

Fastest Local Entanglement Scrambler, Multistage Thermalization, and a Non-Hermitian Phantom

Jaš Bensa and Marko Žnidarič

*Department of Physics, Faculty of Mathematics and Physics,
University of Ljubljana, 1000 Ljubljana, Slovenia*

 (Received 19 January 2021; revised 21 April 2021; accepted 25 May 2021; published 23 July 2021)

We study random quantum circuits and their rate of producing bipartite entanglement, specifically with respect to the choice of 2-qubit gates and the order (protocol) in which these are applied. The problem is mapped to a Markovian process, and we prove that there are large spectral equivalence classes—different configurations have the same spectrum. Optimal gates and the protocol that generate entanglement with the fastest theoretically possible rate are identified. Relaxation towards the asymptotic thermal entanglement proceeds via a series of phase transitions in the local relaxation rate, which is a consequence of non-Hermiticity. In particular, non-Hermiticity can cause the rate to be either faster or, even more interestingly, slower than predicted by the matrix eigenvalue gap. This result is caused by expansion coefficients that grow exponentially with system size, resulting in a “phantom” eigenvalue and is due to nonorthogonality of non-Hermitian eigenvectors. We numerically demonstrate that the phenomenon also occurs in random circuits with nonoptimal generic gates, random $U(4)$ gates, and also without spatial or temporal randomness, suggesting that it could be of wide importance in other non-Hermitian settings, including correlations.

DOI: [10.1103/PhysRevX.11.031019](https://doi.org/10.1103/PhysRevX.11.031019)

Subject Areas: Quantum Physics,
Quantum Information,
Statistical Physics

I. INTRODUCTION

Entanglement is one of the key properties that can make quantum systems different than classical ones, which is reflected in quantum information—large entanglement is a necessary resource to gain an advantage over classical computation, and many of the new phases discovered in recent decades can be distinguished by different patterns of entanglement [1]. Because entanglement, and the related concept of quantum information, plays such a fundamental role, it is also instrumental in the quest to push the boundaries of present-day physics, for instance, trying to figure out essential rules that quantum gravity should obey.

An elementary question is, how can one efficiently generate this resource? We focus on the so-called random quantum circuits [2], where quantum gates are chosen randomly from a certain set of gates. What set one uses might foremost depend on the available resources; while one wants to generate entanglement as quickly as possible, one must use the resources as efficiently as possible.

What is meant by efficient will depend on the context; however, there are some common conditions. Richness of nature emerges from two ingredients: innate properties of constituent objects (particles) and local interactions between them. Locality, being intimately related to causality, is rather important and is also typically the costly resource in quantum computation. Local transformations, i.e., one-site unitary operations, are faster to perform and typically have higher fidelity, while interactions in the form of two-site gates are expensive. We focus on random circuits in which one-site resources are random (1-qubit unitary from the Haar measure) while the two-site transformations are held fixed. Such a choice makes sense for two reasons: (i) It follows quantum information cost guidelines, and (ii) in some cases, it allows exact solvability. We demonstrate that taking a fixed good entangling 2-qubit gate is actually better than randomly choosing the whole 2-qubit transformation. Optimal random circuits that generate entanglement the fastest should also be of interest in the near-term applications of noisy quantum computers where they have been identified as prime candidates in the quest to demonstrate quantum supremacy [3]. Using the optimal circuit in such a quest is of high relevance as it directly affects the execution speed and therefore the attainable fidelity, which can, in turn, be crucial for an experiment to be on the right side of the supremacy frontier.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Random circuits also have another use—they are believed to correctly describe some of the properties of generic quantum systems, e.g., dynamics of operators [4–7], and can therefore serve as models of chaotic many-body systems [8]. The main advantage over chaotic systems is that their randomness enables analytical simplifications, leading to exact results. Exact results for entanglement evolution have also been obtained for the CFT [9]. Recently, the two pictures, that of random circuits and that of solvable quantum systems, emerged in the form of the so-called dual-unitary circuits, e.g., Ref. [10], which are solvable models having some elements of chaotic systems (as well as of integrable ones). As we shall see, extremal random circuits use the same building blocks as dual-unitary circuits. The interesting phenomena we discuss, though, are not limited to extremal nor to dual-unitary circuits.

Entanglement generation has also been studied in the context of black hole physics [11–13]. A question of intense interest is, in particular, the maximal possible entanglement generation, with various bounds and explicit results [14–16]. The fact that random circuits “scramble” information well can also be used for decoupling protocols [17]—a procedure in which any initial (local) correlations between an observer and a system are spread out globally, such that no local measurement on the system can reveal any correlation anymore; an observer becomes decoupled from a system (under local measurements). Such quantum information decoupling is, in fact, what is effectively going on during the thermalization process [18], so random circuits can also be thought of as being models of ideal thermalization (i.e., lacking any Hamiltonian-specific features).

We obtain a complete class of random protocols that optimally generate entanglement and, more importantly, identify several new and surprising features that can emerge in non-Hermitian matrices describing many-body systems. Specifically, we (i) find random circuits that produce entanglement in the fastest possible way. Our fastest circuit is significantly faster than the best previous random circuits. (ii) We identify a number of phase transitions in time—at certain moments, the convergence rate to the asymptotic “thermal” entanglement of random states suddenly changes. This change shows that an ideal thermalization modeled by a random circuit is a two-stage process rather than relaxation with a constant rate. (iii) We find that the convergence rate may not be given by the transfer matrix gap but can instead be either larger or smaller. This does not occur just in some obscure, unimportant cases but in almost every case we looked at, specifically in the fastest circuits as well as in generic ones, and also in circuits with random $U(4)$ gates much studied in the past. This failure of the spectral gap prediction happens in spite of the spectrum itself being rather innocuous; e.g., the second-largest eigenvalue λ_2 is gapped away from both $\lambda_1 = 1$ and $|\lambda_3| < |\lambda_2|$. We also interestingly observe that in the thermodynamic limit, the phenomenon is also observed in a single realization of a random circuit and

that, in fact, randomness is not necessary neither in space nor in time (one can use the same random 1-qubit transformation at all sites and at all times). The fact that a common “folk theorem,” that the decay is given by the gap, does not hold might have important implications in many areas of physics where one deals with non-Hermitian matrices, e.g., dissipative systems, transfer matrices, etc. Preliminary results show [19] the same phenomenon also in out-of-time-ordered correlations (OTOC).

One particularly intriguing case is where the decay is $\sim (\frac{1}{2})^t$ rather than $\sim (\frac{1}{4})^t$ suggested by $|\lambda_2| = \frac{1}{4}$. We show that this occurs because of a phantom eigenvalue $\frac{1}{2}$ —an eigenvalue that is not in the spectrum but is rather just mimicked by exponentially growing expansion coefficients in front of smaller (true) eigenvalues. We are not aware of any similar observations; the closest is perhaps a recent study [20] of vanishing gaps in Lindblad generators, finding that, because of exponentially growing expansion coefficients, the gap does not always give a physically relevant relaxation timescale [21].

Considering conceptual and practical importance of random circuits, it is not surprising that they have a long history. Let us make a brief overview of existing results, focusing on the speed with which entanglement is generated. One of the earliest works that studied the convergence properties of random circuits include Refs. [2,22]. The first exact results about the convergence rate towards random states were made possible by mapping [23] the average dynamics to a Markovian chain. Using the mapping, the question about the speed of generating entanglement boils down to the question about the gap of a certain transfer matrix. Such a mapping is rather fruitful not just for the specific question of entanglement generation [24–28] but also for a nice systematic treatment of any expectation that involves t copies of the propagator U and t copies of U^\dagger —a so-called unitary t -design [29–33]. The simplest entanglement quantifier purity $I = \text{tr} \rho_A^2$ therefore belongs under the realm of a 2-design. The Markovian mapping, in particular, allows us to get exact results for the purity convergence rate (purity “entanglement speed”) by recognizing that the resulting matrix is equivalent to an (integrable) spin chain [24]. Recently, the Markovian description has been used to describe evolution of purity for all possible bipartitions at once [27]. Exact results are now available for various protocols containing the 2-qubit random Haar $U(4)$ gate (including quantities beyond purity), e.g., choosing a random nearest-neighbor pair [24,25], picking a random permutation [34], or having a brick-wall pattern [4,5,35], or for random single-site unitaries and a global phase [28]. If one allows the coupling between all pairs of qubits, one can also get the exact entanglement speed for some nonrandom gates, in particular, for the important XY gate (as well as for the CNOT) [24]; see Table I. Exact results are also available in the limit of large local Hilbert space dimension q [4,8,25,34–36] (we focus on qubits, $q = 2$). For numerical results, see Refs. [26,33,36–38].

TABLE I. Existing exact results for the purity decay rate r_E , defined as $I(t) \sim \exp(-r_E t)$, in qubit random circuits (cf. Fig. 1). Per unit of time, $\sim n$ gates are applied. Deterministic 1-dimensional (1-dim.) nearest-neighbor (NN) protocol with a brick-wall (BW) pattern of gates is faster than randomly choosing a NN pair on which the gate acts. Allowing coupling between an arbitrary pair of gates (∞ -dim. all-all) is even faster. Results are shown for open boundary conditions (OBC) and periodic boundary conditions (PBC). U(4) denotes a Haar random gate, while the XY gate is also known as the iSWAP or DCNOT gate.

