is symmetric in x and t. As a result

$$\frac{\binom{n-1}{x}}{3^t} K^{(t)}(x) = \sum_{i=0}^{t} \binom{n-1}{x}\binom{x}{i}\binom{n-x-1}{t-i} 3^{-i}(-1)^i. \tag{430}$$

is also symmetric in x and t. □

The second lemma we use here is the orthogonality of the Krawtchouk polynomials

Lemma. (Restatement of Lemma 85) *If we define*

$$k^{(t)}(x) := \sum_{i=0}^{t} \binom{x}{i}\binom{N-x}{t-i} p^{t-i}(-q)^i, \tag{431}$$

for $p, q \in [0,1]$ and $p+q = 1$. Then these Krawtchouk polynomials satisfy the following orthogonality relationship

$$\sum_{x=0}^{n} \binom{N}{x} p^x q^{N-x} k^{(t)}(x) k^{(s)}(x) = \binom{N}{t} (pq)^t \delta_{t,s}. \tag{432}$$

Proof. Consider the generating function

$$\begin{aligned} g_{p,x}(z) &= (1+pz)^{N-x}(1-qz)^x \\ &= \sum_{i=0}^{n-x} \binom{N-x}{i} p^i z^i \sum_{j=0}^{x} \binom{x}{i} (-q)^j z^j \\ &= \sum_{t=0}^{N} z^t \sum_{i=0}^{t} \binom{N-x}{t-i}\binom{x}{i} p^{t-i}(-q)^i \\ &= \sum_{t=0}^{N} z^t k^{(t)}(x). \end{aligned} \tag{433}$$

Define the binomial norm $(\cdot,\cdot) : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$, where $\mathcal{F}$ is the set of functions : $[N] \to \mathbb{R}$.

$$f, g \mapsto (f,g) := \mathop{\mathbb{E}}_{X \sim \mathrm{Bin}(N,P)} [f(X)g(X)] = \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x} f(x) g(x). \tag{434}$$

Now for all real values y and z consider the overlap $(g_p(y), g_p(z))$. On the one hand

$$\begin{aligned}(g_p(y), g_p(z)) &= \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x} g_{p,x}(y) g_{p,x}(z), \\ &= \sum_{t,s=0}^{N} z^{t+s} \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x} k^{(t)}(x) k^{(s)}(x), \\ &= \sum_{t,s=0}^{N} z^{t+s} (k^{(t)}, k^{(s)}). \end{aligned} \tag{435}$$

On the other hand

$$\begin{aligned}(g_p(y), g_p(z)) &= \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x} g_{p,x}(y) g_{p,x}(z), \\ &= \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x} (1+py)^{N-x} (1-qy)^x (1+pz)^{N-x} (1-qz)^x, \\ &= \sum_{x=0}^{N} \binom{N}{x} (q(1+pz)(1+py))^{N-x} (p(1-qy)(1-qz))^x, \\ &= (q(1+pz)(1+py) + p(1-qy)(1-qz))^N, \\ &= (1+qpyz)^N, \\ &= \sum_{t=0}^{N} \binom{N}{t} (pq)^t y^t z^t. \end{aligned} \tag{436}$$

Equating these two for all y and z we obtain

$$(k^{(t)}, k^{(s)}) = \sum_{x=0}^{n} \binom{N}{x} p^x q^{N-x} k^{(t)}(x) k^{(s)}(x) = \delta_{t,s} \binom{N}{t} (pq)^t. \tag{437}$$

□

References

1. Aaronson, S., Arkhipov, A.: The computational complexity of linear optics. In: Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing, pp. 333–342. ACM (2011)
2. Aaronson, S., Bouland, A., Kuperberg, G., Mehraban, S.: The computational complexity of ball permutations. In: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, pp. 317–327 (2017)
3. Aaronson, S., Chen, L.: Complexity-theoretic foundations of quantum supremacy experiments. arXiv:1612.05903 (2016)
4. Ambainis, A., Smith, A.: Small pseudo-random families of matrices: Derandomizing approximate quantum encryption. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pp. 249–260. Springer (2004)
5. Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J.C., Barends, R., Biswas, R., Boixo, S., Brandao, F.G., Buell, D.A., et al.: Quantum supremacy using a programmable superconducting processor. Nature **574**(7779), 505–510 (2019)

