Efficient classical simulation of random shallow 2D quantum circuits: Supplemental Material

John C. Napp,[1] Rolando L. La Placa,[1] Alexander M. Dalzell,[2] Fernando G. S. L. Brandão,[2,3] and Aram W. Harrow[1]

[1] *Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2] *Institute for Quantum Information and Matter, Caltech, Pasadena, California 91125, USA.*
[3] *Amazon Web Services, Pasadena, California, USA.*

(Dated: January 15, 2022)

This document contains appendices of the main text. In Appendix A, we present (and justify) a generalized formalism of the stat mech mapping that allows for weak measurements to be inserted in between the Haar-random unitary gates; then we apply this formalism to a model of 1D random quantum circuits with weak measurements inspired by the `CHR` problem introduced in the main text. In Appendix B we propose a second simulation algorithm (`Patching`) for random shallow 2D quantum circuits which is qualitatively quite different from `SEBD`. In Appendix C we utilize the stat mech mapping to establish evidence that `Patching` is efficient exactly when `SEBD` is efficient. In Appendix D, we discuss implications of `SEBD` and `Patching` for the worst-to-average-case reduction for random circuits of [1] based on truncated Taylor series. Appendix E contains deferred proofs.

Appendix A: General description and justification of the stat mech mapping

In the main text, we described how one maps a random quantum circuit to a classical stat mech model. Our description assumed that the circuit consists of Haar-random unitary gates and that at the end of the circuit, some subset of the qudits experience projective measurements. We only described how to perform the mapping and not the background that is used to justify it. Here, we present (and justify) a generalized formalism that allows for weak measurements to be inserted in between the Haar-random unitary gates; then we apply this formalism to a model of 1D random quantum circuits with weak measurements inspired by the `CHR` problem introduced in the main text. The problem of 1D circuits interspersed with weak measurements was previously studied using this approach in [2, 3] (however, we analyze a different weak measurement).

1. Generalized mapping procedure

a. Setup. Let our system consist of n qudits of local dimension q. The circuits we consider are specified by a sequence of pairs of qudits (indicating where unitary gates are applied) and single-qudit weak measurements; this sequence can be assembled into a quantum circuit diagram. The single-qudit measurements are each described by a set $\mathcal{M}$ of measurement operators along with a probability distribution μ over the set $\mathcal{M}$. These sets are normalized such that $\mathrm{tr}\big(M^\dagger M\big)$ is constant for all $M \in \mathcal{M}$ and $\mathbb{E}_{M\leftarrow\mu} M^\dagger M = \mathbb{I}_q$ where $\mathbb{I}_q$ is the $q \times q$ identity matrix. Thus we have $\mathrm{tr}\big(M^\dagger M\big) = q$ for all M. The introduction of a probability measure over $\mathcal{M}$ in our notation, which was also used in [3], is not conventional, but it is equivalent to the standard formulation and will be important for later definitions.

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

Thus, the (non-normalized) output state of the circuit with l unitaries acting on the initial state $|0\ldots0\rangle$ can be expressed as

$$\rho = E\,|0\ldots0\rangle\langle 0\ldots0|\,E^\dagger \tag{A1}$$

with

$$E = (M'_l M_l U_l)\ldots(M'_2 M_2 U_2)(M'_1 M_1 U_1) \tag{A2}$$

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

b. Goal. As discussed in the main text, the objective of the stat mech mapping is to learn something about the entanglement entropy for the output state ρ on some subset A of the qudits. The k-Rényi entanglement entropy for the state ρ on region A is defined as

$$S_k(A)_\rho = \frac{1}{1-k}\log\left(\frac{Z_{k,A}}{Z_{k,\emptyset}}\right) \tag{A3}$$

where

$$Z_{k,\emptyset} = \mathrm{tr}(\rho)^k \tag{A4}$$
$$Z_{k,A} = \mathrm{tr}\big(\rho_A^k\big). \tag{A5}$$

and ρ_A is the reduced density matrix of ρ on region A. The von Neumann entropy $S(A)_\rho = -\mathrm{tr}\left(\frac{\rho_A}{\mathrm{tr}(\rho)}\log\left(\frac{\rho_A}{\mathrm{tr}(\rho)}\right)\right)$ represents the $k\to 1$ limit.

For the purposes of assessing the efficiency of our algorithms, we would like to be able to calculate the average value of the k-Rényi entropy over the random choice of the unitaries in the circuit and for measurement outcomes drawn randomly according to the Born rule (for a rigorous proof to follow, we would need to examine $0<k<1$). Mathematically, we let the notation $\mathbb{E}_U(Q)$ represent the expectation of a quantity Q when the unitaries of the circuit are drawn uniformly at random from the Haar measure and the measurement outcomes are drawn at random from the distribution over their respective sets of measurement operators

$$\mathop{\mathbb{E}}_U(Q) = \mathop{\mathbb{E}}_{M_1\leftarrow\mu_1}\mathop{\mathbb{E}}_{M_1'\leftarrow\mu_1'}\cdots\mathop{\mathbb{E}}_{M_l\leftarrow\mu_1}\mathop{\mathbb{E}}_{M_l'\leftarrow\mu_1'}\int_{U(q^2)}dU_1\ldots\int_{U(q^2)}dU_l\; Q \tag{A6}$$

Here $\int_{U(q^2)}$ denotes integration over the Haar measure of the unitary group with dimension q^2. To take into account the Born rule, a certain choice of unitaries and measurement outcomes leading to the output state ρ, as in Eq. (A1), should be weighted by $\mathrm{tr}(\rho)$, i.e. the product of the relative likelihoods of all the measurement outcomes. Thus, the relevant average k-Rényi entropy values are given by

$$\langle S_k(A)_\rho\rangle := \frac{\mathbb{E}_U(\mathrm{tr}(\rho)S_k(A)_\rho)}{\mathbb{E}_U(\mathrm{tr}(\rho))} \tag{A7}$$
$$= \frac{1}{1-k}\frac{\mathbb{E}_U\left(\mathrm{tr}(\rho)\log\frac{Z_{k,A}}{Z_{k,\emptyset}}\right)}{\mathbb{E}_U(\mathrm{tr}(\rho))}. \tag{A8}$$

However, the quantity naturally computed by the stat mech model is not $\langle S_k(A)_\rho\rangle$, but rather the "quasi-entropy" given by

$$\tilde{S}_k(A) := \frac{1}{1-k}\log\left(\frac{\mathbb{E}_U(\mathrm{tr}(\rho)^k\frac{Z_{k,A}}{Z_{k,\emptyset}})}{\mathbb{E}_U(\mathrm{tr}(\rho)^k)}\right) \tag{A9}$$
$$= \frac{1}{1-k}\log\left(\frac{\mathbb{E}_U(Z_{k,A})}{\mathbb{E}_U(Z_{k,\emptyset})}\right) \tag{A10}$$
$$= \frac{F_{k,\emptyset}-F_{k,A}}{1-k} \tag{A11}$$

where $F_{k,\emptyset/A} := -\log\big(\mathbb{E}_U(Z_{k,\emptyset/A})\big)$. When clear, we abbreviate $\tilde{S}_k(A)$ by $\tilde{S}_k$ and $\langle S_k(A)_\rho\rangle$ by $\langle S_k\rangle$. It is apparent that $\langle S_k\rangle$ is not equal to the quasi-entropy $\tilde{S}_k$: the definition of $\langle S_k\rangle$ weights circuit instances by $\mathrm{tr}(\rho)$ and takes the log before the expectation, while the definition of $\tilde{S}_k$ weights by $\mathrm{tr}(\rho)^k$ and takes the log afterward. Indeed, it is possible for $\tilde{S}_k$ to be smaller than some constant independent of L (area law), while $\langle S_k\rangle$ scales extensively with L (volume law) due to fluctuations of the random variable $Z_{k,A}$ away from its average value toward 0.

Importantly, as noted in the main text, $\tilde{S}_k\to\langle S\rangle$ as $k\to 1$, where $\langle S\rangle$ is the expected von Neumann entropy. This conclusion is justified by L'Hospital's rule and noting that $Z_{k,A}/Z_{k,\emptyset}\to 1$ as $k\to 1$. We will see that the stat mech mapping can only be applied for integers $k\geq 2$, so unfortunately it does not allow for direct access to formula $\tilde{S}_k$ in this limit. As discussed in the main text, even though computation of this limit is out of reach, we still take $\tilde{S}_k$ to be an informative proxy for the entanglement properties of the system, which determines the efficiency of our algorithms.

c. Generalized interaction weights. The method of forming an interaction graph from the random quantum circuit diagram is discussed in the main text and captured by the example in Figure 6. The quantities $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$ are then given by a partition function on this graph, as in Eq. (25). In the main text, we gave equations for the edge weights for edges $\langle s_u t_u\rangle$, $\langle s_{u_1}t_{u_2}\rangle$, and $\langle s_u x_a\rangle$ in Eq. (26), Eq. (27), and Eq. (28). Here we first generalize these to the case where we allow weak measurements, and justify the formulas in the next subsection. The former equation, which gives the interaction weight between the incoming node t_u and the outgoing node s_u for the same unitary u, is unchanged, and still reads

$$\mathrm{weight}(\langle s_u t_u\rangle) = \mathrm{wg}(\tau_u\sigma_u^{-1}, q^2) \tag{A12}$$

where $\mathrm{wg}(\pi, q^2)$ is the Weingarten function. The Weingarten function arises from performing the integrals over the Haar measure in Eq. (A6), and one formula for it is given in the next subsection in Eq. (A19). Note that there exist permutations π for which $\mathrm{wg}(\pi, q^2) < 0$, so the overall weight of a configuration can be negative and our stat mech model would only correspond to a physical model with complex-valued energy.