Gate	Protocol		Rate r_E	Reference
	Configuration	Boundary conditions		
1-dim.(NN):				
U(4)	Random	PBC	$\frac{2}{5} = 0.4$	[24]
U(4)	Random	OBC	$\frac{1}{5} = 0.2$	[24]
U(4)	Permutation	OBC	$\ln \frac{3}{2} \approx 0.40$	[34]
U(4)	BW	PBC	$4 \ln \frac{5}{4} \approx 0.89$	
U(4)	BW	OBC	$2 \ln \frac{3}{4} \approx 0.45$	[4,5,35]
∞ -dim. (all-all):				
U(4)	Random		$\frac{6}{5} = 1.2$	[24,25]
XY, CNOT	Random		$\frac{4}{3} \approx 1.33$	[24]

Summarizing these results in Table I, we see that using NN qubit gates, the fastest known protocol is the BW configuration with a U(4) gate for which the entanglement rate defined via a long-time purity decay $I(t) \asymp \exp(-r_E t)$ is $r_E \approx 0.45$ [39] per boundary link. We find a protocol that is more than 3 times faster and has the maximal possible entanglement speed. The optimal 2-qubit gate will turn out to be the so-called XXZ gate (also known as the pSWAP), a special case of which is the XY gate (iSWAP). The same gate has appeared before in the context of random circuits: It has been observed numerically that the XY gate is the fastest of all gates for random NN, or the all-all protocol [38] (which, though, are much slower than our best BW protocol). The same type of gate has recently emerged in dual-unitary circuits [10]. Because of their special properties, such circuits (see also related concepts [41] of 2-unitaries [42] and perfect tensors [43]) allow exact results, for instance, of Rényi entropies for a nonsymmetric bipartition in a disordered kicked Ising model [10] (this system has a maximal entanglement rate provided one counts time in an optimal way); see also Ref. [44], and for a symmetric half-infinite bipartition, see Ref. [45]. One can also calculate the tripartite information [46] or find circuits with maximal butterfly velocity [47] in the dual-unitary context.

We characterize entanglement generation through the decay of average purity $I(t) = \text{tr} \rho_A^2(t)$. Because we want to understand entanglement generation on a global scale, we focus on a symmetric bipartition [48] of our system

of n qubits into two equal subsystems A and B, each with $n_A = n_B = n/2$ qubits. In other words, for a chain geometry of qubits, we split the chain in two equal halves of consecutive qubits. We calculate the average purity behavior by mapping its dynamics to a Markovian process [23,24,27], thus reducing the problem to that of properties of a particular transfer matrix M . At long times, purity will decay exponentially with a rate r_E determining how fast entanglement is produced. Because the purity entanglement rate r_E will be proportional to the area \mathcal{A} of the boundary between A and B, one often considers the entanglement speed v_E (also called the purity speed, or tsunami velocity [15]) defined as

$$I(t) \asymp \exp(-r_E t), \quad r_E \equiv v_E \mathcal{A} s_\infty. \quad (1)$$

For random circuits, the asymptotic state is an infinite-temperature state having maximal entropy density $s_\infty = \ln q$, while \mathcal{A} is simply the number of 2-qudit gates applied across the boundary between subsystems A and B per unit of time. If one uses natural units of time such that one applies one gate on each bond per unit of time, then \mathcal{A} is just the number of bonds cut by a bipartition, which, for the symmetric bipartition we use, is $\mathcal{A} = 1$ for OBC and $\mathcal{A} = 2$ for PBC. Because one can generate two maximally entangled 2-qudit states per one application of a two-site gate [51], the maximal v_E one can achieve is

$$v_E \leq 2. \quad (2)$$

In quantum circuits, one therefore has the upper bound $r_E \leq 4 \ln q$ for PBC and $r_E \leq 2 \ln q$ for OBC.

One can also study other quantities measuring entanglement, like the von Neumann entanglement entropy $S(t) = -\text{tr}(\rho_A \ln \rho_A)$ or the Rényi entropies $S_r(t) = \ln \rho_A^r / (1 - r)$, and use them to define corresponding rates and speeds. In general, all of those entanglement speeds can be different [35], an example being a circuit with random U(4) gates. The reason for the difference is the variation in entanglement properties of different circuit realizations; for instance, some rare realizations could use very inefficient entanglers from the U(4) set of gates. For the circuits we focus on, we expect such effects to be much less pronounced. In particular, the fastest scramblers that we find have the maximal speed allowed by causality, so all Rényi entropies should have the same maximal speed (similarly as, e.g., in Ref. [10]). Furthermore, even for nonmaximal random circuits with fixed 2-qubit gates, the variation in entanglement between different realizations should be much smaller than, e.g., for the random U(4) case because the entangling 2-qubit gates are the same at every step [even for random U(4), the difference between the speed from the average purity and the average second Rényi entropy is likely very small [35]]. The circuits we study will, at not too short times, produce typical states that have good

TABLE II. New results for the purity decay rate r_E in random circuits (cf. Fig. 1). The fastest entanglement generation for local couplings is achieved for the XXZ gate in a BW configuration (bold). Such circuits are optimal and much faster than, e.g., using random U(4) gates or all-all coupling (Table I), or using the CNOT gate. The staircase configuration S is, on the other hand, the slowest. The rate is, in general, not given by the second-largest eigenvalue $|\lambda_2|$ of the associated transfer matrix.

Gate	Protocol		Rate r_E	$-\ln \lambda_2 $ (for XY)
	Configuration	Boundary conditions		
1-dim.(NN):				
XXZ	BW	PBC	$\ln 16 \approx 2.77$	$\ln 9 \approx 2.20$
XXZ	BW	OBC	$\ln 4 \approx 1.39$	$\ln 4 \approx 1.39$
XXZ	S	PBC	$\ln 4 \approx 1.39$	$\ln 3 \approx 1.10$
XXZ	S	OBC	$\ln 2 \approx 0.69$	$\ln 4 \approx 1.39$
CNOT	BW	PBC	≈ 0.591	
CNOT	BW	OBC	≈ 0.284	
CNOT	S	PBC	≈ 0.570	
CNOT	S	OBC	≈ 0.296	

self-averaging properties, and we expect von Neumann entropy and Rényi entropy to behave essentially the same as the logarithm of the average purity, $S(t) \approx S_2(t) = -\ln I(t)$. We leave possible differences for future work. In the remainder of the paper, we always discuss only purity rates and speeds.

Our results for r_E are summarized in Table II and in Fig. 1. We can see that the rates are, in general, not equal to the ones suggested by the second-largest eigenvalue λ_2 of the transfer matrix M and that, in the two optimal cases, namely, the BW configuration with PBC, or the OBC, the entanglement velocity is $v_E = 2$ and therefore saturates the bound (2). The optimal gate is, in both cases, found to be the XXZ gate for any value of the anisotropy $0 \leq a_z < 1$ in front of the $\sigma_j^z \sigma_{j+1}^z$ coupling [see Eq. (4) for its definition]. In particular, using the XXZ gate is much faster than using random U(4) gates, as well as, for instance, the CNOT gate. We do not find any transition with a_z in the relevant entanglement rate r_E , though we do find it in the leading eigenvalue λ_2 of the transfer matrix

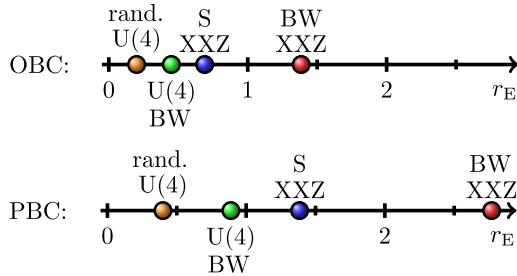


FIG. 1. Overview of purity entanglement generation rates r_E for random circuits with different gates and ordering of NN gates: random (rand.) vs BW vs S. See Fig. 2 for the BW and S structure.

(which, in most cases, determines the late-time entanglement rate after the phase transition at the thermalization time). For dual-unitary systems with transitions with a_z in transfer matrices, see Refs. [46,53].

II. METHODS

A. Random quantum circuits

We study random quantum circuits in which 2-qubit unitary transformations are drawn from some set of unitary gates. One elementary gate—a 2-qubit transformation $U_{i,j}$ acting on qubits i and j —is a product of independent single-qubit random unitaries V_i acting on the i th qubit and V_j acting on the j th qubit, drawn according to the Haar measure on group U(2), and a two-site unitary $W_{i,j}$. Therefore, the whole gate is $U_{i,j} = W_{i,j} V_i V_j$. The two-site $W_{i,j}$ will be the same for all steps, while the single-site V_i are independent for each step and each qubit. The formalism that we use can also accommodate the situation where $W_{i,j}$ would be random from U(4) (we briefly comment on that case in Sec. IV C); however, such circuits produce entanglement more slowly (see Table I) than the ones we study.

A given random circuit is specified by a fixed two-site W and a sequence of qubit pairs, called a protocol or configuration, on which $U_{i,j}$ are applied. For instance, one could choose W to be the CNOT gate and a sequence to be a brick-wall pattern (Fig. 2). We find a gate W and a protocol that generate entanglement the fastest.