6. Barends, R., Kelly, J., Megrant, A., Veitia, A., Sank, D., Jeffrey, E., White, T., Mutus, J., Fowler, A., Campbell, B., Chiaro, B., Dunsworth, A., Neill, C., O'Malley, P., Roushan, P., Vainsencher, A., Wenner, J., Korotkov, A.N., Cleland, A.N., Martinis, J.M.: Superconducting quantum circuits at the surface code threshold for fault tolerance. Nature **508**(7497), 500–503 (2014). arXiv:1402.4848
7. Ben-Hamou, A., Peres, Y.: Cutoff for a stratified random walk on the hypercube. Electron. Commun. Probab. **23**, 10 arXiv:1705.06153 (2018)
8. Boixo, S., Isakov, S.V., Smelyanskiy, V.N., Babbush, R., Ding, N., Jiang, Z., Bremner, M.J., Martinis, J.M., Neven, H.: Characterizing quantum supremacy in near-term devices. Nat. Phys. **14**(6), 595–600 (2018)
9. Boixo, S., Smelyanskiy, V.N., Neven, H.: Fourier analysis of sampling from noisy chaotic quantum circuits. arXiv:1708.01875 (2017)
10. Bouland, A., Fefferman, B., Nirkhe, C., Vazirani, U.: On the complexity and verification of quantum random circuit sampling. Nat. Phys. **15**(2), 159–163 (2019)
11. Bourgain, J., Gamburd, A.: A spectral gap theorem in SU(d). J. Eur. Math. Soc. **14**(5), 1455–1511 (2012). arXiv:1108.6264
12. Brandão, F.G., Chemissany, W., Hunter-Jones, N., Kueng, R., Preskill, J.: Models of quantum complexity growth. arXiv preprint arXiv:1912.04297 (2019)
13. Brandão, F.G.S.L., Harrow, A.W., Horodecki, M.: Local random quantum circuits are approximate polynomial-designs. Commun. Math. Phys. **346**(2), 397–434 (2016). arXiv:1208.0692
14. Bravyi, S., Gosset, D., König, R.: Quantum advantage with shallow circuits. Science **362**(6412), 308–311 (2018)
15. Bremner, M.J., Jozsa, R., Shepherd, D.J.: Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, p. rspa20100301. The Royal Society (2010)
16. Bremner, M.J., Montanaro, A., Shepherd, D.J.: Average-case complexity versus approximate simulation of commuting quantum computations. Phys. Rev. Lett. **117**(8), 080501 (2016). arXiv:1610.01808
17. Brown, W., Fawzi, O.: Scrambling speed of random quantum circuits. arXiv:1210.6644 (2012)
18. Brown, W., Fawzi, O.: Decoupling with random quantum circuits. Commun. Math. Phys. **340**(3), 867–900 (2015)
19. Brown, W.G., Viola, L.: Convergence rates for arbitrary statistical moments of random quantum circuits. Phys. Rev. Lett. **104**, 250501 (2010). arXiv:0910.0913
20. Cerezo, M., Sone, A., Volkoff, T., Cincio, L., Coles, P.J.: Cost-function-dependent barren plateaus in shallow quantum neural networks. arXiv preprint arXiv:2001.00550 (2020)
21. Chao, R., Reichardt, B.W.: Fault-tolerant quantum computation with few qubits. NPJ Quantum Inf. **4**(1), 1–8 (2018)
22. Cleve, R., Leung, D., Liu, L., Wang, C.: Near-linear constructions of exact unitary 2-designs. Quantum Inf. Comput. **16**(9&10), 0721–0756 (2016). arXiv:1501.04592
23. Collins, B., Śniady, P.: Integration with respect to the haar measure on unitary, orthogonal and symplectic group. Commun. Math. Phys. **264**(3), 773–795 (2006)
24. Dalzell, A.M., Hunter-Jones, N., Brandão, F.G.: Random quantum circuits anti-concentrate in log depth. arXiv preprint arXiv:2011.12277 (2020)
25. Dankert, C., Cleve, R., Emerson, J., Livine, E.: Exact and approximate unitary 2-designs and their application to fidelity estimation. Phys. Rev. A **80**, 012304 (2009). arXiv:quant-ph/0606161
26. Debnath, S., Linke, N., Figgatt, C., Landsman, K., Wright, K., Monroe, C.: Demonstration of a small programmable quantum computer with atomic qubits. Nature **536**(7614), 63–66 (2016). arXiv:1603.04512
27. Diniz, I., Jonathan, D.: Comment on "random quantum circuits are approximate 2-designs". cmp, **304**, 281–293 (2011). arXiv:1006.4202
28. Eden, M.: A two-dimensional growth process. Dyn. Fractal Surf. **4**, 223–239 (1961)
29. Farhi, E., Harrow, A.W.: Quantum supremacy through the quantum approximate optimization algorithm. arXiv:1602.07674 (2016)
30. Goodman, R., Wallach, N.: Representations and Invariants of the Classical Groups. Cambridge University Press, Cambridge (1998)
31. Haferkamp, J.: Random quantum circuits are approximate unitary t-designs in depth $o(nt^{5+o(1)})$. arXiv preprint arXiv:2203.16571 (2022)
32. Hangleiter, D., Bermejo-Vega, J., Schwarz, M., Eisert, J.: Anticoncentration theorems for schemes showing a quantum speedup. Quantum **2**, 65 (2018)
33. Harrow, A.W.: The church of the symmetric subspace. arXiv:1308.6595 (2013)
34. Harrow, A.W., Low, R.: Random quantum circuits are approximate 2-designs. Commun. Math. Phys. **291**, 257–302 (2009). arXiv:0802.1919
35. Harrow, A.W., Montanaro, A.: Quantum computational supremacy. Nature **549**(7671), 203–209 (2017)
36. Hastings, M.B., Harrow, A.W.: Classical and quantum tensor product expanders. Q. Inf. Comput. **9**(3&4), 336–360 (2009). arXiv:0804.0011