Meanwhile, the weight of interactions between the outgoing node for one unitary u_1 and the incoming node for a successive unitary u_2 (or between a unitary u and an auxiliary node a) must be updated to account for the case that a weak measurement occurs in between the unitaries. This weight, generalizing Eq. (27) and Eq. (28), is given by

$$\mathrm{weight}(\langle s_{u_1} t_{u_2} \rangle) = \mathop{\mathbb{E}}_{M \leftarrow \mu} \mathrm{tr}\Big((M^\dagger M)^{\otimes k} W_{\sigma_{u_1} \tau_{u_2}^{-1}} \Big) \tag{A13}$$

$$\mathrm{weight}(\langle s_u x_a \rangle) = \mathop{\mathbb{E}}_{M \leftarrow \mu} \mathrm{tr}\Big((M^\dagger M)^{\otimes k} W_{\sigma_u \chi_a^{-1}} \Big) \tag{A14}$$

where W_π is the operator acting on a k-fold tensor product space that performs the permutation π of the registers. Later, in Appendix C, we will be interested in expressing entropies of the *classical* output distribution of the circuit in terms of partition functions and to handle this case we will update Eq. (A14). Note that the quantity $\mathrm{tr}\big(X^{\otimes k} W_\pi\big)$ is equal for all π with the same cycle structure, which corresponds to some partition $\lambda = (\lambda_1, \ldots, \lambda_r)$ of k, where $\sum_i \lambda_i = k$ and $\lambda_1 \geq \ldots \geq \lambda_r > 0$. Then we have

$$\mathrm{tr}\big(X^{\otimes k} W_\pi\big) = \prod_{i=1}^{r} \mathrm{tr}\big(X^{\lambda_i}\big) \tag{A15}$$

This formula allows us to simplify the weight formulas (A13) and (A14) in a few special cases. If no measurement is made, then $\mathcal{M} = \{I\}$ and $\mathrm{weight}(\langle s_{u_1} t_{u_2} \rangle) = q^{C(\sigma_{u_1} \tau_{u_2}^{-1})}$, where $C(\pi)$ is the number of cycles r in the permutation π, recovering the weight equations (27) and (28) from the main text. On the other hand, if a projective measurement onto one of the q basis states is made, then $\mathcal{M} = \{\sqrt{q}\Pi_m\}_{m=0}^{q-1}$ and μ is the uniform distribution, where $\Pi_m = |m\rangle\langle m|$. Since in this case $\mathrm{tr}\big((M^\dagger M)^w\big) = q^w$ for any power w and any $M \in \mathcal{M}$, we have $\mathrm{weight}(\langle s_{u_1} t_{u_2} \rangle) = q^{k-1}$ for any pair σ_{u_1}, τ_{u_2}.

2. Justification of stat mech mapping

In this subsection, our goal is to provide a justification for (1) the mapping procedure that allows $\mathbb{E}_U(Z_{k,\emptyset})$ and $\mathbb{E}_U(Z_{k,A})$ to be expressed as a partition function as in Eq. (25) of the main text and (2) the formulas for the interaction weights for this partition function given in the previous subsection in Eqs. (A12), (A13), and (A14).

To begin, for any integer $k \geq 2$, we rewrite $Z_{k,\emptyset/A}$ from Eqs. (A4) and (A5) as

$$Z_{k,\emptyset} = \mathrm{tr}[(\rho \otimes \ldots \otimes \rho)] \tag{A16}$$

$$Z_{k,A} = \mathrm{tr}\Big[(\rho \otimes \ldots \otimes \rho) W^{(A)}_{(1\ldots k)}\Big] \tag{A17}$$

where each trace includes k copies of ρ and $W^{(A)}_{(1\ldots k)}$ is the linear operator that performs a k-cycle permutation, denoted $(1 \ldots k)$ in cycle notation, of the copies for qudits within region A while leaving the copies of the qudits outside of A unpermuted. When $k = 2$ there are two copies of ρ and $W^{(A)}_{(12)}$ is the swap operator for qudits in A.

After substituting Eq. (A1) for each copy of ρ that appears in the equations above, we obtain an expression with k copies of each unitary U_u and k copies of its adjoint $U_u^\dagger$, as well as k copies each of M_u, $M_u^\dagger$, M'_u, and ${M'_u}^\dagger$. Taking the expectation $\mathbb{E}_U(Z_{k,\emptyset/A})$ introduces integrals over U_u and expectations over M_u and M'_u drawn from distributions μ_u and μ'_u, for each u. To perform the integrals, we rely on techniques for integration over the Haar measure, invoking the formula [4, 5]

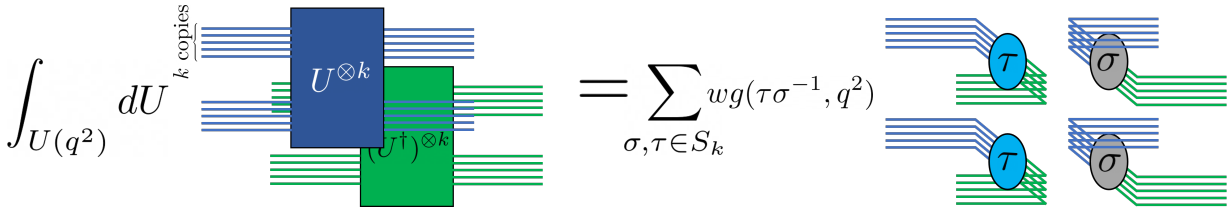
$$\begin{aligned} &\int_{U(q^2)} dU\; U_{i_1 j_1} \ldots U_{i_k j_k} U^\dagger_{i'_1 j'_1} \ldots U^\dagger_{i'_k j'_k} \\ &= \sum_{\sigma,\tau \in S_k} \delta_\sigma(\vec{i}, \vec{j}\,') \delta_\tau(\vec{i}\,', \vec{j})\, \mathrm{wg}(\tau\sigma^{-1}, q^2) \end{aligned} \tag{A18}$$

where on the left hand side U_{ij} is the (i,j) matrix element of the unitary U and on the right hand side S_k is the symmetric group, $\delta_\sigma(\vec{i}, \vec{j}\,')$ is shorthand for $\prod_{a=1}^k \delta_{i_a j'_{\sigma(a)}}$ and $\mathrm{wg}(\tau\sigma^{-1}, q^2)$ is the Weingarten function, which can be defined in several ways, for example by the following expansion [4, 5] over irreducible characters of the symmetric group S_k

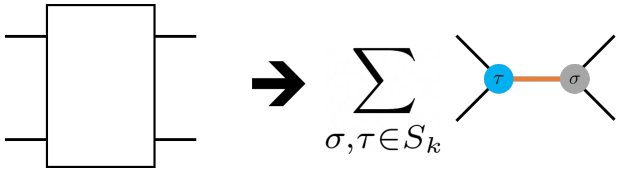
$$\mathrm{wg}(\pi, q^2) = \frac{1}{(q^2)!^2} \sum_\lambda \frac{\chi^\lambda(e)^2}{s_{\lambda,q^2}(1)} \chi^\lambda(\pi) \tag{A19}$$

where the sum is over all partitions λ of the integer k, χ^λ is the irreducible character of S_k associated with the partition λ, e is the identity permutation, and $s_{\lambda,q^2}(1)$ is the Schur polynomial evaluated at 1 which is equal to the dimension of the representation of $U(q^2)$ associated with λ. Note that there exist permutations π for which $\mathrm{wg}(\pi, q^2)$ is negative.

In words, formula (A18) states that Haar integration can be performed by summing over all ways of pairing up the incoming index for each of the k copies of U with an outgoing index of a copy of $U^\dagger$, and the incoming index of each copy of $U^\dagger$ with an outgoing index of a copy of U. The permutations $\sigma, \tau \in S_k$ encode which copies are paired with each other, and each permutation pair (σ, τ) is weighted by $\mathrm{wg}(\tau\sigma^{-1}, q^2)$ in the sum. It is helpful to think of this formula graphically, as in Figure A.1, where we have depicted how the indices pair up after integration over a two-qudit Haar-random unitary.



(a) Haar integration formula applied to k copies of two-qudit gate



(b) Mapping of unitary in circuit diagram

FIG. A.1. (a) Graphical depiction of Haar integration formula given in Eq. (A18). (b) Haar integration formula allows us to replace Haar-random unitaries from circuit diagram with sums over configurations on a graph with nodes taking values in S_k, and edges between graphs contributing a factor to the weight of a configuration.

By applying this formula to all of the Haar-random gates, all of the integrals are eliminated and the tensor network representation of $Z_{k,\emptyset/A}$ can be expressed as a weighted sum over many networks, where each network in this sum corresponds to some choice of (σ_u, τ_u) for every unitary u in the original circuit and some choice of M_u and M'_u from $\mathcal{M}_u$ and $\mathcal{M}'_u$. Furthermore, each network in this weighted sum is itself composed of many disjoint parts that can be individually evaluated. We can see this by observing that when unitary u_2 succeeds unitary u_1 and shares a qudit, Haar integration forces the k tensor indices representing that qudit at that place in the circuit diagram to pair up with the k dual indices for the qudit at the same place, according to some permutation. This happens both at the output of unitary u_1 (corresponding to permutation σ_{u_1}) and at the input of unitary u_2 (corresponding to permutation τ_{u_2}) yielding a set of closed loops in the tensor network diagram. If the weak measurement acting on that qudit between unitaries u_1 and u_2 is M, then k copies of M and k copies of $M^\dagger$ appear among these loops. An example of such a subdiagram is shown in Figure Figure A.2.

This observation justifies the partition function Eq. (25), as we have expressed $\mathbb{E}_U(Z_{k,\emptyset/A})$ as a weighted sum, with each term labelled by pairs of permutations at the locations of each unitary, where the weight is given by a product of factors that depend only on two of these permutations. These factors are the weights given by Eqs. (A12), (A13), and (A14). Eq. (A12) accounts for the factor $\mathrm{wg}(\tau\sigma^{-1}, q^2)$ in Eq. (A18). Meanwhile, we can derive Eq. (A13) by performing the expectation in Figure Figure A.2. We can graphically see that each term in the expansion of this expectation is given by $\mu(M)\,\mathrm{tr}\left(W_\sigma M^{\otimes k} W_{\tau^{-1}} (M^\dagger)^{\otimes k}\right)$, and then simply note that W_π commutes with $X^{\otimes k}$ for any X.

Our final task is to justify the introduction of the auxiliary nodes and corresponding weights in Eq. (A14). The

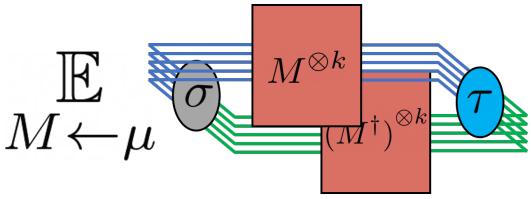


FIG. A.2. Disjoint part that forms tensor network representation of $\mathbb{E}_U(Z_{k,\emptyset/A})$ after performing integrals over Haar-random gates. The weight given by Eqs. (A13) and (A14) is derived by evaluating this diagram and taking the expectation with M drawn from $\mathcal{M}$ according to the distribution μ.

tr in the definition of $Z_{k,\emptyset}$ implies that the indices of the qudits at the circuit output are paired up with their dual indices without permutation. This creates a disjoint closed diagram for each qudit at the circuit output. To evaluate it, we may use the same formula as Eq. (A13) taking $\tau_{u_2} = e$, the identity permutation. This is equivalent to introducing auxiliary nodes, as we have done, that are fixed to e for all qudits and across all terms in the partition function. The same follows for $Z_{k,A}$ with the exception that the operator $W^{(A)}_{(1\ldots k)}$ is applied to the circuit output, which permutes the output indices of any qudit $a \in A$ prior to connecting them with their dual indices. This is equivalent to introducing an auxiliary node and fixing it to the value $(1 \ldots k)$.