In the main body of the paper, we limit ourselves to protocols in one dimension (1D), where the gates are allowed only between NN qubits with either OBC or PBC; i.e., qubits are on a line or on a circle. We only briefly mention the 2D case of the Sycamore [3] quantum processor, and in the Appendix G, we discuss the all-to-all coupling (NN protocols are faster than those). Focusing on the 1D setting, the space of quantum circuits over which we need to optimize is still large (infinite); therefore, simplification is necessary. First, we focus on periodic protocols in which the geometry of applied gates repeats after each period. One period consists of exactly one gate applied on each allowed bond; that is, for OBC, there are $T = n - 1$ gates in a period, whereas there are $T = n$ for

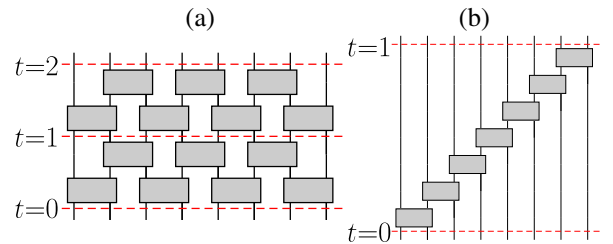


FIG. 2. Illustration of different quantum circuit configurations (here, for OBC): (a) brick-wall (BW) configuration and (b) staircase (S) configuration.

PBC. Optimization therefore needs to be performed over all $T!$ permutations describing different orderings of gates within one period, and over all choices of the gate W . Our time t that we use throughout the paper measures the number of periods (see Fig. 2). In Sec. III, we prove that one does not need to consider all $n!$ permutations but only $\sim n$ of them.

One can also reduce the number of W that need to be checked. Because we have a single-site invariance due to random single-qubit gates, it suffices to consider W in the canonical form parametrized by only three parameters [instead of 16 for general $W \in U(4)$]. Namely, every 2-qubit unitary can be written in the following canonical form [54,55],

$$W_{j,k} = V_j^{(1)} V_k^{(2)} w_{j,k}(\mathbf{a}) V_j^{(3)} V_k^{(4)},$$

$$w_{j,k}(\mathbf{a}) = \exp \left[i \frac{\pi}{4} (a_x \sigma_j^x \sigma_k^x + a_y \sigma_j^y \sigma_k^y + a_z \sigma_j^z \sigma_k^z) \right], \quad (3)$$

where $V_k^{(\alpha)}$ are one-site unitary operators, $\sigma^{x,y,z}$ are Pauli matrices, and $\mathbf{a} = (a_x, a_y, a_z)$ has three real parameters, which can be constrained to $0 \leq a_z \leq a_y \leq a_x \leq 1$. A detailed explanation on how to calculate this decomposition from an arbitrary unitary two-site gate W can be found in Ref. [56]. Fixed single-site $V_k^{(\alpha)}$ for a specific gate W can be lumped together with random unitaries V_k , so we need to explore only canonical gates $w(\mathbf{a})$. For instance, the CNOT gate has $\mathbf{a} = (1, 0, 0)$; the gate that turns out to be the fastest is the XY gate with $\mathbf{a} = (1, 1, 0)$ (also called the iSWAP or DCNOT gate). More generally, the optimal set will consist of gates

$$W_{\text{XXZ}} = \exp \left[i \frac{\pi}{4} (\sigma_j^x \sigma_k^x + \sigma_j^y \sigma_k^y + a_z \sigma_j^z \sigma_k^z) \right], \quad (4)$$

i.e., with parameters $\mathbf{a} = (1, 1, a_z)$ and $a_z < 1$, which we call the XXZ gate (also known as the pSWAP—parametric SWAP gate). The SWAP gate reached at $\mathbf{a} = (1, 1, 1)$ generates no entanglement and therefore has completely different entanglement properties than the XXZ family of gates. For our best protocols, one therefore has, in the thermodynamic limit (TDL), a discontinuous transition from the maximal rate at $a_z < 1$ (XXZ gate) to a zero rate at $a_z = 1$ (SWAP gate).

To quantify bipartite entanglement of pure states, we use purity $I(t)$,

$$I(t) = \text{tr}_A \rho_A^2(t), \quad (5)$$

where $\rho_A(t) = \text{tr}_B |\psi(t)\rangle\langle\psi(t)|$ and $|\psi(t)\rangle = U(t)|\psi(0)\rangle$, with $U(t) = \prod_{(i,j)} U_{i,j}$ a product of all 2-qubit gates up to time t [because we have T gates per unit of time, the total number of 2-qubit gates in $U(t)$ is therefore $T \cdot t$]. The whole system of n qubits is bipartitioned into a subsystem

A with n_A qubits and a complement B with n_B qubits, $n = n_A + n_B$. In explicit numerical demonstrations, we always use a symmetric half-half bipartition, where A is composed of the first $n/2$ consecutive qubits while B is composed of the rest. However, the results do not depend on the specifics of the bipartition, provided one has an extensive $n_A \propto n$ connected subsystem A. The initial state $|\psi(0)\rangle$ is always chosen to be fully separable with respect to any bipartition such that $I(0) = 1$. Purity will therefore decay with time from its initial 1 to the asymptotic value being equal to the purity of random states [57],

$$I_\infty = \frac{2^{n_A} + 2^{n_B}}{1 + 2^n}. \quad (6)$$

For large t , our random circuit uniformly samples U from the unique unitarily invariant Haar measure. Observe that this asymptotic purity is not exactly equal to the purity of the infinite-temperature state $\rho_A \propto \mathbb{1}$, which is $I(\beta = 0) = (1/2^{n/2})$, whereas $I_\infty = (2/2^{n/2})/(1 + 2^{-n})$ for the symmetric bipartition. For easier understanding, we often show

$$S_2(t) = -\log_2 I(t), \quad (7)$$

which is a quantity that will grow from 0 at $t = 0$ to its maximal value $S_2(\infty) \approx (n/2) - 1$ (for a half-half bipartition), which is reached when the state $|\psi(t)\rangle$ converges towards a random state. Note that $S_2(t)$ will reach its saturation $\sim n/2$ at time t_∞ , which is given by

$$\frac{t_\infty}{n_A} = \frac{\ln 2}{r_E} = \frac{1}{v_E \mathcal{A}}. \quad (8)$$

For instance, for the BW PBC random circuit, the scaled thermalization time will be $t_\infty/n_A = 1/4$. To more clearly show the relaxation process towards I_∞ , we also study

$$\Delta S_2(t) = -\ln |I(t) - I_\infty|. \quad (9)$$

Looking at $\Delta S_2(t)$ instead of just $S_2(t)$ will allow us to discuss relaxation (thermalization) on timescales beyond t_∞ , which will reveal interesting new phase transitions. Because the state reached (at nonsmall times) is akin to a random state, one has good self-averaging properties in the thermodynamic limit; thus, other measures of entanglement, like the von Neumann entropy of the Rényi entropies, will behave similarly to the above $S_2(t)$; see Appendix F 1 for an explicit numerical demonstration. Also, because the variance of $I(t)$ between different circuit realizations is small in the TDL, one can also study the average $I(t)$, i.e., averaging over different realizations of single-qubit unitaries V_j . Such averaging is crucial to get an analytically tractable setting, and next, we describe the formalism used for that purpose.

B. Markov chain description

As mentioned, one can describe [23] the average dynamics of ρ_A^2 , i.e., of squares of the expansion coefficients of ρ_A , by a Markovian matrix $M_{i,j}$ that describes the averaged action of one 2-qubit gate $U_{i,j}$. Such a description has been widely used, e.g., Refs. [24,25,30,31,33], for random $U(4)$ gates or for Clifford gates (gates that map a product of Pauli matrices to a single product of Pauli matrices, examples being the CNOT or the XY gate). Recently, a different approach has been taken [27] in which one directly describes the evolution of purity for all possible bipartitions rather than of ρ_A^2 . This approach allows one to describe purity evolution for a protocol with an arbitrary 2-qubit gate $W_{i,j}$. As we shall see, despite being more general and formally different, for the cases where the old method works (Clifford gates), this approach will result in the very same transfer matrix $M_{i,j}$ as obtained in Ref. [24].

One starts by formally defining a Hilbert space with basis $|\mathbf{s}\rangle$, where $\mathbf{s} = (s_1, s_2, \dots, s_n)$ with $s_i \in \{\uparrow, \downarrow\}$. A given \mathbf{s} encodes a bipartition; that is, if $s_i = \uparrow$, the i th qubit is in subsystem A, and otherwise, it is in B. A state vector whose evolution we want to describe is then defined as

$$\Phi'(t) = \sum_{\mathbf{s}} I_{\mathbf{s}}(t) |\mathbf{s}\rangle, \quad (10)$$

where $I_{\mathbf{s}}(t)$ is the purity of a state $|\psi(t)\rangle$ for a given bipartition labeled by \mathbf{s} . The state $\Phi'(t)$ therefore implicitly depends on the initial state—for the most interesting, fully separable, initial state $|\psi(0)\rangle$, the initial $\Phi'(0) \equiv \Phi'_0$ is simply $\Phi'_0 = (|\uparrow\rangle + |\downarrow\rangle)^{\otimes n}$. Averaging over two 1-qubit random matrices V_i and V_j , it has been shown [27] that the vector Φ' is mapped to $M'_{i,j}\Phi'$ (see also Appendix A). Repeating this step for all T 2-qubit unitaries (each involving independent 1-qubit random unitaries) that are applied per unit of time, we get the transformation

