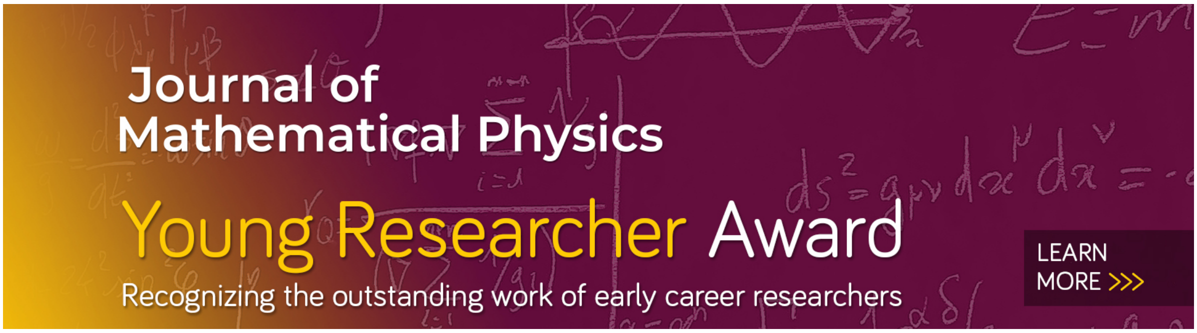
Journal of Mathematical Physics

Mixing and localization in random time-periodic quantum circuits of Clifford unitaries

Cite as: J. Math. Phys. **63**, 032201 (2022); https://doi.org/10.1063/5.0054863
Submitted: 22 April 2021 • Accepted: 04 March 2022 • Published Online: 25 March 2022

Tom Farshi, Daniele Toniolo, Carlos E. González-Guillén, et al.



Journal of Mathematical Physics ARTICLE scitation.org/journal/jmp

Mixing and localization in random time-periodic quantum circuits of Clifford unitaries

Cite as: J. Math. Phys. 63, 032201 (2022); doi: 10.1063/5.0054863
Submitted: 22 April 2021 • Accepted: 4 March 2022 •
Published Online: 25 March 2022

Tom Farshi,[1,2] Daniele Toniolo,[1,2] Carlos E. González-Guillén,[3] Álvaro M. Alhambra,[4,5] and Lluis Masanes[1,6,a)]

AFFILIATIONS

[1] Department of Computer Science, University College London, London, United Kingdom
[2] Department of Physics and Astronomy, University College London, London, United Kingdom
[3] Dept Matemática Aplicada a la Ingeniería Industrial, Universidad Politécnica de Madrid, Madrid, Spain
[4] Max-Planck-Institut für Quantenoptik, Garching, Germany
[5] Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
[6] London Centre for Nanotechnology, University College London, London, United Kingdom

[a)] Author to whom correspondence should be addressed: l.masanes@ucl.ac.uk

ABSTRACT

How much do local and time-periodic dynamics resemble a random unitary? In the present work, we address this question by using the Clifford formalism from quantum computation. We analyze a Floquet model with disorder, characterized by a family of local, time-periodic, and random quantum circuits in one spatial dimension. We observe that the evolution operator enjoys an extra symmetry at times that are a half-integer multiple of the period. With this, we prove that after the scrambling time, namely, when any initial perturbation has propagated throughout the system, the evolution operator cannot be distinguished from a (Haar) random unitary when all qubits are measured with Pauli operators. This indistinguishability decreases as time goes on, which is in high contrast to the more studied case of (time-dependent) random circuits. We also prove that the evolution of Pauli operators displays a form of mixing. These results require the dimension of the local subsystem to be large. In the opposite regime, our system displays a novel form of localization, produced by the appearance of effective one-sided walls, which prevent perturbations from crossing the wall in one direction but not the other.

I. INTRODUCTION

The distinction between chaotic and integrable quantum dynamics[1] plays a central role in many areas of physics, such as the study of equilibration,[2] thermalization,[3] and related topics, such as the eigenstate thermalization hypothesis,[4,5] quantum scars,[6,7] and the generalized Gibbs ensemble.[8] This distinction is also important in the characterization of many-body localization,[9] in the holographic correspondence between gravity and conformal field theory,[10] and in arguments concerning the black-hole information paradox.[11,12] Despite all this, the precise definitions of quantum chaos and integrability are still being debated.[13–15] However, it is well established that the dynamics of quantum chaotic systems shares important features with random unitaries.[16] These are the unitaries obtained with high probability when sampling from the unitary group of the total Hilbert space of the many-body system according to the uniform distribution (Haar measure[17]).

In order to find signatures of quantum chaos in physically relevant systems, it is a common practice to identify in them aspects of random unitaries. Some of these are as follows: the presence of eigenvalue repulsion in the Hamiltonian,[18,19] fast decay of out-of-time order correlators (OTOCs),[20–22] entanglement spreading,[23] operator entanglement,[24] entanglement spectrum,[25,26] and Loschmidt echo.[27] In this work, we take a more operational approach and analyze setups in which the evolution operator of a system is physically indistinguishable from a random unitary. We quantify this indistinguishability with a variant of the quantum information notion of unitary 2-design.[28] A set of unitaries $\mathcal{U} \subset \mathrm{SU}(2^n)$ forms a 2-design if, despite having access to two copies of a given unitary U, we cannot discriminate between the cases where U is

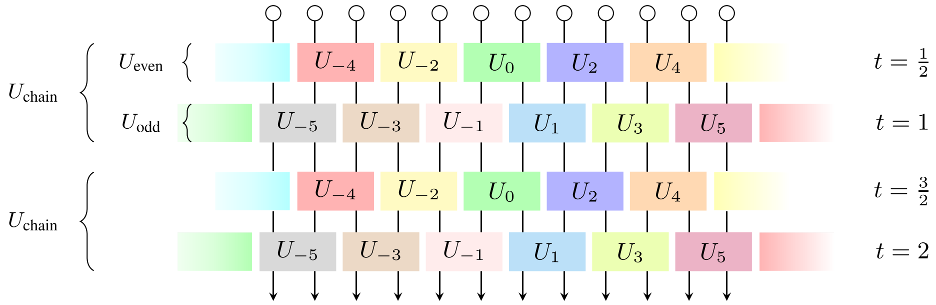


FIG. 1. Time-periodic local dynamics. The physical model analyzed in this work. The circles on top represent lattice sites, each consisting of *N* qubits. Colored blocks represent two-site unitaries, and different colors stand for independently and identically distributed Clifford unitaries, representing the spatial disorder. After the first two half time steps, the dynamics repeats itself.

sampled from $\mathcal{U}$ or from $\mathrm{SU}(2^n)$. Because of this, unitaries sampled from $\mathcal{U}$ are called pseudo-random. In our weaker variant of 2-design, we restrict the class of measurements available for this discrimination process to multi-qubit Pauli operators. To define our set $\mathcal{U}$, we consider a model with (spatial) disorder, where each element of $\mathcal{U}$ is the evolution operator $W(t)$ at a fixed time t generated by a particular configuration of the disorder [see Fig. 1 for $W(2)$].

In this work, we consider a spin chain with L sites and periodic boundary conditions, where each site contains N modes or qubits. The first dynamical period consists of two half steps. In the first half-step, each even site interacts with its right neighbor with a random Clifford unitary (for the definition of the Clifford group, see Appendix A or Ref. 29), and in the second half-step, each odd site interacts with its right neighbor with a random Clifford unitary. These L Clifford unitaries are independent and uniformly sampled from the $2N$-qubit Clifford group. The subsequent periods of the dynamics are repetitions of the first period, as illustrated in Fig. 1. If we denote by U_x the above-mentioned unitary action on sites x and $x+1$ (modulo L due to periodic boundary conditions), then the evolution operator after an *integer* time t is

$$\begin{aligned} W(t) &= \left[(U_1 \otimes U_3 \otimes \cdots \otimes U_{L-1})(U_0 \otimes U_2 \otimes \cdots \otimes U_{L-2})\right]^t \\ &= (U_{\text{odd}} U_{\text{even}})^t = (U_{\text{chain}})^t \end{aligned} \tag{1}$$

and the evolution operator after a *half-integer* time t is

$$W(t) = U_{\text{even}} \, (U_{\text{chain}})^{t-1/2}. \tag{2}$$

This evolution operator can also be generated by a time-periodic Hamiltonian $H(t)$ with nearest-neighbor interactions,

$$W(t) = \mathcal{T} e^{-\mathrm{i}\int_0^t d\tau H(\tau)}, \tag{3}$$

where $\mathcal{T}$ is the time-ordering operator. This type of dynamics is called Floquet. Floquet dynamics in relation with the phenomenon of quantum chaos has been studied, among others, by Prosen and co-workers,[18,23,24,30] with the review Ref. 13. A general review on quantum Floquet systems is given in Ref. 31. In the quantum information community, the term QCA, quantum cellular automaton, commonly denotes such periodic systems,[32] a review is given in Ref. 33. The authors of Ref. 34 considered a QCA, with the same structure as ours but with Haar-distributed unitaries gates instead of Cliffords. (Another Floquet–Clifford model was studied in Ref. 35.) It is important to stress that this time-periodic model is very different from the much more studied time-dependent “random circuits,”[36–42] depicted in Fig. 2. Time-periodic circuits are more difficult to analyze but more relevant to physics since they enjoy a (discrete) time translation symmetry.

We show that the ensemble of evolution operators at half-integer time (2) has a larger symmetry than that at integer times (1). This allows us to prove approximate Pauli mixing:[43] each Pauli operator evolving with the random dynamics (2) reaches any other Pauli operator inside its light cone with approximately equal probability. In the integer time case, this only holds for a restricted class of initial operators, which includes local ones. We also prove that at any half-integer time after the scrambling time t_{scr}, the ensemble of evolution operators (2) cannot be operationally distinguished from Haar-random unitaries (in the sense specified above). We define the scrambling time t_{scr} as the smallest time, allowing for any local perturbation to reach the entire system (in our model, $t_{\text{scr}} = L/4$). In all these results, the degree of approximation increases with N and decreases with time t.

Besides many-body physics, our results are relevant to the field of quantum information. The authors of Ref. 44 designed a protocol to generate pseudo-random unitaries. Many quantum information tasks make use of unitary designs (entanglement distillation,[45] quantum error correction,[39,46] randomized benchmarking,[47] quantum process tomography,[48] quantum state decoupling,[49] and data-hiding[50]). In most current implementations of quantum information, processing qubits are measured in a fixed basis, a particular case of our Pauli measurements. Hence, we expect that our variant of 2-designs restricted to Pauli measurements will be useful in some of these applications, in particular on architectures where a time-periodic drive is more feasible than a time-dependent drive. It is worth mentioning that Google’s quantum

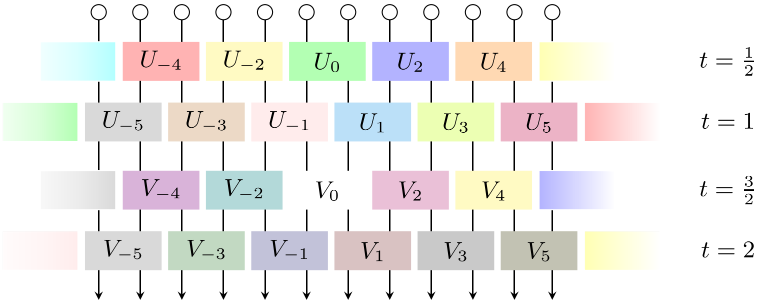

U_{-4} U_{-2} U_0 U_2 U_4 $t=\frac{1}{2}$
U_{-5} U_{-3} U_{-1} U_1 U_3 U_5 $t=1$
V_{-4} V_{-2} V_0 V_2 V_4 $t=\frac{3}{2}$
V_{-5} V_{-3} V_{-1} V_1 V_3 V_5 $t=2$

FIG. 2. Time-dependent local dynamics. In contrast to Fig. 1, the pictured circuit is not periodic in time: different time steps are independently sampled.

supremacy demonstration[51] is based on the statistics of multi-qubit Pauli measurements after pseudo-random unitary dynamics and that their random circuit consists of time-dependent single-qubit gates and time-periodic two-qubit gates.

In the model under consideration, the number of modes per site N is a free parameter that controls the behavior of the system. In the large-N regime ($N \gg \log L$), we obtain the above-mentioned indistinguishability between the evolution operator (2) and a Haar-random unitary, which increases with N. (We recall that large-N is the relevant regime in holographic quantum gravity.) In the opposite regime ($N \ll \log L$), our model displays a novel form of localization produced by the appearance of effective one-sided walls that prevent perturbations from crossing the wall in one direction but not the other. Interestingly, this localized phase seems to challenge the existing classification. On the one hand, our model is not a system of free or interacting particles, so it does not fit in the framework of Anderson localization. On the other hand, the evolution of each local operator is strictly confined to a finite region, so it does not behave as many-body localized. See Ref. 9 for the description of the differences among Anderson and many-body localization and Ref. 52 for a recent physicists' review on localization phenomena.

This model also challenges the classification of integrable and chaotic quantum systems. On the one hand, it has a phase space description of the dynamics such as that of quasi-free bosons and fermions, and it can be efficiently simulated on a classical computer.[53,54] See Ref. 29 for an algorithmic classification of the elements of the Clifford group and Appendix A for the description of the Clifford group phase space. On the other hand, this model does not have anything close to local (or low-weight) integrals of motions, and it behaves like Haar-random dynamics in a way that quasi-free systems do not. Therefore, we believe that Clifford dynamics is valuable for mapping the landscape of many-body phenomena. It is also important to recall that we live in the age of synthetic quantum matter; see Ref. 55, and models similar to ours have actually been implemented on quantum simulators, such as in the recent Google experiment.[51]

In Sec. II, we present our results on mixing (Secs. II B and II C), pseudo-random unitaries (Sec. II D), and strong localization (Sec. II E). In order to do so, we introduce a few mathematical notions beforehand (Sec. II A). The proofs of the theorems presented in Sec. II are presented in Secs. V, V B–V D, VI, and VII, together with several other lemmas. In Sec. VIII A, we discuss the physical significance of the scrambling time. In Sec. VIII B, we discuss the difficulties with classifying our model as integrable or chaotic. We compare time-periodic and time-independent circuits in Sec. VIII C. In Sec. IX, we provide the conclusions of our work. An introduction to the Clifford formalism, Appendix A, and some additional lemmas, Appendices B and C, are also included.

II. RESULTS

A. Preliminaries

An n-qubit *Pauli operator* is a tensor product of Pauli sigma matrices and one-qubit identities times a global phase $\lambda \in \{1, \mathrm{i}, -1, -\mathrm{i}\}$. In what follows, we ignore this global phase λ, so each Pauli operator is represented by a binary vector $\mathbf{u} = (q_1, p_1, q_2, p_2, \ldots, q_n, p_n) \in \{0,1\}^{2n}$ as

$$\sigma_{\mathbf{u}} = \bigotimes_{i=1}^{n} (\sigma_x^{q_i} \sigma_z^{p_i}). \tag{4}$$

Ignoring the global phase λ allows us to write the product of Pauli operators as the simple rule $\sigma_{\mathbf{u}}\sigma_{\mathbf{u}'} = \lambda\, \sigma_{\mathbf{u}+\mathbf{u}'}$, where addition in the vector space $\{0,1\}^{2n}$ is modulo 2. This defines the Pauli group, which is the discrete analog of the Weyl group, or the displacement operators used in quantum optics.

The n-qubit *Clifford* group $\mathcal{C}_n \subseteq \mathrm{SU}(2^n)$ is the set of unitaries U, which map each Pauli operator onto another Pauli operator $U\sigma_{\mathbf{u}}U^\dagger = \lambda\, \sigma_{\mathbf{u}'}$. Each Clifford unitary U can be represented by a $2n \times 2n$ symplectic matrix S with entries in $\{0,1\}$ such that its action on Pauli operators can be calculated in phase space,

$$U\sigma_{\mathbf{u}}U^\dagger = \lambda\, \sigma_{S\mathbf{u}}, \tag{5}$$

where the matrix product $S\mathbf{u}$ is defined modulo 2. We call the binary vectors $\mathbf{u} \in \{0,1\}^{2n}$ the *phase space* representation of the Pauli operator $\sigma_{\mathbf{u}}$ because of its analogy with quasi-free bosons, where dynamics is also linear and symplectic. A detailed introduction to the discrete phase space and Clifford and Pauli groups is provided in Appendix A; see also Refs. 29, 53, and 54. Note that Clifford unitaries are easy to implement in several quantum computation and simulation architectures.

Our model is an L-site lattice with even L and periodic boundary conditions. The corresponding phase space can be written as

$$\mathcal{V}_{\text{chain}} = \bigoplus_{x \in \mathbb{Z}_L} \mathcal{V}_x, \tag{6}$$

where $\mathcal{V}_x \cong \mathbb{Z}_2^{2N}$ is the phase space of site $x \in \mathbb{Z}_L$, which represents N qubits. A local Pauli operator $\sigma_{\mathbf{u}}$ at site x is represented by a phase space vector contained in the corresponding subspace $\mathbf{u} \in \mathcal{V}_x \subseteq \mathcal{V}_{\text{chain}}$. The identity operator corresponds to the zero vector. (See Sec. III for a collated description of the model and its phase space description). In the following, we will denote $S(t)$ as the symplectic matrix acting on the space $\mathcal{V}_{\text{chain}}$ associated with the evolution operator $W(t)$, as defined in Eqs. (1) and (2).

B. Approximate Pauli mixing

A set of unitaries $\mathcal{U}$ is Pauli mixing if a uniformly sampled unitary $U \in \mathcal{U}$ maps any non-identity Pauli operator $\sigma_{\mathbf{u}}$ to any other $\sigma_{\mathbf{u}'} = U\sigma_{\mathbf{u}}U^\dagger$ with uniform distribution.[43] Let us see that our model displays an approximate form of this property.

Each sequence of two-site Clifford unitaries $U_0, \ldots, U_{L-1}$ defines an evolution operator $W(t)$ via Eqs. (1) and (2), which maps each Pauli operator $\sigma_{\mathbf{u}}$ to another Pauli operator $\sigma_{\mathbf{u}'} = \lambda W(t)\sigma_{\mathbf{u}}W(t)^\dagger$. This deterministic map $\mathbf{u} \mapsto \mathbf{u}'$ becomes probabilistic when we let $U_0, \ldots, U_{L-1}$ be random. In this case, the probability that $\mathbf{u}$ evolves onto $\mathbf{u}'$ after a time t is

$$P_t(\mathbf{u}'|\mathbf{u}) = \mathop{\mathbb{E}}_{\{U_x\}} \left|2^{-NL}\mathrm{tr}\Big(\sigma_{\mathbf{u}'} W(t)\sigma_{\mathbf{u}} W(t)^\dagger\Big)\right|, \tag{7}$$

where we use the orthogonality of Pauli matrices $\mathrm{tr}(\sigma_{\mathbf{u}'}\sigma_{\mathbf{u}}) = 2^{NL}\delta_{\mathbf{u}'\mathbf{u}}$. The locality of the dynamics (see Fig. 1) implies that only operators inside the light cone of the initial operator $\sigma_{\mathbf{u}}$ have non-zero probability (7). For example, if the initial operator $\sigma_{\mathbf{u}}$ is supported at the origin $x = 0$, then after a time t, the evolved operator $\sigma_{\mathbf{u}'}$ must be fully supported inside the light cone $-2t + 1 \le x \le 2t$. This means that the corresponding phase space vector $\mathbf{u}'$ is in the causal subspace,

$$\mathbf{u}' \in \bigoplus_{x \in [-2t+1, 2t]} \mathcal{V}_x. \tag{8}$$

The time t at which the causal subspace becomes the whole system is the scrambling time $t_{\text{scr}} = L/4$. Section VIII A discusses the physical significance of this time scale.

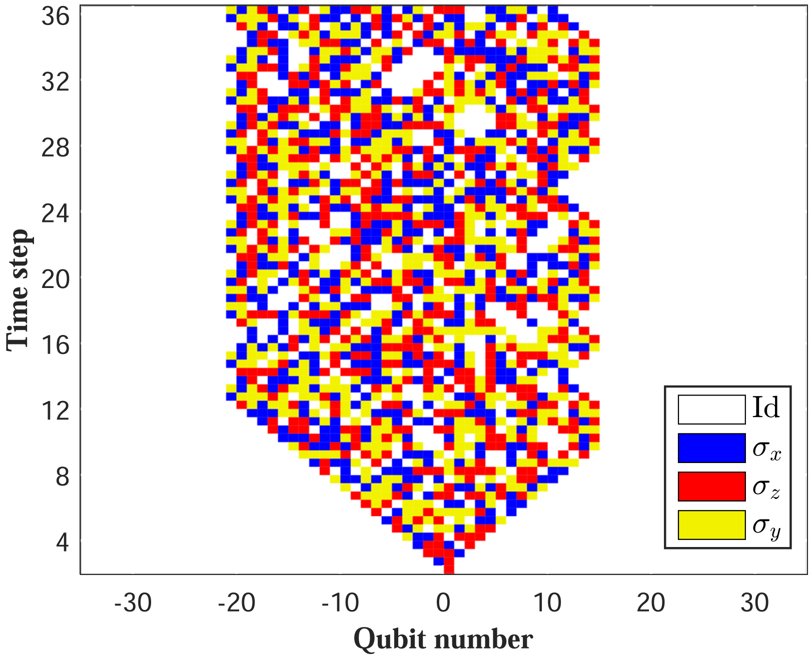


FIG. 3. Strong localization. The Heisenberg evolution of the initial operator σ_z at site $x = 1$. Each lattice site consists of one qubit ($N = 1$) with first-neighbor interactions. After a phase of linear growth, the lateral wings collide with left- and right-sided walls with penetration length $l = 1$, which confine the evolution for all times. This confinement affects all (not necessarily local or Pauli) operators between the two walls. Inside the confined region, evolution seems to be mixing.

When the distribution (7) is approximately uniform inside the light cone, we say that the random dynamics displays approximate Pauli mixing. Let $Q_t(\mathbf{u}')$ denote the uniform distribution over all non-zero vectors $\mathbf{u}'$ in the causal subspace (8); therefore, after the scrambling time $t \geq t_{\text{scr}}$, $Q_t(\mathbf{u}')$ is the uniform distribution over all non-zero vectors in the total phase space $\mathcal{V}_{\text{chain}}$. The following theorem from Sec. V C proves approximate Pauli mixing for initially local operators.

Theorem 18 (approximate Pauli mixing). If the initial Pauli operator $\sigma_{\mathbf{u}}$ is supported at site $x = 0$, then the probability distribution (7) for its evolution $\sigma_{\mathbf{u}'}$ is close to uniform inside the light cone, that is,

$$\sum_{\mathbf{u}'} \left| P_t(\mathbf{u}'|\mathbf{u}) - Q_t(\mathbf{u}') \right| \ \leq \ 130 \times t^2\, 2^{-N} \tag{9}$$

for any integer or half-integer time $t \in [1/2, 2t_{\text{scr}}]$. An analogous statement holds for any other initial location $x \neq 0$.

The above bound is useful in the large-N limit ($N \gg \log L$). In the opposite regime ($N \ll \log L$), mixing cannot take place since the system displays a strong form of localization, in which local operators are mapped onto quasi-local operators. This phenomenon is illustrated in Fig. 3 and detailed in Sec. II E.

The upper bound (9) increases with time due to time correlations and dynamical recurrences (see Sec. VIII A). Hence, as time goes on, the character of the system is less mixing, which is the opposite of what happens in time-dependent dynamics (see Sec. VIII C). Also note that at integer times t, our model can be considered to be time-independent (instead of time-periodic) with discrete time.

The above mixing result only applies to local initial operators. Next, we present a different result that applies to a large class of non-local initial operators. However, due to the complexity of the problem, we only analyze their evolution inside a region $\mathcal{R} = \{1, \ldots, L_{\text{s}}\} \subset \mathbb{Z}_L$.

Theorem 20. Consider an initial vector $\mathbf{u}^0 \in \mathcal{V}_{\text{chain}}$ with non-zero support in all lattice sites ($\mathbf{u}_x^0 \neq \mathbf{0}$ for all $x \in \mathbb{Z}_L$). Consider the evolved vector $\mathbf{u}^t = S(t)\mathbf{u}^0$ inside a region $x \in \{1, \ldots, L_{\text{s}}\} \subseteq \mathbb{Z}_L$, where L_{s} is even and the time is $t \leq \frac{L-L_{\text{s}}}{4}$. If $\mathbf{u}^t_{[1,Ls]}$ is the projection of $\mathbf{u}^t$ in the subspace $\oplus_{x=1}^{L_{\text{s}}} \mathcal{V}_x$, then

$$\sum_{\mathbf{v} \in \mathbb{Z}_2^{2NL_{\text{s}}}} \left| \text{prob}\{\mathbf{v} = \mathbf{u}^t_{[1,Ls]}\} - \frac{1}{2^{2NL_{\text{s}}}} \right| \leq 32\, t\, 2^{-N} \left(2L_{\text{s}} + 3^{\frac{L_{\text{s}}}{2}+1} \right) + 4L 2^{-2N}. \tag{10}$$

C. Pauli mixing at half-integer time

The evolution operator of our model $W(t)$ has an extra symmetry at half-integer time t (see Sec V A). This allows us to prove a mixing result that is stronger than Theorem 23. Specifically, the following theorem (from Sec. V B) applies to any initial Pauli operator instead of only local ones.

Theorem 14. Let $\sigma_{\mathbf{u}'} = \lambda W(t) \sigma_{\mathbf{u}} W(t)^\dagger$ be the evolution of any initial Pauli operator $\sigma_{\mathbf{u}} \neq \mathbb{1}$. At any half-integer time t larger than the scrambling time, in the interval $t \in [t_{\text{scr}}, 2t_{\text{scr}}]$, the probability distribution (7) for the evolved operator $\sigma_{\mathbf{u}'}$ is close to uniform, that is,

$$\sum_{\mathbf{u}'} \left| P_t(\mathbf{u}'|\mathbf{u}) - Q_t(\mathbf{u}') \right| \ \leq \ 33 \times t\, L\, 2^{-N}. \tag{11}$$

The fact that mixing is more prominent at half-integer multiples of the period is not restricted to Clifford dynamics since it applies to a large class of periodic random quantum circuit or Floquet dynamics with disorder. In particular, it holds in any circuit where the two-site random interaction U_x includes one-site random gates V_x that are a 1-design. That is, when the random variable U_x follows the same statistics than the random variable, $U_x' = U_x(V_x \otimes V_{x+1})$. This fact could be useful for implementing pseudo-random unitaries in quantum circuits with a periodic driving.

D. Pseudo-random unitaries

In this section, we prove a consequence of the previous result: the evolution operator $W(t)$ at half-integer times t is hard to physically distinguish from a Haar-random unitary $U \in \text{SU}(2^{NL})$ when the available measurements are Pauli operators. More precisely, imagine that it is given a unitary transformation V, which has been sampled from either the set of evolution operators $\{W(t)\}$ or the full unitary group $\text{SU}(2^{NL})$. The task is to choose a state ρ, process it with the given transformation $\rho \mapsto V\rho V^\dagger$, measure the result with a Pauli operator $\sigma_{\mathbf{u}}$, and guess whether V has been sampled from the set of evolution operators $\{W(t)\}$ or from the full unitary group $\text{SU}(2^{NL})$. In order to sharpen this discrimination procedure, two uses of the transformation V are permitted, which allows for feeding each of them with half of an entangled state ρ (describing two copies of the system). The following result tells us that, in the large-N limit, the optimal guessing probability for the above task is almost as good as a random guess. (Recall that a random guess gives $p_{\text{guess}} = 1/2$). The proof is given in Sec. VI.