There is no need to introduce auxiliary nodes at the beginning of the circuit because we are assuming the circuit acts on the pure product state $|0\ldots0\rangle\langle0\ldots0|$. Thus, the k copies of the index that feeds into the first unitary of the circuit are forced to be 0 and regardless of the permutation value of the incoming node for that unitary, this part of the circuit will contribute a factor of 1. If we had considered circuits that act initially on the maximally

mixed state, we could have handled this by introducing a layer of auxiliary nodes at the beginning of the circuit and fixing their value to e.

3. Mapping applied to 1D circuits with weak measurements

In Section III C, we discussed the connection between the effective 1D dynamics of our SEBD algorithm and previous work (originating from [6–8]) on 1D Haar-random circuits with some form of measurements in between each layer of unitaries.

In this subsection, we apply the stat mech mapping to the 1D with weak measurement model and explain the connection between the area-law-to-volume-law transition that has been observed in numerical simulations and the disorder-to-order thermal transition in the classical stat mech model, which occurs at a non-zero critical temperature T_c. This analysis was first performed in [2] and independently in [3]. The results presented in this section are essentially a reproduction of their analysis but for a different weak measurement, chosen to be relevant for the dynamics of the SEBD algorithm acting on the CHR problem. We include this analysis for two purposes: first, to shed light on the behavior of SEBD acting on CHR, and second, to serve as a more complete example of the stat mech mapping in action, complementing the more heuristic analysis we give in Section VI of the main text.

a. Mapping to the honeycomb lattice. Let us assume our circuit has n qudits of local dimension q arranged on a line with open boundary conditions. A circuit of depth d acts on the qudits where each layer consists of nearest-neighbor two-qudit Haar-random unitaries. In between each layer of unitaries, a weak measurement is performed on every qudit, described by the set $\mathcal{M}$ of measurement operators and a probability distribution μ over $\mathcal{M}$. The first step of the stat mech mapping is to replace each Haar-random unitary with a pair of nodes and connect these nodes according to the order of the unitaries acting on the qudits. The second step is to introduce a new auxiliary node for each qudit and connect each outgoing node within the final layer of unitaries to the corresponding pair of auxiliary nodes. The resulting graph is the honeycomb lattice, as shown in Figure A.3(b). We now review what the interactions are on this graph. The horizontal links in Figure A.3(b) host interactions that contribute a weight equal to the Weingarten function. When $k = 2$, the interaction depends only on if the pair of nodes agree ($\sigma_u \tau_u^{-1} = e$) or if they disagree ($\sigma_u \tau_u^{-1} = (12)$). In this case the interactions are given explicitly by

$$\text{weight}(\langle s_u t_u \rangle) = \text{wg}(\sigma_u \tau_u^{-1}, q^2) \tag{A20}$$

$$= \begin{cases} \frac{1}{q^4-1} & \text{if } \sigma_u \tau_u^{-1} = e \\ -\frac{1}{q^2(q^4-1)} & \text{if } \sigma_u \tau_u^{-1} = (12). \end{cases} \tag{A21}$$

Meanwhile, the diagonally oriented links in Figure A.3(b) host interactions that depend on the details of the weak measurement being applied in between each layer of unitaries, which we now define.

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

$$M_U^{(m)} := \sqrt{q} \cdot \text{diag}(U_{m,\cdot}) \tag{A22}$$

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

$$\sum_{m=0}^{q-1} \mu_U(m) (M_U^{(m)})^\dagger M_U^{(m)} = \sum_{m=0}^{q-1} \text{diag}(|U_{m,\cdot}|^2) = \mathbb{I}_q \tag{A23}$$

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

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

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

$$M^{(0)} := \begin{pmatrix} \cos(\theta/2) & 0 \\ 0 & e^{-i\phi} \sin(\theta/2) \end{pmatrix} \tag{A24a}$$

$$M^{(1)} := \begin{pmatrix} \sin(\theta/2) & 0 \\ 0 & e^{i\phi} \cos(\theta/2) \end{pmatrix}. \tag{A24b}$$

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

To compute the weights on the edges of the stat mech model for $k = 2$, we apply the formula in Eqs. (A13) and

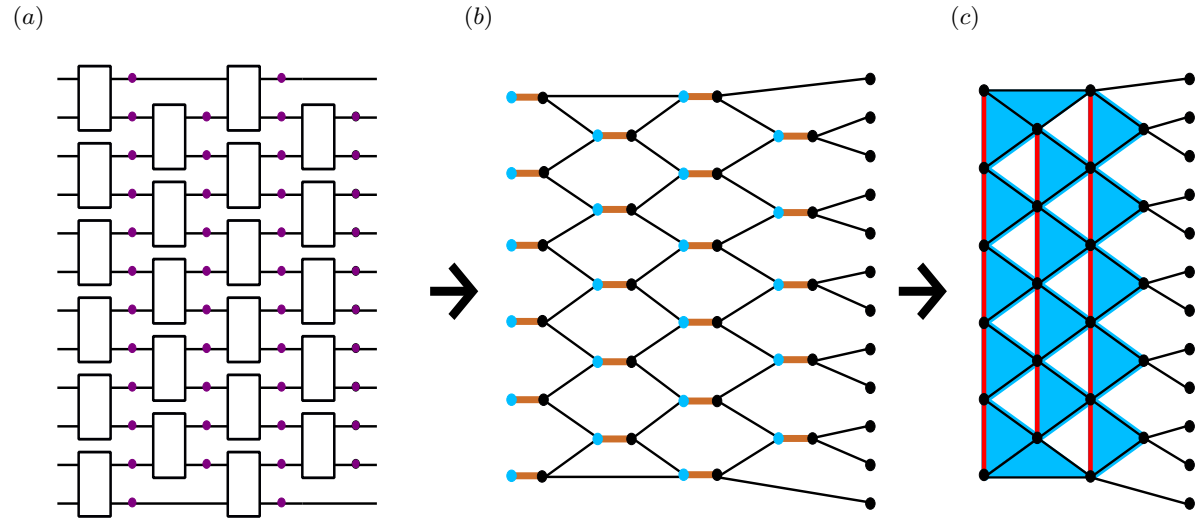


FIG. A.3. Summary of series of maps for Haar-random 1D circuits with weak measurements. (a) The quantum circuit diagram for the unitary plus weak measurement model consists of layers of Haar-random two-qudit gates followed by layers of weak measurements on every qudit, indicated by purple dots. (b) The stat mech mapping results in a model on the honeycomb lattice, where horizontal orange links have weight given by the Weingarten function and diagonal black links have weight that depends on the weak measurement. Blue dots and black dots represent incoming and outgoing nodes, respectively. (c) By decimating the incoming (blue) nodes in the honeycomb lattice, we reduce the number of nodes by half and generate a model with three-body interactions living on rightward-pointing triangles, shaded in blue. When $k = 2$ the weights are all positive, and the three-body interaction can be decomposed into an anti-ferromagnetic interaction along vertical (red) links and ferromagnetic interactions along diagonal (dark blue) links.

(A14).

$$
\begin{aligned}
&\text{weight}(\langle s_{u_1} t_{u_2}\rangle) \\
&= \int_{U(q)} dU \sum_{m=0}^{q-1} \frac{1}{q} \operatorname{tr}\left(\left((M_U^{(m)})^\dagger M_U^{(m)} \right)^{\otimes 2} W_{\sigma_{u_1}\tau_{u_2}^{-1}} \right) \\
&= \int_{U(q)} dU q \sum_{m=0}^{q-1} \begin{cases} \operatorname{tr}\left(\operatorname{diag}(|U_{m,\cdot}|^2)\right)^2 & \text{if } \sigma_{u_1}\tau_{u_2}^{-1} = e \\ \operatorname{tr}(\operatorname{diag}\left(|U_{m,\cdot}|^4\right)) & \text{if } \sigma_{u_1}\tau_{u_2}^{-1} = (12) \end{cases} \\
&= \begin{cases} q^2 & \text{if } \sigma_{u_1}\tau_{u_2}^{-1} = e \\ q^2 \cdot w & \text{if } \sigma_{u_1}\tau_{u_2}^{-1} = (12) \end{cases} \qquad \text{(A25)}
\end{aligned}
$$

where

$$
\begin{aligned}
w &:= \int_{U(q)} dU \sum_m \frac{1}{q} \operatorname{tr}\big(\operatorname{diag}(|U_{m,\cdot}|^4)\big) && \text{(A26)} \\
&= q \int_{U(q)} dU |U_{0,0}|^4 && \text{(A27)} \\
&= q \sum_{\sigma,\tau \in S_2} \operatorname{wg}(\sigma\tau^{-1}, q) && \text{(A28)} \\
&= 2q \sum_{\sigma \in S_2} \operatorname{wg}(\sigma, q) && \text{(A29)} \\
&= 2q \left(\frac{1}{q^2-1} - \frac{1}{q(q^2-1)} \right) && \text{(A30)} \\
&= \frac{2}{q+1}, && \text{(A31)}
\end{aligned}
$$

where in the second line we have invoked the Haar integration formula that appears in Eq. (A18), and then substituted the explicit values for the Weingarten function when $k = 2$. The formula for $\text{weight}(\langle s_u x_a\rangle)$ is given similarly.

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

c. Eliminating negative weights via decimation when $k = 2$. The possibility of a negative weight on the horizontal edges of the honeycomb lattice in Figure A.3(b) appears to impede further progress in the analysis since the classical model cannot be viewed as a physical system with real interaction energies at a real temperature. As discussed in the main text, for $k = 2$, this problem may be circumvented by decimating half of the spins; that is, we explicitly perform the sum over $\{\tau_u\}_u$ in the partition function in Eq. (25), yielding a new stat mech model involving only the outgoing nodes s_u. Since the decimated incoming nodes (except for those in the first layer) each have three neighbors, all three of which are undecimated outgoing nodes, the new model will have a three-body interaction between each such trio of nodes.