37. Hunter-Jones, N.: Unitary designs from statistical mechanics in random quantum circuits. arXiv preprint arXiv:1905.12053 (2019)
38. Janson, S.: Tail bounds for sums of geometric and exponential variables (2014)
39. Kac, M.: Random walk and the theory of Brownian motion. Am. Math. Mon. **54**(7), 369–391 (1947)
40. Kitaev, A.Y., Shen, A.H., Vyalyi, M.N.: Classical and Quantum Computation. Graduate Studies in Mathematics, vol. 47. AMS (2002)
41. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. American Mathematical Society, Providence (2009)
42. Low, R.A.: Pseudo-randomness and Learning in Quantum Computation. PhD thesis, University of Bristol. arXiv:1006.5227 (2010)
43. McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., Neven, H.: Barren plateaus in quantum neural network training landscapes. Nat. Commun. **9**(1), 1–6 (2018)
44. Movassagh, R.: Efficient unitary paths and quantum computational supremacy: a proof of average-case hardness of random circuit sampling. arXiv preprint arXiv:1810.04681 (2018)
45. Movassagh, R.: Cayley path and quantum computational supremacy: a proof of average-case #∂ hardness of random circuit sampling with quantified robustness. arXiv preprint arXiv:1909.06210 (2019)
46. Nahum, A., Ruhman, J., Vijay, S., Haah, J.: Quantum entanglement growth under random unitary dynamics. Phys. Rev. X **7**(3), 031016 (2017)
47. Nahum, A., Ruhman, J., Vijay, S., Haah, J.: Simple heuristics for quantum entanglement growth. Bull. Am. Phys. Soc. **62**, 031016D (2017)
48. Nakata, Y., Hirche, C., Morgan, C., Winter, A.: Unitary 2-designs from random X- and Z-diagonal unitaries. J. Math. Phys. **58**(5), 052203 (2017). arXiv:1502.07514
49. Napp, J.C., La Placa, R.L., Dalzell, A.M., Brandão, F.G.S.L., Harrow, A.W.: Efficient classical simulation of random shallow 2D quantum circuits. Phys. Rev. X **12**, 021021 (2022). arXiv:2001.00021
50. Ofek, N., Petrenko, A., Heeres, R., Reinhold, P., Leghtas, Z., Vlastakis, B., Liu, Y., Frunzio, L., Girvin, S., Jiang, L., et al.: Extending the lifetime of a quantum bit with error correction in superconducting circuits. Nature **536**(7617), 441–445 (2016). arXiv:1602.04768
51. Oliveira, R., Dahlsten, O.C., Plenio, M.B.: Efficient generation of generic entanglement. Phys. Rev. Lett. 98. arXiv:quant-ph/0605126 (2007)
52. Preskill, J.: Quantum computing and the entanglement frontier. arXiv:1203.5813 (2012)
53. See the footnote in https://www.boazbarak.org/sos/prev/files/hw0.pdf
54. Sen, P.: Random measurement bases, quantum state distinction and applications to the hidden subgroup problem. In: Twenty-First Annual IEEE Conference on Computational Complexity, 2006. CCC 2006, pp. 14–pp. IEEE (2005)
55. Susskind, L.: Computational complexity and black hole horizons. Fortschr. Phys. **64**(1), 24–43 (2016)
56. Terhal, B.M., DiVincenzo, D.P.: Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games. arXiv:quant-ph/0205133 (2012)
57. Yung, M.-H., Gao, X.: Can chaotic quantum circuits maintain quantum supremacy under noise? arXiv:1706.08913 (2017)
58. Žnidarič, M.: Exact convergence times for generation of random bipartite entanglement. Phys. Rev. A **78**(3), 032324 (2008)

Communicated by M. Christandl