Theorem 21. Consider the task of discriminating between two copies of $W(t)$ and two copies of a Haar-random unitary U with measurements restricted to Pauli operators, when $t \in [t_{\text{scr}}, 2t_{\text{scr}}]$ is half-integer. The success probability for correctly guessing the given pair of unitaries satisfies

$$\begin{aligned} p_{\text{guess}} &= \frac{1}{2} + \frac{1}{4} \max_{\rho,\mathbf{u},\mathbf{v}} \operatorname{tr}\!\left(\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{v}} \left[\mathop{\mathbb{E}}_{W(t)} W(t)^{\otimes 2} \rho\, W(t)^{\otimes 2\dagger} - \int_{SU(d)} dU\, U^{\otimes 2} \rho\, U^{\otimes 2\dagger} \right]\right) \\ &\leq 1/2 + 9\, tL2^{-N}. \end{aligned} \tag{12}$$

The proof that the optimal guessing probability is given by formula (12) can be found in Ref. 56. If in Theorem 21, measurements were not restricted, then $W(t)$ would be an $(8tL2^{-N})$-approximate unitary 2-design. The precise definition of approximate 2-design allows for using an ancillary system in the discrimination process.[28] However, we have not included this ancillary system in Theorem 21 because it does not provide any advantage.

E. Strong localization ($N \ll \log L$)

The model under consideration has the property that certain combinations of gates in consecutive sites (e.g., $U_x, U_{x+1}, \ldots, U_{x+l}$) generate right- or left-sided walls. These are defined as follows: a right-sided wall at site x stops the growth toward the right of any operator that arrives at x from the left, but it does not necessarily stop the growth toward the left of any operator that arrives at x from the right. The analogous thing happens for left-sided walls (see Fig. 3).

These gate configurations have a non-zero probability; hence, they will appear in a sufficiently long chain with a typical circuit. Below, we provide bounds to this probability. The inverse of this probability is the average distance between walls, which can be understood as the *localization length scale*, and it quantifies the width of the light cones, displayed in Fig. 3.

Each one-sided wall has some penetration length $l \geq 1$ into the forbidden region. Suppose that a realization of U_{chain} contains a right-sided wall at site $x = 0$ with penetration length l. Then, any operator with support on the sites $x \leq 0$ (and identity on $x > 0$) is mapped by $(U_{\text{chain}})^t$ to an operator with support on $x \leq l$ contained in a specific subspace within the interval $x \in [1, l]$ such that entering into region $x > l$ is impossible for all $t \geq 1$. [The restriction to this subspace within the forbidden region can be seen in Fig. 3 (with $l = 1$) by the fact that the right-most points are either yellow followed by red or white followed by white. In addition, the left-most points are either blue followed by yellow or white followed by white.] An initial operator with support on the interval $x \in [1, l]$ that does not have the specific structure mentioned above can pass through and reach the side $x > l$.

Now, let us characterize the pairs of gates U_0, U_1 (which act on sites $\{0,1\}$ and $\{1,2\}$, respectively) that generate a right-sided wall at $x = 0$ with penetration length $l \leq 1$. Let S_0, S_1 be the phase space representation of U_0, U_1. Next, we use the fact that in phase space, subsystems decompose with the direct sum (not the tensor product) rule, which allows us to decompose S_0, S_1 into $2N$-dimensional blocks,

$$S_x = \begin{pmatrix} A_x & B_x \\ C_x & D_x \end{pmatrix}. \tag{13}$$

The flow of information caused by S_x is easily seen by the action of S_x on the vector $(\mathbf{u}_x, \mathbf{u}_{x+1})^T$,

$$\begin{pmatrix} A_x & B_x \\ C_x & D_x \end{pmatrix} \begin{pmatrix} \mathbf{u}_x \\ \mathbf{u}_{x+1} \end{pmatrix} = \begin{pmatrix} A_x \mathbf{u}_x + B_x \mathbf{u}_{x+1} \\ C_x \mathbf{u}_x + D_x \mathbf{u}_{x+1} \end{pmatrix}. \tag{14}$$

Block A_0 (D_0) represents the local dynamics at site $x = 0$ $(x = 1)$ in the first half step. Block C_0 represents the flow from $x = 0$ to $x = 1$ in the first half step, and block C_1 represents the flow from $x = 1$ to $x = 2$ in the second half step.

Imposing that nothing arrives at $x = 2$ after the first whole step amounts to $C_1 C_0 = 0$. This is illustrated in Fig. 4. Imposing that nothing arrives at $x = 2$ after the first two whole steps amounts to

$$C_1 D_0 A_1 C_0 = 0 \ \text{ and } \ C_1 C_0 = 0. \tag{15}$$

This is illustrated in Fig. 5. Finally, imposing that nothing arrives at $x = 2$ after any number $(t+1)$ of whole steps amounts to

$$C_1 (D_0 A_1)^t C_0 = 0 \tag{16}$$

for all integers $t \geq 0$. However, it is proven in Lemma 24 (Sec. VII) that this infinite family of conditions (16) is implied by the cases $t = 0, 1, \ldots, (2^{4N} - 1)$. In addition, for the simplest case $N = 1$, Theorem 25 shows that all conditions (16) follow from the two conditions (15).

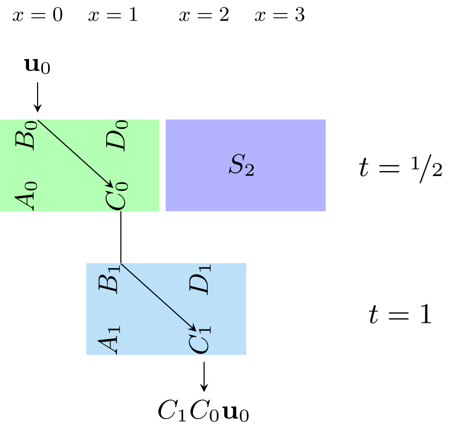


FIG. 4. The flow of information starting from $x = 0$ at $t = 0$ shows that with $C_1C_0 = 0$, there is no information reaching $x = 2$ at $t = 1$.

Equations (15) and (16) can be understood as characterizing a pattern of destructive interference due to disorder that causes localization. The following pair of Clifford unitaries U_0, U_1:

$$U_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} i & 0 & 0 & -i \\ 0 & i & -i & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \quad U_1 = \frac{1}{2}\begin{pmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{pmatrix}, \tag{17}$$

has a phase space representation,

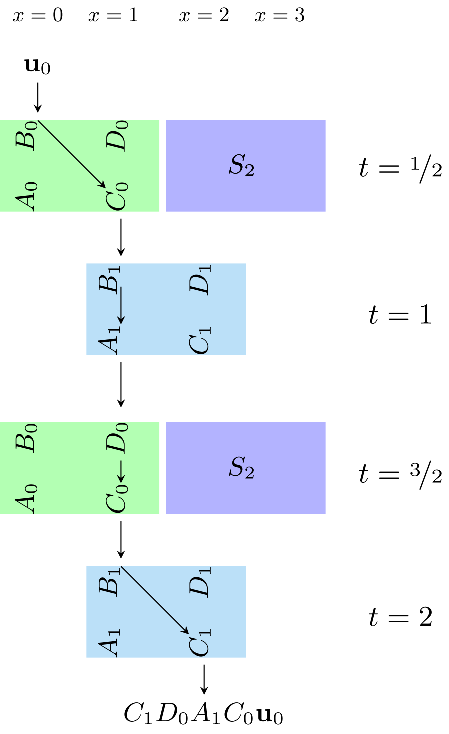


FIG. 5. As illustrated in Fig. 4, the condition $C_1C_0 = 0$ stops the information flow from $x = 0$ at $t = 0$ to $x = 2$ at $t = 1$. To also prevent this from happening at $t = 2$, we demand that $C_1D_0A_1C_0 = 0$.

$$S_0 = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_1 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{18}$$

It can be checked that this pair of matrices satisfies conditions (15), which implies (16).

The following theorem provides the exact value of the probability for the appearance of a one-sided wall with penetration length $l = 1$ in the case $N = 1$.

Theorem 25. For $N = 1$, the conditions

$$C_{x+1}(D_x A_{x+1})^k C_x = 0 \tag{19}$$

for $k \in \{0, 1, 2, \ldots\}$ are implied by the following two conditions:

$$C_{x+1} C_x = 0 \quad \text{and} \quad C_{x+1} D_x A_{x+1} C_x = 0. \tag{20}$$

Furthermore, the probability of this is given exactly by

$$\text{prob}\{C_{x+1} C_x = 0, C_{x+1} D_x A_{x+1} C_x = 0\} = 0.12, \tag{21}$$

which includes trivial localization.

The probability given in Eq. (21) is obtained numerically. It is worth mentioning that this model also displays walls with zero penetration length $(l = 0)$, which are necessarily two-sided. These walls happen when a two-site gate U_x is of the product form $U_x = V_x \otimes V_{x+1}$. This prevents the interaction between the two sides of the gate, and hence, it produces a trivial type of localization. The following theorem (proven in Sec. VII) shows that the probability of these trivial walls is very small.

Theorem 23. The probability that a Clifford unitary $U \in \mathcal{C}_{2N}$ is of the product form is

$$\frac{1}{2}\, 2^{-4N^2} \le \text{prob}\{U \text{ is product}\} \le 2^{-4N^2}. \tag{22}$$

We expect that $(l = 0)$-walls are much less likely, except in the case $N = 1$, than $(l \ge 1)$-walls. This would allow for a regime of (L, N) where the system displays non-trivial localization.

F. Absence of localization ($N \gg \log L$)

The following theorem provides an upper bound for the probability that one-sided walls appear at a particular location. This upper bound implies that when $N \gg \log L$, a typical circuit has no localization.

Theorem 27. The conditions

$$C_{x+1}(D_x A_{x+1})^k C_x = 0 \quad \text{for all} \quad k \in \{0, 1, 2, \ldots\} \tag{23}$$

are sufficient to prevent all right-ward propagation past position x at any time. The probability that this family of constrains holds is upper-bounded by

$$\begin{aligned} &\text{prob}\{C_{x+1}(D_x A_{x+1})^k C_x = 0,\ \forall k \in \mathbb{N}\} \\ &\le\ \text{prob}\{C_{x+1} C_x = 0\}\ \le\ \frac{2N + 1}{(1 - 2^{-2N})^{2N}}\, 2^{2N - 2N^2}. \end{aligned} \tag{24}$$

By symmetry, left-sided walls have the same probabilities.

If the system is finite $(L < \infty)$, a sufficiently large N will eliminate the presence of localization in most realizations of the dynamics U_{chain}. This fact is crucial in the mixing results of Theorems 18, 20, and 21. Our previous results showing the mixing property in the regime $N \gg \log L$ suggest that in this regime, the probability that the whole system has a wall of any type vanishes.

III. DESCRIPTION OF THE MODEL

In this section, we further specify the model analyzed in this work.

A. Locality, time-periodicity, and disorder

Consider a spin chain with an even number L of sites and periodic boundary conditions. Each site is labeled by $x \in \mathbb{Z}_L$ and contains N qubits (Clifford modes), so the Hilbert space of each site has dimension 2^N. The dynamics of the chain is discrete in time, and hence, it is characterized by a unitary U_{chain}, not a Hamiltonian. Locality is imposed by the fact that U_{chain} is generated by first-neighbor interactions in the following way:

$$U_{\text{chain}} = \left(\bigotimes\nolimits_{x\,\text{odd}} U_x\right)\left(\bigotimes\nolimits_{x\,\text{even}} U_x\right), \tag{25}$$

where the unitary U_x only acts on sites x and $x+1$ (mod L is understood). Expression (25) tells us that each time step decomposes into two half steps: in the first half, each even site interacts with its right neighbor, and in the second half, each even site interacts with its left neighbor. This is illustrated in Fig. 1.

We define the evolution operator at integer and half-integer times $t \in \mathbb{Z}/2$ in the following way:

$$W(t) = \begin{cases} (U_{\text{chain}})^t & \text{integer } t, \\ \left(\bigotimes\nolimits_{x\,\text{even}} U_x\right)(U_{\text{chain}})^{t-1/2} & \text{half} - \text{integer } t. \end{cases} \tag{26}$$

We understand that t is half-integer when $t - 1/2 \in \mathbb{Z}$.

Translation invariance amounts to imposing that all U_x with even x are identical and all U_x with odd x are identical too. However, in this work, we are interested in disordered systems, where the translation invariance is broken. In fact, here, we break the translation invariance in the strongest possible form since each two-site unitary U_x is independently sampled from the uniform distribution over the Clifford group.

B. Phase space description

The phase space of the whole chain is

$$\mathcal{V}_{\text{chain}} = \bigoplus\nolimits_x \mathcal{V}_x \cong \mathbb{Z}_2^{2NL}, \tag{27}$$

where $\mathcal{V}_x \cong \mathbb{Z}_2^{2N}$ is the phase space of site x. The phase space representation of U_x is the symplectic matrix $S_x \in \mathcal{S}_{2N}$, where S_x acts on the subspace $\mathcal{V}_x \oplus \mathcal{V}_{x+1}$. Using this direct-sum decomposition, we can write

$$S_x = \begin{pmatrix} A_x & B_x \\ C_x & D_x \end{pmatrix}, \tag{28}$$

where A_x, B_x, C_x, D_x are $2N \times 2N$ matrices, with $A_x : \mathcal{V}_x \to \mathcal{V}_x$, $B_x : \mathcal{V}_{x+1} \to \mathcal{V}_x$, $C_x : \mathcal{V}_x \to \mathcal{V}_{x+1}$, and $D_x : \mathcal{V}_{x+1} \to \mathcal{V}_{x+1}$. The phase space representation of U_{chain} given in (25) is

$$S_{\text{chain}} = \left(\bigoplus\nolimits_{x\,\text{odd}} S_x\right)\left(\bigoplus\nolimits_{x\,\text{even}} S_x\right). \tag{29}$$

Note that the tensor product becomes a direct sum, in analogy with the quantum optics formalism. Using the single-site decomposition (27) and (28), we can write the two half steps in (29) as

$$\bigoplus_{x\,\text{even}} \mathcal{S}_x = \begin{pmatrix} A_0 & B_0 & & & & & \\ C_0 & D_0 & & & & & \\ & & A_2 & B_2 & & & \\ & & C_2 & D_2 & & & \\ & & & & \ddots & & \\ & & & & & A_{L-2} & B_{L-2} \\ & & & & & C_{L-2} & D_{L-2} \end{pmatrix}, \tag{30}$$

$$\bigoplus_{x \text{ odd}} \mathcal{S}_x = \begin{pmatrix} D_{L-1} & & & & & & C_{L-1} \\ & A_1 & B_1 & & & & \\ & C_1 & D_1 & & & & \\ & & & \ddots & & & \\ & & & & A_{L-3} & B_{L-3} & \\ & & & & C_{L-3} & D_{L-3} & \\ B_{L-1} & & & & & & A_{L-1} \end{pmatrix}, \tag{31}$$

where the blank spaces represent blocks with zeros. The "phase space evolution operator" is then

$$S(t) = \begin{cases} (\mathcal{S}_{\text{chain}})^t & \text{integer } t, \\ (\bigoplus_{x \text{ even}} \mathcal{S}_x)(\mathcal{S}_{\text{chain}})^{t-1/2} & \text{half}-\text{integer } t. \end{cases} \tag{32}$$

IV. RANDOM SYMPLECTIC MATRICES

In this section, we discuss results relating to (uniformly) random symplectic matrices. These results will then be used in Sec. V to show that the random local circuit model we consider approximately satisfies the requirement of Pauli mixing: it maps any initial Pauli operator to the uniform distribution over all Pauli operators.

An equivalent way to write the symplectic condition $S^T J S = J$ is that the columns of the matrix $S = (\mathbf{u}_1, \mathbf{v}_1, \mathbf{u}_2, \mathbf{v}_2, \ldots, \mathbf{u}_n, \mathbf{v}_n)$ satisfy

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0, \tag{33}$$

$$\langle \mathbf{u}_i, \mathbf{v}_j \rangle = \delta_{ij}. \tag{34}$$

We recall the notation $\langle \mathbf{r}, \mathbf{s} \rangle \equiv \mathbf{r}^T J \mathbf{s}$ for all $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{Z}_2^{2n}$. Using this, we can uniformly sample from $\mathcal{S}_n$ by sequentially generating the columns of S.

Lemma 1. The following algorithm allows us to uniformly sample from the symplectic group $\mathcal{S}_n$.

1. Generate $\mathbf{u}_1$ by picking any of the $(2^{2n} - 1)$ non-zero vectors in $\mathbb{Z}_2^{2n}$.
2. Generate $\mathbf{v}_1$ by picking any of the 2^{2n-1} vectors satisfying $\langle \mathbf{u}_1, \mathbf{v}_1 \rangle = 1$.
3. Generate $\mathbf{u}_2$ by picking any of the $(2^{2n-2} - 1)$ non-zero vectors satisfying $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \langle \mathbf{v}_1, \mathbf{u}_2 \rangle = 0$.
4. Generate $\mathbf{v}_2$ by picking any of the 2^{2n-3} vectors satisfying $\langle \mathbf{u}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$ and $\langle \mathbf{u}_2, \mathbf{v}_2 \rangle = 1$.
5. Continue generating $\mathbf{u}_3, \mathbf{v}_3, \ldots, \mathbf{u}_n, \mathbf{v}_n$ in an analogous fashion, completing the matrix $S = (\mathbf{u}_1, \mathbf{v}_1, \mathbf{u}_2, \mathbf{v}_2, \ldots, \mathbf{u}_n, \mathbf{v}_n)$.

Proof. We first discuss the number of vectors $(\mathbf{u}_1, \mathbf{v}_1, \mathbf{u}_2, \mathbf{v}_2, \ldots, \mathbf{u}_n, \mathbf{v}_n)$, as stated above, that ensures S symplectic.

$\mathbf{v}_1$ has $2n$ components, since there is one constraint, the number of independent components is $2n - 1$, and each component belongs to $\mathbb{Z}_2$; therefore, the number of vectors $\mathbf{v}_1$ equals 2^{2n-1}. Note that since $\mathbf{u}_1$ is non-vanishing, a vanishing $\mathbf{v}_1$ cannot solve $\langle \mathbf{u}_1, \mathbf{v}_1 \rangle = 1$.

$\mathbf{u}_2$ must satisfy two constraints therefore the number of independent components is $2n - 2$ and there are 2^{2n-2} solutions. This includes also the case that $\mathbf{u}_2$ is vanishing, but since S must be full rank we need to exclude it, therefore there are $2^{2n-2} - 1$ admissible $\mathbf{u}_2$ vectors and so on.

We now prove that the distribution of simplectic matrices S generated with the algorithm is uniform. Let us see that the number of symplectic matrices whose first column is the non-zero vector $\mathbf{u}_1$ is independent of $\mathbf{u}_1$.

1. The number of vectors $\mathbf{v}_1$ satisfying $\langle \mathbf{u}_1, \mathbf{v}_1 \rangle = 1$ is independent of which non-zero $\mathbf{u}_1$ we choose.
2. The number of vectors $\mathbf{u}_2$ satisfying $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = 1$ and $\langle \mathbf{u}_2, \mathbf{v}_1 \rangle = 0$ is independent of the pair $\mathbf{u}_1, \mathbf{v}_1$ (being both non-zero and $\langle \mathbf{u}_1, \mathbf{v}_1 \rangle = 1$) that we choose.
3. And analogously for $\mathbf{v}_2, \mathbf{u}_3$, and so on.

This shows that the number of matrices having a fixed first column $\mathbf{u}_1$ is independent of $\mathbf{u}_1$. Therefore, all first columns $\mathbf{u}_1$ need to have the same probability. Using steps 1, 2, and 3 in a similar fashion, we can analogously conclude that all second columns $\mathbf{v}_1$ need to have the same probability. In addition, analogously, all vectors for column k (compatible with columns $1, 2, \ldots, k - 1$) need to have the same probability. This shows the uniformity provided by the sampling algorithm of Lemma 6. □

To obtain the above numbers, we use the fact that when $\langle \mathbf{u}, \mathbf{v} \rangle = 1$, both $\mathbf{u}$ and $\mathbf{v}$ are non-zero. From these same numbers, the next result follows.

Lemma 2. The order of the symplectic group is

$$|\mathcal{S}_n| = (2^{2n} - 1)2^{2n-1}(2^{2n-2} - 1)2^{2n-3} \cdots (2^2 - 1)2^1, \tag{35}$$

and it satisfies

$$a(n)\,2^{2n^2+n} \;\le\; |\mathcal{S}_n| \;\le\; b(n)\,2^{2n^2+n} \tag{36}$$

with $0.64 < a(n) < b(n) < 0.78$.

The proof of (35) is a classic result to be found, for example, in Ref. 57, and it also directly follows from Lemma 1. The proof of Eq. (36) is in Appendix B. Finally, the next lemma shows that uniformly distributed symplectic matrices have random outputs.

Lemma 3 (uniform output). If $S \in \mathcal{S}_n$ is uniformly distributed, then for any pair of non-zero vectors $\mathbf{u}, \mathbf{u}' \in \mathbb{Z}_2^{2n}$, we have

$$\mathrm{prob}\{\mathbf{u}' = S\mathbf{u}\} = (2^{2n} - 1)^{-1}. \tag{37}$$

Proof. Let us first consider the case $\mathbf{u} = (1, 0, \ldots, 0)^T$. If we follow the algorithm of Lemma 1, then the image of $(1, 0, \ldots, 0)^T$ is uniformly distributed over the $(2^{2n} - 1)$ non-zero vectors, and hence, it follows (37). To show (37) for any given $\mathbf{u}$, take any $S_0 \in \mathcal{S}_n$ such that $S_0\mathbf{u} = (1, 0, \ldots, 0)^T$, and note that if S is uniformly distributed, then so is SS_0. □

A. Rank of sub-matrices of S

Lemma 4. Any given $S \in \mathcal{S}_{2n}$ can be written in the block form

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{38}$$

according to the local decomposition $\mathbb{Z}_2^{4n} = \mathbb{Z}_2^{2n} \oplus \mathbb{Z}_2^{2n}$. If S is uniformly distributed, this then induces a distribution on the sub-matrices A, B, C, D. For each of them $(E = A, B, C, D)$, the induced distribution satisfies

$$\mathrm{prob}\{\mathrm{rank}\, E \le 2n - k\} \;\le\; \min\{2^k, 4\}\,\frac{2^{-k^2}}{(1 - 2^{-2n})^k} \approx 4 \times 2^{-k^2}. \tag{39}$$

Proof. We proceed by studying the rank of C and later generalizing the results to A, B, D. Equation (39) is trivial for $k = 0$, so in what follows, we assume that $k \ge 1$. Let us start by counting the number of matrices $S \in \mathcal{S}_{2n}$ with a sub-matrix C satisfying $C\mathbf{u} = \mathbf{0}$ for a given (arbitrary) non-zero vector $\mathbf{u} \in \mathbb{Z}_2^{2n}$. Let r denote the position of the last "1" in $\mathbf{u}$ so that it can be written as follows:

$$\mathbf{u} = \left(\underbrace{\mathbf{u}^1, \ldots, \mathbf{u}^{r-1}}_{r-1}, 1, \underbrace{0, \ldots, 0}_{2n-r}\right)^T, \tag{40}$$

where $\mathbf{u}^1, \ldots, \mathbf{u}^{r-1} \in \{0, 1\}$. Then, the constraint $C\mathbf{u} = \mathbf{0}$ can be written as

$$\begin{cases} C_{i,1} = 0 \text{ if } r = 1, \text{ with } 1 \le i \le 2n, \\ C_{i,r} = \sum_{j=1}^{r-1} C_{i,j}\, \mathbf{u}^j \text{ if } r > 1, \text{ with } 1 \le i \le 2n, \end{cases} \tag{41}$$

where $C_{i,j}$ are the components of C. (41) reads as a constraint on the rth column of the matrix C.

Next, we follow the algorithm introduced in Lemma 1 for generating a matrix $S \in \mathcal{S}_{2n}$ column by column, from left to right, and in addition to the symplectic constraints we include (41). Constraint (41) can be imposed by ignoring it during the generation of columns $1, \ldots, r - 1$, completely fixing the rows $2n < i \le 4n$ of the r column, that corresponds to the rth column of the matrix C and again ignoring it during the generation of columns $r + 1, \ldots, 4n$. By counting as in Lemma 2, we obtain that the number of matrices $S \in \mathcal{S}_{2n}$ satisfying $C\mathbf{u} = \mathbf{0}$ is as follows:

$$\begin{aligned} &|\{S \in \mathcal{S}_{2n} : C\mathbf{u} = \mathbf{0}\}| \\ &\le \begin{cases} (2^{4n} - 1)(2^{4n-1}) \cdots (2^{4n-(r-2)} - 1)2^{2n-(r-1)}(2^{4n-r} - 1) \cdots 2^1, \text{ r even,} \\ (2^{4n} - 1)(2^{4n-1}) \cdots 2^{4n-(r-2)} 2^{2n-(r-1)} 2^{4n-r} \cdots 2^1, \text{ r odd.} \end{cases} \end{aligned} \tag{42}$$

Equation (42) is an inequality because, for some values of the first $r - 1$ columns of S and the rth column of C, it is impossible to complete the rth column of A satisfying the symplectic constraints (33) and (34).

The probability that a random S satisfies $C\mathbf{u} = \mathbf{0}$ is

$$\mathrm{prob}\{C\mathbf{u} = \mathbf{0}\} = \frac{|\{S \in \mathcal{S}_{2n} : C\mathbf{u} = \mathbf{0}\}|}{|\mathcal{S}_{2n}|}. \tag{43}$$

By noting that all factors in (42) are the same as in (35) except for the factor at position r, we obtain

$$\begin{aligned} \mathrm{prob}\{C\mathbf{u} = \mathbf{0}\} &\le \frac{2^{2n-(r-1)}}{2^{4n-(r-1)} - \alpha'} \le \frac{2^{2n-(r-1)}}{2^{4n-(r-1)} - 1} \\ &= \frac{2^{-2n}}{1 - 2^{(r-1)-4n}} \le \frac{2^{-2n}}{1 - 2^{-2n}}, \end{aligned} \tag{44}$$

where $\alpha' = 1$ if r is odd and $\alpha' = 0$ otherwise. The last inequality above follows from $r \le 2n$. Bound (44) is also correct for $r = 1$. The fact that bound (44) is independent of r is crucial for the rest of the proof.