We may furthermore observe that, for our choice of weak measurement when $k = 2$, the three-body weight may be re-expressed as the product of three two-body weights acting on the three edges of the triangle. Below

we give formulas for the two-body weights; our formulas are a unique decomposition of the three-body interaction up to a shifting of overall constant factors from one link to another. Thus, via decimation we have moved from the honeycomb lattice with two-body interactions to the triangular lattice with two-body interactions, as illustrated in Figure A.3(c). There are two kinds of two-body interactions on this triangular lattice. Vertically oriented links between nodes s_{u_1} and s_{u_2} host anti-ferromagnetic interactions

$$\text{weight}(\langle s_{u_1} s_{u_2}\rangle) \tag{A32}$$

$$= \begin{cases} \frac{1}{q^4-1} & \text{if } \sigma_{u_1}\sigma_{u_2} = e \\ \frac{w}{1+q^2}\left((q^2-w^2)(q^2w^2-1)\right)^{-1/2} & \text{if } \sigma_{u_1}\sigma_{u_2} = (12) \end{cases} \tag{A33}$$

and diagonally oriented links host ferromagnetic interactions, where

$$\text{weight}(\langle s_{u_1} s_{u_2}\rangle) = \begin{cases} q\sqrt{q^2-w^2} & \text{if } \sigma_{u_1}\sigma_{u_2} = e \\ q\sqrt{w^2q^2-1} & \text{if } \sigma_{u_1}\sigma_{u_2} = (12). \end{cases} \tag{A34}$$

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

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

$$E/kT = -J_1 \sum_{\langle ij\rangle_1} g_i g_j - J_2 \sum_{\langle ij\rangle_2} g_i g_j - J_3 \sum_{\langle ij\rangle_3} g_i g_j \tag{A35}$$

where $g_i \in \{+1, -1\}$ are Ising spin variables and the three sums are over links along each of the three triangular axes. This model has been studied and its phase diagram is well understood [9, 10]. In the setting where along two of the axes the interaction strength is equal in magnitude and ferromagnetic, while along the third axis it is weaker in magnitude and antiferromagnetic, the model is known to experience a phase transition as the temperature is varied. At high temperatures, it is in the disordered phase; in other words, samples drawn from the thermal distribution exhibit exponentially decaying correlations between spin values σ_u with a constant correlation length of ξ. At low temperatures, it is in an ordered phase where samples exhibit long-range correlation. At the critical point, the interaction strengths satisfy the equation [9, 10]

$$\sinh(2J_1)\sinh(2J_2) + \sinh(2J_2)\sinh(2J_3) + \sinh(2J_1)\sinh(2J_3) = 1. \tag{A36}$$

For us, parameter q plays the role of the temperature, and the interaction strengths, derived from Eqs. (A32) and (A34), are given by

$$J_1 = J_2 = \frac{1}{4}\log\left(\frac{q^2-w^2}{w^2q^2-1}\right), \tag{A37}$$

$$J_3 = -\frac{1}{2}\log\left(\frac{w(q^2-1)}{\sqrt{(q^2-w^2)(q^2w^2-1)}}\right). \tag{A38}$$

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

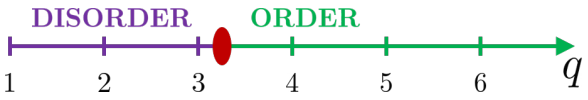


FIG. A.4. Phase diagram showing for which values of q the anisotropic Ising model on the triangular lattice is ordered and disordered. The critical point, indicated by the red dot, occurs at $q_c = 3.249$.

e. Connection between (dis)order and scaling of entanglement entropy. We expect the scaling of the quantity $\tilde{S}_2 = F_{2,A} - F_{2,\emptyset} = -\log\big(\mathbb{E}_U(Z_{2,A})/\,\mathbb{E}_U(Z_{2,\emptyset})\big)$ to be related to the order or disorder of the model by the following argument. For $\mathbb{E}_U(Z_{2,\emptyset})$, the auxiliary spins are all set to $\chi_a = e$, biasing the bulk spins nearby to prefer e over (12). For $\mathbb{E}_U(Z_{2,A})$, the spins within the region A are twisted so that $\chi_a = (12)$, introducing a domain wall at the boundary. In the ordered phase, the bias introduced at the boundary extends throughout the whole bulk since there is no decay of correlation with distance. The domain wall at the boundary in the calculation of $\mathbb{E}_U(Z_{2,A})$ forces the bulk to separate into two regions with distinct phases separated by a domain wall that cuts through the bulk. The domain wall has length of order $\min(|A|, d)$ where $|A|$ is the number of sites in region A and d is the depth. In the calculation of $\mathbb{E}_U(Z_{2,\emptyset})$, there is no domain wall. The addition of one additional unit of domain wall within a configuration leads the weight of the configuration to decrease by a constant factor, so in the ordered phase we expect $-\log\big(\mathbb{E}_U(Z_{2,A})/\,\mathbb{E}_U(Z_{2,\emptyset})\big) = O(\min(|A|, d))$. Meanwhile, in the disordered phase, there is a natural length scale ξ that boundary effects will penetrate into the bulk. The domain wall at the boundary due to twisted boundary conditions will be washed out by the bulk disorder after a distance on the order of $\xi = O(1)$. Thus we expect $-\log\big(\mathbb{E}_U(Z_{2,A})/\,\mathbb{E}_U(Z_{2,\emptyset})\big) = O(1)$. A cartoon illustrating this logic appears in Figure 9 of the main text. For further discussion of the connection between order-disorder properties of the stat mech model and entropic properties of the underlying quantum objects, see [2, 3, 11].

This logic suggests that, if we take the scaling of $\tilde{S}_2$ to be a good proxy for the scaling of $\langle S_2\rangle$, the disorder-to-order phase transition in the classical model would be

accompanied by an area-law-to-volume-law phase transition in the Rényi-2 entropy of the output of random circuits.

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

Due to the similarity between the dynamics studied in this section and that of SEBD running on CHR, our conclusion provides a partial explanation for the numerical observation presented in Section V that the average entanglement entropy $\langle S_k \rangle$ satisfies an area law when SEBD runs on CHR for $q = 2$ and various values of k.

g. Additional observations appearing in previous work. The above analysis is essentially a restatement of what appears in recent works by Bao, Choi, and Altman [2] and separately Jian, You, Vasseur, and Ludwig [3], except that here we analyzed a different weak measurement. In particular, [2] considered the case where a projective measurement occurs with some probability p on each qudit after each layer of unitaries, and otherwise there is no measurement. They made the observation that we describe above that the $k = 2$ mapping can be written as a 2-body anisotropic Ising model on the triangular lattice with an exact solution. Both of these papers went beyond what we have presented here to analyze the $q \to \infty$ limit directly, where they observed that the stat mech model becomes a standard ferromagnetic Potts model on the square lattice for all integers k. For $k = 2$ this is exactly the square lattice Ising model and indeed, we can see from Eq. (A38) that when $q \to \infty$, $J_3 \to 0$; the anti-ferromagnetic links along one axis vanish leaving a square lattice with exclusively ferromagnetic interactions. The fact that the model becomes tractable for all integers $k \geq 2$ allows these papers to invoke analytic continuation and make sense of the $k \to 1$ limit, where the quasi-entropy $\tilde{S}_k$ exactly becomes the expected von Neumann entropy $\langle S \rangle$.

Appendix B: Patching

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

The algorithm we describe is an adaptation and simplification of the Gibbs state preparation algorithm of [12]. In that paper, the authors essentially showed that a quantum Gibbs state defined on a lattice can be prepared by a quasipolynomial time quantum algorithm, if the Gibbs state satisfies two properties: (1) exponential decay of correlations and (2) exponentially decaying quantum conditional mutual information for shielded regions. Our situation is simpler than the one considered in that paper, due to the fact that sufficiently separated regions of the lattice are causally disconnected as a result of the fact that the circuit inducing the distribution is constant-depth and therefore has a constant-radius lightcone. The structure of our algorithm is very similar to theirs, except we can make some simplifications and substantial improvements as a result of the constant-radius lightcone and the fact that we are sampling from a classical distribution rather than a quantum Gibbs state.

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

Lemma B.1 (see e.g. [13]). *Let X, Y, Z be discrete random variables, and let p_{XYZ} denote their joint distribu-*

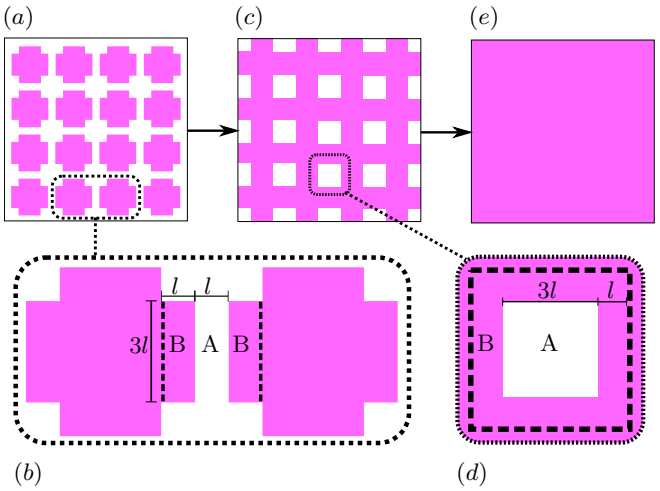


FIG. B.1. `Patching`. Pink represents marginals of the output distribution that have been approximately sampled, while white represents unsampled regions. In (a), the algorithm has sampled from disconnected patches. Figure (b) depicts how the algorithm transitions from configuration (a) to (c). Namely, the algorithm generates a sample from the conditional distribution on A, conditioned on the configuration of region B. Similarly, figure (d) depicts how the "holes" of configuration (c) are filled in. The end result is shown in (e), an approximate sample from the global distribution on the full lattice.

tion. Then

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

Following [12], we also formally define a notion of CMI decay.