$$\Phi'(t+1) = M'\Phi'(t), \quad M' = \prod_{(i,j)}^T M'_{i,j}. \quad (11)$$

The dynamics of purity—the expansion coefficients of $\Phi'(t)$ —is therefore given by iterating a single-step transfer matrix M' describing the Markovian average purity evolution. To get purity for the most interesting, symmetric, half-half bipartition, one just has to project $\Phi'(t)$ to the basis state $|\uparrow\dots\uparrow\downarrow\dots\downarrow\rangle$. Slightly abusing notation and defining the initial vector $\Phi'_0 = (1, 1, \dots, 1)$ and the bipartition basis vector Φ'_{half} with components $[\Phi'_{\text{half}}]_k = \delta_{k,2^{n/2}}$, we have the average purity after t steps,

$$I(t) = \Phi'_{\text{half}} (M')^t \Phi'_0. \quad (12)$$

The 4×4 matrix $M'_{i,j}$ can be calculated for an arbitrary $W_{i,j}$ in its canonical form $w(\mathbf{a})$ [Eq. (3)]. Following the procedure in Ref. [27], we obtain (Appendix A)

$$M'_{i,j} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ h & b_+ & b_- & h \\ h & b_- & b_+ & h \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad b_{\pm} = \frac{1}{36}(3 \pm 6u + 5v), \quad (13)$$

where the basis is ordered as $s_i s_j = \{\uparrow\uparrow, \uparrow\downarrow, \downarrow\uparrow, \downarrow\downarrow\}$, and $h = \frac{1}{9}(3 - v)$, $u = \cos(\pi a_x) + \cos(\pi a_y) + \cos(\pi a_z)$ and $v = \cos(\pi a_x) \cos(\pi a_y) + \cos(\pi a_x) \cos(\pi a_z) + \cos(\pi a_y) \cos(\pi a_z)$. For instance, for the XY gate, we have $u = v = -1$, resulting in

$$M'_{i,j}(\text{XY}) = \frac{1}{9} \begin{pmatrix} 9 & 0 & 0 & 0 \\ 4 & -2 & 1 & 4 \\ 4 & 1 & -2 & 4 \\ 0 & 0 & 0 & 9 \end{pmatrix}. \quad (14)$$

Not surprisingly, because of local unitary averaging, the parameters h , u , v can be related to known 2-qubit gate objects. Specifically, one has $h = 2e(W)$, where $e(U)$ is the entangling power [58] of unitary U , while v itself is equal to the local invariant G_2 [59,60]. Note, though, that knowing the two-site properties of the gate W does not help us understand the entanglement generating properties of the whole circuit. As we will see, the latter is a full many-body property (the interesting effects that we discuss only arise in the TDL) and also crucially depends on the configuration, not just on the choice of a 2-qubit gate. For instance, the entangling power [61] of W is maximal for the XY as well as for the CNOT gate, and continuously decreases with a_z for the XXZ gate. On the other hand, our results show that, in the many-body setting of random circuits, the CNOT gate is, for instance, more than 4 times slower than the XY gate (Table II) and that, for the XXZ gate, the rate does not depend on a_z .

We observe that the matrix $M'_{i,j}$ is not symmetric. Often, it is easier to work with symmetric matrices, and it turns out that it is possible to transform $M'_{i,j}$ to a symmetric form. Such transformation will make symmetries more transparent and facilitate comparison with previous Markovian descriptions of random circuits. We can obtain the symmetric form with a similarity transformation $M_{i,j} = A_i^{-1} A_j^{-1} M'_{i,j} A_i A_j$, where

$$A_i = \begin{pmatrix} 1 & -1/\sqrt{3} \\ 1 & 1/\sqrt{3} \end{pmatrix}, \quad (15)$$

arriving at

$$M_{i,j} = \begin{pmatrix} \frac{1}{36}(33 + v) & 0 & 0 & \frac{1}{12}(3 - v) \\ 0 & u_+ & u_- & 0 \\ 0 & u_- & u_+ & 0 \\ \frac{1}{12}(3 - v) & 0 & 0 & \frac{1}{4}(1 + v) \end{pmatrix}, \quad (16)$$

where $u_{\pm} = (3 \pm u)/6$. This elementary two-site matrix can be equivalently written in terms of Pauli matrices,

$$M_{i,j} = d\mathbb{1} + J_x \sigma_i^x \sigma_j^x + J_y \sigma_i^y \sigma_j^y + J_z \sigma_i^z \sigma_j^z + \frac{h}{2} (\sigma_i^z + \sigma_j^z), \quad (17)$$

where $d = (39 + 6u + 5v)/72$, $J_x = (9 - 2u - v)/24$, $J_y = (3 - 2u + v)/24$, $J_z = (3 - 6u + 5v)/72$, and $h = (3 - v)/9$. For the XY gate, it is equal to

$$M_{i,j}(\text{XY}) = \frac{1}{9} \begin{pmatrix} 8 & 0 & 0 & 3 \\ 0 & 3 & 6 & 0 \\ 0 & 6 & 3 & 0 \\ 3 & 0 & 0 & 0 \end{pmatrix}. \quad (18)$$

Observe that even if one starts with a noninteracting gate like the XY, the resulting M is interacting ($J_z \neq 0$). For Clifford gates, e.g., CNOT and XY, this is, in fact, the same transfer matrix as the one already derived in Ref. [24]. The new mapping [27] that we use therefore generalizes the Markovian mapping to non-Clifford gates. As a side note, the transfer matrix $M_{i,j}$ for a random U(4) gate [24] is obtained by formally using $u = 0$ and $v = -3/5$ in Eq. (17).

The similarity transformation A preserves the spectrum—the spectrum of M is the same as that of M' ; only the vectors have to be transformed. Similarly as in Eq. (11), we use M to denote a product of T 2-qubit $M_{i,j}$ appearing in one unit of time. We use a prime to denote matrices and vectors written in the (unrotated) basis $|\mathbf{s}\rangle$, like $M'_{i,j}$ or $\Phi'(t)$, whereas unprimed objects, like $M_{i,j}$ or $\Phi(t)$, are in the basis transformed by A (where needed, we denote the basis vectors of this space by \mathbf{e}_k). For instance, the initial vector corresponding to a product initial state Φ'_0 is transformed to $\Phi_0 = (A^{-1})^{\otimes n} \Phi'_0$ and has components $\Phi_0 = (1, 0, \dots, 0)$, while $\Phi_{\text{half}} = (A^T)^{\otimes n} \Phi'_{\text{half}}$. A boldface notation is used to stress that a vector (components) is written in a given basis, like Φ_0 , while Φ_0 is used for a basis-independent ket notation. The average purity after t periods of our random circuits in this new basis is $I(t) = \Phi_{\text{half}} M^t \Phi_0$. For a symmetric bipartition and a product initial state, because of the structure of Φ_0 and Φ_{half} , purity is determined by one column of M^t or, equivalently, by one row of $(M')^t$. Because of a block structure of $M_{i,j}$ [Eq. (16)], it is also clear that M conserves the parity $Z = (-1)^{N_{\downarrow}} = \prod_j \sigma_j^z$ of a vector $|\Phi(t)\rangle$, i.e., the number of down spins N_{\downarrow} . Because the initial state in the primed basis $|\Phi'_0\rangle$ is invariant under the particle-hole transformation (i.e., spin-flip), $X|\Phi'_0\rangle = |\Phi'_0\rangle$, where $X = \prod_j \sigma_j^x$ (this is because the purity is invariant under exchanging subsystems $A \leftrightarrow B$), the parity of the initial $|\Phi_0\rangle$ is always even (even N_{\downarrow}). Indeed, if $X|\Phi'_0\rangle = |\Phi'_0\rangle$, then $XA^{\otimes n}|\Phi_0\rangle = A^{\otimes n}|\Phi_0\rangle$, leading to $(A^{-1})^{\otimes n}XA^{\otimes n}|\Phi_0\rangle = |\Phi_0\rangle$. Because

$A_j^{-1} \sigma_j^x A_j = \sigma_j^z$, one immediately gets $(-1)^{N_{\downarrow}} |\Phi_0\rangle = |\Phi_0\rangle$. The relevant parity sector of M describing purity evolution of physical $|\Phi_0\rangle$ is therefore the one with an even number of down spins. We note that, in some cases, other additional symmetries are present. For instance, for BW protocol with PBC and n divisible by 4 one also has a reflection symmetry about the site $n/2$ and the translation symmetry by two sites.

Another useful observation is that $M_{i,j}$ has eigenvectors that are independent of gate parameters u and v , namely,

$$\begin{aligned} \mathbf{v}_1 &= (3, 0, 0, 1), & \lambda &= 1, \\ \mathbf{v}_2 &= (0, 1, 1, 0), & \lambda &= 1, \\ \mathbf{v}_3 &= (0, -1, 1, 0), & \lambda &= u/3, \\ \mathbf{v}_4 &= (-1, 0, 0, 3), & \lambda &= (3 + 5v)/18. \end{aligned} \quad (19)$$

One consequence of this is that M always has (at least) two eigenvalues equal to $\lambda_1 = 1$. The corresponding eigenvectors are steady states and are also independent of the gate W , which is consistent with the asymptotic convergence to purity of random states. The physically relevant steady-state vector Φ'_{∞} , i.e., the even-parity eigenvector of M' with $\lambda_1 = 1$, has components equal to I_{∞} [Eq. (6)],

$$|\Phi'_{\infty}\rangle = \sum_{\mathbf{s}} I_{\mathbf{s}}(\infty) |\mathbf{s}\rangle. \quad (20)$$

One can indeed use the above one-site eigenvectors $\mathbf{v}_{1,2}$ to construct the steady state on n qubits and explicitly verify I_{∞} (see Appendix B). The steady state (20) can also be written compactly as a matrix product state of rank 2; see Appendix C.