Next, we generalize bound (44) to the case where $C\mathbf{u}_i = 0$ for k given linearly independent vectors $\mathbf{u}_i \in \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. To do this, we take the $2n \times k$ matrix $[\mathbf{u}_1, \ldots, \mathbf{u}_k]$ and perform Gauss–Jordan elimination, operating on the columns, to obtain a matrix $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ having a column-echelon form. $\{C\mathbf{u}_1 = \mathbf{0}, \ldots, C\mathbf{u}_k = \mathbf{0}\}$ is equivalent to $\{C\mathbf{v}_1 = \mathbf{0}, \ldots, C\mathbf{v}_k = \mathbf{0}\}$, in fact only two operations are performed on the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ to obtain the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$: changing the order of the vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and replacing a vector $\mathbf{u}_j$ with the sum of $\mathbf{u}_j$ with another vector $\mathbf{u}_l$. If we denote by r_i the position of the last "1" of $\mathbf{v}_i$, then the column-echelon form amounts to $r_1 < r_2 < \cdots < r_k$. Now, we proceed as above to generate each column of S satisfying the symplectic and the $C\mathbf{v}_i = \mathbf{0}$ constraints. This gives

$$\mathrm{prob}\{C\mathbf{u}_1 = \mathbf{0}, \ldots, C\mathbf{u}_k = \mathbf{0}\} \le \frac{2^{2n-(r_1-1)}}{2^{4n-(r_1-1)} - \alpha'_1} \frac{2^{2n-(r_2-1)}}{2^{4n-(r_2-1)} - \alpha'_2} \cdots \frac{2^{2n-(r_k-1)}}{2^{4n-(r_k-1)} - \alpha'_k}, \tag{45}$$

where $\alpha'_i \in \{0, 1\}$. Similarly, as in (44), we obtain

$$\mathrm{prob}\{C\mathbf{u}_1 = \mathbf{0}, \ldots, C\mathbf{u}_k = \mathbf{0}\} \le \frac{2^{-2nk}}{(1 - 2^{-2n})^k}. \tag{46}$$

If we multiply the above bound by the number $\mathcal{N}_k^{2n}$ of k-dimensional subspaces of $\mathbb{Z}_2^{2n}$ (see Appendix C), then we obtain

$$\begin{aligned} \mathrm{prob}\{\mathrm{rank}(C) \le 2n - k\} &= \mathcal{N}_k^{2n}\ \mathrm{prob}\{C\mathbf{u}_1 = \mathbf{0}, \ldots, C\mathbf{u}_k = \mathbf{0}\} \\ &\le \min\{2^k, 4\} \frac{2^{2nk}}{2^{k^2}} \frac{2^{-2nk}}{(1 - 2^{-2n})^k} \\ &= \min\{2^k, 4\} \frac{2^{-k^2}}{(1 - 2^{-2n})^k}, \end{aligned} \tag{47}$$

where in the last inequality, we used Lemma 34. Using Lemma 36 (Appendix C), the above argument applies to any of the four sub-matrices A, B, C, D. The proof of Eq. (39) is then completed. □

B. Rank of product of sub-matrices

Lemma 5. Let the random matrices $S_1, S_2, \ldots, S_r \in \mathcal{S}_{2n}$ be independent and uniformly distributed, which induces a distribution for the sub-matrices,

$$S_i = \begin{pmatrix} A_i & B_i \\ C_i & D_i \end{pmatrix}. \tag{48}$$

For any choice $E_i \in \{A_i, B_i, C_i, D_i\}$, for each $i \in \{1, \ldots, r\}$, we have

$$\mathrm{prob}\{\mathrm{rank}(E_r \cdots E_1) \le 2n - k\} \le \frac{2^k}{(1 - 2^{-2n})^k} \binom{k + r - 1}{k} 2^{-\frac{1}{2}k^2}. \tag{49}$$

Proof. Before analyzing the rank of the product of r independent random matrices $C_r \cdots C_1$, we start by a much simpler problem. Analyzing the rank of the product CF where C follows the usual C-distribution and F is a fixed $2n \times 2n$ matrix with $\mathrm{rank}(F) = 2n - k_1$. Noting that the input space of C has dimension $2n - k_1$, from (47), with $k_2 \equiv k - k_1 \ge 0$, we obtain

$$\begin{aligned}\operatorname{prob}\{\operatorname{rank}(CF) \le 2n-k\} &\le \mathcal{N}_{k_2}^{2n-k_1}\operatorname{prob}\{C\mathbf{u}_1 = \mathbf{0},\ldots,C\mathbf{u}_{k_2} = \mathbf{0}\}\\ &\le \min\{2^{k_2},4\}\,\frac{2^{(2n-k_1)k_2}}{2^{k_2^2}}\,\frac{2^{-2nk_2}}{(1-2^{-2n})^{k_2}}\\ &\le \frac{2^{k_2-k_1k_2-k_2^2}}{(1-2^{-2n})^{k_2}}.\end{aligned} \tag{50}$$

Proceeding in a similar fashion, we can analyze the product of two independent C-matrices. To do so, we multiply two factors (50) and sum over all possible intermediate kernel sizes k_1, obtaining

$$\begin{aligned}\operatorname{prob}\{\operatorname{rank}(C_2C_1) \le 2n-k\} &\le \sum_{k_1=0}^{k}\frac{2^{k_2-k_2k_1-k_2^2}}{(1-2^{-2n})^{k_2}}\,\frac{2^{k_1-k_1^2}}{(1-2^{-2n})^{k_1}}\\ &= \sum_{k_1=0}^{k}\frac{2^{k-k_2k_1-k_1^2-k_2^2}}{(1-2^{-2n})^{k}},\end{aligned} \tag{51}$$

where again $k_2 = k - k_1$.

Equation (51) works as follows: the matrix F in (50) has a fixed rank equal to $2n - k_1$, that is, the dimension of the input space of C. In (51), the input space of C_1 is the full space $\mathbb{Z}_2^{2n}$ that has dimension $2n$; therefore, the factor $\frac{2^{k_1-k_1^2}}{(1-2^{-2n})^{k_1}}$ in (51) equals the upper bound in (50), that is, $\frac{2^{k_2-k_1k_2-k_2^2}}{(1-2^{-2n})^{k_2}}$ with $k_1 = 0$ and then with k_2 replaced by k_1. Moreover, the input space of C_2 has dimension $2n - k_1$, that is, such as in Eq. (50), that explains the first factor in (51).

Analogously, we can bound the rank of a product of r independent random C-matrices as

$$\begin{aligned}\operatorname{prob}\{\operatorname{rank}(C_r\cdots C_1) \le 2n-k\} &\le \sum_{\{k_i\}}\prod_{i=1}^{r}\frac{2^{k_i-k_i\sum_{j=1}^{i}k_j}}{(1-2^{-2n})^{k_i}}\\ &= \frac{2^k}{(1-2^{-2n})^k}\sum_{\{k_i\}}2^{-\sum_{i=1}^{r}k_i\sum_{j=1}^{i}k_j},\end{aligned} \tag{52}$$

where the sum $\sum_{\{k_i\}}$ runs over all sets of r non-negative integers $\{k_1,\ldots,k_r\}$ such that $\sum_{i=1}^{r}k_i = k$. These are all ways of sharing out k units among r distinguishable parts. The number of all these sets equals

$$\sum_{\{k_i\}}1 = \binom{k+r-1}{r-1} = \binom{k+r-1}{k}. \tag{53}$$

Finally, for any set $\{k_1,\ldots,k_r\}$, we have

$$\begin{aligned}k^2 = \sum_{i=1}^{r}\sum_{j=1}^{r}k_ik_j &\le \sum_{i=1}^{r}\sum_{j=1}^{i}k_ik_j + \sum_{i=1}^{r}\sum_{j=i}^{r}k_ik_j\\ &= 2\sum_{i=1}^{r}\sum_{j=1}^{i}k_ik_j.\end{aligned} \tag{54}$$

Substituting (53) and (54) back into (52), we obtain

$$\operatorname{prob}\{\operatorname{rank}(C_r\cdots C_1) \le 2n-k\} \;\le\; \frac{2^k}{(1-2^{-2n})^k}\binom{k+r-1}{k}2^{-\frac{1}{2}k^2}. \tag{55}$$

Once again, by using Lemma 36, this proof applies to all products of sub-matrices $E \in \{A,B,C,D\}$. □

Lemma 6. If the random variables $S_1, S_2, \ldots, S_r \in \mathcal{S}_{2n}$ and $\mathbf{u} \in \mathbb{Z}_2^{2n}$ are independent and uniformly distributed, it follows that

$$\operatorname{prob}\{E_r\cdots E_1\mathbf{u} = \mathbf{0}\} \le 8\,r\,2^{-n}. \tag{56}$$

with $E_j \in \{A_j, B_j, C_j, D_j\}$ being the sub-blocks of the symplectic matrices $S_1,\ldots,S_r$.

Proof. If M is a fixed $2n \times 2n$ matrix with rank $M = 2n - k$ and $\mathbf{u} \in \mathbb{Z}_2^{2n}$ is uniformly distributed, then

$$\operatorname{prob}\{M\mathbf{u} = \mathbf{0}\} = \frac{2^k}{2^{2n}}. \tag{57}$$

In addition, if rank $M > 2n - k$, then

$$\operatorname{prob}\{M\mathbf{u} = \mathbf{0}\} \leq \frac{2^{k-1}}{2^{2n}}. \tag{58}$$

This inequality is useful for the following bound:

$$\begin{aligned}
&\operatorname{prob}\{C_r \cdots C_1\mathbf{u} = \mathbf{0}\} \\
&\quad = \operatorname{prob}\{C_r \cdots C_1\mathbf{u} = \mathbf{0} \text{ and } \operatorname{rank}(C_r \cdots C_1) > 2n - k\} \\
&\quad + \operatorname{prob}\{C_r \cdots C_1\mathbf{u} = \mathbf{0} \text{ and } \operatorname{rank}(C_r \cdots C_1) \leq 2n - k\} \\
&\quad \leq \operatorname{prob}\{C_r \cdots C_1\mathbf{u} = \mathbf{0} \,|\, \operatorname{rank}(C_r \cdots C_1) > 2n - k\} \\
&\quad + \operatorname{prob}\{\operatorname{rank}(C_r \cdots C_1) \leq 2n - k\} \\
&\quad \leq 2^{k-1-2n} + \frac{2^k}{(1 - 2^{-2n})^k}(1 + r)^k 2^{-\frac{1}{2}k^2},
\end{aligned} \tag{59}$$

where the last inequality uses (58) and Lemma 5 (and additional Lemma 35 in Appendix C).

Using

$$\begin{aligned}
\frac{1}{(1 - 2^{-2n})^k} &\leq \frac{1}{(1 - 2^{-2n})^{2n}} = \left(1 + \frac{1}{2^{2n} - 1}\right)^{2n} = \left(1 + \frac{1}{2\left(2^{2n-1} - \frac{1}{2}\right)}\right)^{2n} \\
&\leq \left(1 + \frac{1}{4n}\right)^{2n} \leq \sqrt{e} < 2,
\end{aligned} \tag{60}$$

we obtain

$$\operatorname{prob}\{C_r \cdots C_1\mathbf{u} = \mathbf{0}\} \leq 2^{k-2n} + 2(4r)^k 2^{-\frac{1}{2}k^2} = \epsilon, \tag{61}$$

where the last equality defines ϵ. Note that the left-hand side above is independent of k. Hence, for each value of k, we have a different upper bound. We are interested in the tightest one of them. Therefore, we need to find a value of $k \in [1, 2n]$ that makes the upper bound (61) have a small enough value. This can be done by equating each of the two terms to $\epsilon/2$ as

$$2^{k-2n} = 2(4r)^k 2^{-\frac{1}{2}k^2} = \frac{\epsilon}{2}. \tag{62}$$

Isolating k from the first and second terms gives

$$k = 2n - \log_2 \frac{2}{\epsilon}, \tag{63}$$

$$k = \log_2 4r + \sqrt{\log_2^2 4r + \log_2 \frac{2}{\epsilon} + 1}, \tag{64}$$

where we only keep the positive solution. Equating the above two identities for k, we obtain

$$\begin{aligned}
n &= \frac{1}{2}\left(\log_2 4r + \log_2 \frac{2}{\epsilon} + \sqrt{\log_2^2 4r + \log_2 \frac{2}{\epsilon} + 1}\right) \\
&\leq \frac{1}{2}\left(\log_2 4r + \log_2 \frac{2}{\epsilon} + \log_2 4r + \sqrt{\log_2 \frac{2}{\epsilon} + 1}\right) \\
&\leq \log_2 4r + \log_2 \frac{2}{\epsilon},
\end{aligned} \tag{65}$$

which implies

$$\epsilon \leq 8\, r\, 2^{-n}. \tag{66}$$

Substituting this into (61), we finish the proof of this lemma. □

V. LOCAL DYNAMICS IS PAULI MIXING

In this section, using the results from Sec. IV, we will prove that in the regime $N \gg \log L$, the random dynamics of the model that we are considering maps any Pauli operator to any other Pauli operator with approximately uniform probability.

The time evolution of an initial vector $\mathbf{u}^0 \in \mathcal{V}_{\text{chain}}$ at time t is denoted by $\mathbf{u}^t = S(t)\mathbf{u}^0$. If the initial vector is supported only at the origin $\mathbf{u}^0 \in \mathcal{V}_0$, then, as time t increases, the evolved vector $\mathbf{u}^t$ is supported on the light cone,

$$x \in \{-(2t-1), -(2t-2), \ldots, 2t\} \subseteq \mathbb{Z}_L. \tag{67}$$

This leads to the definition of scrambling time: the length of the chain, L, is taken to be an integer multiple of 4, and the system goes from $-L/2$ to $L/2$ with periodic boundary conditions. The scrambling time is the smallest time such that a perturbation supported at $x = 0$ at $t = 0$ evolves spreading its support to $\{-L/2+1, \ldots, L/2\}$; therefore,

$$t_{\text{scr}} \equiv \frac{L}{4}. \tag{68}$$

The definition equally applies to the evolution of a vector $\mathbf{u}^t = S(t)\mathbf{u}^0$ as above.

Finally, we denote the projection of $\mathbf{u}$ on the local subspace $\mathcal{V}_x$ by $\mathbf{u}_x$.

Lemma 7. Consider a vector $\mathbf{u}^0$ supported at the origin $\mathcal{V}_0$ and its time evolution $\mathbf{u}^t$ for any $t \in \{1/2, 1, 3/2, \ldots, 2t_{\text{scr}}\}$. The projection of $\mathbf{u}^t$ at the rightmost site of the light cone $x = 2t$ follows the probability distribution,

$$P(\mathbf{u}_{2t}^t) = \begin{cases} \dfrac{1-q_t}{2^{2N}-1} & \text{if } \mathbf{u}_{2t}^t \neq \mathbf{0}, \\ q_t & \text{if } \mathbf{u}_{2t}^t = \mathbf{0}, \end{cases} \tag{69}$$

where $q_t \leq 2t2^{-2N}$. The projection onto the second rightmost site $\mathbf{u}_{2t-1}^t$ also obeys distribution (69).

Proof. After half a time step, the evolved vector $\mathbf{u}^{1/2}$ is supported on sites $x \in \{0, 1\}$ and it is determined by

$$\mathbf{u}_0^{1/2} \oplus \mathbf{u}_1^{1/2} = S_0(\mathbf{u}_0^0 \oplus \mathbf{0}). \tag{70}$$

Lemma 3 tells us that the vector $\mathbf{u}_0^{1/2} \oplus \mathbf{u}_1^{1/2}$ is uniformly distributed over all non-zero vectors in $\mathcal{V}_0 \oplus \mathcal{V}_1$. This implies that the vector $\mathbf{u}_0^{1/2}$ (and the same for $\mathbf{u}_1^{1/2}$) satisfies

$$\text{prob}\{\mathbf{u}_0^{1/2} = \mathbf{0}\} = \frac{2^{2N}-1}{2^{4N}-1} \leq 2^{-2N} \tag{71}$$

and has probability distribution of the form (69) with $t = 1/2$.

In the next time step, we have

$$\mathbf{u}_1^1 \oplus \mathbf{u}_2^1 = S_1(\mathbf{u}_1^{1/2} \oplus \mathbf{0}). \tag{72}$$

Hence, if $\mathbf{u}_1^{1/2} = \mathbf{0}$, then $\mathbf{u}_1^1 = \mathbf{u}_2^1 = \mathbf{0}$. In addition, applying again Lemma 3, we see that if $\mathbf{u}_1^{1/2} \neq \mathbf{0}$, then $\mathbf{u}_1^1 \oplus \mathbf{u}_2^1$ is uniformly distributed over all non-zero values. Putting these things together, we conclude that $\mathbf{u}_1^1$ (and the same for $\mathbf{u}_2^1$) satisfies

$$\begin{aligned} \text{prob}\{\mathbf{u}_1^1 = \mathbf{0}\} &= \text{prob}\{\mathbf{u}_1^{1/2} = \mathbf{0}\} + \text{prob}\{\mathbf{u}_1^{1/2} \neq \mathbf{0}\}\,\text{prob}\{\mathbf{u}_x^1 = \mathbf{0} | \mathbf{u}_1^{1/2} \neq \mathbf{0}\} \\ &\leq \text{prob}\{\mathbf{u}_1^{1/2} = \mathbf{0}\} + \text{prob}\{\mathbf{u}_1^{1/2} \neq \mathbf{0}\}\, 2^{-2N} \\ &\leq 2 \times 2^{-2N} \end{aligned} \tag{73}$$

and has probability distribution of the form (69) with $t = 1$.

We can proceed as above, applying Lemma 3 to each evolution step

$$\mathbf{u}_{2t-1}^t \oplus \mathbf{u}_{2t}^t = S_{2t-1}(\mathbf{u}_{2t-1}^{t-1/2} \oplus \mathbf{0}) \tag{74}$$

for $t = 1/2, 1, 3/2, 2, \ldots$ This gives us the following recursive equation:

$$\begin{aligned} \text{prob}\{\mathbf{u}_{2t}^t = \mathbf{0}\} &= \text{prob}\{\mathbf{u}_{2t-1}^{t-1/2} = \mathbf{0}\} + \text{prob}\{\mathbf{u}_{2t-1}^{t-1/2} \neq \mathbf{0}\}\,\text{prob}\{\mathbf{u}_{2t}^t = \mathbf{0} | \mathbf{u}_{2t-1}^{t-1/2} \neq \mathbf{0}\} \\ &\leq 2t \times 2^{-2N}. \end{aligned} \tag{75}$$

In addition, the same for $\mathbf{u}_{2t-1}^{t}$. In addition, Lemma 3 implies that $\mathbf{u}_{2t-1}^{t}$ and $\mathbf{u}_{2t}^{t}$ follow the probability distribution (69) for all $t = 1/2, 1, 3/2, 2, \ldots, 2t_{\mathrm{scr}}$.

For $t > 2t_{\mathrm{scr}}$, the recursion relation (74) includes repeated matrices S_x. Hence, the argument is no longer valid. □

Lemma 8. If the initial vector $\mathbf{u}^0 \in \mathcal{V}_{\mathrm{chain}}$ is supported on all lattice sites ($\mathbf{u}_x^0 \neq \mathbf{0}$ for all x), then the projection of its evolution $\mathbf{u}^t$ onto any site $x \in \mathbb{Z}_L$ satisfies

$$\mathrm{prob}\{\mathbf{u}_x^t \neq \mathbf{0}\} \geq 1 - 16\, t\, 2^{-N} \tag{76}$$

for all $t \in \{1/2, 1, 3/2, \ldots, 2t_{\mathrm{scr}}\}$.

Proof. To prove this lemma, we proceed similarly as in Lemma 7. However, here, the recursive equation [Eq. (74)] need not have a $\mathbf{0}$-input in the right system,

$$\mathbf{u}_{2t-1}^{t} \oplus \mathbf{u}_{2t}^{t} = S_{2t-1}\big(\mathbf{u}_{2t-1}^{t-1/2} \oplus \mathbf{u}_{2t}^{t-1/2}\big). \tag{77}$$

This difference in the premises does not change conclusion (71) due to the fact that bound (37) is independent of $\mathbf{u}_1^0$ being zero or not. This gives (69) for $t = 1/2$. In addition, using

$$\begin{aligned}\mathrm{prob}\{\mathbf{u}_2^1 = \mathbf{0}\} &= \mathrm{prob}\{\mathbf{u}_1^{1/2} \oplus \mathbf{u}_2^{1/2} = \mathbf{0}\} + \mathrm{prob}\{\mathbf{u}_1^{1/2} \oplus \mathbf{u}_2^{1/2} \neq \mathbf{0} \text{ and } \mathbf{u}_2^1 = \mathbf{0}\}\\ &\leq \mathrm{prob}\{\mathbf{u}_1^{1/2} = \mathbf{0}\} + \mathrm{prob}\{\mathbf{u}_x^1 = \mathbf{0} \,|\, \mathbf{u}_1^{1/2} \oplus \mathbf{u}_2^{1/2} \neq \mathbf{0}\}\\ &\leq 2^{-2N+1},\end{aligned} \tag{78}$$

we obtain the same probability distribution as in (69) for $t = 1$, but under different premises. However, here, there is a very delicate point. As can be seen in Fig. 6, the vector $\mathbf{u}_2^1$ is partly determined by S_2, and hence, it is not independent of S_2. Crucially, the bound (73) for $\mathbf{u}_2^1$ holds regardless of the right input $\mathbf{u}_2^{1/2}$, and hence, it is independent of S_2. This fact can be summarized with the following bound:

$$P(\mathbf{u}_2^1|S_2) = \begin{cases} \dfrac{1-q_1}{2^{2N}-1} & \text{if } \mathbf{u}_2^1 \neq \mathbf{0},\\ q_1 & \text{if } \mathbf{u}_2^1 = \mathbf{0},\end{cases} \tag{79}$$

for any S_2, where $q_1 \leq 22^{-2N}$. That is, the correlation between $\mathbf{u}_2^1$ and S_2 can only happen through small variations of q_1.

For $t > 1$, the inputs in (77) are not independent of the matrix S_{2t-1}, as illustrated in Fig. 6, and hence, Lemma 3 cannot be applied. If we restrict Eq. (77) to the rightmost output ($x = 2t$), then we obtain

$$\begin{aligned}\mathbf{u}_{2t}^{t} &= C_{2t-1}\mathbf{u}_{2t-1}^{t-1/2} + D_{2t-1}\mathbf{u}_{2t}^{t-1/2}\\ &= C_{2t-1}\mathbf{u}_{2t-1}^{t-1/2} + \mathbf{v}^{t-1/2},\end{aligned} \tag{80}$$

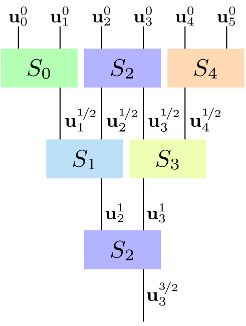


FIG. 6. The causal past of $\mathbf{u}_2^1$ is partly determined by S_2. Hence, at $t = 3/2$, the input $\mathbf{u}_2^1$ of S_2 is not independent of S_2. This makes the exact probability distribution of $\mathbf{u}_3^{3/2}$ very complicated. To overcome this problem, we exploit the fact that S_1 appears only once in the past of $\mathbf{u}_3^{3/2}$. This allows us to map the randomness of S_1 to $\mathbf{u}_3^{3/2}$ for most of the values of the other gates (S_1, S_2, S_3, S_4). More concretely, we can apply Lemma 6 to the cases $r = 1$, $E_1 = C_2$, and $\mathbf{u} = \mathbf{u}_2^1$, resulting in that $\mathbf{u}_3^{3/2}$ is approximately uniform.

where the vector $\mathbf{v}^{t-1/2} = D_{2t-1}\mathbf{u}_{2t}^{t-1/2} \in \mathbb{Z}_2^{2N}$ is not independent of C_{2t-1}. Expanding this recursive relation, we obtain

$$\begin{aligned} \mathbf{u}_{2t}^{t} &= C_{2t-1}C_{2t-2}\mathbf{u}_{2t-2}^{t-1} + C_{2t-1}\mathbf{v}^{t-1} + \mathbf{v}^{t-1/2} \\ &= C_{2t-1}\cdots C_2\mathbf{u}_2^1 + \mathbf{w}^t, \end{aligned} \tag{81}$$

where the random vector

$$\mathbf{w}^t = C_{2t-1}\cdots C_3\mathbf{v}^1 + \cdots + C_{2t-1}C_{2t-2}\mathbf{v}^{t-3/2} + C_{2t-1}\mathbf{v}^{t-1} + \mathbf{v}^{t-1/2} \tag{82}$$

is not independent of the matrices $C_{2t-1}, \ldots, C_2$. Crucially, the bound (79) for the distribution of $\mathbf{u}_2^1$ is independent of all these matrices.

Let us introduce the uniformly distributed random variable $\mathbf{u} \in \mathbb{Z}_2^{2N}$, which is independent of all gates S_x. According to (79), the random variable $\mathbf{u}_2^1$ is close to uniform; hence, it has a small statistical distance with $\mathbf{u}$,

$$\begin{aligned} \mathrm{d}(\mathbf{u}_2^1, \mathbf{u}) &= \sum_{\mathbf{u}_2^1} \left|P(\mathbf{u}_2^1) - 2^{-2N}\right| \\ &= \left|q_1 - 2^{-2N}\right| + \left(2^{2N} - 1\right)\left|\frac{1-q_1}{2^{2N}-1} - 2^{-2N}\right| \\ &= 2\left|q_1 - 2^{-2N}\right| \leq 2\,2^{-2N}. \end{aligned}$$

For any event $\mathcal{E} \subseteq \mathbb{Z}_2^{2N}$, we have that

$$\begin{aligned} \mathrm{prob}\{\mathbf{u}_2^1 \in \mathcal{E}\} &= \sum_{\mathbf{u}_2^2 \in \mathcal{E}} P(\mathbf{u}_2^1) \leq \sum_{\mathbf{u}_2^2 \in \mathcal{E}} \left(2^{-2N} + \left|P(\mathbf{u}_2^1) - 2^{-2N}\right|\right) \\ &\leq \mathrm{prob}\{\mathbf{u} \in \mathcal{E}\} + \mathrm{d}(\mathbf{u}_2^1, \mathbf{u}) \\ &\leq 2\,2^{-2N} + \mathrm{prob}\{\mathbf{u} \in \mathcal{E}\}. \end{aligned} \tag{83}$$

Next, we apply this bound to the particular event $\mathcal{E}$ defined by $\mathbf{u}_{2t}^t = \mathbf{0}$, which can be written as $C_{2t-1}\cdots C_2\mathbf{u}_2^1 = \mathbf{w}^t$ by using (81). Putting all this together, we obtain

$$\mathrm{prob}\{\mathbf{u}_{2t}^t = \mathbf{0}\} \leq 2\,2^{-2N} + \mathrm{prob}\{C_{2t-1}\cdots C_2\mathbf{u} = \mathbf{w}^t\}, \tag{84}$$

where the random variable $\mathbf{u} \in \mathbb{Z}_2^{2N}$ is uniformly distributed and independent of $\mathbf{w}^t$ and C_i for all $i \in \{2t-1, \ldots, 2\}$. This has the advantage that now we can invoke Lemma 6. Let us start by rewriting

$$\mathrm{prob}\{C_{2t-1}\cdots C_2\mathbf{u} = \mathbf{w}^t\} = \mathop{\mathbb{E}}_{C_i, \mathbf{w}^t} \mathop{\mathbb{E}}_{\mathbf{u}} \delta\left(C_{2t-1}\cdots C_2\mathbf{u} - \mathbf{w}^t, \mathbf{0}\right),$$

and let us consider the average $\mathop{\mathbb{E}}_{\mathbf{u}} \delta\left(C_{2t-1}\cdots C_2\mathbf{u} - \mathbf{w}^t, \mathbf{0}\right)$ for a fixed value of the variables $\mathbf{w}^t$ and C_i. If the vector $\mathbf{w}^t$ is not in the range of the matrix $(C_{2t-1}\cdots\ C_2)$, then the average is zero. If the vector $\mathbf{w}^t$ is in the range of the matrix $(C_{2t-1}\cdots\ C_2)$, then there is a vector $\tilde{\mathbf{w}}$ such that $\mathbf{w}^t = (C_{2t-1}\cdots C_2)\tilde{\mathbf{w}}$. Then, we can write the average as

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{u}} \delta\left(C_{2t-1}\cdots C_2\mathbf{u} - \mathbf{w}^t, \mathbf{0}\right) &= \mathop{\mathbb{E}}_{\mathbf{u}} \delta(C_{2t-1}\cdots C_2(\mathbf{u} + \tilde{\mathbf{w}}), \mathbf{0}) \\ &= \mathop{\mathbb{E}}_{\mathbf{u}} \delta(C_{2t-1}\cdots C_2\mathbf{u}, \mathbf{0}), \end{aligned}$$

where the last equality follows from the fact that the random variable $\mathbf{u} + \tilde{\mathbf{w}}$ is uniform and independent of C_i, likewise $\mathbf{u}$. Combining together the two cases for $\mathbf{w}^t$, we can write

$$\begin{aligned} \mathrm{prob}\{C_{2t-1}\cdots C_2\mathbf{u} = \mathbf{w}^t\} &\leq \mathop{\mathbb{E}}_{C_i} \mathop{\mathbb{E}}_{\mathbf{u}} \delta(C_{2t-1}\cdots C_2\mathbf{u}, \mathbf{0}) \\ &= \mathrm{prob}\{C_{2t-1}\cdots C_2\mathbf{u} = \mathbf{0}\} \\ &\leq 8(2t-2)2^{-N}, \end{aligned} \tag{85}$$

where the last step follows from Lemma 6. Substituting this back into (84), we obtain

$$\mathrm{prob}\{\mathbf{u}_{2t}^t = \mathbf{0}\} \leq 2\,2^{-2N} + 16\,(t-1)2^{-N} \leq 16\,t\,2^{-N}. \tag{86}$$

If we repeat all the steps of this proof since (80) substituting $\mathbf{u}_{2t}^t$ for $\mathbf{u}_{2t-1}^t$, then we arrive at

$$\mathrm{prob}\{\mathbf{u}_{2t-1}^t = \mathbf{0}\} \leq 2\,2^{-2N} + \mathrm{prob}\{A_{2t-1}C_{2t-2}\cdots C_2\mathbf{u} = \mathbf{w}^t\},$$

instead of (84). However, Lemma 6 also applies in this case, giving the bound

$$\operatorname{prob}\{A_{2t-1}C_{2t-2}\cdots C_2\mathbf{u} = \mathbf{0}\} \le 8(2t-2)2^{-N},$$

which implies that

$$\operatorname{prob}\{\mathbf{u}^t_{2t-1} = \mathbf{0}\} \le 16\,t\,2^{-N}.$$

In addition, since the premises of this lemma are invariant under translations in the chain $\mathbb{Z}_L$, then the conclusions hold for all $x \in \mathbb{Z}_L$. □

In order to prove the next theorem, it is important to note the following remark. The bound (86) requires that either $\mathbf{u}^0_0 \neq \mathbf{0}$ or $\mathbf{u}^1_0 \neq \mathbf{0}$, but does not require $\mathbf{u}^0_x \neq \mathbf{0}$ for $x > 1$.