Definition 1 (Markov property)**.** *Let p denote a probability distribution supported on Λ. Then p is said to satisfy the* $\delta(l)$-Markov condition *if, for any tripartition of a subregion X of the lattice into subregions $X = A \cup B \cup C$ such that* $\mathrm{dist}(A, C) \geq l$*, we have*

$$I(A:C|B)_p \leq \delta(l). \tag{B1}$$

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

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

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

Proof. The algorithm proceeds in three steps, illustrated in Figure B.1. First, for each square subregion R_i shaded in Figure B.1(a) with $i \in [O(n/l^2)]$, sample from $\mathcal{D}_{R_i}$, the marginal distribution of $\mathcal{D}$ on subregion R_i. To do this, first restrict to the qudits and gates in the lightcone of R_i. Sampling from the output distribution on R_i produced by this restricted version of the circuit is equivalent to sampling from the marginal on R_i of the true distribution produced by the full circuit. Since $l > 2d$, this restriction of the circuit is contained in a sublattice of dimensions $O(l) \times O(l)$. Using standard tensor network methods [14], sampling from the output distribution of this restricted circuit on R_i can be performed in time $q^{O(dl)}$. Since there are $O(n/l^2)$ patches, this step can be performed in time $nq^{O(dl)}$. After performing this step, we have prepared the state $\mathcal{D}_{R_1} \otimes \cdots \otimes \mathcal{D}_{R_k} = \mathcal{D}_{R_1,\ldots,R_k}$ where the equality holds because the patches are separated by $l > 2d$ and are therefore mutually independent.

In the second step, we apply "recovery maps" to approximately prepare a sample from the larger, connected lattice subregion S shaded in Figure B.1(c). The prescription for these recovery maps is given in Figure B.1(b). Referring to this figure, a recovery map $\mathcal{R}_{B\to AB}$ is applied to generate a sample from subregion A, conditioned on the state of region B. Explicitly, the mapping is given by linearly extending the map $\mathcal{R}_{B\to AB}(|b\rangle\langle b|_B) = \sum_a \mathcal{D}_{A|B}(a|b)\,|a\rangle\langle a|_A \otimes |b\rangle\langle b|_B$. Note that, for a tripartite distribution $\mathcal{D}_{ABC}$, $\mathcal{R}_{B\to AB}(\mathcal{D}_{BC}) = \mathcal{D}_{A|B}\mathcal{D}_B\mathcal{D}_{C|B}$. To implement this recovery map, one can again restrict to gates in the lightcone of region AB and utilize standard tensor network simulation algorithms to generate a sample from the marginal distribution on A, conditioned on the (previously sampled) state of B. The time complexity for this step is again $q^{O(dl)}$. After applying this and $O(n/l^2)$ similar recovery maps, we obtain a sample from a distribution $\mathcal{D}'_S$. By Lemma B.1, the triangle inequality, and Definition 1, the error of this step is bounded as

$$\|\mathcal{D}'_S - \mathcal{D}_S\|_1 \leq O(1)(n/l^2)\sqrt{\delta(l)} = O(1)(n/l^2)\sqrt{\delta(l)}. \tag{B2}$$

Note that the fact that the errors caused by recovery maps acting on disjoint regions accumulate at most linearly has been referred to previously [12] as the "union property" for recovery maps. The final step is very similar to the previous step. We now apply recovery maps, described by Figure B.1(d), to fill in the "holes" of the subregion S and approximately obtain a sample from the full distribution $\mathcal{D} = \mathcal{D}_\Lambda$. By a similar analysis, we find that the error incurred in this step is again $O(1)(n/l^2)\sqrt{\delta(l)}$, and therefore the procedure samples from a distribution $\mathcal{D}'_\Lambda$ for which $\|\mathcal{D}'_\Lambda - \mathcal{D}_\Lambda\|_1 \leq O(1)(n/l^2)\sqrt{\delta(l)}$.

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

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

Corollary 1. *Let $\mathcal{C}$ be a random circuit distribution. Define $\mathcal{C}$ to be $\delta(l)$-Markov if, for any tripartition of a subregion X of the lattice into subregions $X = A \cup B \cup C$ such that dist$(A, C) \geq l$, we have*

$$\langle I(A : C|B)_{\mathcal{D}} \rangle \leq \delta(l) \tag{B3}$$

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

Thus, proving that some uniform worst-case-hard circuit family $\mathcal{C}$ is $\operatorname{poly}(n)e^{-\Omega(l)}$-Markov provides another route to proving the part of Conjecture 1 about sampling with small total variation distance error. In Section VI, we will give analytical evidence that if $\mathcal{C}$ is a random circuit distribution of sufficiently low depth and small qudit dimension, then $\mathcal{C}$ is indeed $\operatorname{poly}(n)e^{-\Omega(l)}$-Markov, and if the depth or qudit dimension becomes sufficiently large, then $\mathcal{C}$ is not $\operatorname{poly}(n)f(l)$-Markov for any $f(l) = o(1)$, supporting Conjecture 2, which states that our algorithms exhibit computational phase transitions.

Finally, we note that `Patching` can also be used to estimate specific output probabilities of a random circuit instance C with high probability if C is drawn from a distribution $\mathcal{C}$ that is $\operatorname{poly}(n)e^{-\Omega(l)}$-Markov. This shows that the Markov condition could also be used to prove the second part of Conjecture 1 regarding computing output probabilities with small error. This is similar to how `SEBD` can also be used to compute output probabilities, as discussed in Section III B.

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

$$|\mathcal{D}'(\mathbf{x}) - \mathcal{D}(\mathbf{x})| \leq q^{-n}/r(n) \tag{B4}$$

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

Proof. With probability $1 - 1/\operatorname{poly}(n)$ over the circuit instance C, `Patching` with some choice of lengthscale $l = \Theta(\log n)$ efficiently samples from a distribution $\mathcal{D}'_C$ that is $1/\operatorname{poly}(n)$-close in variational distance to $\mathcal{D}_C$ for any choice of polynomials. Hence, for an output probability $\mathbf{y}$ chosen uniformly at random and a circuit C drawn from $\mathcal{C}$, it holds that

$$\mathop{\mathbb{E}}_{\mathbf{y}} \mathop{\mathbb{E}}_{C} |\mathcal{D}'(\mathbf{y}) - \mathcal{D}(\mathbf{y})| \leq q^{-n}/\operatorname{poly}(n) \tag{B5}$$

if $l = c \log n$ and c is a sufficiently large constant. By a nearly identical argument to that used in the proof of Corollary 3, due to the invariance of $\mathcal{C}$ under application of a final layer of single qudit gates, for some fixed $\mathbf{x} \in [q]^n$ we also have

$$\mathop{\mathbb{E}}_{C} |\mathcal{D}'(\mathbf{x}) - \mathcal{D}(\mathbf{x})| \leq q^{-n}/\operatorname{poly}(n) \tag{B6}$$

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

Appendix C: Efficiency of `Patching` algorithm from stat mech

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

We have previously applied the stat mech model to study expected entropies of quantum states. However, we now wish to study expected entropies of the classical output distribution. To this end, we now consider the non-unitary quantum circuit consisting of the original, unitary circuit followed by a layer of dephasing channels applied to every qudit. The resulting mixed state is classical (i.e., diagonal in the computational basis) and is exactly equal to the output distribution we want to study. That is, the state after application of the dephasing channels is $\sum_{\mathbf{x}} \mathcal{D}(\mathbf{x}) |\mathbf{x}\rangle\langle\mathbf{x}|$ where $\mathcal{D}$ is the output distribution of the circuit. Note that the application of the dephasing channel is not described in the formalism we have discussed previously, but is easily incorporated. In particular, we need to compute the weights between the auxiliary node x_a and the corresponding outgoing node s_u associated with the unitary u that is the last in the circuit to act on qudit a. We may update Eq. (A14) (whose original form was derived in Appendix A 2) and compute the following, letting $|\Phi_k\rangle \equiv \left(\sum_{i=0}^{q-1} |i\rangle \otimes |i\rangle\right)^{\otimes k}$.

$$\begin{aligned}\text{weight}(\langle s_u x_a\rangle) &= \langle\Phi_k|(I\otimes W_{\sigma_u^{-1}})\left(\sum_{i=0}^{q-1}|i\rangle\langle i|\otimes|i\rangle\langle i|\right)^{\otimes k}(I\otimes W_{\chi_a})|\Phi_k\rangle \\ &= \sum_{i_1,\ldots,i_k}\langle i_1,\ldots,i_k|W_{\sigma_u^{-1}}|i_1,\ldots,i_k\rangle\ \langle i_1,\ldots,i_k|W_{\chi_a}|i_1,\ldots,i_k\rangle\,.\end{aligned} \tag{C1}$$

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

$$\text{weight}(\langle s_u x_a\rangle)=\begin{cases}q^{C(\sigma_u)}, & \chi_a=e\\ q, & \chi_a=(1\ldots k).\end{cases} \tag{C2}$$

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

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

a. Disordered stat mech model suggests `Patching` *is successful.* We consider the quasi-CMI defined by

$$\tilde{I}_2(A:C|B):=\tilde{S}_2(AB)+\tilde{S}_2(BC)-\tilde{S}_2(B)-\tilde{S}_2(ABC), \tag{C3}$$

where all quasi-entropies are taken with respect to the collection of classical output distributions that arise from the quantum circuit architecture. This definition is in analogy to the definition of CMI as $I(A:C|B)=S(AB)+S(BC)-S(B)-S(ABC)$ [13]. Note that we may define the quasi-k-CMI $\tilde{I}_k(A:C|B)$ analogously for any nonnegative k, and it holds that $\langle I(A:C|B)_{\mathcal{D}}\rangle=\lim_{k\to1}\tilde{I}_k(A:C|B)$ where the angle brackets denote an expectation over random circuit instances.

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

$$\tilde{I}_2(A:C|B)=(F_{2,AB}-F_{2,B})-(F_{2,ABC}-F_{2,BC}). \tag{C4}$$

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

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

$$\tilde{I}_2(A:C|B)\le\text{poly}(n,q)e^{-\text{dist}(A,C)/\xi} \tag{C5}$$

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

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

As before, consider the stat mech model obtained by applying dephasing channels to all qudits after the application of all gates. Consider some connected, strict subset A of qudits on the original grid. Suppose we are interested in the quasi-entropy $\tilde{S}_k(A)=(F_{k,A}-F_{k,\emptyset})/(k-1)$ of the output distribution on this region. This quantity is given by the free energy cost of twisting the boundary conditions (auxiliary spins) associated with region A from e to $(1\ldots k)$. The auxiliary spins associated with qudits in the complement of A are fixed to be in the identity permutation configuration, e. For both sets of boundary conditions, all non-auxiliary spins will order in the configuration e. This is because the configuration e maximizes the weights in Eq. (C2) for spins connected to auxiliary spins in the configuration e, and the weight of a spin connected to an auxiliary spin in the configura-

tion $(1 \dots k)$ is independent of that spin's configuration. Hence, regardless of the configuration of the auxiliary spins, all bulk spins are in the identity permutation configuration in the $q \to \infty$ limit of infinitely strong couplings.