Time evolution of the average purity is therefore Markovian [Eq. (11)], i.e., there is no memory. The transition matrix, either M' or M [Eqs. (13) and (16)], is, however, not stochastic (rows do not sum to 1) and therefore does not describe transition probabilities. The probabilistic interpretation is recovered if one works with squares of the expansion coefficients of the state ρ on the original full 16-dimensional two-site Pauli basis [23]; however, it is lost if one uses reduction [24] down to four nontrivial basis states that results in M [Eq. (18)] or directly works with purity evolution [27].

Note that the transfer matrix for one step M is, in general, not symmetric, even though the individual gates $M_{i,j}$ are. The spectrum of M is therefore complex and contained within the unit circle. Provided the spectrum has a gap (which is always the case for nontrivial gates), the next-largest eigenvalue $|\lambda_2|$ should determine the asymptotic decay of purity and therefore the growth of entanglement [decay of $I(t)$]. Asymptotically, one should have

$$|I(t) - I_{\infty}| \asymp |\lambda_2|^t, \quad (21)$$

where we denote by λ_2 the largest eigenvalue that is smaller than 1 (in absolute value). The corresponding entanglement rate (1) would then be $r_E = -\ln |\lambda_2|$.

In the next section, we therefore study the gap of M for various gates W and configurations and, in particular, prove that there are large classes of configurations with the same spectrum, which enables us to then find circuits with the largest gap, i.e., the smallest $|\lambda_2|$. However, it turns out, surprisingly, that the entanglement rate is, in fact, not given by $|\lambda_2|$ because of the nonsymmetric nature of M . As we will show, one can have two configurations with exactly the same spectrum but with different entanglement rates (for example, taking XXZ gates and comparing the S configuration with OBC, and BW with OBC, see Table II). Nevertheless, the spectral analysis suggests good candidates for optimal circuits. Furthermore, the two theorems on spectral equivalence are completely general and hold for products of any matrices acting on nearest-neighbor sites (for PBC, the matrices need to be symmetric) and could therefore be of use in other situations.

Combining spectral results on λ_2 from Sec. III and numerically calculated true entanglement rates in Sec. IV will result in the main message of our work: In a many-body setting involving non-Hermitian transfer matrices, spectral analysis can give the wrong results.

III. SPECTRAL OPTIMIZATION

In this section, we determine the protocol and the gate with the smallest second-largest eigenvalue $|\lambda_2|$, i.e., the largest gap of M , in a 1D geometry where only local gates between nearest-neighbor sites are allowed. Specifically, if we have T distinct NN gates, i.e., $T = n - 1$ for OBC and $T = n$ for PBC, we find a sequence (configuration) of those T gates among all $T!$ different permutations as well as among all possible choices of a 2-qubit gate $M_{i,i+1}$ (held fixed for all gates) that has the smallest $|\lambda_2|$.

As mentioned, we treat chains with periodic boundary conditions, where the gate $M_{n,1}$ is also allowed, and chains with open boundary conditions. In both cases, we first prove that, among all factorially many $T!$ configurations, there are only a few with different gaps. The outline of proofs is to show that the spectrum of a product of gates $M_{i,i+1}$ does not change under certain rearrangements (holding the two-site gate, i.e., $M_{i,j}$, fixed), which then allows us to focus on equivalence classes defined as an arrangement with the same spectrum. We use the equivalence notation $A \simeq B$ if the matrices A and B have the same spectrum. It turns out that, for OBC, there is only one equivalence class, while for PBC, there are $\lfloor n/2 \rfloor$ ($\lfloor n/2 \rfloor$ is the largest integer smaller than or equal to $n/2$). This greatly reduces the complexity and allows us to numerically find the optimal configuration and the gate.

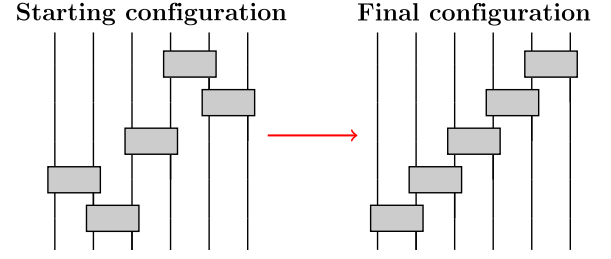


FIG. 3. Transforming the initial configuration $M = M_{4,5} \times M_{5,6} M_{3,4} M_{1,2} M_{2,3}$ to the final S configuration preserves the spectrum of M . For OBC, any configuration can be transformed to S.

A. OBC protocols

For OBC, we have $(n - 1)!$ possible configurations; however, all of them are spectrally equivalent. Namely, one can prove the following theorem.

Theorem 1: Any product of $n - 1$ two-site matrices acting on $n - 1$ distinct NN sites on a line (OBC protocol) is spectrally equivalent to the canonical staircase configuration (denoted, briefly, by S):

$$S_1^{(n-1)} = M_{n-1,n} \dots M_{2,3} M_{1,2}. \quad (22)$$

The superscript in $S_j^{(k)}$ indicates the number of consecutive gates in such canonical order, and the subscript j is for the first site on which they act. For the proof that uses the spectral equivalence under cyclic permutations, $ABC \simeq CAB$, see Appendix D 1.

As a consequence, one can transform any configuration with OBC to the canonical form S (see Fig. 3 for an illustration) without affecting the spectrum of M . For OBC, there is thus a single spectral equivalence class, and we can focus only on one representative S configuration; therefore, we only need to find the optimal 2-qubit gate $M_{i,i+1}$ [Eq. (16)]. Staircase configurations have been considered before in, e.g., the context of operator spreading [5] or complexity [64].

We have calculated the second-largest eigenvalue $|\lambda_2|$ of $M = S_1^{(n-1)}$ for all of the different 2-qubit gates in the canonical form. Results (see Fig. 20 in Appendix D 4) show that the smallest eigenvalues $|\lambda_2|$ come from the region around $a_x = a_y = 1, a_z = 0$, which is the XY gate. Looking more precisely at the n dependence for the XY gate (Fig. 21), we find that, in the TDL, the S configuration with OBC has

$$|\lambda_2| = \frac{1}{4}. \quad (23)$$

Exploring the region around the XY gate, we find that the same largest eigenvalue (23) is also achieved for sufficiently small nonzero values of a_z and $a_x = a_y = 1$. In Fig. 4, we can see that at $a_z = a_c \approx 0.32$, the transition

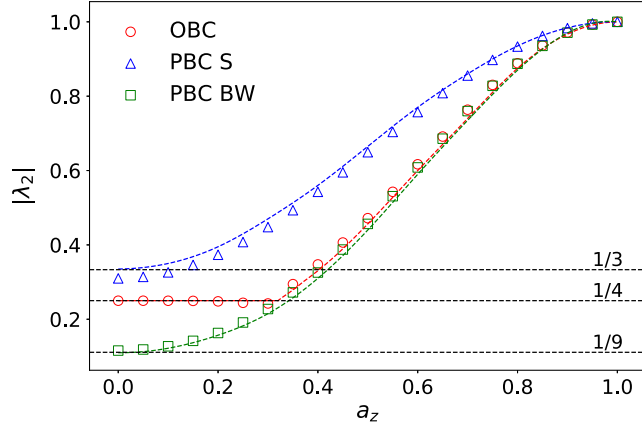


FIG. 4. $|\lambda_2|$ for the XXZ gate [Eq. (4)]. Symbols are numerical data for $n = 16$, while dashed curves are results in the TDL (obtained by extrapolation, like in Fig. 21). For $a_z = 1$, all three cases converge to a nonentangling SWAP gate with $\lambda_2 = 1$. For the BW configuration with PBC (“fastest” class $p = \lfloor n/2 \rfloor$), and for the S configuration with PBC (“slowest” PBC class $p = 1$), the dependence is smooth. For the OBC (all configurations are equivalent), the gap does not change until a transition at $a_z = a_c \approx 0.32$.

happens in $|\lambda_2|$, so the optimal $|\lambda_2| = \frac{1}{4}$ is obtained for all XXZ gates with $a_z \leq a_c$.

B. PBC protocols

The PBC case is a bit more complicated, resulting in $\lfloor n/2 \rfloor$ spectral equivalence classes. Each class can be labeled by an integer $p = 1, \dots, \lfloor n/2 \rfloor$ denoting the maximal number of consecutive commuting gates in the canonical representative configuration

$$S_{2p}^{(n-2p+1)} B_1^{(2p-1)}, \quad (24)$$

where $B_1^{(2p-1)}$ is a brick-wall configuration made out of $2p - 1$ gates (see Fig. 5 for an example),

$$B_1^{(2p-1)} = M_{2p-2,2p-1} \cdots M_{4,5} M_{2,3} M_{2p-1,2p} \cdots M_{3,4} M_{1,2}. \quad (25)$$

We note that $B_1^{(n)}$ denotes the full BW configuration with PBC, i.e., $B_1^{(n)} = M_{n,1} B_1^{(2\lfloor n/2 \rfloor - 1)}$.

More precisely, the following theorem holds.