Lemma 9. After the scrambling time $t \in [t_{\text{scr}}, 2t_{\text{scr}}]$, with t integer or half-integer, the evolved vector $\mathbf{u}^t = S(t)\mathbf{u}^0$ is non-zero at each lattice site with probability

$$\operatorname{prob}\{\mathbf{u}^t_x \neq \mathbf{0}, \forall\, x \in \mathbb{Z}_L\} \ge 1 - 16\,t\,L\,2^{-N} \tag{87}$$

for any initial non-zero vector $\mathbf{u}^0 \in \mathcal{V}_{\text{chain}}$.

Proof. Let $\mathcal{F}(\mathbf{u}^0) \subseteq \mathbb{Z}_L \times \mathbb{N}$ be the set of space-time points consisting of the causal future of the sites $x' \in \mathbb{Z}_L$ where the initial vector $\mathbf{u}^0$ has support ($\mathbf{u}^0_{x'} \neq \mathbf{0}$). For example, if the initial vector is supported in the origin of the chain $\mathbf{u}^0 \in \mathcal{V}_0$, then the causal future is given by the light cone (67).

The main objective in this proof is to bound the probability of $\mathbf{u}^t_x \neq \mathbf{0}$ for any fixed site $x \in \mathbb{Z}_L$ and time $t \in [t_{\text{scr}}, 2t_{\text{scr}}]$. For the sake of simplicity, let us start by considering the case of x odd and t integer. In this case, the left-most space-time points in the causal past of (x,t) that are also contained in $\mathcal{F}(\mathbf{u}^0)$ are

$$(x-1, t-1/2), \ldots, (x-n, t-n/2), \ldots, (x_e, t_e). \tag{88}$$

We have that either $t_e = 0$ or $t_e > 0$. In the first case ($t_e = 0$), we have that $\mathbf{u}^0$ has support on x_e or $x_e + 1$. In addition, we can prove

$$\operatorname{prob}\{\mathbf{u}^t_x = \mathbf{0}\} \le 16\,t\,2^{-N} \tag{89}$$

by applying the same procedure as in Lemma 8. Note that the possibility that $\mathbf{u}^0_{x'} = \mathbf{0}$ for $x' > x_e + 1$ does not affect the argument (see last paragraph in the Proof of Lemma 8).

In the second case ($t_e > 0$), the sequence (88) can be continued by including the following points from $\mathcal{F}(\mathbf{u}^0)$:

$$(x_e, t_e - 1/2), \ldots, (x_e + n, t_e - 1/2 - n/2), \ldots, (x_0 - 1, 1/2), \begin{cases} (x_0 - 1, 0), \\ (x_0, 0), \end{cases} \tag{90}$$

where the last element is chosen so that it belongs to $\mathcal{F}(\mathbf{u}^0)$. If $\mathbf{u}^0$ has support on both $(x_0 - 1, 0)$ and $(x_0, 0)$, then the choice is arbitrary. Here, for the sake of concreteness, we assume that $\mathbf{u}^0_{x_0} \neq \mathbf{0}$ and take $(x_0, 0)$ as the last point of the sequence. The subindex e stands for "elbow," because it labels the point where the sequence (88) changes direction to (90) (see Fig. 7).

Now, we can write our chosen vector $\mathbf{u}^t_x$ as

$$\begin{aligned} \mathbf{u}^{t_e - 1/2}_{x_e} &= B_{x_e}\cdots B_{x_0-2}B_{x_0-1}\mathbf{u}^0_{x_0}, \\ \mathbf{u}^{t_e}_{x_e} &= D_{x_e-1}\mathbf{u}^{t_e-1/2}_{x_e}, \\ \mathbf{u}^t_x &= C_{x-1}\cdots C_{x_e+1}C_{x_e}\mathbf{u}^{t_e}_{x_e} + \mathbf{w}, \end{aligned}$$

where the random vector $\mathbf{w}$ is correlated with $B_{x_e}, \ldots, B_{x_0-1}$ and $C_{x-1}, \ldots, C_{x_e}$ but not with D_{x_e-1}. Vector $\mathbf{w}$ is analogous to $\mathbf{w}^t$, defined in (82). Note also that the random matrices $B_{x_e}, \ldots, B_{x_0-1}$ are not independent of $C_{x-1}, \ldots, C_{x_e}$, but that D_{x_e-1} is independent of all the rest. (Figure 7 shows an example where the gates associated with $B_{x_e}, \ldots, B_{x_0-1}$ are in blue, those of $C_{x-1}, \ldots, C_{x_e}$ in red, and that of D_{x_e-1} in yellow.)

Now, we can start constructing our bound as

$$\begin{aligned} \operatorname{prob}\{\mathbf{u}^t_x = \mathbf{0}\} &= \operatorname{prob}\{\mathbf{u}^t_x = \mathbf{0} \text{ and } \mathbf{u}^{t_e-1/2}_{x_e} = \mathbf{0}\} + \operatorname{prob}\{\mathbf{u}^t_x = \mathbf{0} \text{ and } \mathbf{u}^{t_e-1/2}_{x_e} \neq \mathbf{0}\} \\ &\le \operatorname{prob}\{\mathbf{u}^{t_e-1/2}_{x_e} = \mathbf{0}\} + \operatorname{prob}\{\mathbf{u}^t_x = \mathbf{0} \text{ and } \mathbf{u}^{t_e-1/2}_{x_e} \neq \mathbf{0}\}. \end{aligned} \tag{91}$$

The first term can be bounded with the recursive equation [Eq. (75)] as

$$\operatorname{prob}\{\mathbf{u}^{t_e-1/2}_{x_e} = \mathbf{0}\} \le 2(t_e - 1/2)2^{-2N}.$$

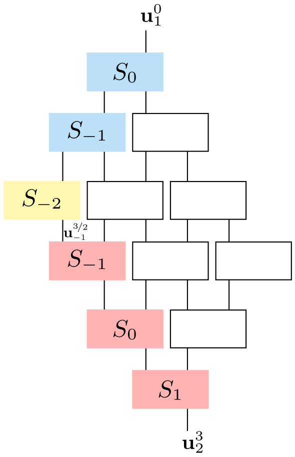


FIG. 7. The evolution of an initially local operator $\mathbf{u}^0 \in \mathcal{V}_1$ at site $x = 1$. The figure only displays gates S_x that are in the intersection of the causal future of the initial location $x = 1$ and the causal past of the chosen point $\mathbf{u}_2^3$. The probability of $\mathbf{u}_2^3 = \mathbf{0}$ is bounded by analyzing the sequence of colored gates, which has an "elbow" at location $(x_\mathrm{e}, t_\mathrm{e}) = (-1, 3/2)$. The analysis of blue gates uses Lemma 7, and that of red gates uses Lemma 8. The key feature of the bound is that the yellow gate S_{-2} appears only once.

The second term can be bounded by using the independence of $D_{x_\mathrm{e}-1}$, the fact that $\mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}-1/2}$ is not zero, and proceeding in a manner similar to (83) and (84). Therefore, we again introduce the uniformly distributed random vector $\mathbf{u} \in \mathbb{Z}_2^{2N}$, which is independent of all gates $S_{x_\mathrm{e}}, S_{x_\mathrm{e}+1}, \ldots, S_{x-1}$. The statistical distance between $\mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}}$ and $\mathbf{u}$, conditioned on $\mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}-1/2} \neq \mathbf{0}$, is

$$\begin{aligned} \mathrm{d}(\mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}}, \mathbf{u}) &= \sum_{\mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}}} \left| P(\mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}} | \mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}-1/2} \neq \mathbf{0}) - 2^{-2N} \right| \\ &= \left(2^{2N} - 1\right) \left| \frac{2^{2N}}{2^{4N} - 1} - 2^{-2N} \right| + \left| \frac{2^{2N} - 1}{2^{4N} - 1} - 2^{-2N} \right| \\ &\leq 2^{1-4N}, \end{aligned} \tag{92}$$

where we have used Lemma 3. Proceeding in a manner similar to (83) and (84), we obtain

$$\begin{aligned} &\mathrm{prob}\{\mathbf{u}_x^t = \mathbf{0} \,\mathrm{and}\, \mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}-1/2} \neq \mathbf{0}\} \\ &\quad = \mathrm{prob}\{C_{x-1} \cdots C_{x_\mathrm{e}} \mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}} = \mathbf{w} \,\mathrm{and}\, \mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}-1/2} \neq \mathbf{0}\} \\ &\quad \leq \mathrm{d}(\mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}}, \mathbf{u}) + \mathrm{prob}\{C_{x-1} \cdots C_{x_\mathrm{e}} \mathbf{u} = \mathbf{w} \,\mathrm{and}\, \mathbf{u}_{x_\mathrm{e}}^{t_\mathrm{e}-1/2} \neq \mathbf{0}\} \\ &\quad \leq 2^{1-4N} + \mathrm{prob}\{C_{x-1} \cdots C_{x_\mathrm{e}} \mathbf{u} = \mathbf{w}\}. \end{aligned}$$

The bound (85) exploits the fact that that $\mathbf{w}$ and $\mathbf{u}$ are independent, giving

$$\mathrm{prob}\{C_{x-1} \cdots C_{x_\mathrm{e}} \mathbf{u} = \mathbf{w}\} \leq 8\,(x - x_\mathrm{e})\,2^{-N} = 16\,(t - t_\mathrm{e})\,2^{-N}.$$

Putting all things together, we obtain

$$\mathrm{prob}\{\mathbf{u}_x^t = \mathbf{0}\} \leq 2(t_\mathrm{e} - 1/2)2^{-2N} + 2^{1-4N} + 16\,(t - t_\mathrm{e})\,2^{-N} \leq 16\,t\,2^{-N}. \tag{93}$$

Finally, we use the union bound to conclude that

$$\mathrm{prob}\left\{\exists\, x \in \mathbb{Z}_L : \mathbf{u}_x^t = \mathbf{0}\right\} \leq 8\,t\,L\,2^{-N},$$

which is equivalent to the statement (87). □

A. Twirling technique and Pauli invariance

This section, see Fig. 8, illustrates the fact that, at integer time t, the probability distribution of $W(t)$ is invariant under the transformation

$$W(t) \;\mapsto\; \left(\bigotimes_x V_x'\right)^\dagger W(t) \left(\bigotimes_x V_x'\right) \tag{94}$$

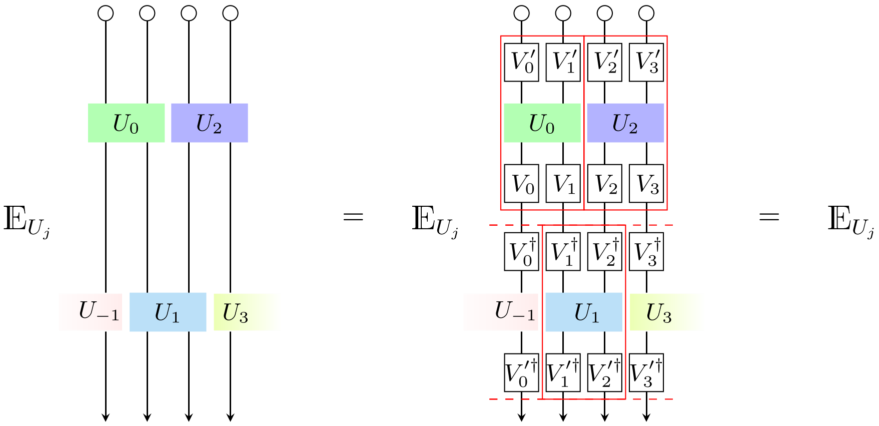

$\mathbb{E}_{U_j}$
U_0
U_2
U_{-1}
U_1
U_3
=
V'_0 V'_1 V'_2 V'_3
V_0 V_1 V_2 V_3
$V_0^\dagger$ $V_1^\dagger$ $V_2^\dagger$ $V_3^\dagger$
$V_0'^\dagger$ $V_1'^\dagger$ $V_2'^\dagger$ $V_3'^\dagger$

FIG. 8. Twirling technique. The probability distribution of the evolution operator is invariant under local transformations. On the left, we have a section of the circuit of Fig. 1. On the middle, we use the fact that, for any pair of local Clifford unitaries V_x, V_{x+1}, the random two-site Clifford unitary $(V_x \otimes V_{x+1})U_x(V'_x \otimes V'_{x+1})$ has the same probability distribution than U_x. On the right, we see that all local unitaries get canceled except for those of the initial and final times. Note that if the time t is integer or half-integer, the invariance property of $W(t)$ is different according to Eqs. (94) and (96).

for any string of local Clifford unitaries $V'_1, \ldots, V'_L \in \mathcal{C}_N$. This property translates to distribution (7) as

$$P_t([\textstyle\bigoplus_x S_x^{-1}]\mathbf{u}' | [\bigoplus_x S_x]\mathbf{u}) = P_t(\mathbf{u}'|\mathbf{u}) \tag{95}$$

for any list of local symplectic matrices $S_1, \ldots, S_L$. In order to prove Theorem 21, we exploit the fact that, at half-integer time t, the evolution operator displays a higher degree of symmetry. The probability distribution of $W(t)$ is invariant under the transformation

$$W(t) \;\mapsto\; (\textstyle\bigotimes_x V_x) W(t) (\bigotimes_x V'_x) \tag{96}$$

for any string of local Clifford unitaries $V_1, V'_1, \ldots, V_L, V'_L \in \mathcal{C}_N$. This translates onto $P_t(\mathbf{u}'|\mathbf{u})$ in a way analogous to (95).

In this section, we will present what is referred to as the twirling technique (Fig. 8) in the work[34] and discuss how it applies to the random Clifford circuit model that we are considering.

We recollect that the definition of the evolution operator after an *integer* time t is

$$\begin{aligned} W(t) &\equiv [(U_1 \otimes U_3 \otimes \cdots \otimes U_{L-1})(U_0 \otimes U_2 \otimes \cdots \otimes U_{L-2})]^t \\ &= (U_{\text{odd}} U_{\text{even}})^t = (U_{\text{chain}})^t, \end{aligned}$$

and the definition of the evolution operator after a *half-integer* time t is

$$W(t) \equiv U_{\text{even}}\,(U_{\text{chain}})^{t-1/2}.$$

Lemma 10. Consider a set of $2L$ single-site Clifford unitaries $V_x, V'_x \in \mathcal{C}_N$; these unitaries are fixed. At integer time t, the random evolution operator $W(t)$, as defined above, has the same probability distribution as

$$\left(\bigotimes_{x=0}^{L-1} V_x'^\dagger\right) W(t) \left(\bigotimes_{x=0}^{L-1} V'_x\right). \tag{97}$$

Similarly, at half-integer time t, the evolution operator $W(t)$ has the same probability distribution as

$$\left(\bigotimes_{x=0}^{L-1} V_x\right) W(t) \left(\bigotimes_{x=0}^{L-1} V'_x\right). \tag{98}$$

Proof. First, we note that any uniformly distributed two-site Clifford unitary $U_x \in \mathcal{C}_{2N}$ has the same probability distribution as the unitary $(V_x \otimes V_{x+1})U_x(V'_x \otimes V'_{x+1})$ for any arbitrary choice of $V_x, V_{x+1}, V'_x, V'_{x+1} \in \mathcal{C}_N$; this is denoted as single-site Haar invariance. Hence, we introduce the primed notation for the random two-site Clifford unitary U_x,

$$U'_x \equiv (V_x \otimes V_{x+1})U_x(V'_x \otimes V'_{x+1}) \text{ for even } x \in \mathbb{Z}_L, \tag{99}$$

$$U'_x \equiv (V'_x \otimes V'_{x+1})^{-1}U_x(V_x \otimes V_{x+1})^{-1} \text{for odd } x \in \mathbb{Z}_L, \tag{100}$$

where $V_x, V_{x+1}, V'_x, V'_{x+1} \in \mathcal{C}_N$ are any arbitrary choice of single-site Clifford unitary. Consequently, the primed version of the global dynamics for integer t becomes

$$W'(t) = \left(\bigotimes_{x=0}^{L-1} V_x'^{\dagger}\right) W(t) \left(\bigotimes_{x=0}^{L-1} V'_x\right), \tag{101}$$

and for half-integer t,

$$W'(t) = \left(\bigotimes_{x=0}^{L-1} V_x\right) W(t) \left(\bigotimes_{x=0}^{L-1} V'_x\right). \tag{102}$$

The single-site Haar invariance of the probability distributions of the primed and not-primed evolution operators is identical; this proves the result. □

Next, we will define the Pauli invariance and state when it applies to our model.

Definition 11. An n-qubit random unitary $U \in \mathrm{SU}(2^n)$ with probability distribution $P(U)$ is Pauli invariant if $P(U\sigma) = P(U)$ for all $\sigma \in \mathcal{P}_n$ and $U \in \mathrm{SU}(2^n)$.

Lemma 12. At half-integer time t, the random evolution operator $W(t)$ is Pauli invariant.

Proof. The proof of this lemma follows from Lemma 10. When t is half-integer, $W(t)$ and $(\otimes_{x=0}^{L-1} V_x) W(t) (\otimes_{x=0}^{L-1} V'_x)$ have identical probability distributions, where $V_x, V'_x \in \mathcal{C}_n$. Since $\mathcal{P}_n \subset \mathcal{C}_n$, we can choose $(\otimes_{x=0}^{L-1} V_x)$ to be any element of the Pauli group. Hence, $W(t)$ is Pauli invariant. □

B. Half-integer times

Lemma 13. At half-integer $t \geq t_{\mathrm{scr}}$, the probability distribution of the evolved vector $\mathbf{u}^t = S(t)\mathbf{u}^0$ conditioned on it being non-zero at every site is uniform,

$$\mathrm{prob}\{\mathbf{u}^t = \mathbf{v} | \mathbf{u}_x^t \neq \mathbf{0}, \forall x \in \mathbb{Z}_L\} = \frac{1}{(2^{2N}-1)^L} \tag{103}$$

for all vectors $\mathbf{v}$ that are non-zero at every site $\mathbf{v}_x \neq \mathbf{0}$, $\forall x \in \mathbb{Z}_L$.

Proof. The proof of this lemma follows from the twirling technique discussed in Sec. V A, Lemma 10. The probability distribution of the evolved vector $\mathbf{u}^t = S(t)\mathbf{u}^0$ is identical to

$$\mathbf{u}^t = \left(\bigoplus_{x=0}^{L-1} X_x\right) S(t) \left(\bigoplus_{x=0}^{L-1} Y_x\right) \mathbf{u}^0,$$

where $X_x, Y_x \in \mathcal{S}_N$ are arbitrary single-site matrices. Hence, since the choice of each X_x is arbitrary, each X_x is independent and uniformly distributed over all single-site symplectic matrices. Therefore, imposing the condition that the evolved vector is non-zero on every site, then, since the twirling matrices X_x are independent and uniform, the probability distribution of the evolved vector at each site is independent and uniformly distributed over all non-zero vectors. The application of Lemma 3 eventually provides the conditional probability (103). □

Theorem 14. Let $\sigma_{\mathbf{u}'} = \lambda W(t) \sigma_{\mathbf{u}} W(t)^\dagger$ be the evolution of any initial Pauli operator $\sigma_{\mathbf{u}} \neq \mathbb{1}$. At any half-integer time t larger than the scrambling time, in the interval $t \in [t_{\mathrm{scr}}, 2t_{\mathrm{scr}}]$, the probability distribution (7) for the evolved operator $\sigma_{\mathbf{u}'}$ is close to uniform, namely,

$$\sum_{\mathbf{u}'} \left| P_t(\mathbf{u}'|\mathbf{u}) - Q_t(\mathbf{u}') \right| \;\leq\; 33 \times t\, L\, 2^{-N}. \tag{104}$$

Remark. The following is an equivalent statement to Theorem 14 formulated in phase space; we then present a proof.

Theorem 14 (alternative form). For any initial non-zero vector $\mathbf{u}^0 \in \mathcal{V}_{\mathrm{chain}}$, the probability distribution of the time evolved vector $\mathbf{u}^t = S(t)\mathbf{u}^0$, at any half-integer time in the interval $t \in [t_{\mathrm{scr}}, 2t_{\mathrm{scr}}]$, is approximately uniformly distributed over all non-zero vectors of the total system and bounded by

$$\sum_{\mathbf{v}} \left| \mathrm{prob}\{\mathbf{u}^t = \mathbf{v}\} - \frac{1}{2^{2NL}-1} \right| \leq 32\, tL2^{-N} + L2^{-2N}.$$

Proof. Below, we make use of

$$\begin{aligned} \mathrm{prob}(A) &= \mathrm{prob}(A \wedge B) + \mathrm{prob}(A \wedge \bar{B}) \\ &= \mathrm{prob}(A|B)\mathrm{prob}(B) + \mathrm{prob}(A|\bar{B})\mathrm{prob}(\bar{B}), \end{aligned}$$

where A and B are events in a probability space.

Defining $q \equiv \text{prob}\{\mathbf{u}_x^t \neq \mathbf{0}, \forall x \in \mathbb{Z}_L\}$, we rewrite $\text{prob}\{\mathbf{u}^t = \mathbf{v}\}$ as follows:

$$\begin{aligned}\text{prob}\{\mathbf{u}^t = \mathbf{v}\} = q\,\text{prob}\{\mathbf{u}^t = \mathbf{v} | \mathbf{u}_x^t \neq \mathbf{0}, \forall x \in \mathbb{Z}_L\} \\ + (1-q)(\text{prob}\{\mathbf{u}^t = \mathbf{v} | \exists y \in \mathbb{Z}_L : \mathbf{u}_y^t = \mathbf{0}\}).\end{aligned}$$

Adding and subtracting $q\frac{1}{2^{2NL}-1}$ in the sum and then applying the triangular inequality, we find that

$$\begin{aligned}\sum_{\mathbf{v}}\left|\text{prob}\{\mathbf{u}^t = \mathbf{v}\} - \frac{1}{2^{2NL}-1}\right| \leq \sum_{\mathbf{v}}\left|q\,\text{prob}\{\mathbf{u}^t = \mathbf{v}|\mathbf{u}_x^t \neq \mathbf{0}\ \forall x\} - \frac{1}{2^{2NL}-1}\right| \\ + (1-q)\sum_{\mathbf{v}}\left|\text{prob}\{\mathbf{u}^t = \mathbf{v}|\exists y \in \mathbb{Z}_L : \mathbf{u}_y^t = \mathbf{0}\} - \frac{1}{2^{2NL}-1}\right|.\end{aligned}$$

We can upper bound the first term with $q \leq 1$ and apply Lemma 13 to find that

$$q\sum_{\mathbf{v}}\left|\text{prob}\{\mathbf{u}^t = \mathbf{v}|\mathbf{u}_x^t \neq \mathbf{0}\ \forall x\} - \frac{1}{2^{2NL}-1}\right| \leq L2^{-2N}.$$

To bound the second term, we note that the maximum value of the sum is 2; in fact,

$$\begin{aligned}&(1-q)\sum_{\mathbf{v}}\left|\text{prob}\{\mathbf{u}^t = \mathbf{v}|\exists y \in \mathbb{Z}_L : \mathbf{u}_y^t = \mathbf{0}\} - \frac{1}{2^{2NL}-1}\right| \\ &\leq (1-q)\sum_{\mathbf{v}}\left(\text{prob}\{\mathbf{u}^t = \mathbf{v}|\exists y \in \mathbb{Z}_L : \mathbf{u}_y^t = \mathbf{0}\} + \frac{1}{2^{2NL}-1}\right) = 2(1-q),\end{aligned}$$

and we use the result of Lemma 9 to find that

$$(1-q) \leq 16tL2^{-N}. \tag{105}$$

This gives the stated result. □

C. Integer times

In this section, we will consider only initial vectors, which are supported (i.e., non-zero) on a single site, $\mathbf{u}^0 \in \mathcal{V}_0 \subseteq \mathcal{V}_{\text{chain}}$, and their time evolution at integer times only. The validity of the following lemma is not restricted to integer times or even quantum circuits.

Lemma 15. Let $\mathbf{u}$ be a fixed non-zero element of $\mathbb{Z}_2^{2N}$. Let the probability distribution $P(\mathbf{v})$ over $\mathbf{v} \in \mathbb{Z}_2^{2N}$ have the property that $P(S\mathbf{v}) = P(\mathbf{v})$ for any $S \in \mathcal{S}_N$ such that $S\mathbf{u} = \mathbf{u}$. Then, it must be of the form

$$P(\mathbf{v}) = \begin{cases} q_1 & \text{if } \mathbf{v} = \mathbf{0}, \\ q_2 & \text{if } \mathbf{v} = \mathbf{u}, \\ q_3 & \text{if } \langle \mathbf{v}, \mathbf{u}\rangle = 0 \text{ and } \mathbf{v} \neq \mathbf{0}, \mathbf{u}, \\ q_4 & \text{if } \langle \mathbf{v}, \mathbf{u}\rangle = 1, \end{cases} \tag{106}$$

where the positive numbers q_i are constrained by the normalization of $P(\mathbf{v})$.