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

$$\tilde{S}_k(A) = \frac{F_{k,A} - F_{k,\emptyset}}{k-1} = |A|\log(q). \tag{C6}$$

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

$$\langle S(A) \rangle = \lim_{k \to 1} |A|\log(q) = |A|\log(q). \tag{C7}$$

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

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

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

As before, for each auxiliary spin associated with region X that we "twist", the weight of the configuration is decreased by a factor of q^{1-k} relative to the configuration in which all auxiliary spins are set to e. However, in this case, as opposed to our previous calculation, *all* of the auxiliary spins are twisted. Recall from Eq. (C2) that the weight between a twisted auxiliary spin and a bulk spin is independent of the value of the bulk spin. Hence, if all auxiliary spins are twisted, the lowest energy state in the bulk is no longer just the configuration in which all spins take the value e – in the absence of a symmetry-breaking boundary condition, there is now a global spin-flip symmetry and the ground space is $k!$-fold degenerate, consisting of all configurations in which all bulk spins are aligned. This symmetry contributes a factor of $k!$ to the partition function and $-\log(k!)$ to the free energy. We hence calculate

$$\tilde{S}_k(X) = |A|\log(q) - \frac{\log(k!)}{k-1}. \tag{C8}$$

We now perform the analytic continuation to the Shannon entropy:

$$\langle S(X) \rangle = \lim_{k \to 1} \tilde{S}_k(X) \tag{C9}$$

$$= |A|\log(q) - \lim_{k \to 1} \frac{\log(k!)}{k-1} \tag{C10}$$

$$= |A|\log(q) - \frac{1-\gamma}{\ln(2)} \tag{C11}$$

$$\approx |A|\log(q) - 0.61, \tag{C12}$$

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

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

$$\langle I(A:C|B)_{\mathcal{D}} \rangle \tag{C13}$$

$$\equiv \langle S(AB)_{\mathcal{D}} + S(BC)_{\mathcal{D}} - S(B)_{\mathcal{D}} - S(ABC)_{\mathcal{D}} \rangle \tag{C14}$$

$$= [(|A| + |B|)\log(q)] + [(|B| + |C|)\log(q)] \tag{C15}$$

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

$$= \frac{1-\gamma}{\ln(2)} \approx 0.61. \tag{C16}$$

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

Appendix D: Relation to worst-to-average-case reductions based on truncated Taylor series

Recently, it was shown [15] that for any constant-depth random circuit family with Haar-random gates acting on n qubits for which it is #P-hard to compute output probabilities in the worst case, there does not exist a poly(n)-time algorithm for computing the output probability of some arbitrary output string $\mathbf{x}$ up to additive error $2^{-\tilde{\Theta}(n^3)}$ with high probability over the circuit realization, unless there exists a poly(n)-time randomized algorithm for computing a #P-hard function. (Note: in even more recent work using the same technique, the error robustness has been improved from $2^{-\tilde{\Theta}(n^3)}$ to $2^{-\Theta(n\log(n))}$ [16, 17].) Essentially, for Haar-random circuits, near-exact average-case computation of output probabilities is as hard as worst-case computation of output probabilities. Our complexity separation in Section IV shows that the error tolerance for this hardness result cannot be improved to $2^{-n}/2^{n^c}$ for any $c < 1$.

This hardness result builds on and improves other prior work [1] on the average-case hardness of random circuit simulation. In particular, the original paper [1] uses a

different interpolation scheme than that used in [15] to perform the worst-to-average-case reduction. Interestingly, as discussed below, we find that the interpolation scheme of [1] cannot be used to prove hardness results about our algorithms' performance on a shallow random 2D quantum circuit possessing worst-case hardness for computing output probabilities; this essentially is a consequence of how `SEBD` and `Patching` exploit the unitarity of the circuit to be simulated. While this observation may be of technical interest for future work on worst-to-average-case reductions for quantum circuit simulations, the alternative interpolation scheme of [15] does not suffer from this limitation.

While [1, 15] prove hardness results for the near-exact computation of output probabilites of random circuits, it is ultimately desirable to prove hardness for the Random Circuit Sampling (RCS) problem of sampling from the output distribution of a random circuit with small error in variational distance, as this is the computational task corresponding to the problem that the quantum computer solves. *A priori*, one might hope that such a result could be proved via such a worst-to-average-case reduction. In particular, it was pointed out in these works that improving the error tolerance of the hardness result to $2^{-n}/\operatorname{poly}(n)$ would be sufficient to prove hardness of RCS. Our work rules out such a proof strategy working in general by showing that even improving the error tolerance to $2^{-n}/2^{n^c}$ for any constant $c < 1$ is unachievable. In particular, any proof of the hardness of RCS should be sensitive to the depth and should not be applicable to the worst-case-hard shallow random circuit ensembles that admit approximate average-case classical simulations.

Implications for reductions based on truncated Taylor series

In this section, we discuss the relation between our algorithms (`SEBD` and `Patching`) applied to the computation of output probabilities and the recent result [1] on the hardness of average-case simulation of random circuits based on polynomial interpolation via truncated Taylor series. In particular, we discuss how this polynomial interpolation argument is insufficient to show that the task of even *exactly* computing output probabilities and sampling from the output distribution of a constant-depth Haar-random circuit instance with high probability using our algorithms is classically hard, even though these circuits possess worst-case hardness. We first briefly review their technique before discussing a limitation in the robustness of the polynomial interpolation scheme. We then discuss how this robustness limitation makes the interpolation scheme inapplicable to our algorithms.

The main point is that our algorithms exploit unitarity (via the fact that gates outside of the lightcone of the qudits currently under consideration are ignored), but the hardness result of [1] holds with respect to circuit families that are non-unitary, albeit very close to unitary in some sense. Our algorithms are unable to simulate these slightly non-unitary circuits to the precision required for the worst-to-average case reduction, regardless of how well they can simulate Haar-random circuit families. While it is true that in this scheme there is an adjustable parameter K which, when increased, brings the non-unitary circuit family closer to approximating the true Haar-random family, increasing K also increases the degree of the interpolating polynomial. This makes the interpolation more sensitive to errors in such a way that, for any choice of K, the robustness that the interpolation can tolerate is not large enough to overcome the inherent errors that our algorithms make when trying to simulate these non-unitary families. The existence of simulation algorithms like `SEBD` and `Patching`, which exploit the unitarity of the circuit, may present an obstruction to applying worst-to-average-case reduction techniques that obtain a polynomial structure at the expense of unitarity. Note that, as discussed previously, a very recent alternative worst-to-average case reduction [15] based on "Cayley paths" rather than truncated Taylor series does not suffer from this same limitation.

Background: truncated Haar-random circuit ensembles and polynomial interpolation

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

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

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

as $\mathcal{D}(C,\theta,K)$. Define the quantity

$$p_0(A,\theta,K) := |\langle 0|A(C,\theta,K)|0\rangle|^2. \tag{D1}$$

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

Now, assume that there exists some classical algorithm $\mathcal{A}$ and some $\epsilon = 1/\operatorname{poly}(n)$ such that, for some fixed $K \leq \operatorname{poly}(n)$ and for all $0 \leq \theta \leq \epsilon$, $\mathcal{A}$ can compute $p_0(A,\theta,K)$ up to additive error $\delta \leq 2^{-n^c}$ for some constant c, with probability $1-1/\operatorname{poly}(n)$ over $A(C,\theta,K) \sim \mathcal{D}(C,\theta,K)$. Then, $\mathcal{A}$ may evaluate $p_0(A,\theta,K)$ for $2mK+1$ evenly spaced values of θ in the range $[0,\epsilon]$ (up to very small error), and construct an interpolating polynomial $q_0(A,\theta,K)$. By a result of Rakhmanov [18], there is some interval $[a,b] \subset [0,\epsilon]$ such that $b-a \geq 1/\operatorname{poly}(n)$ and $|p_0(A,\theta,K) - q_0(A,\theta,K)| \leq 2^{-n^{c'}}$ for $\theta \in [a,b]$ where c' depends on c. One then invokes the following result of Paturi.

Lemma D.1 ([19]). *Let $p : \mathbb{R} \to \mathbb{R}$ be a real polynomial of degree d, and suppose $|p(x)| \leq \delta$ for all $|x| \leq \epsilon$. Then $|p(1)| \leq \delta e^{2d(1+1/\epsilon)}$.*

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

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

Limitation of the interpolation argument

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

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

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

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

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

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

Inapplicability to `SEBD` and `Patching`

To summarize the findings above, the argument of [1] for the hardness of computing output probabilities of random circuits applies not directly to Haar-random circuit distributions, but rather to distributions over slightly non-unitary circuits that are exponentially close to the corresponding Haar distributions in some sense. We argued that the interpolation scheme cannot be straight-

[1] A simplified and slightly weaker version of our argument was also reported in [15].

forwardly applied to circuits that are truly Haar-random, and therefore it cannot be used to conclude that simulating truly Haar-random circuits, even exactly, is classically hard.

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

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

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

Appendix E: Deferred proofs

Lemma 1. *Let ϵ_i denote the sum of the squares of all singular values discarded in the compression during iteration i of the simulation of a circuit C with output distribution $\mathcal{D}_C$ by `SEBD` with no bond dimension cutoff, and let Λ denote the sum of all singular values discarded over the course of the algorithm. Then the distribution $\mathcal{D}'_C$ sampled from by `SEBD` satisfies*

$$\frac{1}{2}\|\mathcal{D}'_C - \mathcal{D}_C\|_1 \leq \mathbb{E}\sum_{i=1}^{L_2}\sqrt{2\epsilon_i} \leq \sqrt{2}\,\mathbb{E}\,\Lambda, \tag{E1}$$

where the expectations are over the random measurement outcomes.

Proof. We rely upon a well-known fact about the error caused by truncating the bond dimension of a MPS, which we state in Lemma E.1.