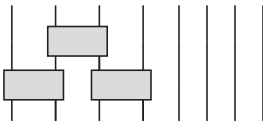


FIG. 5. Brick-wall configuration $B_1^{(3)}$ with three gates, i.e., $B_1^{(2p-1)}$ with $p = 2$.

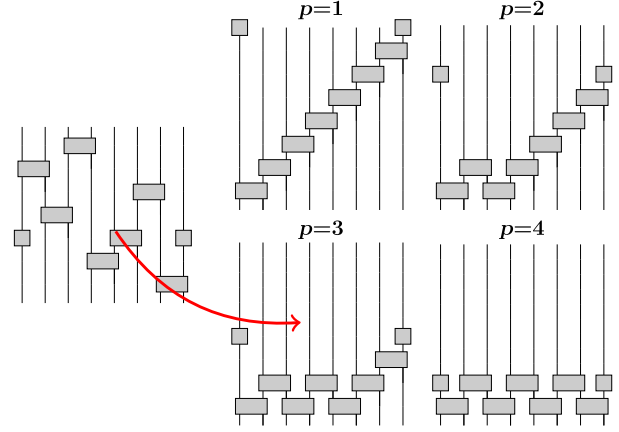


FIG. 6. Spectrally equivalent configurations for PBC protocols. In the example with $n = 8$, there are just four classes labeled by $p = 1, \dots, 4$. The specific configuration shown in the left can be transformed to $S_6^{(3)} B_1^{(5)}$ with $p = 3$ (red arrow).

Theorem 2. Any product of n symmetric two-site matrices acting on distinct NN on a circle (PBC configuration) is spectrally equivalent to one of $\lfloor n/2 \rfloor$ canonical configurations of the form (Fig. 6)

$$S_{2p}^{(n-2p+1)} B_1^{(2p-1)}. \quad (26)$$

The proof that uses cyclic permutations and the fact that the spectra of A and A^T are the same can be found in Appendix D 2 [observe that our $M_{i,j}$ in Eq. (16) are symmetric matrices].

For PBC protocols, we therefore have to find the optimal configuration among the $\lfloor n/2 \rfloor$ different ones, as well as the optimal 2-qubit gate. We have numerically computed $|\lambda_2|$ for small sizes n (see Appendix D 4 and Fig. 22). One observes that fixing a 2-qubit gate (canonical parameters a_x, a_y, a_z) and n , the gap always monotonically increases as we change the equivalence class from $p = 1$ to $p = \lfloor n/2 \rfloor$. In the TDL, configurations in the class $p = 1$ have the smallest gap, and those in $p = \lfloor n/2 \rfloor$ have the largest. Focusing on the region around the XY gate, we see that, in the TDL and for $p = 1$ as well as for $p = \lfloor n/2 \rfloor$, the fastest gate is obtained for $a_z = 0, a_x = a_y = 1$ (Figs. 23 and 24 in Appendix D 4), which is the XY gate. Analyzing the dependence on n (Fig. 21 in Appendix D 4), we get that, in the TDL, the eigenvalue is

$$|\lambda_2| = \frac{1}{9}, \quad \text{BW with PBC}, \quad (27)$$

for the fastest class $p = \lfloor n/2 \rfloor$, and

$$|\lambda_2| = \frac{1}{3}, \quad \text{S with PBC}, \quad (28)$$

for the slowest $p = 1$, which we simply call S PBC. Contrary to the OBC case, the smallest eigenvalue $|\lambda_2|$ is achieved only at a single point $a_x = a_y = 1, a_z = 0$. For the XXZ gate, the

gap smoothly decreases as one moves away from $a_z = 0$; see Fig. 4.

Among all OBC and PBC protocols, the one with the largest gap, and therefore the fastest expected entanglement generation, is the BW with PBC and the XY gate for which $|\lambda_2| = 1/9$. While we use numerics to obtain this value, we can, in fact, rigorously show (see Appendix D 3) that $\lambda = 1/9$ is in the spectrum of M for $p = \lfloor n/2 \rfloor$, which implies an exact relation $|\lambda_2| \geq 1/9$. Proving analytically that $|\lambda_2|$ is $1/3$, $1/4$, and $1/9$ for the XY gate and PBC S, OBC, and PBC BW, respectively (Fig. 4), is an interesting open problem.

Note that the same eigenvalues $\lambda_2 = 1/3, 1/4, 1/9$ are also the leading nontrivial eigenvalues of the Markovian matrix propagating OTOCs [19].

IV. ENTANGLEMENT GENERATION SPEED

In the previous section, we identified gates and configurations that have the smallest $|\lambda_2|$ and are therefore expected to have the fastest asymptotic entanglement generation rate r_E . Comparing the spectral gap for the S configuration between PBC and OBC (Table II), we see that the gap is larger for the OBC than for the PBC. If the gap was the end of the story, that would mean that the entanglement generation rate would be larger for the OBC than for the PBC, which would be odd as one would expect exactly the opposite. For symmetric bipartition, one has two cuts for the PBC and only one for the OBC; therefore, the rate should be twice as large for PBC than for OBC. A faster rate for the OBC would mean that applying less gates across the cut would produce more entanglement. In this section, we see that numerically analyzing the true entanglement rate will rectify this odd spectral situation—for OBC, the true rate will be slower than suggested by $|\lambda_2|$; for PBC, it will be faster than $|\lambda_2|$, such that, at the end, the ratio of the two is recovered at the expected value 2.

For all extremal random circuits, either the fastest (BW) or the slowest in its class (the S configuration), the relevant entanglement rate r_E on times smaller than the thermalization time t_∞ [Eq. (8)] is, surprisingly, not necessarily given by $-\ln|\lambda_2|$ despite $|\lambda_2|$ being gapped away from other eigenvalues. In other words, the time t_∞ at which S_2 reaches its random-state saturation value of $\sim n/2$ is not equal to $t_\infty = n/(-2 \log_2 |\lambda_2|)$. It can be either larger because of a phantom eigenvalue—a phenomenon where exponentially large expansion coefficients mimic a fake eigenvalue larger than $|\lambda_2|$ —or smaller, again because of exponentially large coefficients in front of eigenvalues that are smaller than $|\lambda_2|$. As we shall see, the culprit lies in the non-Hermiticity of the transfer matrix M . A quick overview of our results for extremal circuits that we focus on can be found in Table III.

Non-Hermiticity is important for the following reason. The spectral decomposition of a non-Hermitian operator M has the form (assuming diagonalizability)

TABLE III. Origin of the entanglement rate. “Phantom” denotes a phantom eigenvalue resulting in a slower entanglement generation than given by $|\lambda_2|$, “faster” denotes cases with faster generation than given by $|\lambda_2|$, and λ_2 is for the one case where the rate is actually determined by $|\lambda_2|$.

Configuration	Gate	
	XY, $a_z = 0$	XXZ, $a_z \in (0, 1)$
BW OBC	λ_2	λ_2 [$a_z \in (0, a_c)$], faster
BW PBC	Faster	Faster
S OBC	Phantom	Phantom, faster
S PBC	Faster	Faster

$$M = \sum_j \lambda_j |R_j\rangle\langle L_j|, \quad \langle L_j | R_k \rangle = \delta_{j,k}, \quad (29)$$

where λ_j are, in general, complex eigenvalues, while $|R_j\rangle$ and $\langle L_j|$ are the associated right and left eigenvectors. Importantly, left and right eigenvectors are mutually orthogonal; however, left and right eigenvectors are not orthogonal between themselves. Expanding the initial state $|\Phi_0\rangle$ over a nonorthogonal basis $|R_j\rangle$, we have

$$|\Phi(t)\rangle = \sum_j c_j \lambda_j^t |R_j\rangle, \quad c_j = \langle L_j | \Phi_0 \rangle. \quad (30)$$

If M were Hermitian (having orthogonal $|R_j\rangle$), the triangle inequality would bound $|c_j|^2 \leq \langle \Phi_0 | \Phi_0 \rangle$. For non-Hermitian M , no such bound exists, and the expansion coefficient c_j can be arbitrarily big (see, e.g., Ref. [65] for a simple Ising chain in a tilted field with an imaginary component, where the eigenvectors effectively span a subspace of lower dimension and thus the expansion coefficients can become large; see also the Lindblad case in Ref. [20]). This is precisely what will happen for our random circuits. Exponentially large (in n) coefficients $c_{j>2}$ will dominate over the term $c_2 \lambda_2^t$ for times $t < t_\infty$, causing the entanglement rate r_E to be larger than $-\log |\lambda_2|$. In one case, the entanglement rate will be smaller than $-\log |\lambda_2|$, i.e., like there is a “phantom” eigenvalue larger than $|\lambda_2|$, which we explain below.

To calculate the correct entanglement rate, we use numerical simulations to get $I(t)$, or closely related $\Delta S_2(t)$ [Eq. (9)]. We use two methods with the goal of simulating as large a system as possible, which turns out to be important as the multistage thermalization reveals itself clearly only in relatively large systems—simulating systems with, e.g., just $n = 16$ qubits, one could easily miss it. The first method works by directly iterating the full state $|\Phi(t)\rangle$, obtaining

$$I(t) - I_\infty = \Phi_{\text{half}} \cdot M^t \cdot (\Phi_0 - \Phi_\infty), \quad (31)$$

where $\Phi_\infty = (A^{-1})^{\otimes n} \Phi'_\infty$ [Eq. (20)]. We use such iteration to obtain the exact $\Delta S_2(t)$ for $n \leq 34$. For larger circuits,

memory requirements of the exact method become prohibitive, and we use the matrix product state (MPS) ansatz for a more memory-efficient representation of $|\Phi'(t)\rangle$. While the MPS method [66] allows us to study large systems, it has its limitation beyond times when S_2 becomes too large. Namely, when, e.g., $S_2 \sim 50$, the floating-point double-precision “noise” of size $10^{-15} \approx 2^{-50}$ becomes comparable to $I(t)$, making further simulation hard despite using large MPS bond dimensions up to 1500.