Proof. We initially consider that $\mathbf{u} = (1, 0, \ldots, 0)^T$ and the subgroup of $\mathcal{S}_N$ that leaves $\mathbf{u}$ unchanged. If $\mathbf{v} = \mathbf{0}$ or$\mathbf{u}$, then the action of this subgroup has no effect, and hence, we require a parameter for each in the distribution, q_1 and q_2, respectively. This is not the case for all other choices of $\mathbf{v}$ since the action of the subgroup will transform $\mathbf{v}$ into some other vector in $\mathbb{Z}_2^{2N}$. This transformation is constrained by the symplectic form:

$$\langle \mathbf{v}, \mathbf{u}\rangle = \langle S\mathbf{v}, S\mathbf{u}\rangle = \langle S\mathbf{v}, \mathbf{u}\rangle, \tag{107}$$

and hence, the subgroup is composed of two subgroups, which transform $\mathbf{v}$ into another vector in $\mathbb{Z}_2^{2N}$ that has the same value for the symplectic form. Furthermore, the two subgroups are such they can map any vector to any other vector with the same value for the symplectic form.

This can be seen by considering the case where $\mathbf{v} = (0, 1, \ldots, 0)^T$, so $\langle \mathbf{u}, \mathbf{v}\rangle = 1$. The subgroup that keeps $\mathbf{u} = (1, 0, \ldots, 0)^T$ unchanged consists of all the elements of $\mathcal{S}_N$ with $\mathbf{u}$ as the first column of the matrix. Hence, by Lemma 1, we can select the second column of the matrix to be any vector, which has the symplectic form of 1 with the first column, which is $\mathbf{u}$. Thus, we can map $\mathbf{v}$ to any other vector with the symplectic form one with $\mathbf{u}$, which is also unchanged. Then, by noting that the product of symplectic matrices is a symplectic matrix, the subgroup can map any vector with the symplectic form of one with $\mathbf{u}$ to any other. Similarly, this argument applies to the other case where the symplectic form has a value of zero.

Then, since $P(S\mathbf{v}) = P(\mathbf{v})$, all vectors that give the same value for $\langle \mathbf{v}, \mathbf{u}\rangle$ have the same probability. Thus, we get the probability distribution in (106).

Finally, we note that since via a symplectic transformation $\mathbf{u}$ can be mapped to any other vector in $\mathbb{Z}_2^{2N}$ and that the product of two symplectic matrices is symplectic, this result applies for any $\mathbf{u} \in \mathbb{Z}_2^{2N}$. □

Lemma 16. For an initial vector $\mathbf{u}^0 \in \mathcal{V}_0$ supported on location $x = 0$, the probability that the value of the symplectic form between the evolved vector $\mathbf{u}^t = S_{\text{chain}}^t \mathbf{u}^0$, with integer t, and the initial vector $\langle \mathbf{u}^t, \mathbf{u}^0\rangle = \langle \mathbf{u}_0^t, \mathbf{u}_0^0\rangle$ is equal to s and has an s-independent upper bound given by

$$\text{prob}\{\langle \mathbf{u}_0^t, \mathbf{u}_0^0\rangle = s\} \le \frac{1}{2} + 8\, t\, 2^{-N}. \tag{108}$$

Furthermore, this result is independent of the location of the support of $\mathbf{u}^0$, provided that it is a single site.

Proof. To prove this lemma, we proceed similarly as in Lemma 9. That is, we consider a sequence of gates in the causal past of $\mathbf{u}_0^t$ with an elbow shape (see example in Fig. 7). More concretely, we write $\mathbf{u}_0^t$ as

$$\mathbf{u}_{1-t}^{t/2} = B_{1-t} \cdots B_{-2} B_{-1} A_0 \mathbf{u}_0^0, \tag{109}$$

$$\mathbf{u}_{1-t}^{t/2+1/2} = D_{-t} \mathbf{u}_{1-t}^{t/2}, \tag{110}$$

$$\mathbf{u}_0^t = C_{-1} \cdots C_{2-t} C_{1-t} \mathbf{u}_{1-t}^{t/2+1/2} + \mathbf{w}, \tag{111}$$

where, crucially, the random vector $\mathbf{w}$ is independent of the random matrix D_{-t}. This vector $\mathbf{w}$ is defined in a way similar to (82).

Next, we follow a sequence of steps similar to those from (91)–(93). First, we write

$$\begin{aligned}
&\text{prob}\Big\{\langle \mathbf{u}_0^0, \mathbf{u}_0^t\rangle = s\Big\}\\
&= \text{prob}\Big\{\langle \mathbf{u}_0^0, \mathbf{u}_0^t\rangle = s \text{ and } \mathbf{u}_{1-t}^{t/2} = \mathbf{0}\Big\} + \text{prob}\Big\{\langle \mathbf{u}_0^0, \mathbf{u}_0^t\rangle = s \text{ and } \mathbf{u}_{1-t}^{t/2} \ne \mathbf{0}\Big\}\\
&\le \text{prob}\Big\{\mathbf{u}_{1-t}^{t/2} = \mathbf{0}\Big\} + \text{prob}\Big\{\langle \mathbf{u}_0^0, \mathbf{u}_0^t\rangle = s \text{ and } \mathbf{u}_{1-t}^{t/2} \ne \mathbf{0}\Big\}.
\end{aligned} \tag{112}$$

Second, we bound the first term by using the recursive relation (75) as

$$\text{prob}\Big\{\mathbf{u}_{1-t}^{t/2} = \mathbf{0}\Big\} \le 2\, t\, 2^{-2N}. \tag{113}$$

Third, we introduce the uniformly distributed random vectors $\mathbf{u}, \mathbf{u}' \in \mathbb{Z}_2^{2N}$, which are independent of the gates $S_{1-t}, S_{2-t}, \ldots, S_{-2}$, and write

$$\begin{aligned}
&\text{prob}\Big\{\langle \mathbf{u}_0^0, \mathbf{u}_0^t\rangle = s \text{ and } \mathbf{u}_{1-t}^{t/2} \ne \mathbf{0}\Big\}\\
&= \text{prob}\Big\{\mathbf{u}_0^{0\,T} J C_{-1} \cdots C_{1-t} \mathbf{u}_{1-t}^{t/2+1/2} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle \text{ and } \mathbf{u}_{1-t}^{t/2} \ne \mathbf{0}\Big\}\\
&\le \text{d}\Big(\mathbf{u}_{1-t}^{t/2+1/2}, \mathbf{u}\Big) + \text{prob}\Big\{\mathbf{u}_0^{0\,T} J C_{-1} \cdots C_{1-t} \mathbf{u} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle \text{ and } \mathbf{u}_{1-t}^{t/2} \ne \mathbf{0}\Big\}\\
&\le \text{d}\Big(\mathbf{u}_{1-t}^{t/2+1/2}, \mathbf{u}\Big) + \text{prob}\Big\{\mathbf{u}_0^{0\,T} J C_{-1} \cdots C_{1-t} \mathbf{u} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle\Big\}\\
&\le \text{d}\Big(\mathbf{u}_{1-t}^{t/2+1/2}, \mathbf{u}\Big) + \text{d}\Big(J C_{-1}^T J \mathbf{u}_0^0, \mathbf{u}'\Big) + \text{prob}\Big\{\mathbf{u}'^T J C_{-2} \cdots C_{1-t} \mathbf{u} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle\Big\}.
\end{aligned} \tag{114}$$

Fourth, using (92), we can write the bounds

$$\begin{aligned}
\text{d}\Big(\mathbf{u}_{1-t}^{t/2+1/2}, \mathbf{u}\Big) &\le 2^{1-4N},\\
\text{d}\Big(J C_{-1}^T J \mathbf{u}_0^0, \mathbf{u}'\Big) &\le 2^{1-4N}.
\end{aligned} \tag{115}$$

Fifth, in order to bound the third term in (114), we note that for any non-zero $\mathbf{a} \in \mathbb{Z}_2^{2N}$, we have $\text{prob}\{\mathbf{u}'^T \mathbf{a} = s\} = 1/2$ for both $s = 0, 1$; therefore,

$$\begin{aligned}
&\text{prob}\Big\{\mathbf{u}'^T J C_{-2} \cdots C_{1-t} \mathbf{u} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle\Big\}\\
&= \text{prob}\Big\{\mathbf{u}'^T J C_{-2} \cdots C_{1-t} \mathbf{u} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle \text{ and } C_{-2} \cdots C_{1-t} \mathbf{u} \ne \mathbf{0}\Big\}\\
&+ \text{prob}\Big\{\mathbf{u}'^T J C_{-2} \cdots C_{1-t} \mathbf{u} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle \text{ and } C_{-2} \cdots C_{1-t} \mathbf{u} = \mathbf{0}\Big\}.
\end{aligned}$$

Next we bound the first term by using the fact that the uniformly distributed vector $\mathbf{u}'$ is independent of $\mathbf{a} := JC_{-2}\cdots C_{1-t}\mathbf{u}$ and $\langle \mathbf{u}_0^0, \mathbf{w}\rangle$, as

$$\text{prob}\Big\{\mathbf{u}'^T\mathbf{a} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle \text{ and } \mathbf{a} \neq \mathbf{0}\Big\} \leq \text{prob}\Big\{\mathbf{u}'^T\mathbf{a} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle | \, \mathbf{a} \neq \mathbf{0}\Big\} = \frac{1}{2}.$$

The second term can be easily bounded as

$$\begin{aligned} &\text{prob}\Big\{\mathbf{u}'^T JC_{-2}\cdots C_{1-t}\mathbf{u} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle \text{ and } C_{-2}\cdots C_{1-t}\mathbf{u} = \mathbf{0}\Big\} \\ &\leq \text{prob}\{C_{-2}\cdots C_{1-t}\mathbf{u} = \mathbf{0}\} \;\leq\; 8(t-2)2^{-N}, \end{aligned}$$

where the last inequality follows from Lemma 6. Combining the above two bounds, we obtain

$$\text{prob}\Big\{\mathbf{u}'^T JC_{-2}\cdots C_{1-t}\mathbf{u} = s + \langle \mathbf{u}_0^0, \mathbf{w}\rangle\Big\} \;\leq\; \frac{1}{2} + 8(t-2)2^{-N}.$$

Sixth, putting everything together back from (112), we arrive at

$$\begin{aligned} \text{prob}\Big\{\langle \mathbf{u}_0^0, \mathbf{u}_0^t\rangle = s\Big\} &\leq 2\,t\,2^{-2N} + 4\,2^{-4N} + \frac{1}{2} + 8(t-2)2^{-N} \\ &\leq \frac{1}{2} + 8\,t\,2^{-N}, \end{aligned}$$

as we wanted to show. □

Lemma 17. For an initial vector $\mathbf{u}^0 \in \mathcal{V}_0$ supported at $x = 0$, the probability distribution of the evolved vector $\mathbf{u}^t = S_{\text{chain}}^t\mathbf{u}^0$ at integer times, conditioned on the evolved vector being non-zero at every site and different from the initial single-site non-zero vector, $\mathbf{u}_x^t \neq 0 \;\forall x \in \mathbb{Z}_L$ and $\mathbf{u}_0^t \neq \mathbf{u}_0^0$, after the scrambling time t_{scr} is of the form

$$\text{prob}\{\mathbf{u}^t|\mathbf{u}_x^t \neq 0 \;\forall x \in \mathbb{Z}_L, \mathbf{u}_0^t \neq \mathbf{u}_0^0\} \leq \frac{1}{(2^{2N}-1)^{L-1}}\begin{cases} \dfrac{8t2^{-N} + 1/2}{2^{2N-1}-2} & \text{if } \langle \mathbf{u}_0^t, \mathbf{u}_0^0\rangle = 0, \\ \dfrac{8t2^{-N} + 1/2}{2^{2N-1}} & \text{if } \langle \mathbf{u}_0^t, \mathbf{u}_0^0\rangle = 1, \end{cases}$$

and before the scrambling time for all sites within the causal light cone, the probability distribution is of the form

$$\text{prob}\{\mathbf{u}^t|\mathbf{u}_x^t \neq 0 \;\forall x \in [-2t+1, 2t], \mathbf{u}_0^t \neq \mathbf{u}_0^0\} \leq \frac{1}{(2^{2N}-1)^{4t-1}}\begin{cases} \dfrac{8t2^{-N} + 1/2}{2^{2N-1}-2} & \text{if } \langle \mathbf{u}_0^t, \mathbf{u}_0^0\rangle = 0, \\ \dfrac{8t2^{-N} + 1/2}{2^{2N-1}} & \text{if } \langle \mathbf{u}_0^t, \mathbf{u}_0^0\rangle = 1. \end{cases} \tag{116}$$

Furthermore, this result holds for any choice of the single site at which the initial vector is non-zero.

Proof. The proof of this lemma uses the twirling technique discussed in Sec. V A, Lemma 10. The probability distribution of the evolved vector $\mathbf{u}^t = (S_{\text{chain}})^t\mathbf{u}^0$ at integer times is identical to

$$\begin{aligned} \mathbf{u}^t &= \Big(\bigoplus_{x=0}^{L-1} X_x\Big) S_{\text{chain}}^t \Big(\bigoplus_{x=0}^{L-1} X_x^{-1}\Big)\mathbf{u}^0 \\ &= \Big(\bigoplus_{x=0}^{L-1} X_x\Big) S_{\text{chain}}^t \Big(X_0^{-1}\mathbf{u}_0^0 \bigoplus_{x=1}^{L-1}\mathbf{0}\Big), \end{aligned} \tag{117}$$

where $X_x \in \mathcal{S}_N$ are arbitrary single-site symplectic matrices. Equation (117) follows from the fact that $\mathbf{u}^0$ has been assumed supported at $x = 0$; therefore, $\big(\oplus_{x=0}^{L-1} X_x^{-1}\big)\mathbf{u}^0$ is supported at $x = 0$ as well. If we restrict X_0 to the elements of $\mathcal{S}_N$ that satisfy $X_0\mathbf{u}_0^0 = \mathbf{u}_0^0$, then the probability distribution of $\mathbf{u}^t$ is identical to $\big(\oplus_{x=0}^{L-1} X_x\big)\mathbf{u}^t$. Since the choice of symplectic matrices $\oplus_{x=0}^{L-1} X_x$ to twirl is arbitrary, we can take each single-site matrix to be independent and uniformly distributed over all single-site symplectic matrices, except for X_0 which is uniformly distributed over the restricted set satisfying $X_0\mathbf{u}_0^0 = \mathbf{u}_0^0$. Then, we condition on the evolved vector being non-zero at all sites x and different for the initial single-site non-zero vector, $\mathbf{u}_x^t \neq 0 \;\forall x \in \mathbb{Z}_L$ and $\mathbf{u}_0^t \neq \mathbf{u}_0^0$. Therefore, under this condition, the evolved vector at each site is independent and uniformly distributed over all non-zero vectors (Lemma 3) apart from the initial vector $\mathbf{u}_0^0$. On the vector space $\mathcal{V}_0$, we invoke Lemma 15, and hence, the evolved vector $\mathbf{u}_0^t$ $(\neq \mathbf{0}, \mathbf{u}_0^0)$ at $x = 0$ is uniformly distributed over all the vectors with the same symplectic form with $\mathbf{u}_0^0$, $\langle \mathbf{u}_0^t, \mathbf{u}_0^0\rangle$. Hence, using Lemma 16, which gives an upper bound for the probability of $\langle \mathbf{u}_0^t, \mathbf{u}_0^0\rangle \in \{0, 1\}$, we get the stated result. □

The following theorem establishes approximate Pauli mixing: the probability that $\mathbf{u}$ evolves onto $\mathbf{u}'$ after a time t, given by

$$P_t(\mathbf{u}'|\mathbf{u}) = \mathop{\mathbb{E}}_{\{U_x\}} \left|2^{-NL}\mathrm{tr}\Big(\sigma_{\mathbf{u}'} W(t)\sigma_{\mathbf{u}} W(t)^\dagger\Big)\right|, \tag{118}$$

is close to the uniform distribution over all non-zero vectors $\mathbf{u}'$ in the causal subspace (8) denoted by $Q_t(\mathbf{u}')$. After the scrambling time $t \geq t_{\mathrm{scr}}$, $Q_t(\mathbf{u}')$ is the uniform distribution over all non-zero vectors in the total phase space $\mathcal{V}_{\mathrm{chain}}$.

Theorem 18 (approximate Pauli mixing). If the initial Pauli operator $\sigma_{\mathbf{u}}$ is supported at site $x = 0$, then the probability distribution (7) for its evolution $\sigma_{\mathbf{u}'}$ is close to uniform inside the light cone,

$$\sum_{\mathbf{u}'}\left|P_t(\mathbf{u}'|\mathbf{u}) - Q_t(\mathbf{u}')\right| \ \leq\ 130 \times t^2\, 2^{-N}, \tag{119}$$

for any integer or half-integer time $t \in [1/2, 2t_{\mathrm{scr}}]$. An analogous statement holds for any other initial location $x \neq 0$.

Remark. The following is an alternative enunciation of Theorem 18 formulated in phase space rather than Hilbert space. A proof of this alternative form then follows.

Theorem 18 (alternative form). For an initial vector supported at $x = 0$, the evolved vector $\mathbf{u}^t = S^t_{\mathrm{chain}}\mathbf{u}^0$, at integer times, is approximately uniformly distributed over all non-zero vectors within the light cone. For any $t \in [1, t_{\mathrm{scr}}]$ and $x \in [-2t+1, 2t]$, we have

$$\sum_{\mathbf{v}\in\mathbb{Z}_2^{8Nt}}\left|\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}\} - \frac{1}{2^{8Nt}-1}\right| \leq\ 32t(4t+1)2^{-N} + 4t2^{-2N}. \tag{120}$$

For any $t \in [t_{\mathrm{scr}}, 2t_{\mathrm{scr}}]$, it holds that

$$\sum_{\mathbf{v}\in\mathbb{Z}_2^{2NL}}\left|\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}\} - \frac{1}{2^{2NL}-1}\right| \leq 32t(L+1)2^{-N} + L2^{-2N}.$$

Proof. Let us consider the case $t \geq t_{\mathrm{scr}}$ first. Similarly to the Proof of Theorem 14, we employ

$$\begin{aligned}\mathrm{prob}(A) &= \mathrm{prob}(A\wedge B) + \mathrm{prob}(A\wedge \bar{B})\\ &= \mathrm{prob}(A|B)\mathrm{prob}(B) + \mathrm{prob}(A|\bar{B})\mathrm{prob}(\bar{B}),\end{aligned}$$

where A and B are events in a probability space. With $q \equiv \mathrm{prob}\{\mathbf{u}^t_x \neq \mathbf{0}\ \forall x\in\mathbb{Z}_L \wedge \mathbf{u}^t_0 \neq \mathbf{u}^0_0\}$, $\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}\}$ is then rewritten in the following way:

$$\begin{aligned}\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}\} &= q\,\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}|\mathbf{u}^t_x \neq \mathbf{0}\ \forall x\in\mathbb{Z}_L \wedge \mathbf{u}^t_0 \neq \mathbf{u}^0_0\}\\ &\quad + (1-q)(\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}|\ \exists x\in\mathbb{Z}_L \text{ such that } \mathbf{u}^t_x = \mathbf{0} \vee \mathbf{u}^t_0 = \mathbf{u}^0_0\}).\end{aligned}$$

Summing and subtracting $q\frac{1}{2^{2NL}-1}$ into the sum over $\mathbf{v}$ and using the triangular inequality, we find that

$$\begin{aligned}&\sum_{\mathbf{v}}\left|\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}\} - \frac{1}{2^{2NL}-1}\right|\\ &\leq q\sum_{\mathbf{v}}\left|\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}|\mathbf{u}^t_x \neq \mathbf{0}\ \forall x\in\mathbb{Z}_L, \mathbf{u}^t_0 \neq \mathbf{u}^0_0\} - \frac{1}{2^{2NL}-1}\right|\\ &+ (1-q)\sum_{\mathbf{v}}\left|\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}|\ \exists x\in\mathbb{Z}_L \text{ such that } \mathbf{u}^t_x = \mathbf{0} \vee \mathbf{u}^t_0 = \mathbf{u}^0_0\} - \frac{1}{2^{2NL}-1}\right|.\end{aligned}$$

We can bound the first term using $q \leq 1$ and apply Lemma 17 to find that

$$q\sum_{\mathbf{v}}\left|\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}|\mathbf{u}^t_x \neq \mathbf{0}\ \forall x\in\mathbb{Z}_L, \mathbf{u}^t_0 \neq \mathbf{u}^0_0\} - \frac{1}{2^{2NL}-1}\right| \leq 16t2^{-N} + (L+1)2^{-2N}.$$

To evaluate the second term above, we upper bound the sum with its maximum value of 2 and use the result of Lemma 9 to find that

$$(1-q)\sum_{\mathbf{v}}\left|\mathrm{prob}\{\mathbf{u}^t = \mathbf{v}|\ \exists x\in\mathbb{Z}_L \text{ such that } \mathbf{u}^t_x = \mathbf{0} \vee \mathbf{u}^t_0 = \mathbf{u}^0_0\} - \frac{1}{2^{2NL}-1}\right| \leq 32t(L+1)2^{-N}.$$

Combining this gives the stated result for integer times after the scrambling time.

To derive the results for integer times before the scrambling time, we note that the derivation is identical with the substitution $L \to 4t$, which agree when $t = t_{\text{scr}}$ (and after this time). □

D. Approximate mixing with arbitrary initial state

Consider a subsystem of the chain comprising L_{s} consecutive sites, where L_{s} is even. Without loss of generality, we choose this subsystem to be $\{1, 2, \ldots, L_{\text{s}}\} \subseteq \mathbb{Z}_L$. We analyze the state of this subsystem at times

$$t \le \frac{L - L_{\text{s}}}{4}. \tag{121}$$

This condition ensures that the left backward wave front of $\mathbf{u}_1^t$ and the right backward wave front of $\mathbf{u}_{L_{\text{s}}}^t$ do not collide. Without this condition, the analysis becomes very complicated.

Lemma 19. Consider an initial vector $\mathbf{u}^0 \in \mathcal{V}_{\text{chain}}$ supported on all lattice sites ($\mathbf{u}_x^0 \ne \mathbf{0}$ for all $x \in \mathbb{Z}_L$), and its evolution at time t, $\mathbf{u}^t$. Define the random variable $s_x = \langle \mathbf{u}_x^t, \mathbf{u}_x^0 \rangle$ at each site of the region $x \in \{1, \ldots, L_{\text{s}}\} \subseteq \mathbb{Z}_L$, where L_{s} is even. Then, we have

$$P(s_1, \ldots, s_{L_{\text{s}}}) \le 2^{-L_{\text{s}}} + 32\, t\, 3^{\frac{L_{\text{s}}}{2}+1}\, 2^{-N}, \tag{122}$$

as long as $t \le (L - L_{\text{s}})/4$.

Proof. The value of the random vectors $\mathbf{u}_1^t, \ldots, \mathbf{u}_{L_{\text{s}}}^t$ is only determined by the random matrices $S_{2-2t}, \ldots, S_{L_{\text{s}}+2t-2}$. The rest of matrices S_x are not contained in the causal past of the region under consideration $\{1, 2, \ldots, L_{\text{s}}\}$. In order to simplify this proof, we will replace $S_{2-2t}, \ldots, S_{L_{\text{s}}+2t-2}$ by a new set of random variables defined in what follows.

Let us label by $y \in \{1, \ldots, L_{\text{s}}/2\}$ the pair of neighboring sites $\{2y-1, 2y\} \subseteq \{1, \ldots, L_{\text{s}}\}$. For each pair y, we consider a given non-zero vector $\mathbf{a}_y \in \mathbb{Z}_2^{4N}$ and define the random variables

$$\mathbf{b}_y = S_{2y-1}^{-1} \mathbf{a}_y, \tag{123}$$

$$h_y = \left\langle \mathbf{a}_y, \mathbf{u}_{2y-1}^t \oplus \mathbf{u}_{2y}^t \right\rangle = \left\langle \mathbf{b}_y, \mathbf{u}_{2y-1}^{t-1/2} \oplus \mathbf{u}_{2y}^{t-1/2} \right\rangle. \tag{124}$$

The left-most random contribution to h_y is the matrix S_{2y-2t} or equivalently the vector $\mathbf{w}_y$, defined through

$$\tilde{\mathbf{w}}_y \oplus \mathbf{w}_y = S_{2y-2t}(\mathbf{u}_{2y-2t}^0 \oplus \mathbf{u}_{2y-2t+1}^0). \tag{125}$$

We note that $\mathbf{w}_y \in \mathcal{V}_{2y-2t+1}$. This contribution and others are illustrated in Fig. 9. The contribution of the vector $\mathbf{w}_y$ to h_y (and $\mathbf{u}_{2y-1}^{t-1/2}$) is "transmitted through" the matrices $S_{2y-2}, S_{2y-3}, \ldots, S_{2y-2t+2}, S_{2y-2t+1}$. More precisely, $\mathbf{w}_y$ is mapped via the matrix product

$$F_y = C_{2y-2} C_{2y-3} \cdots C_{2y-2t+2} C_{2y-2t+1}, \tag{126}$$

where we have used decomposition (28). We denote by $\mathbf{v}_y$ all contributions to $\mathbf{u}_{2y-1}^{t-1/2}$ that are not $F_y \mathbf{w}_y$,

$$\mathbf{v}_y = (\mathbf{u}_{2y-1}^{t-1/2} + F_y \mathbf{w}_y) \oplus \mathbf{u}_{2y}^{t-1/2}. \tag{127}$$

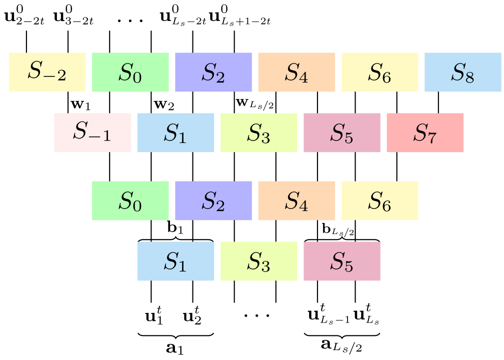


FIG. 9. The region $\{1, 2, \ldots, 6\}$ at time $t = 2$, and its causal past back to $t = 0$. (Hence, $L_{\text{s}} = 6$.) All the random matrices $S_{-2}, \ldots, S_8$ contribute to the value of the vectors $\mathbf{u}_1^t, \ldots, \mathbf{u}_6^t$. The left-most contribution to the vector $\mathbf{u}_1^t$ is the matrix S_{-2} or equivalently the vector $\mathbf{w}_1$. The given vector $\mathbf{a}_y$ associated with the pair of neighboring sites y and its 1/2-step backward time translations $\mathbf{b}_y$ are also represented.

We remark that $\mathbf{v}_y \in \mathcal{V}_{2y-1} \oplus \mathcal{V}_{2y}$. The last random variable that we need to define is $g_y = \langle \mathbf{b}_y, \mathbf{v}_y \rangle$, which together with (124) allows us to write

$$h_y = \langle \mathbf{b}_y, F_y \mathbf{w}_y + \mathbf{v}_y \rangle = \langle \mathbf{b}_y, F_y \mathbf{w}_y \rangle + g_y. \tag{128}$$

Note the slight abuse of notation in that we write $F_y \mathbf{w}_y$ instead of $F_y \mathbf{w}_y \oplus \mathbf{0}$.