Lemma E.1 (follows from [20])**.** *Suppose the MPS $|\psi\rangle$ is compressed via truncation of small singular values, and ϵ is the sum of the squares of the discarded singular values. Then if $|\psi^{(t)}\rangle$ is the truncated version of the MPS after normalization,*

$$\||\psi\rangle\langle\psi| - |\psi^{(t)}\rangle\langle\psi^{(t)}|\|_1 \leq \sqrt{8\epsilon}. \tag{E2}$$

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

$$\mathcal{N}_t(\rho) = \sum_{\mathbf{x}_t} \Pi_t^{\mathbf{x}_t} V_t \rho V_t^\dagger \Pi_t^{\mathbf{x}_t}, \tag{E3}$$

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

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

$$\sigma_t := (T_t \circ \mathcal{N}_{t-1})(\sigma_{t-1}), \tag{E4}$$

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

We now have

$$\|\mathcal{D}_C - \mathcal{D}'_C\|_1 = \|\rho_{L_2+1} - \sigma_{L_2+1}\|_1 \tag{E5}$$

$$\leq \|\rho_{L_2+1} - \mathcal{N}_{L_2}(\sigma_{L_2})\|_1 + \|\mathcal{N}_{L_2}(\sigma_{L_2}) - \sigma_{L_1+1}\|_1 \tag{E6}$$

$$\leq \|\rho_{L_2} - \sigma_{L_2}\|_1 + \|\mathcal{N}_{L_2}(\sigma_{L_2}) - \sigma_{L_2+1}\|_1, \tag{E7}$$

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

$$\|\mathcal{D}_C - \mathcal{D}'_C\|_1 \leq \sum_{i=1}^{L_2} \|\mathcal{N}_i(\sigma_i) - \sigma_{i+1}\|_1 \tag{E8}$$

$$= \sum_{i=1}^{L_2-1} \|\mathcal{N}_i(\sigma_i) - (T_{i+1} \circ \mathcal{N}_i)(\sigma_i)\|_1 \tag{E9}$$

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

Lemma 5. *Let $\lambda_1 \geq \lambda_2 \geq \cdots$ denote the half-chain Schmidt values after at least $n/2$ iterations of the toy model process. Then with probability at least $1-\delta$ the half-chain Schmidt values indexed by $i \geq i^* = \exp\left(\Theta(\sqrt{\log(n/\delta)})\right)$ obey the asymptotic scaling*

$$\lambda_i \propto \exp\left(-\Theta(\log^2(i))\right). \tag{E10}$$

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

$$r \leq \exp\left(\Theta\left(\sqrt{\log(n/\epsilon\delta)}\right)\right). \tag{E11}$$

Proof. Suppose that an EPR pair is measured $2t$ times, corresponding to each of the two qubits being measured t times. A calculation shows that the probability of obtaining s M_1 outcomes is given by a mixture of two binomial distributions. Letting S be the random variable denoting the number of M_1 outcomes, we find that $\Pr[S=s]$ is given by

$$\frac{1}{2}\Pr\left[B_{2t,\sin^2(\theta/2)} = s\right] + \frac{1}{2}\Pr\left[B_{2t,\cos^2(\theta/2)} = s\right], \tag{E12}$$

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

$$|00\rangle + \tan^{2(t-s)}(\theta/2)\,|11\rangle\,. \tag{E13}$$

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

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

$$= |00\rangle + \tan^{2t\cos(\theta)-2X_{2t,\sin^2(\theta/2)}}(\theta/2)\,|11\rangle \tag{E14}$$

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

$$|00\rangle + \gamma^{t-X'_{2t,\sin^2(\theta/2)}}\,|11\rangle\,. \tag{E15}$$

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

$$|00\rangle + \gamma^{-t-X'_{2t,\cos^2(\theta/2)}}\,|11\rangle\,. \tag{E16}$$

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

$$|00\rangle + \gamma^{t-X'_{2t,\sin^2(\theta/2)}}\,|11\rangle\,, \tag{E17}$$

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

We can then model the final state as

$$\bigotimes_t |00\rangle + \gamma^{t-X'_{2t,\sin^2(\theta/2)}}\,|11\rangle \tag{E18}$$

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

Let $Q(\ell)$ denote the number of "strict partitions" of ℓ, i.e. the number of ways of writing $\ell = t_1 + t_2 + \ldots$ for positive integers $t_1 < t_2 < \ldots$. Precise asymptotics are known for $Q(\ell)$ (see https://oeis.org/A000009 and [21]):

$$Q(\ell) = \exp\left(\Theta(\sqrt{\ell})\right). \tag{E19}$$

By expanding Eq. (E18) as a superposition over computational basis states, we obtain the unnormalized Schmidt coefficients $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots$; each coefficient in the expansion gives an unnormalized Schmidt coefficient. There are $Q(\ell)$ unnormalized Schmidt coefficients that are distributed as $\gamma^{\ell - X'_{2\ell,\sin^2(\theta/2)}}$, where we used

the fact that $X'_{t_1,\sin^2(\theta/2)} + X'_{t_2,\sin^2(\theta/2)}$ is distributed as $X'_{t_1+t_2,\sin^2(\theta/2)}$. We say that these $Q(\ell)$ coefficients live in sector ℓ. For a fixed probability p, let $K_{\ell,p}$ denote the smallest positive integer for which, with probability at least $1-p$, all sector-ℓ coefficients lie in the range $[\gamma^{\ell+K_{\ell,p}}, \gamma^{\ell-K_{\ell,p}}]$. By the union bound, to upper bound $K_{\ell,p}$ it suffices to find an integer a for which

$$\Pr\Big[\Big|X'_{2\ell,\sin^2(\theta/2)}\Big| \geq a\Big] \leq \frac{p}{Q(\ell)} = p\exp\Big(-\Theta(\sqrt{\ell})\Big). \tag{E20}$$

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

$$K_{\ell,p} \leq \Theta\left(\sqrt{\ell\log(1/p) + \ell\sqrt{\ell}}\right). \tag{E21}$$

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

Still working with the unnormalized state of Eq. (E18), we now study the scaling between the Schmidt index i and corresponding coefficient $\tilde{\lambda}_i$ for i in the regime $i \geq \exp\Big(\Theta(\sqrt{\log(1/p)})\Big)$. Note that $\tilde{\lambda}_i = \gamma^\ell$ for some integer ℓ. We first lower bound ℓ. Note that the lower bound is achieved if, for each sector j, all coefficients in that sector are equal to $\gamma^{j-K_{j,p}}$. In this case, the exponent ℓ is equal to $\ell' - K_{\ell',p}$, where ℓ' is the smallest integer such that

$$i \leq \sum_{j=1}^{\ell'} Q(\ell') = \exp\Big(\Theta(\sqrt{\ell'})\Big). \tag{E22}$$

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

$$\tilde{\lambda}_i = \exp\big(\Theta(-\log^2(i))\big),\ i \geq \exp\Big(\Theta(\sqrt{\log(1/p)})\Big). \tag{E23}$$

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

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

$$\epsilon = \sum_{j\geq i} \lambda_j^2 < \sum_{j\geq i} \tilde{\lambda}_j^2 = \exp\big(-\Theta(\log^2(i))\big) \tag{E24}$$

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

$$i \leq \exp\Big(\Theta(\sqrt{\log(1/\epsilon)})\Big). \tag{E25}$$

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

$$r \leq \max\left(\exp\Big(\Theta(\sqrt{\log(1/\epsilon)})\Big), \exp\Big(\Theta(\sqrt{\log(1/p)})\Big)\right) \tag{E26}$$

$$= \exp\left(\Theta\left(\sqrt{\log\left(\frac{n}{\epsilon\cdot\delta}\right)}\right)\right) \tag{E27}$$

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

Lemma 7. *Suppose a 1D random circuit C is applied to qubits $\{1,\ldots,n\}$ consisting of a layer of 2-qubit Haar-random gates acting on qubits $(k,k+1)$ for odd $k \in \{1,\ldots,n-1\}$, followed by a layer of 2-qubit Haar-random gates acting on qubits $(k,k+1)$ for even $k \in \{1,\ldots,n-1\}$. Suppose the qubits of region $B := \{i,i+1,\ldots,j\}$ for $j \geq i$ are measured in the computational basis, and the outcome b is obtained. Then, letting $|\psi_b\rangle$ denote the post-measurement pure state on the unmeasured qubits, and letting $A := \{1,2,\ldots,i-1\}$ denote the qubits to the left of B,*

$$\mathbb{E}\, S(A)_{\psi_b} \leq c^{|B|} \tag{E28}$$

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

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

Lemma E.2. *Let $|\psi\rangle_{AB}$ be some state on subsystems A and B with subsystem B a qubit, and let $|H\rangle_{CD}$ be some two-qubit Haar-random state on subsystems C and D. Suppose a Haar-random two-qubit gate U is applied to subsystems B and C. If subsystem B is measured in the computational basis and outcome b is obtained, then the von Neumann entropy of the post-measurement state $|\psi_b\rangle_{ABCD}$ in subsystem A satisfies*

$$\mathop{\mathbb{E}}_{b,H,U} S(A)_{\psi_b} \leq c\cdot S(A)_\psi \tag{E29}$$

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

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

$$\begin{aligned}&\frac{1}{\sqrt{\Pr(Y=b)}}\left(\sqrt{p\cdot\Pr(Y=b|X=1)}\,|e_1\rangle_A\,|b\rangle_B\,|\phi_{b,1}\rangle_{C,D} + \sqrt{(1-p)\cdot\Pr(Y=b|X=2)}\,|e_2\rangle_A\,|b\rangle_B\,|\phi_{b,2}\rangle_{C,D}\right)\\ =&\sqrt{\Pr(X=1|Y=b)}\,|e_1\rangle_A\,|b\rangle_B\,|\phi_{b,1}\rangle_{C,D} + \sqrt{\Pr(X=2|Y=b)}\,|e_2\rangle_A\,|b\rangle_B\,|\phi_{b,2}\rangle_{C,D}\end{aligned} \tag{E30}$$

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

$$\epsilon := \min_b |\langle\phi_{b,1}|\phi_{b,2}\rangle|^2. \tag{E31}$$

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

$$\mathop{\mathbb{E}}_Y \lambda_{\max}(\rho_{A,Y}) \geq \mathop{\mathbb{E}}_Y [\Pr(X=1|Y) + \epsilon\Pr(X=2|Y)] \tag{E32}$$

$$= p + \epsilon(1-p). \tag{E33}$$

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

$$\begin{aligned}\mathop{\mathbb{E}}_Y S(A)_{\psi_Y} &= \mathop{\mathbb{E}}_Y H_2(\lambda_{\max}(\rho_{A,Y})) &\text{(E34)}\\ &\leq H_2(\mathop{\mathbb{E}}_Y \lambda_{\max}(\rho_{A,Y})) &\text{(E35)}\\ &\leq H_2(p+\epsilon(1-p)). &\text{(E36)}\end{aligned}$$