The main part of this section is devoted to the brick-wall configuration (Sec. IV A) and to the staircase configuration (Sec. IV B), both for the XXZ gate. In Sec. IV C, we briefly demonstrate that the same phenomena are also observed for generic fixed gates as well as for the random U(4) gates. In Sec. IV D, we touch upon the experimentally relevant Sycamore processor that uses a 2D layout of qubits.

A. Fastest scrambler

Let us first look at the fastest random circuits, namely, the BW configuration for which the spectral analysis would predict $r_E = \ln 9$ for the XY gate and PBC, and $r_E = \ln 4$ for the XY with OBC. The BW configuration with an appropriate gate type that we identify turns out to be the fastest possible entanglement scrambler saturating the theoretical upper bound on the entanglement speed in Eq. (2).

In the following, we show a number of figures of the same type, which demonstrate the asymptotic entanglement rate and a multistage thermalization. Let us describe what those figures show. The main quantity is the local entanglement rate r_E , which is equal to the local slope of $\Delta S_2(t) = -\ln |I(t) - I_\infty|$. Specifically, from numerical data, we plot

$$\frac{r_E(t + \frac{1}{2})}{\ln 2} = \log_2 \left| \frac{I(t) - I_\infty}{I(t+1) - I_\infty} \right|. \quad (32)$$

We use base-2 logarithms so that, for qubits, the theoretical upper bound on the rate is an integer (2 or 4 depending on boundary conditions). In addition, we show plots of purity $-\log_2 I(t)$ as well as purity with a subtracted asymptotic saturation value I_∞ , i.e., $-\log_2 |I(t) - I_\infty|$.

In Fig. 7, we show results of numerical simulations for the BW configuration with the XXZ gate and PBC. In frame (a), we show the XY gate ($a_z = 0$), whereas in (b) and (c), a more general XXZ gate with $a_z = 0.5$ is shown. The behavior is similar for any $a_z < 1$ ($a_z = 1$ corresponds to the SWAP gate, which does not generate any entanglement). We can see from the purity plot in Fig. 7(c) that there is a kink—a change in the entanglement growth rate—at the point when $I(t)$ gets close to I_∞ (dashed saturating curves). The change in r_E is also clearly seen in Fig. 7(b). Regardless of a_z , the rate in the TDL increases towards 4 for times smaller than the saturation time t_∞ (note that we rescale the time axis by $n_A = n/2$). However, the

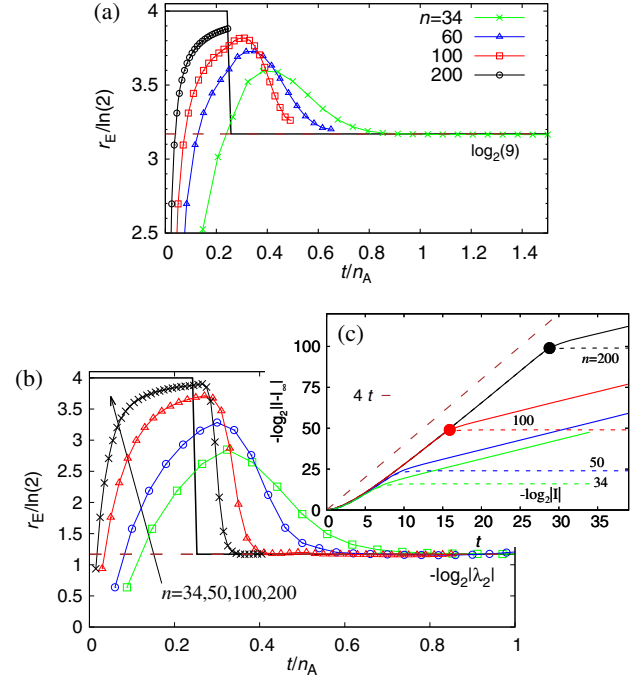


FIG. 7. Local entanglement rate r_E for the BW configuration with PBC. Panel (a) is for $a_z = 0$ (the XY gate), and (b) and (c) are for $a_z = 0.5$. The full black line in panel (a) is a conjectured r_E in the TDL, which has a discontinuous jump from 4 for $t \leq t_c = t_\infty$ to a smaller value given by λ_2 . In panel (c), dotted curves are $-\log_2 |I(t) - I_\infty|$ and full curves $-\log_2 |I(t) - I_\infty|$, while the transition times t_c for $n = 100$ and 200 are marked by full circles. The dashed brown line in panels (a) and (b) agrees with $-\log_2 |\lambda_2|$ from Fig. 4.

convergence with n is rather slow. If one were to look only at small n , one could be misled into thinking that the rate is given by $r_E = -\ln |\lambda_2|$ [$n = 34$ in Fig. 7(a)]; however, at larger n , the initial bump gets higher and moves to smaller times. The rate is therefore not determined by λ_2 and equal to $-\ln |\lambda_2|$ but is instead larger. For instance, for $a_z = 0$, we have $\lambda_2 = \frac{1}{9}$, while the rate converges with n towards $\ln 16$ [Fig. 7(a)]. This case is perhaps even more clear for $a_z = 0.5$ in Fig. 7(b). Based on the data, we therefore conjecture the following scenario in the TDL, which holds for any $a_z < 1$: The rate is $r_E = \ln 16$ for $t \leq t_\infty = \frac{1}{4} n_A$, at which point there is a discontinuous transition in the rate to a smaller value determined by the second-largest eigenvalue $|\lambda_2|$. For any zero or nonzero $a_z < 1$, there is one discontinuous phase transition in the local rate at the critical time $t_c = \frac{1}{4} n_A$, which is reflected in the phase diagram shown in Fig. 9(a). The relaxation process has two phases: In the first faster phase, whose entanglement rate is in the TDL independent of a_z (even if we are very close to a non-entangling SWAP gate), one approaches a random state; in the second, which starts at $t_c = t_\infty = n_A/(2\mathcal{A})$ when the state is already close to random, the relaxation is slower and determined by $-\ln |\lambda_2|$. Therefore, the physically relevant

entanglement generation rate that determines how long it takes to generate all $\sim(n/2)$ bits of entanglement is not determined by the matrix gap, but rather, it is equal to $\ln 16$.

By studying the spectrum of M (in the even sector) and the relevant overlaps with Φ_0 and Φ_{half} [Eq. (31)] that appear in the spectral expansion

$$I(t) = I_\infty + \sum_{j=2}^{\infty} d_j \lambda_j^t, \quad d_j = \langle \Phi_{\text{half}} | R_j \rangle \langle L_j | \Phi_0 \rangle, \quad (33)$$

we can identify the leading terms in the sum. Focusing on $a_z = 0$, there is exactly one with $|\lambda_2| \approx \frac{1}{9}$ and $(n-4)/4$ with $|\lambda_j| \approx \frac{1}{16}$. The weight d_j of terms with $\lambda_j \sim 1/16$ grows with n , albeit first for small n rather slowly [see, e.g., slow convergence with n in Fig. 7(a)]. Nevertheless, for large n , the terms $d_3(\frac{1}{16})^t$ overwhelm

the “leading” $d_2(\frac{1}{9})^t$; this happens roughly when the ratio $d_3/d_2 \sim (16/9)^{n/8} \approx 1.07^n$ gets large.

Data for the BW with OBC are shown in Fig. 8. For $a_z = 0$, the rate is in the TDL $r_E = 2 \ln 2$ for any t , which is in line with the fact that, in this case, $|\lambda_2| = \frac{1}{4}$; therefore, one has $r_E = -\ln |\lambda_2|$. Out of the four cases studied, BW and S with OBC or PBC, this case is the only one where the gap determines the entanglement rate. Apparent singularities in the rate at integer t/n_A , visible in Fig. 8(a), are due to a sign change in the $I(t) - I_\infty$, which is also visible as sharp kinks in $-\log_2 |I(t) - I_\infty|$ plotted in frame (b). The case of nonzero a_z is different. As we have seen in Fig. 4, for $a_z < a_c$, the gap for the OBC does not change with a_z , while for $a_z > a_c \approx 0.32$, it starts to decrease, which, in turn, is also reflected in the local entanglement rate r_E at $t > t_c$. Up to the thermalization time t_∞ , the rate stays at $2 \ln 2$ and is therefore not equal to $-\ln |\lambda_2|$ for $a_z > a_c$, while at $t_c = t_\infty = n_A/2$, the rate jumps to $-\ln |\lambda_2|$ [Fig. 8(c)]. This change in the rate is continuous at $a_z = a_c$ but discontinuous for $a_z > a_c$. Correspondingly, $\Delta S_2(t)$ [Eq. (9)] exhibits a kink at t_c [Fig. 8(d)]. It is worth noting that while in the scaled time t/n_A the rate in the TDL will be $\ln 4$ already at $t/n_A \rightarrow 0$ [Fig. 8(c)], in real time t , there is a “delay” of about $\Delta t \approx 15$ before the rate becomes approximately $\ln 4$ [slower initial growth in Fig. 8(d)]. The dependence of r_E on a_z for BW with OBC is shown in the phase diagram in Fig. 9(b).