In summary, we have replaced the variables $S_{2-2t}, \ldots, S_{L_s+2t-2}$ by the variables $\mathbf{w}_y, \mathbf{b}_y, F_y, g_y$ for $y = 1, \ldots, L_s/2$. (We are not using $\mathbf{v}_y, \tilde{\mathbf{w}}_y$ anymore.) These variables are not all independent, but they satisfy the following independence relations:

- $\mathbf{w}_1, \mathbf{b}_1, \ldots, \mathbf{w}_{L_s/2}, \mathbf{b}_{L_s/2}$ are independent and uniform.
- $\mathbf{w}_y$ is independent of $g_{y'}$ for all $y' \geq y$.
- F_y is independent of $\mathbf{w}_{y'}$ and $\mathbf{b}_{y''}$ for all $y' \leq y$ and $y'' \geq y$.

To continue with the proof, it is convenient to introduce the following notation:

$$\mathbf{u}_{\geq y} = (\mathbf{u}_y, \mathbf{u}_{y+1}, \ldots, \mathbf{u}_{L_s/2}), \tag{129}$$

$$\mathbf{u}_{\leq y} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_y), \tag{130}$$

and analogously for $>, <$ and the rest of variables $\mathbf{b}_y, F_y, g_y$. This allows us to write the joint probability distribution of $h_1, \ldots, h_{L_s/2}$ as

$$P(h_{\geq 1}) = \sum_{\mathbf{w}_{\geq 1}, \mathbf{b}_{\geq 1}, F_{\geq 1}, g_{\geq 1}} P(\mathbf{w}_{\geq 1}, \mathbf{b}_{\geq 1}, F_{\geq 1}, g_{\geq 1}) \prod_y \delta(h_y, \langle \mathbf{b}_y, F_y \mathbf{w}_y \rangle + g_y). \tag{131}$$

Equation (131) follows directly from the definition of the Kronecker-delta. Note that we can write the above distribution $P(\mathbf{w}_{\geq 1}, \mathbf{b}_{\geq 1}, F_{\geq 1}, g_{\geq 1})$ as

$$\begin{aligned} P(\mathbf{w}_{\geq 1}, \mathbf{b}_{\geq 1}, F_{\geq 1}, g_{\geq 1}) = &\sum_{S_0, S_1, \ldots, S_{L-1}} P(S_0) P(S_1) \cdots P(S_{L-1}) \\ &\times \prod_{y=1}^{L_s/2} \delta\Big(\mathbf{w}_y, S_{2y-2t}\big[\boldsymbol{u}^0_{2y-2t} \oplus \boldsymbol{u}^0_{2y-2t+1}\big]\Big) \times \delta\big(\mathbf{b}_y, S^{-1}_{2y-1} \mathbf{a}_y\big) \\ &\times \delta(F_y, C_{2y-2} \cdots C_{2y-2t+1}) \times \delta\Big(g_y, \Big\langle \mathbf{b}_y, (\mathbf{u}^{t-1/2}_{2y-1} + F_y \mathbf{w}_y) \oplus \mathbf{u}^{t-1/2}_{2y} \Big\rangle\Big). \end{aligned} \tag{132}$$

The following sum-rule is repeatedly exploited below, where $\mathbf{0}_{2N}$ denotes the $2N \times 2N$ matrix with all entries equal to 0; instead, $\mathbf{0}$ is the vector with $2N$ components equal to 0,

$$\sum_{\mathbf{w}_1} P(\mathbf{w}_1) \delta(h_1, \langle \mathbf{b}_1, F_1 \mathbf{w}_1 \rangle + g_1) = \begin{cases} \delta(h_1, g_1) & \text{if } (F_1 \oplus \mathbf{0}_{2N})^T J \mathbf{b}_1 = \mathbf{0}, \\ 1/2 & \text{otherwise.} \end{cases} \tag{133}$$

Equation (133) is obtained as follows. We first note that

$$\langle \mathbf{b}_1, F_1 \mathbf{w}_1 \oplus \mathbf{0} \rangle = \langle \mathbf{b}_1, (F_1 \oplus \mathbf{0}_{2N})(\mathbf{w}_1 \oplus \mathbf{0}) \rangle = \mathbf{b}_1^T J (F_1 \oplus \mathbf{0}_{2N})(\mathbf{w}_1 \oplus \mathbf{0}).$$

If $(F_1 \oplus \mathbf{0}_{2N})^T J \mathbf{b}_1 = \mathbf{0}$, the first of Eq. (133) follows from the normalization of the probability $P(\mathbf{w}_1)$. If $(F_1 \oplus \mathbf{0}_{2N})^T J \mathbf{b}_1 \neq \mathbf{0}$ since the values of $\mathbf{w}_1$ are distributed uniformly over all the vectors of $\mathbb{Z}_2^{2N}$, implying that $P(\mathbf{w}_1) = \frac{1}{2^{2N}}$, then $\langle \mathbf{b}_1, F_1 \mathbf{w}_1 \rangle$ takes half of the times the value 0 and half of the times the value 1. Recall that F_1 and $\mathbf{b}_1$ are fixed in (133). The second equation of (133) then follows.

Using $\delta(h, h') \leq 1$ for all h, h' and (133), we can write

$$\begin{aligned} P(h_{\geq 1}) = &\sum_{\mathbf{w}_{\geq 1}, \mathbf{b}_{\geq 1}, F_{\geq 1}, g_{\geq 1}} P(\mathbf{w}_1) P(\mathbf{w}_{\geq 2}, \mathbf{b}_{\geq 1}, F_{\geq 1}, g_{\geq 1}) \prod_y \delta(h_y, \langle \mathbf{b}_y, F_y \mathbf{w}_y \rangle + g_y) \\ &\leq \sum_{\mathbf{w}_{\geq 2}, \mathbf{b}_{\geq 1}, F_{\geq 1}, g_{\geq 1}} \delta\Big((F_1 \oplus \mathbf{0}_{2N})^T J \mathbf{b}_1, \mathbf{0}\Big) P(\mathbf{w}_{\geq 2}, \mathbf{b}_{\geq 1}, F_{\geq 1}, g_{\geq 1}) \prod_{y \geq 2} \delta(h_y, \langle \mathbf{b}_y, F_y \mathbf{w}_y \rangle + g_y) \\ &+ \frac{1}{2} \sum_{\mathbf{w}_{\geq 2}, \mathbf{b}_{\geq 2}, F_{\geq 2}, g_{\geq 2}} P(\mathbf{w}_{\geq 2}, \mathbf{b}_{\geq 2}, F_{\geq 2}, g_{\geq 2}) \prod_{y \geq 2} \delta(h_y, \langle \mathbf{b}_y, F_y \mathbf{w}_y \rangle + g_y) \\ &\leq \text{prob}\{(F_1 \oplus \mathbf{0}_{2N})^T J \mathbf{b}_1 = \mathbf{0}\} \\ &+ \frac{1}{2} \sum_{\mathbf{w}_{\geq 2}, \mathbf{b}_{\geq 2}, F_{\geq 2}, g_{\geq 2}} P(\mathbf{w}_{\geq 2}, \mathbf{b}_{\geq 2}, F_{\geq 2}, g_{\geq 2}) \prod_{y \geq 2} \delta(h_y, \langle \mathbf{b}_y, F_y \mathbf{w}_y \rangle + g_y). \end{aligned}$$

To bound the term associated with the case $(F_1 \oplus \mathbf{0}_{2N})^T J \mathbf{b}_1 \neq \mathbf{0}$, we extended the sum over $\mathbf{b}_1, F_1$ from the values satisfying $(F_1 \oplus \mathbf{0}_{2N})^T J \mathbf{b}_1 \neq \mathbf{0}$ to all values. Since the variables $\mathbf{b}_1, F_1, g_1$ do not appear in any of the remaining δ-functions, we can trace them out. Subsequently, we repeat the above process by summing over $\mathbf{w}_2$, using the analog of (133) for $y = 2$, and summing over $\mathbf{w}_2, F_2, g_2$, obtaining

$$P(h_{\geq 1}) = \epsilon + \frac{1}{2}\left(\epsilon + \frac{1}{2}\sum P(\mathbf{w}_{\geq 3}, \mathbf{b}_{\geq 3}, F_{\geq 3}, g_{\geq 3}) \prod_{y \geq 3} \delta(h_y, \langle \mathbf{b}_y, F_y \mathbf{w}_y \rangle + g_y)\right),$$

where we define $\epsilon = \text{prob}\{(F_1 \oplus \mathbf{0}_{2N})^T J \mathbf{b}_1 = \mathbf{0}\}$. Continuing in this fashion yields

$$\begin{aligned} P(h_1, \ldots, h_{L_s/2}) &= \epsilon \sum_{k=0}^{L_s/2-1} 2^{-k} + 2^{-L_s/2} \\ &\leq 2\epsilon + 2^{-L_s/2}. \end{aligned} \tag{134}$$

We now wish to turn this bound from a distribution of h_y to the distribution of $s_x \equiv \langle \mathbf{u}_x^t, \mathbf{u}_x^0 \rangle$ (recalling that $x \in \{1, 2, \ldots, L_s\}$ and $y \in \{1, \ldots, L_s/2\}$), that is to say, we want to bound $P(s_1, s_2, \ldots, s_{L_s})$.

Let us consider first the simplest case, that is, $L_s = 2$, $h_1 = s_1 + s_2$. The couple (s_1, s_2) has four possible realizations $(0,0), (1,0), (0,1), (1,1)$. We have the following bounds:

$$\begin{aligned} P(h_1 = 0) &= P(0,0) + P(1,1) \leq \frac{1}{2} + 2\epsilon, \\ P(h_1 = 1) &= P(0,1) + P(1,0) \leq \frac{1}{2} + 2\epsilon, \\ P(0,0) + P(1,1) &\geq \frac{1}{2} - 2\epsilon. \end{aligned} \tag{135}$$

The last bound combines normalization $P(0,0) + P(1,1) + P(0,1) + P(1,0) = 1$ and the second bound above. Similarly, it holds that $P(0,1) + P(1,0) \geq \frac{1}{2} - 2\epsilon$.

The bound (134) extends to the case where rather than the values of $h_y \equiv s_{2y-1} + s_{2y}$ being fixed, the values of certain h_y and of certain s_x are fixed. In the case of $L_s = 2$, this amounts to three cases: h_1 fixed, s_1 fixed, and s_2 fixed. It then follows that

$$\begin{aligned} P(s_1 = 0) &= P(0,0) + P(0,1) \leq \frac{1}{2} + 2\epsilon, \\ P(s_1 = 1) &= P(1,0) + P(1,1) \leq \frac{1}{2} + 2\epsilon, \\ P(s_2 = 0) &= P(0,0) + P(1,0) \leq \frac{1}{2} + 2\epsilon, \\ P(s_2 = 1) &= P(0,1) + P(1,1) \leq \frac{1}{2} + 2\epsilon. \end{aligned} \tag{136}$$

The lower bound can be obtained similarly to what we have done above, for example,

$$P(0,1) + P(1,1) \geq \frac{1}{2} - 2\epsilon. \tag{137}$$

Summing (135) with (136) and subtracting (137), we obtain $P(0,0) \leq \frac{1}{4} + 3\epsilon$. With a similar approach, we obtain $P(0,0) \geq \frac{1}{4} - 3\epsilon$. The procedure that we have described holds for all $P(s_1, s_2)$; then,

$$\frac{1}{4} - 3\epsilon \leq P(s_1, s_2) \leq \frac{1}{4} + 3\epsilon.$$

We introduce a matrix formalism to re-obtain the result above; this formalism will allow us to treat the case $L_s > 2$. The set of inequalities (135) and (136), with the respective lower bounds, can be written as

$$\begin{pmatrix} \frac{1}{2} - 2\epsilon \\ \frac{1}{2} - 2\epsilon \\ \frac{1}{2} - 2\epsilon \\ \frac{1}{2} - 2\epsilon \\ \frac{1}{2} - 2\epsilon \\ \frac{1}{2} - 2\epsilon \end{pmatrix} \leq \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} P(0,0) \\ P(0,1) \\ P(1,0) \\ P(1,1) \end{pmatrix} \leq \begin{pmatrix} \frac{1}{2} + 2\epsilon \\ \frac{1}{2} + 2\epsilon \\ \frac{1}{2} + 2\epsilon \\ \frac{1}{2} + 2\epsilon \\ \frac{1}{2} + 2\epsilon \\ \frac{1}{2} + 2\epsilon \end{pmatrix}. \tag{138}$$

We denote the 6×4 matrix above with A. With regard to the system of inequalities above, we have already shown explicitly the solution of it; in particular, we saw that to find the upper bound $P(s_1, s_2) \leq \frac{1}{4} + 3\epsilon$, we need both upper and lower bounds in (135) and (136). The same holds for the lower bound. It is easy to describe a way to obtain a solution of (138) where we get exactly the term of $O(1)$ and we overestimate the correction $O(\epsilon)$. Since in (138) the term of $O(1)$ is the same in both the upper bound and the lower bound, then the term of $O(1)$ in (138) is obtained replacing the inequalities with equality. The solution of the corresponding system is also easily obtained by inspection; in fact, since every row of A has two entries equal to 1, a solution of the system is $P(s_1, s_2) = \frac{1}{4}$ for all (s_1, s_2) since A is a full rank matrix and this is also the only solution. To evaluate the error, we consider the following equality, where $\mathbf{a}_j$ is the jth row of the matrix A and $\mathbf{P}$ is the column vector of probabilities, as in Eq. (138):

$$\frac{1}{2}(\mathbf{a}_3 + \mathbf{a}_5 - \mathbf{a}_2) \cdot \mathbf{P} = P(0,0). \tag{139}$$

The equation above can be generalized to every $P(s_1, s_2)$. This means that, in general, we only need to sum or subtract three among the inequalities in (138) to obtain any $P(s_1, s_2)$; therefore, the maximal error in modulus that can arise is equal to 6ϵ; then, we can rewrite

$$\frac{1}{4} - 6\epsilon \leq P(s_1, s_2) \leq \frac{1}{4} + 6\epsilon. \tag{140}$$

It is easy to understand that each row of the matrix A carries a "label" as specified below; in fact, for example, the product of the first row of A with the vector that has entries given by $P(s_1, s_2)$, outlined in Eq. (138), gives $P(0,0) + P(1,1) \equiv P(h_1 = 0)$; therefore, the first row carries the label $h_1 = 0$,

$$A \equiv \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} h_1 = 0 \\ h_1 = 1 \\ s_1 = 0 \\ s_1 = 1 \\ s_2 = 0 \\ s_2 = 1 \end{pmatrix}. \tag{141}$$

The generalization to the case $L_s = 4$ is given, considering the Kronecker product (i.e., the tensor product in the standard basis) of the matrix A with itself. For example, the first row of the matrix $A \otimes A$ carries the label $h_1 = 0, h_2 = 0$, the second row carries the label $h_1 = 0, h_2 = 1$, the third row carries $h_1 = 0, s_2 = 0$, and so on.

In the case of $L_s = 4$, we want to bound $P(s_1, s_2, s_3, s_4)$, to get them the idea is the same as that exploited in the case $L_s = 2$, namely, taking linear combinations of bounds on $P(h_1, h_2)$, $P(h_1, s_3)$, $P(h_1, s_4)$, and so on. We note that Eq. (139) generalizes to $L_s = 4$ as follows:

$$\frac{1}{4}(\mathbf{a}_3 + \mathbf{a}_5 - \mathbf{a}_2) \otimes (\mathbf{a}_3 + \mathbf{a}_5 - \mathbf{a}_2) \cdot \mathbf{P} = P(0,0,0,0). \tag{142}$$

$\mathbf{a}_j$ denotes row jth of matrix A, and the Kronecker product of two rows is a row with L_s^2 elements, and $\mathbf{P}$ denotes the vector of all possible choices of $P(s_1, s_2, s_3, s_4)$. Equation (142) involves nine bounds because there are nine terms in the tensor product $(\mathbf{a}_3 + \mathbf{a}_5 - \mathbf{a}_2) \otimes (\mathbf{a}_3 + \mathbf{a}_5 - \mathbf{a}_2)$, and so

$$\frac{1}{16} - 54\varepsilon \leq P(0,0,0,0) \leq \frac{1}{16} + 54\varepsilon.$$

Note that the error 54ε arises as the product of the error associated with each bound in (140) and the number of inequalities that is 9. To generalize this to arbitrary L_s, we just consider further tensor products of A, and hence,

$$2^{-L_s} - 2\,3^{\frac{L_s}{2}+1}\varepsilon \le P(s_1,\dots,s_{L_s}) \le 2^{-L_s} + 2\,3^{\frac{L_s}{2}+1}\varepsilon. \tag{143}$$

Using Lemma 6, with $r = 2t$ and $n = N$, we have $\epsilon < 16t2^{-N}$; this implies (122). □

Theorem 20. Consider an initial vector $\mathbf{u}^0 \in \mathcal{V}_{\text{chain}}$ with non-zero support in all lattice sites ($\mathbf{u}_x^0 \neq \mathbf{0}$ for all $x \in \mathbb{Z}_L$). Consider the evolved vector $\mathbf{u}^t = S(t)\mathbf{u}^0$ inside a region $x \in \{1,\dots,L_s\} \subseteq \mathbb{Z}_L$, where L_s is even and the time is $t \le \frac{L-L_s}{4}$. If $\mathbf{u}^t_{[1,Ls]}$ is the projection of $\mathbf{u}^t$ in the subspace $\oplus_{x=1}^{L_s}\mathcal{V}_x$, then

$$\sum_{\mathbf{v}\in\mathbb{Z}_2^{2NL_s}} \left|\text{prob}\{\mathbf{v} = \mathbf{u}^t_{[1,Ls]}\} - \frac{1}{2^{2NL_s}}\right| \le 32\,t\,2^{-N}\left(2L_s + 3^{\frac{L_s}{2}+1}\right) + 4L2^{-2N}. \tag{144}$$

Proof. First, we re-state $\text{prob}\{\mathbf{v} = \mathbf{u}^t\}$ in the following way:

$$\begin{aligned}\text{prob}\{\mathbf{v} = \mathbf{u}^t\} &= q\,\text{prob}\{\mathbf{v} = \mathbf{u}^t | \mathbf{u}_x^t \neq \mathbf{0}, \mathbf{u}_x^0\ \forall x \in \mathbb{Z}_{L_s}\}\\ &\quad + (1-q)(1 - \text{prob}\{\mathbf{v} = \mathbf{u}^t | \mathbf{u}_x^t \neq \mathbf{0}, \mathbf{u}_x^0\ \forall x \in \mathbb{Z}_{L_s}\}),\end{aligned}$$

where $x \in \{1,\dots,L_s\} \subseteq \mathbb{Z}_L$ with L_s being even and q being the probability of distribution $\text{prob}\{\mathbf{u}_x^t \neq \mathbf{0}, \mathbf{u}_x^t\ \forall x \in \mathbb{Z}_{L_s}\}$, and similarly with the complement. Then, using convexity, we find that

$$\begin{aligned}\sum_{\mathbf{v}\in\mathbb{Z}_2^{2NL_s}} \left|\text{prob}\{\mathbf{v} = \mathbf{u}^t\} - \frac{1}{2^{2NL_s}}\right| &\le q \sum_{\mathbf{v}\in\mathbb{Z}_2^{2NL_s}} \left|\text{prob}\{\mathbf{v} = \mathbf{u}^t | \mathbf{u}_x^t \neq \mathbf{0}, \mathbf{u}_x^0\ \forall x\} - \frac{1}{2^{2NL_s}}\right|\\ &\quad +(1-q) \sum_{\mathbf{v}\in\mathbb{Z}_2^{2NL_s}} \left|1 - \text{prob}\{\mathbf{v} = \mathbf{u}^t | \mathbf{u}_x^t \neq \mathbf{0}, \mathbf{u}_x^0\ \forall x\} - \frac{1}{2^{2NL_s}}\right|.\end{aligned}$$

We can evaluate the first term using the upper bound $q \le 1$ and use Lemma 15 combined with Lemma 19 to find that

$$q \sum_{\mathbf{v}\in\mathbb{Z}_2^{2NL_s}} \left|\text{prob}\{\mathbf{v} = \mathbf{u}^t | \mathbf{u}_x^t \neq \mathbf{0}, \mathbf{u}_x^0\ \forall x \in \mathbb{Z}_{L_s}\} - \frac{1}{2^{2NL_s}}\right| \le 32\,t3^{\frac{L_s}{2}+1}2^{-N} + L2^{2-2N}. \tag{145}$$

To evaluate the second term, we can upper bound the sum by its maximum value, 2, and use the result of Lemma 9 to upper bound $(1-q)$ to find that

$$(1-q) \sum_{\mathbf{v}\in\mathbb{Z}_2^{2NL_s}} \left|1 - \text{prob}\{\mathbf{v} = \mathbf{u}^t | \mathbf{u}_x^t \neq \mathbf{0}, \mathbf{u}_x^0\ \forall x \in \mathbb{Z}_{L_s}\} - \frac{1}{2^{2NL_s}}\right| \le 64L_st2^{-N}. \tag{146}$$

Combining these two terms, we get the stated result. □

VI. APPROXIMATE 2-DESIGN AT HALF-INTEGER TIME

In this section, we will combine the results of Secs. V and V A, with the results of Ref. 43, to show that the random circuit model we consider is an approximate 2-design in a weak sense (Theorem 21).

As discussed in the main body, in Ref. 43 (specifically in Appendix A), it is demonstrated that if a Clifford circuit satisfies both Pauli invariance (Sec. V A, Definition 11) and Pauli mixing (Sec. V, Theorem 14), then it is an exact 2-design. In the following theorem, we will demonstrate that when Pauli mixing is only approximate, as in our case, then the random Clifford circuit is instead an approximate 2-design when one has access to Pauli measurements alone.

Theorem 21. Consider the task of discriminating between two copies of $W(t)$ and two copies of a Haar-random unitary U with measurements restricted to Pauli operators, when $t \in [t_{\text{scr}}, 2t_{\text{scr}}]$ is half-integer. The success probability for correctly guessing the given pair of unitaries satisfies

$$\begin{aligned}p_{\text{guess}} &= \frac{1}{2} + \frac{1}{4}\max_{\rho,\mathbf{u},\mathbf{v}} \text{tr}\left(\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{v}}\left[\mathop{\mathbb{E}}_{W(t)} W(t)^{\otimes 2}\rho\, W(t)^{\otimes 2\dagger} - \int_{\text{SU}(d)} dU\, U^{\otimes 2}\rho\, U^{\otimes 2\dagger}\right]\right)\\ &\le 1/2 + 9\,tL2^{-N}.\end{aligned} \tag{147}$$

Proof. Let us consider a general state describing two copies of the system

$$\rho = \sum_{\mathbf{u},\mathbf{v}} \alpha_{\mathbf{u},\mathbf{v}}\, \sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{v}}, \tag{148}$$

where $\alpha_{0,0} = 2^{-2NL}$ by normalization. The coefficients $\alpha_{\mathbf{u},\mathbf{v}}$ must satisfy the following equation:

$$\alpha_{\mathbf{u},\mathbf{v}}\, 2^{2NL} = \operatorname{tr}(\rho\, \sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{v}}) \in [-1, 1]. \tag{149}$$

Applying the average dynamics to ρ, we obtain

$$\mathop{\mathbb{E}}_{W(t)} W(t)^{\otimes 2} \rho\, W(t)^{\otimes 2\dagger} = 2^{-2NL} \mathbb{1} \otimes \mathbb{1} + \sum_{\mathbf{u},\mathbf{v}\neq 0} \alpha_{\mathbf{v},\mathbf{v}} \mathrm{prob}\{\mathbf{v} = S(t)\mathbf{u}\} \sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{u}}. \tag{150}$$

The fact that terms $\alpha_{\mathbf{u},\mathbf{u}'}$ and $\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{u}'}$ with $\mathbf{u} \neq \mathbf{u}'$ are not present in the above expression follows from the fact that $W(t)$ is Pauli invariant (see Appendix A of Ref. 43), which is proven in Lemma 12. Recall that at half-integer t, we have the time-reversal symmetry

$$\mathrm{prob}\{\mathbf{v} = S(t)\mathbf{u}\} = \mathrm{prob}\{\mathbf{u} = S(t)\mathbf{v}\}. \tag{151}$$

Applying the Haar twirling on ρ, we obtain

$$\int_{\mathrm{SU}(d)} dU\, U^{\otimes 2} \rho\, U^{\otimes 2\dagger} = 2^{-2NL} \mathbb{1} \otimes \mathbb{1} + \sum_{\mathbf{u},\mathbf{v}\neq 0} \alpha_{\mathbf{v},\mathbf{v}}\, \gamma\, \sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{u}}, \tag{152}$$

where $\gamma = (2^{2NL} - 1)^{-1}$ is the uniform distribution over non-zero vectors in $\mathcal{V}_{\mathrm{chain}}$. Substituting (150) and (152) into (147), we obtain

$$\begin{aligned} &\operatorname{tr}\left(\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{v}} \left[\mathop{\mathbb{E}}_{W(t)} W(t)^{\otimes 2} \rho\, W(t)^{\otimes 2\dagger} - \int_{\mathrm{SU}(d)} dU\, U^{\otimes 2} \rho\, U^{\otimes 2\dagger}\right]\right) && (153)\\ &= \delta_{\mathbf{u},\mathbf{v}} \sum_{\mathbf{w}\neq 0} \alpha_{\mathbf{w},\mathbf{w}} (\mathrm{prob}\{\mathbf{u} = S(t)\mathbf{w}\} - \gamma) 2^{2NL} && (154)\\ &\leq \delta_{\mathbf{u},\mathbf{v}} \sum_{\mathbf{w}\neq 0} |\mathrm{prob}\{\mathbf{u} = S(t)\mathbf{w}\} - \gamma| \leq 33\, tL2^{-N} \delta_{\mathbf{u},\mathbf{v}}, && (155) \end{aligned}$$

where in the last two inequalities, we use (149), (151), and Theorem 14. This implies that the guessing probability satisfies $p_{\mathrm{guess}} \leq 1/2 + 9tL2^{-N}$; hence, (147). □

The following result is not presented in the main text because it is difficult to interpret. It is important to not confuse the infinite norm between two states with the infinite norm between two maps. What we have here is the first. The second is the definition of quantum tensor-product expander.