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

$$\frac{\mathbb{E}_Y S(A)_{\psi_Y}}{S(A)_\psi} \leq r(p,\epsilon) \leq c(\epsilon) \tag{E37}$$

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

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

By Lemma E.2, after the first iteration we are left with the state $|\psi\rangle_{LR}\,|b_{i_1}\rangle_{i_1}$, such that $\mathbb{E}S(L)_\psi \leq c$ for some constant $c < 1$. In all iterations, we let L denote the current subsystem to the left of the measured qubits, and R denote the subsystem to the right of the measured qubits. Now consider the second iteration. Depending on whether i was even or odd, R may consist of one or two qubits immediately after the measurement of i. In the former case, we may apply Lemma E.2 again, obtaining $\mathbb{E}S(L)_\psi \leq c^2$ after the measurement of qubit $i+1$, and obtaining a two-qubit subsystem to the right of the measured qubits. In the latter case, as a consequence of concavity of von Neumann entropy, we have $\mathbb{E}S(L)_\psi \leq c$ after measurement, and are left with a one-qubit subsystem to the right of the measured qubits. Iterating this process, after all qubits of subregion B have been

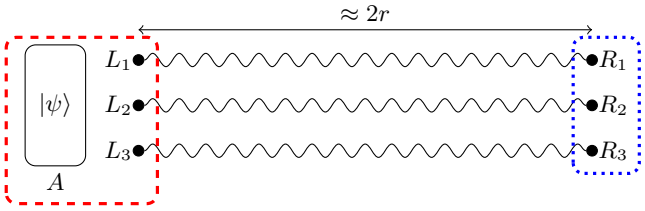


FIG. E.1. Illustration of the state after the qubits of columns $i, i+1, \ldots, j$ have been measured, but before gates in the lightcone of registers A and L have been performed. In each row i, we are left with a post-measurement bipartite state $|\phi_i\rangle_{L_i R_i}$ depicted by a wavy line. The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

measured, we are left with some state $|\psi\rangle_{LR}$ such that $\mathbb{E}\, S(L)_\psi \le c^{|B|/2} \le c'^{|B|}$ where $c' = \sqrt{c} < 1$. Finally, local unitary gates are applied to $|\psi\rangle_{LR}$ to obtain the final post-measurement state on the entire chain. Since each unitary is applied to only the left of region B or only the right of region B, the entanglement entropy across the (A, A^c) cut is unaffected by these gates, and remains bounded by $c^{|B|}$ in expectation. □

Lemma 8. *Let C be an instance of* `Brickwork`(L, r, v). *Then, with probability at least $1 - 2^{-\Theta(r)}$ over the circuit instance,* `SEBD` *running with maximal bond dimension cutoff $D = \Theta(1)$ and truncation error parameter $\epsilon = 2^{-\Theta(r)}$ can be used to (1) sample from the output distribution of C up to error $n2^{-\Theta(r)}$ in variational distance and (2) compute the output probability of an arbitrary output string up to additive error $n2^{-\Theta(r)}/2^n$ in runtime $\Theta(n)$.*

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

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

We next apply all gates in the lightcone of the qubits of columns $\{i, i+1, \ldots, j\}$, before measuring these qubits in the computational basis. Note that in this step, we are effectively performing a set of L one-dimensional depth-2 Haar-random circuits, and then measuring $\Theta(r)$ intermediate qubits for each of the L instances. For each instance, we are left with a (generically entangled) pure state between a "left" and "right" subsystem, as illustrated in Figure E.1. Let L_i (R_i) denote the left (right) subsystem associated with row i, and let $|\phi_i\rangle_{L_i R_i}$ denote the associated post-measurement pure state on these subsystems. By Lemma 7, it follows that the expected entanglement entropy for any 1D instance obeys $\mathbb{E}\, S(L_i)_{\phi_i} \le 2^{-\Theta(r)}$ where the expectation is over random circuit instance and measurement outcomes.

The next step is to apply all gates in the lightcone of the qubits of registers A and $L := \cup_i L_i$ before measuring these registers, inducing a (random) 1D post-measurement pure state on subsystem $R := \cup_i R_i$. It is straightforward to verify that the distribution of the random 1D pure state $|\psi'\rangle_R$ obtained via this procedure is identical to that obtained from repeatedly iterating `SEBD` through column j[2]. Indeed, the procedures are identical up to performing commuting gates and commuting measurements in different orders, which does not affect the measurement statistics or post-measurement states.

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

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

Now,

$$\sum_x \Pr[x] S(R_1 \dots R_k)_{\psi'}$$
$$= \sum_x \Pr[x] S(\rho^x_{R_1,\dots,R_k}) \tag{E38}$$
$$\leq S\left(\sum_x \Pr[x] \rho^x_{R_1,\dots,R_k}\right) \tag{E39}$$
$$= S(\rho_{R_1,\dots,R_k}) \tag{E40}$$
$$= \sum_{i=1}^{k} S(R_i)_{\phi_i} \tag{E41}$$

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

$$\mathbb{E}\, S(R_1 \dots R_k)_{\psi'} \leq L 2^{-\Theta(r)} \tag{E42}$$

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

$$\mathbb{E}\, \lambda_{\max}(R_1 \cdots R_k) \geq \mathbb{E}\, 2^{-S(R_1 \cdots R_k)_{\psi'}} \tag{E43}$$
$$\geq 2^{-\mathbb{E}\, S(R_1 \cdots R_k)_{\psi'}} \tag{E44}$$
$$\geq 2^{-L 2^{-\Theta(r)}} \tag{E45}$$
$$\geq 1 - L 2^{-\Theta(r)}. \tag{E46}$$

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

Therefore, if we run `SEBD` using a *per bond* truncation error of $\epsilon = L 2^{-\Theta(r)}$ and a maximum bond dimension cutoff of $D = O(1)$, the failure probability will be upper bounded by $Lv 2^{-\Theta(r)}$; here we used the union bound to upper bound the probability that any of the $O(Lv)$ bonds over the course of the algorithm becomes larger than the cutoff D. Hence, by Corollary 1, for at least $1 - 2^{-\Theta(r)}$ fraction of random circuit instances, `SEBD` can sample from the output distribution with variational distance error $Lv 2^{-\Theta(r)} < n 2^{-\Theta(r)}$. Similarly, by Corollary 3, for at least $1 - 2^{-\Theta(r)}$ fraction of circuit instances, `SEBD` can compute the probability of the all-zeros output string up to additive error $n 2^{-\Theta(r)}/2^n$.

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

[1] Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani, "On the complexity and verification of quantum random circuit sampling," Nature Physics **15**, 159 (2019), arXiv:1803.04402.

[2] Yimu Bao, Soonwon Choi, and Ehud Altman, "Theory of the phase transition in random unitary circuits with measurements," Phys. Rev. B **101**, 104301 (2020), arXiv:1908.04305.

[3] Chao-Ming Jian, Yi-Zhuang You, Romain Vasseur, and Andreas W. W. Ludwig, "Measurement-induced criticality in random quantum circuits," Phys. Rev. B **101**, 104302 (2020), arXiv:1908.08051.

[4] Benoît Collins, "Moments and cumulants of polynomial random variables on unitary groups, the itzykson-zuber integral, and free probability," International Mathematics Research Notices **2003**, 953–982 (2003), arXiv:math-ph/0205010.

[5] Benoît Collins and Piotr Śniady, "Integration with respect to the Haar measure on unitary, orthogonal and symplectic group," Communications in Mathematical Physics **264**, 773–795 (2006), arXiv:math-ph/0402073.

[6] Yaodong Li, Xiao Chen, and Matthew P. A. Fisher, "Quantum Zeno effect and the many-body entanglement transition," Phys. Rev. B **98**, 205136 (2018), arXiv:1808.06134.

[7] Amos Chan, Rahul M. Nandkishore, Michael Pretko, and Graeme Smith, "Unitary-projective entanglement dynamics," Phys. Rev. B **99**, 224307 (2019), arXiv:1808.05949.

[8] Brian Skinner, Jonathan Ruhman, and Adam Nahum, "Measurement-induced phase transitions in the dynamics of entanglement," Phys. Rev. X **9**, 031009 (2019), arXiv:1808.05953.

[9] Raymond Marie Ferdinand Houtappel, "Order-disorder in hexagonal lattices," Physica **16**, 425–455 (1950).

[10] John Stephenson, "Ising-model spin correlations on the triangular lattice. IV. anisotropic ferromagnetic and antiferromagnetic lattices," Journal of Mathematical Physics **11**, 420–431 (1970).

[11] Romain Vasseur, Andrew C. Potter, Yi-Zhuang You, and Andreas W. W. Ludwig, "Entanglement transitions from

holographic random tensor networks," Phys. Rev. B **100**, 134203 (2019), arXiv:1807.07082.

[12] Fernando GSL Brandão and Michael J Kastoryano, "Finite correlation length implies efficient preparation of quantum thermal states," Communications in Mathematical Physics **365**, 1–16 (2019), arXiv:1609.07877.

[13] Thomas M Cover and Joy A Thomas, *Elements of Information Theory* (1991).

[14] Igor L Markov and Yaoyun Shi, "Simulating quantum computation by contracting tensor networks," SIAM Journal on Computing **38**, 963–981 (2008), arXiv:quant-ph/0511069.

[15] Ramis Movassagh, "Quantum supremacy and random circuits," (2019), arXiv:1909.06210.

[16] Adam Bouland, Bill Fefferman, Zeph Landau, and Yunchao Liu, "Noise and the frontier of quantum supremacy," (2021), arXiv:2102.01738.

[17] Yasuhiro Kondo, Ryuhei Mori, and Ramis Movassagh, "Fine-grained analysis and improved robustness of quantum supremacy for random circuit sampling," (2021), arXiv:2102.01960.

[18] Evguenii A Rakhmanov, "Bounds for polynomials with a unit discrete norm," Annals of mathematics , 55–88 (2007).

[19] Ramamohan Paturi, "On the degree of polynomials that approximate symmetric boolean functions (preliminary version)," in *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing* (ACM, 1992) pp. 468–474.

[20] F. Verstraete and J. I. Cirac, "Matrix product states represent ground states faithfully," Phys. Rev. B **73**, 094423 (2006), arXiv:cond-mat/0505140.

[21] Raymond Ayoub, *Introduction to the analytic theory of numbers* (American Mathematical Society, 1963).