Lemma 22. The dynamics $W(t)$ defined in Eq. (26), with $t \geq t_{\mathrm{scr}}$ half-integer, is closed to an approximate 2-design with respect to the infinity norm, namely, for any state ρ, it holds that

$$\left\| \mathop{\mathbb{E}}_{W(t)} W(t)^{\otimes 2} \rho\, W(t)^{\otimes 2\dagger} - \int_{\mathrm{SU}(2^{NL})} dU\, U^{\otimes 2} \rho\, U^{\otimes 2\dagger} \right\|_{\infty} \leq 33\, tL2^{-N}. \tag{156}$$

Proof. Let $|\phi_0\rangle \equiv \left(\frac{|0,1\rangle - |1,0\rangle}{\sqrt{2}}\right)^{\otimes NL}$ denote the NL-fold tensor product of the singlet state, where each singlet entangles each qubit of the first copy of the system and the corresponding qubit in the second copy of the system. This implies that $(\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{u}})|\phi_0\rangle = (-1)^{|\mathbf{u}|}|\phi_0\rangle$, where $|\mathbf{u}| \equiv \left(\sum_j u_j\right) \mathrm{mod} 2$. $\sigma_{\mathbf{u}}$ is defined as in (A1),

$$\sigma_{\mathbf{u}} = \bigotimes_{i=1}^{n} (\sigma_x^{q_i} \sigma_z^{p_i}) \in \mathrm{U}(2^n)$$

with $\mathbf{u} = (q_1, p_1, q_2, p_2, \ldots, q_n, p_n) \in \mathbb{Z}_2^{2n}$. Any Bell state (as described above) can be written as $|\phi_{\mathbf{v}}\rangle = (\mathbb{1} \otimes \sigma_{\mathbf{v}})|\phi_0\rangle$ for all $\mathbf{v} \in \mathcal{V}_{\mathrm{chain}}$. Note that these form an orthonormal basis for the Hilbert space of two copies of the system $\langle\phi_{\mathbf{u}}|\phi_{\mathbf{v}}\rangle = \delta_{\mathbf{u},\mathbf{v}}$. In addition, using the commutation relations (A6), we obtain

$$\begin{aligned} (\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{u}})|\phi_{\mathbf{v}}\rangle &= (\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{u}})(\mathbb{1} \otimes \sigma_{\mathbf{v}})|\phi_0\rangle \\ &= (-1)^{\langle \mathbf{v},\mathbf{u}\rangle} (\mathbb{1} \otimes \sigma_{\mathbf{v}})(-1)^{|\mathbf{u}|}|\phi_0\rangle \\ &= (-1)^{\langle \mathbf{v},\mathbf{u}\rangle + |\mathbf{u}|}|\phi_{\mathbf{v}}\rangle. \end{aligned} \tag{157}$$

This together with (150) and (152) implies that the argument inside the norm (156) is diagonal in the $|\phi_{\mathbf{v}}\rangle$ basis. Therefore, the following bound for each element of the basis provides the bound for the ∞-norm:

$$\begin{aligned}
&\langle\phi_{\mathbf{v}}|\Bigg(\mathop{\mathbb{E}}_{W(t)} W(t)^{\otimes 2}\rho\, W(t)^{\otimes 2\dagger} - \int_{\mathrm{SU}(d)} dU\, U^{\otimes 2}\rho\, U^{\otimes 2\dagger}\Bigg)|\phi_{\mathbf{v}}\rangle && (158)\\
&\quad = \sum_{\mathbf{u},\mathbf{w}\neq 0} \alpha_{\mathbf{w},\mathbf{w}}(\mathrm{prob}\{\mathbf{u} = S(t)\mathbf{w}\} - \gamma)\langle\phi_{\mathbf{v}}|\sigma_{\mathbf{u}}\otimes\sigma_{\mathbf{u}}|\phi_{\mathbf{v}}\rangle && (159)\\
&\quad = \sum_{\mathbf{u},\mathbf{w}\neq 0} \alpha_{\mathbf{w},\mathbf{w}}(\mathrm{prob}\{\mathbf{u} = S(t)\mathbf{w}\} - \gamma)(-1)^{\langle\mathbf{v},\mathbf{u}\rangle + |\mathbf{u}|} && (160)\\
&\quad \le \sum_{\mathbf{u},\mathbf{w}\neq 0} 2^{-2NL}|\mathrm{prob}\{\mathbf{u} = S(t)\mathbf{w}\} - \gamma| \le 33\, tL2^{-N}. && (161)
\end{aligned}$$

□

VII. LOCALIZATION WITH $N \ll \log L$

In this section, we consider the same spin chain with random local Clifford dynamics, and again, we will work in the phase space description, which is discussed in Appendix A. We will show that in the regime of $N \ll \log L$, the random dynamics, instead of displaying scrambling, results in the localization of all operators in a bounded region.

The most simple case that results in localization is when one of the L two-site gates S_x has $C_x = 0$, so there is no right-ward propagation, and hence, the time-periodic nature of the circuit prevents right-ward propagation for all subsequent times also. A bound on the probability of this happening is given in the following theorem.

Theorem 23. Any given $S \in \mathcal{S}_{2n}$ can be written in the block form

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{162}$$

according to the decomposition $\mathbb{Z}_2^{4n} = \mathbb{Z}_2^{2n} \oplus \mathbb{Z}_2^{2n}$, and if S is uniformly distributed, then this induces a distribution on the sub-matrices A, B, C, D. For each of the sub-matrices $(E = A, B, C, D)$, the induced distribution satisfies

$$\frac{2^{-4N^2}}{2} \le \mathrm{prob}\{E = 0\} = \frac{|\mathcal{S}_n|^2}{|\mathcal{S}_{2n}|} \le 2^{-4N^2}. \tag{163}$$

It also holds that $\mathrm{prob}\{A = 0|D = 0\} = \mathrm{prob}\{D = 0|A = 0\} = \mathrm{prob}\{B = 0|C = 0\} = \mathrm{prob}\{C = 0|B = 0\} = 1$.

Proof. We first consider when $C = 0$. By Lemma 37 in Appendix C, this implies that $B = 0$. Therefore, A and D are both $2n \times 2n$ symplectic matrices, which can be counted independently. Following the counting algorithm in Lemma 1, the number of choices of S with $C = 0$ is given exactly by

$$|\{S \in \mathcal{S}_{2n} : C = 0\}| = |\mathcal{S}_n||\mathcal{S}_n| = |\mathcal{S}_n|^2. \tag{164}$$

Finally, dividing by the total number of choices for S gives the probability. Using Lemmas 36 and 37, this argument applies to any of the four sub-matrices A, B, C, D. The bounds are found using Lemma 35. □

We refer to this as trivial localization as it is equivalent a non-interacting matrix and, hence, results in the spin chain being split into two independent parts. In the rest of this section, we investigate other conditions for localization, which are not trivial and occur as a result of the dynamics.

Lemma 24 shows that the number of powers k that needs to satisfy Eq. (180) is finite.

Lemma 24. The conditions

$$C_{x+1}(D_x A_{x+1})^k C_x = 0 \quad \text{for all} \quad k \in \{0, 1, 2, \ldots, 2^{4N} - 1\} \tag{165}$$

imply that

$$C_{x+1}(D_x A_{x+1})^k C_x = 0 \quad \text{for all} \quad k \in \{0, 1, 2, \ldots\}. \tag{166}$$

Proof. Suppose that the square matrix M has n linearly independent powers

$$M, M^2, M^3, \ldots, M^n \tag{167}$$

and that M^{n+1} is a linear combination of (167). Let us prove that for any integer $m > n$, the matrix M^m is also a linear combination of (167). First, note that our premise $M^{n+1} = \sum_{k=0}^{n} a_k\, M^k$ implies that

$$M^{n+2} = \sum_{k=0}^{n} a_k\, M^{k+1} = \sum_{k=0}^{n-1} a_k\, M^{k+1} + a_n \sum_{k=0}^{n} a_k\, M^k \tag{168}$$

is also a linear combination of (167). Now, we can proceed by induction. For any $m > n$, suppose that the matrix M^m is a linear combination of (167), that is, $M^m = \sum_{k=0}^{n} b_k\, M^k$. Then, proceeding as before, we have

$$M^{m+1} = \sum_{k=0}^{n} b_k\, M^{k+1} = \sum_{k=0}^{n-1} b_k\, M^{k+1} + b_n \sum_{k=0}^{n} a_k\, M^k, \tag{169}$$

which proves our claim.

Finally, we apply this result to $M = D_x A_{x+1}$, and note that since M is a square matrix of dimension 2^{2N}, it can have at most 2^{4N} linearly independent powers. □

Theorem 25. For $N = 1$, the conditions

$$C_{x+1}(D_x A_{x+1})^k C_x = 0 \tag{170}$$

for $k \in \{0, 1, 2, \ldots\}$ are implied by the following two conditions:

$$C_{x+1} C_x = 0 \quad \text{and} \quad C_{x+1} D_x A_{x+1} C_x = 0. \tag{171}$$

Furthermore, the probability of this is given exactly by

$$\text{prob}\{C_{x+1} C_x = 0, C_{x+1} D_x A_{x+1} C_x = 0\} = 0.12, \tag{172}$$

which includes trivial localization.

Proof. We are concerned with the case $N = 1$; then, S_x and S_{x+1} are 4×4 symplectic matrices and the sub blocks A, B, C, D are 2×2 matrices. We first note that if $C_x = 0$ and/or $C_{x+1} = 0$, which is trivial localization, then it is clear that the conditions for all k are satisfied. Hence, we now focus only on the cases where $C_x \neq 0$ and $C_{x+1} \neq 0$. Moreover, we note that we will only focus on the cases where Rank(C_x) = Rank$(C_{x+1}) = 1$ since if either of C_x or C_{x+1} is full rank, then to satisfy $C_{x+1} C_x = 0$, the other of the C matrices must be the zero matrix.

When Rank$(C_{x+1}) = 1$, then $C_{x+1}^T J C_{x+1} = 0$; this follows from the fact that the matrix C_{x+1} has only one distinct column that is non-zero. By the symplectic conditions, Eq. (176), this implies that A_{x+1} is a 2×2 symplectic matrix. This argument also applies to C_x, and so D_x is also a 2×2 symplectic matrix.

Therefore, since the product of symplectic matrices is also a symplectic matrix, for $N = 1$ neglecting the cases of trivial localization ($C_x = 0$ and/or $C_{x+1} = 0$), the conditions for right localization, (171), become

$$C_{x+1} S^k C_x = 0, \tag{173}$$

where S is a generic 2×2 symplectic matrix. For all 2×2 symplectic matrices, there exist $\alpha,\ \beta \in \mathbb{Z}_2$ such that

$$S^2 = \alpha \mathbb{I} + \beta S, \tag{174}$$

which can be verified by a direct check. Hence, if $C_{x+1} C_x = 0$ and $C_{x+1} S C_x = 0$ hold, then $C_{x+1} S^k C_x = 0$ for all $k > 1$.

The exact result for the probability given above for the case of $N = 1$ follows from directly counting, with the aid of a computer program, the number of symplectic matrices that satisfy (171). □

In Lemma 26, we provide an explicit example showing that conditions (171) sufficient to ensure localization in the case of $N = 1$ are not enough to imply (180); therefore, (171) does not imply localization for $N > 1$.

Lemma 26. In the case of $N > 1$, the set of Eq. (180) is sufficient to ensure the presence of a hard wall. For qubits, $N = 1$, Eq. (171) imply Eq. (180). We show that for $N > 1$, (171) does not imply (180) by explicitly constructing an example for $N = 2$ that also generalizes to all $N > 1$. In what follows, to ease the notation, we set $x = 0$. In the following, J_{4N} is the symplectic form of order $4N$. The definition of the symplectic matrix, $S^T J_{4N} S = J_{4N}$, when S is written in the block form

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{175}$$

reads

$$\begin{cases} A^T J_{2N} A + C^T J_{2N} C = J_{2N}, \\ A^T J_{2N} B + C^T J_{2N} D = 0, \\ B^T J_{2N} B + D^T J_{2N} D = J_{2N}, \end{cases} \tag{176}$$

with J_{2N} being the symplectic form of order $2N$. A solution of system (176) is given by

$$\begin{cases} C^T J_{2N} C = C J_{2N} C^T = 0, \\ A^T J_{2N} A = J_{2N}, \\ D^T J_{2N} D = J_{2N}, \\ B = A J_{2N} C^T J_{2N} D. \end{cases} \tag{177}$$

This implies that A and D are symplectic, and B is determined by A, C, D.

Our goal is to build C_0, D_0, A_1, and C_1 such that $C_1 C_0 = 0$, $C_1 D_0 A_1 C_0 = 0$ but $C_1 (D_0 A_1)^2 C_0 \neq 0$ showing that with $N > 1$, the proof given above for qubits fails and the whole set of Eq. (180) must be satisfied.

Let us write straight away the matrices S_0 and S_1 and then discuss their structure,

$$S_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad S_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{178}$$

The blocks C_0 and C_1 are the projection onto $e_1 \equiv (1,0,0,0)^T$ and $e_2 \equiv (0,1,0,0)^T$. They satisfy $C^T J_4 C = C J_4 C^T = 0$ and also $C_1 C_0 = 0$. To ensure $C_1 (D_0 A_1)^2 C_0 \neq 0$, $(D_0 A_1)^2$ must map e_1 into e_2; on the other hand, to ensure $C_1 D_0 A_1 C_0 = 0$, $D_0 A_1$ must not map e_1 into e_2. This is achieved, for example, by

$$D_0 A_1 = \begin{pmatrix} 0_2 & J_2 \\ \mathbb{1}_2 & 0_2 \end{pmatrix}, \quad (D_0 A_1)^2 = \begin{pmatrix} J_2 & 0_2 \\ 0_2 & J_2 \end{pmatrix} = J_4. \tag{179}$$

The matrix $D_0 A_1$ has been written in the block form to show that this construction generalizes to higher dimensions; in fact, in every dimension, J_{2N} maps e_1 to e_2. At the same time, C_0 and C_1 in higher dimensions are still the projection onto e_1 and e_2. With regard to higher powers of $D_0 A_1$, it is easy to see that $(D_0 A_1)^4 = \mathbb{1}$; therefore, $(D_0 A_1)^6 = (D_0 A_1)^2$, and in general, $\forall k \in \mathbb{N}$ $(D_0 A_1)^{4k+2} = (D_0 A_1)^2$.

A. Absence of localization with $N \gg \log L$

The following theorem provides an upper bound for the probability that one-sided walls appear at a particular location. This upper bound implies that when $N \gg \log L$, a typical circuit has no localization.

Theorem 27. The conditions

$$C_{x+1} (D_x A_{x+1})^k C_x = 0 \quad \text{for all} \quad k \in \{0, 1, 2, \ldots\} \tag{180}$$

are sufficient to prevent all right-ward propagation past position x at any time. The probability that this family of constrains holds is upper-bounded by

$$\begin{aligned} &\text{prob}\{C_{x+1} (D_x A_{x+1})^k C_x = 0, \ \forall k \in \mathbb{N}\} \\ &\leq \text{prob}\{C_{x+1} C_x = 0\} \ \leq \ \frac{2N+1}{(1 - 2^{-2N})^{2N}} \, 2^{2N - 2N^2}. \end{aligned} \tag{181}$$

Proof. This proof is clearer with reference to Figs. 4 and 5. The condition $C_{x+1}C_x = 0$ prevents right-ward propagation for a single time step, however (unless $C_x = 0$); then, $A_{x+1}C_x \neq 0$, and hence, in subsequent time steps, there could be right-ward propagation. In the next time step, the only way for possible right-ward propagation to occur, which would not be blocked by the condition $C_{x+1}C_x = 0$, is $C_{x+1}D_xA_{x+1}C_x$, and so the additional requirement $C_{x+1}D_xA_{x+1}C_x = 0$ prevents right-ward propagation. Once again, the same argument applies for subsequent time steps, and hence, we require that $C_{x+1}(D_xA_{x+1})^kC_x = 0$ for $k \geq 2$ ($k \in \mathbb{N}$). The bound given in (181) is obtained from Eq. (50) with $k = 2N$ and $r = 2$. □

Remark. Theorem 27 provides a sufficient condition. There are, of course, other potential conditions and mechanisms by which right-ward propagation is prevented.

VIII. DISCUSSION

A. The scrambling time

In this section, we argue that the time t at which the evolution operator $W(t)$ maximally resembles a Haar unitary (Theorem 21) is around the scrambling time t_{scr}. For this, we note that there are two factors contributing to this resemblance: causality and recurrences.

Causality. If U is a Haar-random unitary, then a local operator A is mapped to a completely non-local operator $UAU^\dagger$ with high probability. However, in our model, the evolution $W(t)AW(t)^\dagger$ of a local operator A is supported in its light cone, which only reaches the whole system at the scrambling time t_{scr}. Hence, for $W(t)AW(t)^\dagger$ to be a completely non-local operator, we need $t \geq t_{\text{scr}}$.

Recurrences. The powers U^t of a Haar-random unitary $U \in \mathrm{SU}(d)$ lose their resemblance to a Haar unitary as t increases. This can be quantified with the spectral form factor, which for a Haar unitary U takes the small value $|\mathrm{tr}\, U|^2 \approx 1$, while for its powers, it takes the larger value $|\mathrm{tr}\, U^t|^2 \approx t$. Specifically, we have

$$K_{\text{Haar}}(t) = \int_{\mathrm{SU}(d)} dU \left|\mathrm{tr} U^t\right|^2 = \begin{cases} t & \text{if } 0 < t < d, \\ d & \text{if } t \geq d. \end{cases} \tag{182}$$

That is, as time t grows, the form factor of U^t tends to that of the Poisson spectrum (integrable system),

$$K_{\text{Poisson}}(t) = d \qquad \text{for all } t > 0. \tag{183}$$

In our model, the evolution operator $W(t)$ is never a Haar unitary, but its resemblance decreases as t increases. In particular, the fact that the Clifford group is finite implies the existence of a recurrence time t_{rec} such that the evolution operator is trivial $W(t_{\text{rec}}) = \mathbb{1}$.

In summary, for $W(t)$ to maximally resemble a Haar unitary, the time t should be the smallest possible to avoid recurrences, but still larger than t_{scr}. This argument explains why the "long-time ensemble" does not resemble a random unitary, as found in Ref. 58. By the long-time ensemble, we mean the set of unitaries $\{e^{-\mathrm{i}Ht} : t \in \mathbb{R}\}$ generated by a fixed Hamiltonian H.

B. Is Clifford dynamics integrable or chaotic?

In this section, we argue that Clifford dynamics has some of the features of quasi-free boson and fermion systems, but at the same time, it displays a stronger chaos. For this reason, we believe that Clifford dynamics is a very interesting setup to understand the landscape of quantum many-body phenomena. Next, we enumerate essential properties of Clifford dynamics: the first two are in common with quasi-free systems and the subsequent four are not.

Phase space description and classical simulability. Clifford unitaries can be represented as symplectic transformations in a phase space (in a similar fashion to quasi-free bosons) of dimension exponentially smaller than the Hilbert space. The phase space structure of the Clifford group is described in Appendix A. This dimensional reduction allows us to efficiently simulate the evolution of any Pauli operator (and many other relevant operators) with a classical computer.

Anderson localization. Clifford dynamics with disorder (meaning that each gate U_x in Fig. 1 is statistically independent and identically distributed) displays a strong form of localization, reminiscent of Anderson's localization. Until now, this strong form of localization has only been observed in free-particle systems. However, Clifford dynamics cannot be understood in terms of free particles.

Discrete time. The Clifford phase space is a vector space over a finite field; hence, evolution cannot be continuous in time. That is, we can have Floquet-type but not Hamiltonian-type dynamics. The dynamical maps are symplectic matrices with $\mathbb{Z}_2$ entries, and these cannot be diagonalized. This lack of eigenmodes prevents us from using many tools and intuitions of quasi-free systems.

No particles. Some specific Clifford dynamics have gliders, which is the discrete-time analog of free particles. However, the typical translation-invariant Clifford dynamics consists of fractal patterns,[59] and in the non-translation-invariant case (i.e., disorder), we see patterns such as those in Fig. 3. None of these patterns can be understood in terms of free or interacting particles.

Signatures of chaos. If we allow for fully non-local dynamics, quasi-free bosons and fermions cannot generate a 1-design. This is because their evolution operators commute with the number operator (bosons) or the parity operator (fermions). On the contrary, in the non-local case, Clifford dynamics generates a 3-design.[43,60] Hence, we see that despite the above-mentioned similarities with quasi-free systems, Clifford matter seems to display stronger chaos. However, chaotic dynamics can be diagnosed by a small (absolute) value of out-of-time order correlators (OTOC),[61] which is not observed in the Clifford case. In fact, for any Clifford unitary W and two Pauli operators $\sigma_{\mathbf{u}}, \sigma_{\mathbf{v}}$, the OTOC at infinite temperature takes the maximum value $|\frac{1}{d}\operatorname{tr}(\sigma_{\mathbf{u}} W \sigma_{\mathbf{v}} W^\dagger \sigma_{\mathbf{u}} W \sigma_{\mathbf{v}} W^\dagger)| = 1$. Incidentally, a small OTOC follows from being a 4-design but not a 3-design.

Absence of local integrals of motion. In the translation-invariant case, some Clifford models[62] with local interactions have fully nonlocal integrals of motion. This means that each operator that commutes with the evolution operator involves couplings, which do not decay with the distance and act on an extensive number of sites (unbounded wight).

C. Time-dependent vs time-independent circuits

Time-dependent local quantum circuits (see Fig. 2) have been used as a model for chaotic dynamics in numerous contributions.[36–40] It has been proven that these circuits generate approximate k-designs where the order increases with time as $k \sim t^{1/10}$ (although the scaling is conjectured to be $k \sim t$[40]). Some authors have attempted to model chaotic systems with conserved quantities by using time-dependent local circuits constrained so that each gate commutes with an operator of the form $Q = \sum_x \sigma_z^{(x)}$, where x labels all sites.[63,64] These Q-conserving circuits also generate approximate k-designs in the operator space orthogonal to Q, with k increasing as time passes.

We argue that the dynamics of Q-conserving circuits is very different from time-independent circuits such as the model we are studying, Fig. 1. Despite the fact that in both cases there are conserved quantities [Q and $W(t = 1)$], Q-conserving time-dependent circuits do not have time correlations nor recurrences (see Sec. VIII A). This implies that they resemble Haar unitaries more and more as time goes on. Instead, as discussed in Sec. VIII A, time-independent dynamics loses its resemblance to Haar unitaries with time.

Previous works[41,42] have constructed unitary designs with "nearly time-independent" dynamics. This consists of an evolution where the Hamiltonian changes a small number of times, and it is time-independent in between changes. A different line of work[13,18,23,24,30] analyses disordered time-periodic dynamics with non-Clifford gates. These more general dynamics make these models more chaotic than ours. However, these works only prove that these models display certain aspects of Haar-random unitaries, instead of indistinguishability as captured in Theorem 21.

D. A variant of our model

We define our model as having L sites, with N qubits per site, and nearest-neighbor interactions. However, this is equivalent to say that it has LN sites, with a single-qubit per site, and $2N$-range interactions. For this, we use the fact that any Clifford gate of $2N$ qubits can be written as a circuit of depth $\mathcal{O}(N^2/\log N)$.[53,54] Hence, a dynamical period in the LN-site circuit decomposes into $\mathcal{O}(N^2/\log N)$ elementary time steps.

IX. CONCLUSION AND OUTLOOK

The dynamics of highly chaotic quantum systems, such as black holes,[11,65] is often modeled with Haar-random unitaries, which allows for the exact calculation of relevant quantities. This model is often justified by the fact that local random circuits[36,38,39] generate 2-designs. However, these circuits are time-dependent, while presumably the dynamics of black holes are not.[66] In this work, we make a step forward toward the justification of the Haar-unitary model of dynamics in quantum chaotic systems by proving that the evolution operator of a time-periodic model cannot be distinguished from a random unitary in some physically relevant setups.

An important question that remains open is whether local and time-independent (or time-periodic) dynamics can generate a 2-design. This amounts to not restricting the measurement in the discrimination process. The results in Refs. 13, 18, 23, 24, and 30 provide some hope in this direction. However, we expect that the 2-design property is best achieved around the scrambling time, and it fades away as time goes on (see discussion in Sec. VIII A and in Ref. 42). More generally, we would like to characterize which further properties of random unitaries are present in naturally occurring dynamics.

ACKNOWLEDGMENTS

We are grateful to Nick Hunter-Jones, Oliver Lunt, and Arijeet Pal for valuable discussions. Tom Farshi acknowledges financial support from the Engineering and Physical Sciences Research Council (Grant No. EP/L015242/1). Daniele Toniolo and Lluis Masanes acknowledge financial support from the UK's Engineering and Physical Sciences Research Council (Grant No. EP/R012393/1). Carlos González acknowledges financial support from the Spanish MINECO (Project No. MTM2017-88385-P), MECD "José Castillejo" program (Grant No. CAS16/00339), and Programa Propio de I+D+i of the Universidad Politécnica de Madrid. Research at Perimeter Institute is supported, in part, by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that give rise to Fig. 3 are available from the corresponding author upon reasonable request.

APPENDIX A: CLIFFORD DYNAMICS AND DISCRETE PHASE SPACE

In this appendix, we first define the Pauli and Clifford groups and then present the phase space description of Clifford dynamics. This description is known from previous works,[29,53,54] and we include it here for clarity of presentation.

The Pauli sigma matrices together with the identity $\{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}$ form a basis of the space of operators of one qubit $\mathbb{C}^2$. In addition, the 16 matrices obtained by multiplying $\{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}$ times the coefficients $\{1, \mathrm{i}, -1, -\mathrm{i}\}$ form a group. This is called the Pauli group of one qubit, and it is denoted by $\mathcal{P}_1$. The generalization to n qubits is as follows.

Definition 28. The **Pauli group** of n qubits $\mathcal{P}_n$ is the set of matrices $\mathrm{i}^u \sigma_{\mathbf{u}}$ where

$$\sigma_{\mathbf{u}} = \bigotimes_{i=1}^{n} (\sigma_x^{q_i} \sigma_z^{p_i}) \in \mathrm{U}(2^n) \tag{A1}$$

for all phases $u \in \mathbb{Z}_4$ and vectors $\mathbf{u} = (q_1, p_1, q_2, p_2, \ldots, q_n, p_n) \in \mathbb{Z}_2^{2n}$. We also define $\bar{\mathcal{P}}_n = \mathcal{P}_n / \{1, \mathrm{i}, -1, -\mathrm{i}\}$, which satisfies $\bar{\mathcal{P}}_n \cong \mathbb{Z}_2^{2n}$.

Here, $\mathbb{Z}_2^{2n}$ stands for a $2n$-dimensional vector space with addition and multiplication operations defined modulo 2. Using the identity $\sigma_z \sigma_x = -\sigma_x \sigma_z$ and the definition $\beta(\mathbf{u}, \mathbf{u}') = \sum_{i=1}^{n} p_i q_i'$, we obtain the multiplication and inverse rules,

$$\sigma_{\mathbf{u}} \sigma_{\mathbf{u}'} = (-1)^{\beta(\mathbf{u},\mathbf{u}')} \sigma_{\mathbf{u}+\mathbf{u}'}, \tag{A2}$$

$$\sigma_{\mathbf{u}}^{-1} = (-1)^{\beta(\mathbf{u},\mathbf{u})} \sigma_{\mathbf{u}}. \tag{A3}$$

The Pauli group (A1) is the discrete version of the Weyl group or the *displacement operators* used in quantum optics. Concretely, if $\hat{Q}$ and $\hat{P}$ are quadrature operators (satisfying the canonical commutation relations $[\hat{Q}, \hat{P}] = \mathrm{i}\mathbb{1}$), then we can write the analogy as

$$\sigma_x^q \sigma_z^p \quad \longleftrightarrow \quad e^{\mathrm{i}\hat{P}q} e^{\mathrm{i}\hat{Q}p}, \tag{A4}$$

where the phase space variables (q, p) take values in $\mathbb{Z}_2^2$ on the left of (A4) and in $\mathbb{R}^2$ on the right. This analogy also extends to the set of transformations that preserve the phase space structure. Before characterizing these transformations, let us define the phase space associated with the Pauli group.

Definition 29. The **discrete phase space** of n qubits $\mathbb{Z}_2^{2n}$ is the $2n$-dimensional vector space over the field $\mathbb{Z}_2$, endowed with the symplectic (antisymmetric) bilinear form

$$\langle \mathbf{u}, \mathbf{u}' \rangle = \mathbf{u}^T J \mathbf{u}', \text{ where } J = \bigoplus_{i=1}^{n} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{A5}$$

for all $\mathbf{u}, \mathbf{u}' \in \mathbb{Z}_2^{2n}$. Note that the form is indeed antisymmetric $\langle \mathbf{u}, \mathbf{u}' \rangle = \langle \mathbf{u}', \mathbf{u} \rangle = -\langle \mathbf{u}', \mathbf{u} \rangle \mathrm{mod}\, 2$, which implies that $\langle \mathbf{u}, \mathbf{u} \rangle = 0$.

Using the symplectic form (A5) and the rules [(A2) and (A3)], we can write the commutation relations of the Pauli group as

$$\sigma_{\mathbf{u}}\, \sigma_{\mathbf{u}'}\, \sigma_{\mathbf{u}}^{-1}\, \sigma_{\mathbf{u}'}^{-1} = (-1)^{\langle \mathbf{u}, \mathbf{u}' \rangle}. \tag{A6}$$

In analogy with the continuous (bosonic) phase space, in the following two definitions, we introduce the transformations that preserve the symplectic form (A5) and the Pauli group, respectively.

Definition 30. The **symplectic group** $\mathcal{S}_n$ is the set of matrices $S : \mathbb{Z}_2^{2n} \to \mathbb{Z}_2^{2n}$ such that

$$\langle S\mathbf{u}, S\mathbf{u}' \rangle = \langle \mathbf{u}, \mathbf{u}' \rangle \tag{A7}$$

for all $\mathbf{u}, \mathbf{u}' \in \mathbb{Z}_2^{2n}$. This is equivalent to the condition $S^T J S = J \bmod 2$.

Definition 31. A unitary $U \in \mathrm{U}(2^n)$ is **Clifford** if it maps each Pauli operator $\sigma \in \mathcal{P}_n$ to a Pauli operator $U\sigma U^\dagger \in \mathcal{P}_n$. Two Clifford unitaries U, V are equivalent if there is a complex phase λ with $|\lambda| = 1$ such that $U = \lambda V$. The **Clifford group** of n qubits $\mathcal{C}_n$ is the set of equivalence classes of Clifford unitaries in $\mathrm{U}(2^n)$.

In this work, we only consider the adjoint representation $\sigma \mapsto U\sigma U^\dagger$; hence, the phase λ does not play any role.

Lemma 32 (structure of $\mathcal{C}_n$). Each Clifford transformation $U \in \mathcal{C}_n$ is characterized by a symplectic matrix $S \in \mathcal{S}_n$ and a vector $\boldsymbol{s} \in \mathbb{Z}_2^{2n}$ so that

$$U\sigma_{\mathbf{u}}U^{\dagger} = \mathrm{i}^{\alpha[S,\mathbf{u}]}\,(-1)^{\langle s,\mathbf{u}\rangle}\,\sigma_{S\mathbf{u}}, \tag{A8}$$

where the function α takes values in $\mathbb{Z}_4$. More precisely, we have $\mathcal{C}_n \cong \bar{\mathcal{P}}_n \rtimes \mathcal{S}_n$.

In this work, the function α does not play any role; hence, we do not provide a characterization.

Proof. For each $U \in \mathcal{C}_n$, there are two functions

$$s : \mathbb{Z}_2^{2n} \to \mathbb{Z}_4, \tag{A9}$$

$$S : \mathbb{Z}_2^{2n} \to \mathbb{Z}_2^{2n} \tag{A10}$$

such that

$$U\sigma_{\mathbf{u}}U^{\dagger} = \mathrm{i}^{s[\mathbf{u}]}\,\sigma_{S[\mathbf{u}]}. \tag{A11}$$

Note that, at this point, we do not make any assumption about these functions, such as linearity. Using (A2), we obtain the equality between the following two expressions:

$$\begin{aligned} U\sigma_{\mathbf{u}}\sigma_{\mathbf{u}'}U^{\dagger} &= (-1)^{\beta(\mathbf{u},\mathbf{u}')}\,U\sigma_{\mathbf{u}+\mathbf{u}'}U^{\dagger} \\ &= (-1)^{\beta(\mathbf{u},\mathbf{u}')}\,\mathrm{i}^{s[\mathbf{u}+\mathbf{u}']}\,\sigma_{S[\mathbf{u}+\mathbf{u}']}, \end{aligned} \tag{A12}$$

$$\begin{aligned} U\sigma_{\mathbf{u}}U^{\dagger}U\sigma_{\mathbf{u}'}U^{\dagger} &= \left(\mathrm{i}^{s[\mathbf{u}]}\,\sigma_{S[\mathbf{u}]}\right)\left(\mathrm{i}^{s[\mathbf{u}']}\,\sigma_{S[\mathbf{u}']}\right) \\ &= (-1)^{\beta(S\mathbf{u},S\mathbf{u}')}\,\mathrm{i}^{s[\mathbf{u}]+s[\mathbf{u}']}\,\sigma_{S[\mathbf{u}]+S[\mathbf{u}']}, \end{aligned} \tag{A13}$$

which implies the $\mathbb{Z}_2$-linearity of the S function. Hence, from now on, we write its action as a matrix $S[\mathbf{u}] = S\mathbf{u}$. Next, if we impose the commutation relations of the Pauli group (A6) as follows:

$$\begin{aligned} (-1)^{\langle \mathbf{u},\mathbf{u}'\rangle} &= U\sigma_{\mathbf{u}}\,\sigma_{\mathbf{u}'}\,\sigma_{\mathbf{u}}^{-1}\,\sigma_{\mathbf{u}'}^{-1}\,U^{-1} \\ &= \left(\mathrm{i}^{s[\mathbf{u}]}\sigma_{S\mathbf{u}}\right)\left(\mathrm{i}^{s[\mathbf{u}']}\sigma_{S\mathbf{u}'}\right)\left(\mathrm{i}^{s[\mathbf{u}]}\sigma_{S\mathbf{u}}\right)^{-1}\left(\mathrm{i}^{s[\mathbf{u}']}\sigma_{S\mathbf{u}'}\right)^{-1} \\ &= (-1)^{\langle S\mathbf{u},S\mathbf{u}'\rangle}, \end{aligned} \tag{A14}$$

we find that the matrices S are symplectic. Conversely, it has been proven[29,67–69] that for each symplectic matrix $S \in \mathcal{S}_n$, there is $U \in \mathcal{C}_n$ such that $U\sigma_{\mathbf{u}}U^{\dagger} \propto \sigma_{S\mathbf{u}}$ for all $\mathbf{u}$.

Now, let us obtain the set of pairs (S, s) associated with the subgroup $\bar{\mathcal{P}}_n \subseteq \mathcal{C}_n$. Using (A6), we see that the Clifford transformation $\sigma_{\mathbf{v}} \in \bar{\mathcal{P}}_n$ has $S = \mathbb{1}$ and $s[\mathbf{u}] = 2\langle \mathbf{v},\mathbf{u}\rangle$ for any $\mathbf{v} \in \mathbb{Z}_2^{2n}$. Next, let us prove the converse. By equating (A12) and (A13) with $S = \mathbb{1}$, we see that any Clifford transformation U with $S = \mathbb{1}$ has a phase function s satisfying

$$s[\mathbf{u}+\mathbf{u}'] = s[\mathbf{u}] + s[\mathbf{u}'] \tag{A15}$$

for all pairs $\mathbf{u},\mathbf{u}'$. In addition, since the map $\sigma_{\mathbf{u}} \to U\sigma_{\mathbf{u}}U^{\dagger}$ preserves the Hermiticity or anti-Hermiticity of $\sigma_{\mathbf{u}}$, the phase function in $U\sigma_{\mathbf{u}}U^{\dagger} = \mathrm{i}^{s[\mathbf{u}]}\sigma_{\mathbf{u}}$ has to satisfy $s[\mathbf{u}] \in \{0, 2\}$ for all $\mathbf{u}$. Combining this with (A15), we deduce that if $S = \mathbb{1}$, then $s[\mathbf{u}] = 2\langle \mathbf{v},\mathbf{u}\rangle$ for some vector $\mathbf{v} \in \mathbb{Z}_2^{2n}$. In summary, an element of the Clifford group belongs to the Pauli group if, and only if, there is a vector $\mathbf{v} \in \mathbb{Z}_2^{2n}$ such that $S = \mathbb{1}$ and $s[\mathbf{u}] = 2\langle \mathbf{v},\mathbf{u}\rangle$.

Now, let us show that $\mathcal{C}_n/\bar{\mathcal{P}}_n \cong \mathcal{S}_n$. By definition, any Clifford element $U\bar{\mathcal{P}}_nU^{\dagger} \subseteq \bar{\mathcal{P}}_n$ satisfies $U\bar{\mathcal{P}}_n = \bar{\mathcal{P}}_nU$; hence, $\bar{\mathcal{P}}_n \subseteq \mathcal{C}_n$ is a normal subgroup. This allows us to allocate each element $U \in \mathcal{C}_n$ into an equivalence class $U\bar{\mathcal{P}}_n \subseteq \mathcal{C}_n$ and define a group operation between classes. In order to prove the isomorphism $\mathcal{C}_n/\bar{\mathcal{P}}_n \cong \mathcal{S}_n$, we need to check that two transformations U, U' are in the same equivalence class ($\exists \mathbf{v} : U = U'\sigma_{\mathbf{v}}$) if and only if they have the same symplectic matrix $S = S'$. Identity (A6) tells us that $U = U'\sigma_{\mathbf{v}}$ implies that $S = S'$. To prove the converse, let us assume that U, U' have symplectic matrices $S = S'$. Due to the fact U^{-1} has symplectic matrix S^{-1}, the product $U^{-1}U'$ has a symplectic matrix $S^{-1}S = \mathbb{1}$. As proven above, this implies that $U^{-1}U' \in \bar{\mathcal{P}}_n$, and therefore, both are in the same class.

Finally, for each symplectic matrix S, we define $\alpha[S,\mathbf{u}] = s[\mathbf{u}]$, where s is the phase function of an arbitrarily chosen element in the equivalence class defined by S. The phase function of the other elements in the class S is $s[\mathbf{u}] = \alpha[S,\mathbf{u}] + 2\langle \mathbf{v},\mathbf{u}\rangle$ for all $\mathbf{v} \in \mathbb{Z}_2^{2n}$. □

APPENDIX B: PROOF OF LEMMA 2

Proof. We start considering $\ln|\mathcal{S}_n|$,

$$
\begin{aligned}
\ln|\mathcal{S}_n| &= \ln\left[\prod_{i=1}^{n}(2^{2i}-1)\prod_{j=1}^{n}2^{2j-1}\right] \\
&= \sum_{i=1}^{n}\ln(2^{2i}-1)+\sum_{j=1}^{n}\ln 2^{2j-1} \\
&= \sum_{i=1}^{n}\ln\bigl[2^{2i}(1-2^{-2i})\bigr]+\sum_{j=1}^{n}(2j-1)\ln 2 \\
&= \sum_{i=1}^{n}2i\ln 2+\sum_{i=1}^{n}\ln(1-2^{-2i})+\sum_{j=1}^{n}(2j-1)\ln 2 \\
&= n(n+1)\ln 2+\sum_{i=1}^{n}\ln(1-2^{-2i})+n^2\ln 2 \\
&= n(2n+1)\ln 2+\sum_{i=1}^{n}\ln(1-2^{-2i}). \qquad \text{(B1)}
\end{aligned}
$$

We use $\frac{x}{1+x} < \ln(1+x) < x$, with $x \neq 0$ and $x > -1$, to upper and lower bound the logarithm in (B1). The corresponding bounds on $|\mathcal{S}_n|$ are obtained after exponentiating

$$\sum_{i=1}^{n}\ln(1-2^{-2i}) < -\sum_{i=1}^{n}2^{-2i} = -\sum_{i=1}^{n}\frac{1}{4^i} = -\frac{1}{3}\left(1-\frac{1}{4^n}\right). \qquad \text{(B2)}$$

$b(n)$ is defined to be $b(n) \equiv e^{-\frac{1}{3}(1-\frac{1}{4^n})}$; moreover, $b(n) < b(1) = e^{-\frac{1}{4}} < 0.78$.

To obtain the lower bound of $|S_n|$, from (B1), we have

$$\sum_{i=1}^{n}\ln(1-2^{-2i}) > -\sum_{i=1}^{n}\frac{2^{-2i}}{1-2^{-2i}}$$

and $a(n) \equiv e^{-\sum_{i=1}^{n}\frac{2^{-2i}}{1-2^{-2i}}} < b(n)$. We observe that

$$a(n) \equiv e^{-\sum_{i=1}^{n}\frac{1}{2^{2i}-1}} \geq e^{-\frac{1}{3}\sum_{i=0}^{n-1}\frac{1}{2^{2i}}} > e^{-\frac{4}{9}} > 0.64. \qquad \text{(B3)}$$

□

APPENDIX C: ADDITIONAL LEMMAS

In this section, we include additional lemmas that are used in the proof of other results.

Lemma 33. The number of k-dimensional subspaces of $\mathbb{Z}_2^n$ is

$$\mathcal{N}_k^n = \prod_{i=0}^{k-1}\frac{2^n-2^i}{2^k-2^i}. \qquad \text{(C1)}$$

Proof. Let us start by counting how many lists of k linearly independent vectors $(\mathbf{u}_1,\ldots,\mathbf{u}_k)$ are in $\mathbb{Z}_2^n$. The first vector $\mathbf{u}_1$ can be any element of $\mathbb{Z}_2^n$ except the zero vector $\mathbf{0}$, giving a total of (2^n-1) possibilities. Following that, $\mathbf{u}_2$ can be any element of $\mathbb{Z}_2^n$ that is not contained in the subspace generated by $\mathbf{u}_1$, which is $\{\mathbf{0},\mathbf{u}_1\}$, giving (2^n-2) possibilities. Analogously, $\mathbf{u}_3$ can be any element of $\mathbb{Z}_2^n$ that is not contained in the subspace generated by $\{\mathbf{u}_1,\mathbf{u}_2\}$, which is $\{\mathbf{0},\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_1+\mathbf{u}_2\}$, giving (2^n-2^2) possibilities. Following in this fashion, we arrive at the following conclusion. The number of lists of k linearly independent vectors is

$$\mathcal{L}_k^n = (2^n-2^0)(2^n-2^1)(2^n-2^2)\cdots(2^n-2^{k-1}). \qquad \text{(C2)}$$

It is important to note that many lists $(\mathbf{u}_1,\ldots,\mathbf{u}_k)$ generate the same subspace. Hence, in order to obtain $\mathcal{N}_k^n$, we have to divide $\mathcal{L}_k^n$ by the number of lists, which generate that same subspace.

First, we note that a list $(\mathbf{u}_1,\ldots,\mathbf{u}_n)$ is a basis of $\mathbb{Z}_2^n$ with its vectors in a particular order. Hence, $\mathcal{L}_n^n$ is the number of basis (in particular order) of $\mathbb{Z}_2^n$. Second, we use the fact that the subspace of $\mathbb{Z}_2^n$ generated by the list $(\mathbf{u}_1,\ldots,\mathbf{u}_k)$ is isomorphic to $\mathbb{Z}_2^k$ so that the number of basis (in a particular order) generating that subspace is $\mathcal{L}_k^k$. Putting things together, we obtain $\mathcal{N}_k^n = \mathcal{L}_k^n/\mathcal{L}_k^k$, as in (C1). □

Lemma 34. Let $\mathcal{N}_k^n$ be the number of k-dimensional subspaces of $\mathbb{Z}_2^n$; then, we have

$$2^{(n-k)k}(1-2^{k-n})^k \;\leq\; \mathcal{N}_k^n \;\leq\; 2^{(n-k)k}\min\{2^k,4\}. \qquad \text{(C3)}$$

Proof. Taking Lemma 33 and neglecting the negative terms in the numerator give

$$\mathcal{N}_k^n = \prod_{i=0}^{k-1} \frac{2^n - 2^i}{2^k - 2^i} \le \prod_{i=0}^{k-1} \frac{2^n}{2^k - 2^i} \tag{C4}$$

$$= \frac{2^{nk}}{2^{k^2}} \prod_{i=0}^{k-1} \frac{1}{1 - 2^{i-k}} = 2^{(n-k)k} \prod_{j=1}^{k} \frac{1}{1 - 2^{-j}} \tag{C5}$$

$$\le 2^{(n-k)k} \prod_{j=1}^{\infty} \frac{1}{1 - 2^{-j}}, \tag{C6}$$

where in the last inequality, we have extended the product to infinity. It turns out that this infinite product is the inverse of Euler's function ϕ evaluated at $1/2$, which has the value

$$\phi(1/2) = \prod_{j=1}^{\infty} (1 - 2^{-j}) \approx .28 \ge \frac{1}{4}. \tag{C7}$$

Combining the two above inequalities, we obtain

$$\mathcal{N}_k^n \le 2^{(n-k)k} 4. \tag{C8}$$

For the cases where $k = 0, 1$, we can improve this bound. When $k = 0$, the coefficient is 1 by definition, and when $k = 1$, the product $\prod_{i=0}^{k-1} (1 - 2^{i-k})^{-1}$ evaluates to 2. Hence, for $k = 0, 1$, we can replace 4 by 2^k, and therefore, this improvement is captured concisely by changing 4 to $\min\{2^k, 4\}$.

We obtain the lower bound by instead neglecting the negative terms in the denominator,

$$\mathcal{N}_k^n \ge \prod_{i=0}^{k-1} \frac{2^n - 2^i}{2^k} = \frac{2^{nk}}{2^{k^2}} \prod_{i=0}^{k-1} (1 - 2^{i-n}). \tag{C9}$$

The remaining product can be bounded using by

$$\prod_{i=0}^{k-1} (1 - 2^{i-n}) \ge \prod_{i=0}^{k-1} (1 - 2^{k-n}) \ge (1 - 2^{k-n})^k \tag{C10}$$

since $n \ge k > i$, and hence, we get the final lower bound. □

Lemma 35. The binomial coefficient can be bounded by

$$\binom{k + r - 1}{k} < (1 + r)^k \le (2r)^k. \tag{C11}$$

Proof. We start with the bound

$$\binom{k + r - 1}{k} = \prod_{i=1}^{k} \frac{r + k - i}{i} < \prod_{i=1}^{k} \left(1 + \frac{r}{i}\right). \tag{C12}$$

This follows from

$$\prod_{i=1}^{k} \frac{r + k - i}{i} = \frac{1}{k!} \prod_{i=1}^{k} (r + k - i) = \frac{1}{k!} (r + k - 1)(r + k - 2) \dots r, \tag{C13}$$

$$\prod_{i=1}^{k} \frac{r + i}{i} = \frac{1}{k!} \prod_{i=1}^{k} (r + i) = \frac{1}{k!} (r + k)(r + k - 1) \dots (r + 1). \tag{C14}$$

The order of the factors in the product in (C14) has been inverted. It is easy to see by inspection that (C13) lower bounds (C14). Further bounding, we get

$$\binom{k + r - 1}{k} < \prod_{i=1}^{k} \left(1 + \frac{r}{i}\right) \le (1 + r)^k \le (2r)^k. \tag{C15}$$

□

Lemma 36. For any given $S \in \mathcal{S}_{2n}$ written in the block form

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{C16}$$

according to the decomposition $\mathbb{Z}_2^{4n} = \mathbb{Z}_2^{2n} \oplus \mathbb{Z}_2^{2n}$,

$$\begin{pmatrix} B & A \\ D & C \end{pmatrix}, \quad \begin{pmatrix} C & D \\ A & B \end{pmatrix}, \text{ and } \quad \begin{pmatrix} D & C \\ B & A \end{pmatrix} \tag{C17}$$

are all also symplectic matrices.

Proof. Using the symplectic matrix

$$M = \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}, \tag{C18}$$

we show, using the result for the product of symplectic matrices, that the three permuted versions of S are also symplectic matrices via MS, SM, and MSM. □

Lemma 37. For any given $S \in \mathcal{S}_{2n}$ written in the block form

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{C19}$$

according to the decomposition $\mathbb{Z}_2^{4n} = \mathbb{Z}_2^{2n} \oplus \mathbb{Z}_2^{2n}$, the following two properties hold:

$$B = 0 \iff C = 0, \tag{C20}$$
$$A = 0 \iff D = 0. \tag{C21}$$

Proof. First, we consider the single case where $C = 0$. Following the algorithm for generating a symplectic matrix in Lemma 1, we see that A must be a $(2n \times 2n)$ symplectic matrix. Hence, any choice for the columns of B will have a symplectic form of one with at least one column of the matrix A. Therefore, to fulfill the symplectic constraints for the entire matrix S, the corresponding column of D must have a symplectic form of one with a column of C. However, this is not possible since $C = 0$; therefore, $B = 0$. Finally, by Lemma 36, this argument applies to each block. □

REFERENCES

[1] *Chaos and Quantum Physics: Proceedings of the Les Houches Summer School 1989*, edited by M. J. Giannoni, A. Voros, and J. Zinn-Justin (Elsevier, 1991).

[2] A. J. Short and T. C. Farrelly, New J. Phys. **14**, 013063 (2012).

[3] M. Rigol, V. Dunjko, and M. Olshanii, Nature **452**, 854 (2008).

[4] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, Adv. Phys. **65**, 239 (2016).

[5] J. M. Deutsch, H. Li, and A. Sharma, Phys. Rev. E **87**, 042135 (2013).

[6] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić, Nat. Phys. **14**, 745 (2018).

[7] S. Moudgalya, B. A. Bernevig, and N. Regnault, "Quantum many-body scars and Hilbert space fragmentation: A review of exact results," arXiv:2109.00548 [cond-mat.str-el] (2021).

[8] F. H. L. Essler and M. Fagotti, J. Stat. Mech.: Theory Exp. **2016**, 064002.

[9] J. Z. Imbrie, J. Stat. Phys. **163**, 998 (2016).

[10] J. Maldacena, Int. J. Theor. Phys. **38**, 1113 (1999).

[11] P. Hayden and J. Preskill, J. High Energy Phys. **2007**, 120.

[12] J. Maldacena and L. Susskind, Fortschr. Phys. **61**, 781 (2013).

[13] T. Prosen, J. Phys. A: Math. Theor. **40**, 7881 (2007).

[14] J.-S. Caux and J. Mossel, J. Stat. Mech.: Theory Exp. **2011**, P02023.

[15] J. A. Scaramazza, B. S. Shastry, and E. A. Yuzbashyan, Phys. Rev. E **94**, 032106 (2016).

[16] F. Haake, *Quantum Signatures of Chaos*, 3rd ed. (Springer, 2010).

[17] B. Simon, *Representations of Finite and Compact Groups* (AMS, 1996).

[18] P. Kos, M. Ljubotina, and T. Prosen, Phys. Rev. X **8**, 021062 (2018).

[19] T. Mondal and P. Shukla, Phys. Rev. E **99**, 022124 (2019).
[20] J. Maldacena, S. H. Shenker, and D. Stanford, J. High Energy Phys. **2016**, 106.
[21] Y. Gu and A. Kitaev, J. High Energy Phys. **2019**, 75.
[22] A. Chan, A. De Luca, and J. T. Chalker, Phys. Rev. X **8**, 041019 (2018).
[23] B. Bertini, P. Kos, and T. Prosen, Phys. Rev. X **9**, 021033 (2019).
[24] B. Bertini, P. Kos, and T. Prosen, SciPost Phys. **8**, 67 (2020).
[25] C. Chamon, A. Hamma, and E. R. Mucciolo, Phys. Rev. Lett. **112**, 240501 (2014).
[26] S. Zhou, Z.-C. Yang, A. Hamma, and C. Chamon, SciPost Phys. **9**, 87 (2020).
[27] B. Yan, L. Cincio, and W. H. Zurek, Phys. Rev. Lett. **124**, 160603 (2020).
[28] D. Gross, K. Audenaert, and J. Eisert, J. Math. Phys. **48**, 052104 (2007).
[29] R. Koenig and J. A. Smolin, J. Math. Phys. **55**, 122202 (2014).
[30] B. Bertini, P. Kos, and T. Prosen, Phys. Rev. Lett. **121**, 264101 (2018).
[31] M. Bukov, L. D'Alessio, and A. Polkovnikov, Adv. Phys. **64**, 139 (2015).
[32] D.-M. Schlingemann, H. Vogts, and R. F. Werner, J. Math. Phys. **49**, 112104 (2008).
[33] T. Farrelly, Quantum **4**, 368 (2020); arXiv:1904.13318 [quant-ph].
[34] C. Sünderhauf, D. Pérez-García, D. A. Huse, N. Schuch, and J. I. Cirac, Phys. Rev. B **98**, 134204 (2018).
[35] A. Chandran and C. R. Laumann, Phys. Rev. B **92**, 024301 (2015).
[36] A. W. Harrow and R. A. Low, Commun. Math. Phys. **291**, 257 (2009).
[37] A. Harrow and S. Mehraban, "Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates," arXiv:1809.06957 [quant-ph] (2018).
[38] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Commun. Math. Phys. **346**, 397 (2016).
[39] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Phys. Rev. Lett. **116**, 170502 (2016).
[40] N. Hunter-Jones, "Unitary designs from statistical mechanics in random quantum circuits," arXiv:1905.12053 [quant-ph] (2019).
[41] Y. Nakata, C. Hirche, M. Koashi, and A. Winter, Phys. Rev. X **7**, 021006 (2017).
[42] F. G. Brandão, W. Chemissany, N. Hunter-Jones, R. Kueng, and J. Preskill, PRX Quantum **2**, 030316 (2021).
[43] Z. Webb, Quantum Inf. Comput. **16**, 1379 (2016); arXiv:1510.02769 [quant-ph].
[44] J. Emerson, Y. S. Weinstein, M. Saraceno, S. Lloyd, and D. G. Cory, Science **302**, 2098 (2003).
[45] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[46] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter, Proc. R. Soc. London, Ser. A **465**, 2537 (2009).
[47] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504 (2011).
[48] A. J. Scott, J. Phys. A: Math. Theor. **41**, 055308 (2008).
[49] W. Brown and O. Fawzi, Commun. Math. Phys. **340**, 867 (2015).
[50] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal, IEEE Trans. Inf. Theory **48**, 580 (2002).
[51] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Nature **574**, 505 (2019).
[52] R. Nandkishore and D. A. Huse, Annu. Rev. Condens. Matter Phys. **6**, 15 (2015).
[53] S. Aaronson and D. Gottesman, Phys. Rev. A **70**, 052328 (2004).
[54] D. Gottesman, "The Heisenberg representation of quantum computers," arXiv:quant-ph/9807006 [quant-ph] (1998).
[55] I. Bloch, J. Dalibard, and S. Nascimbène, Nat. Phys. **8**, 267 (2012).
[56] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
[57] E. Artin, *Geometric Algebra* (Interscience Publishers, 1957).
[58] Y. Huang, F. G. S. L. Brandão, and Y.-L. Zhang, Phys. Rev. Lett. **123**, 010601 (2019).
[59] J. Gütschow, S. Uphoff, R. F. Werner, and Z. Zimborás, J. Math. Phys. **51**, 015203 (2010).
[60] H. Zhu, Phys. Rev. A **96**, 062336 (2017).
[61] D. A. Roberts and B. Yoshida, J. High Energy Phys. **2017**, 121.
[62] Z. Zimborás, T. Farrelly, S. Farkas, and L. Masanes, "Does causal dynamics imply local interactions?," arXiv:2006.10707 [quant-ph] (2020).
[63] V. Khemani, A. Vishwanath, and D. A. Huse, Phys. Rev. X **8**, 031057 (2018).
[64] N. Hunter-Jones, "Operator growth in random quantum circuits with symmetry," arXiv:1812.08219 [quant-ph] (2018).
[65] D. N. Page, Phys. Rev. Lett. **71**, 1291 (1993).
[66] E. Witten, Adv. Theor. Math. Phys. **2**, 253–291 (1998).
[67] D. Gross, S. Nezami, and M. Walter, Commun. Math. Phys. **385**, 1325–1393 (2021).
[68] A. R. Calderbank, E. M. Rains, P. M. Shor, and N. J. A. Sloane, IEEE Trans. Inf. Theory **44**, 1369 (1998).
[69] D. Gross, J. Math. Phys. **47**, 122107 (2006).