The BW configuration with the XXZ gate is the fastest entanglement scrambler. Considering that λ_2 does not give the correct rate, one might wonder whether some other gate, which did not look optimal according to λ_2 in Sec. III, could be even better. The answer is no because the rates 4 and 2 for PBC and OBC, respectively, saturate the bound in Eq. (2). There are no better scramblers. What could, in principle, happen (but we think is unlikely) is that for some other canonical gate with parameters \mathbf{a} different from XXZ, one would get the same maximal rate. For instance, for $\mathbf{a} = (0.9, 0.8, 0.5)$ with BW PBC, one again has a faster rate than given by $|\lambda_2|$, though at approximately 2.77 it is smaller than 4 (see Sec. IV C for more details).

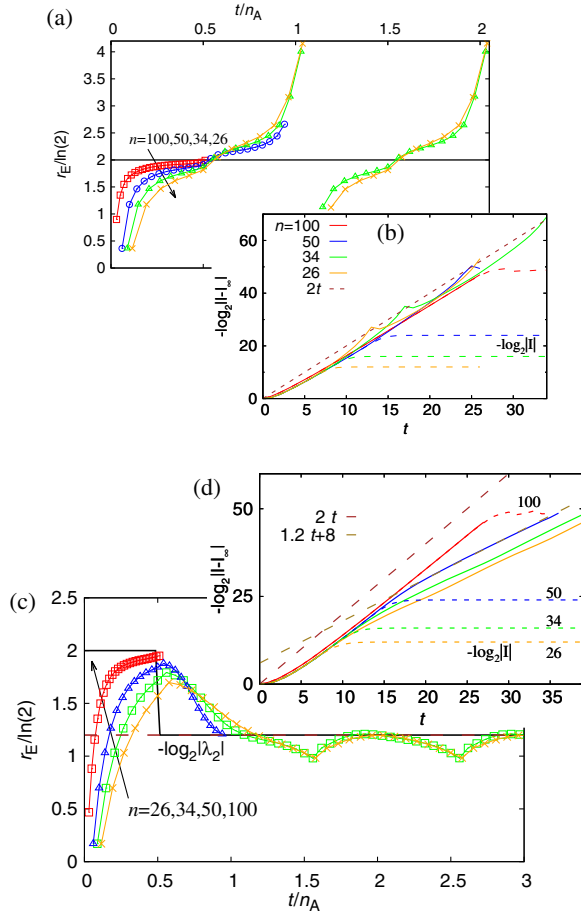


FIG. 8. Entanglement rate for the BW and OBC. Panels (a) and (b) are for $a_z = 0$ (the XY gate), and (c) and (d) are for $a_z = 0.5$. For $a_z = 0$, the local rate is always equal to $\ln 4$ in the TDL. For $a_z = 0.5$, the conjectured rate in the TDL [full black line in (c)] is $\ln 4$ for $t < t_c = t_\infty = n_A/2$ and smaller $-\ln |\lambda_2|$ for $t > t_c$. The dashed brown line in panel (c) is at 1.2 and is approximately equal to $-\ln_2 |\lambda_2|$ from Fig. 4.

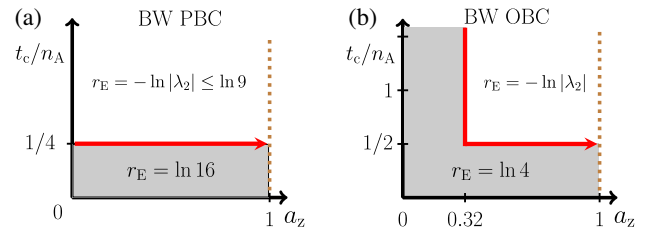


FIG. 9. Phase diagram of the local entanglement rate r_E for the BW configuration. (a) For PBC, there is one phase transition; (b) for OBC, there is, again, one discontinuous transition for $a_z > a_c \approx 0.32$. The brown dashed vertical line at $a_z = 1$ indicates a nonentangling SWAP gate.

B. Slow scramblers

Here, we discuss the S configuration [Fig. 2(b)], which is also an extremal configuration in the spectral sense. Namely, for PBC, its $|\lambda_2|$ is the largest among all configurations with PBC (spectral equivalence class $p = 1$). Note that we do not discuss the slowest scramblers in the absolute sense here because we still use the fastest gate within the class $p = 1$ (XXZ). For instance (Table II), XXZ with S and OBC is still faster than CNOT (“slow gate”) with BW and PBC (fastest configuration). We expect the same behavior that we reveal here for $p = 1$ to also hold in the TDL for any finite p , i.e., for a spectral class that consists of a finite BW section and an extensive S part.

Let us first have a look at the S configuration with PBC. Results are shown in Fig. 10. We can see that, similarly to the BW with PBC, the entanglement rate r_E is larger than $-\ln|\lambda_2|$. Specifically, for $a_z = 0$, one has $-\ln|\lambda_2| = \ln 3$ while the rate is $r_E = \ln 4$ [Figs. 10(a) and 10(b)]. While the numerical results are not very clear, perhaps for $a_z = 0$ the rate also converges in the TDL to $r_E = \ln 4$ for times $t > t_\infty$ so that there is no phase transition. On the other hand, for

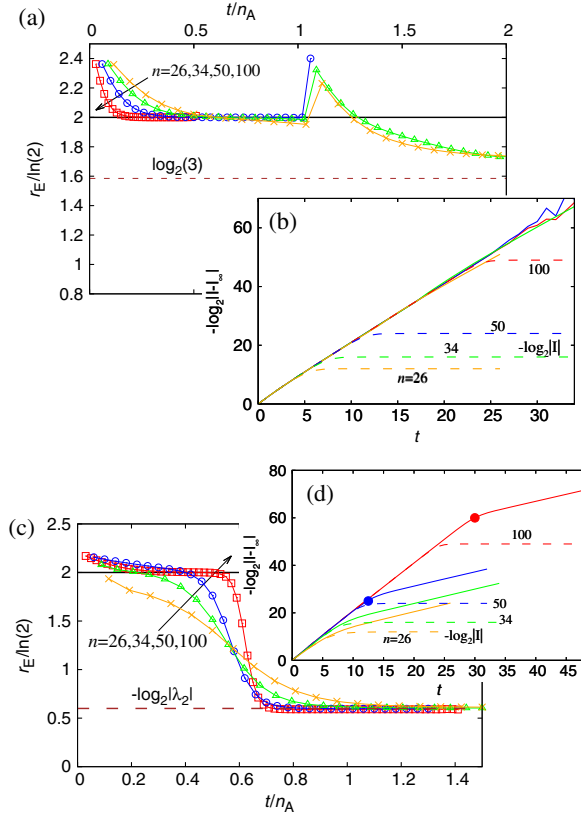


FIG. 10. (a,b) Entanglement rate for the S configuration with PBC and $a_z = 0$, and (c,d) $a_z = 0.5$. At $a_z = 0$, the rate is always $r_E = 2 \ln 2$, and it is larger than $-\ln|\lambda_2|$. At nonzero a_z , there is a phase transition from $r_E = \ln 4$ for $t < t_c$ to $-\ln|\lambda_2|$ at larger times. The transition appears to happen at larger t/n_A than 0.5 (here, around $t_c/n_A \approx 0.6$). The brown dashed line at 0.60 is in line with $|\lambda_2| \approx 0.66$ from Fig. 4.

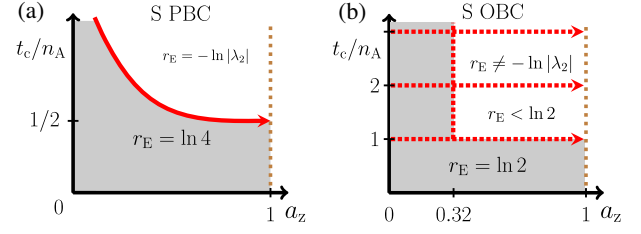


FIG. 11. Phase diagram of r_E for the staircase configuration.

nonzero a_z , one has a phase transition in the rate from the initial $r_E = \ln 4$ to a smaller rate given by the spectral gap, Fig. 10(c). For $a_z = 0.5$, the phase transition seems to happen at a distinctively larger time t_c than the thermalization time t_∞ [see Fig. 10(d)], while at larger a_z —for instance, $a_z = 0.7$ (data not shown)—it is very close to $t_\infty = n_A/2$. As opposed to the BW configuration, here the convergence with n towards the TDL rate $\ln 4$ is much faster, so there is almost no delay in the time when purity starts to follow the asymptotics [Figs. 10(b) and 10(d)]. A conjectured phase diagram is sketched in Fig. 11(a).

The explanation for such larger entanglement rates is similar to the BW configuration with PBC, though some details are different. The even-parity spectrum of M for $n = 12$ is shown in Fig. 12. Besides the steady state giving I_∞ (blue point), one has $n - 1$ nondegenerate eigenvalues (λ_2) distributed around the circle with radius $\frac{1}{3}$ (in the TDL, Fig. 4) and $(n - 1)(n - 4)/2$ eigenvalues around the circle of smaller radius $\frac{1}{4}$ [those are grouped into $n - 1$ in the TDL degenerate clusters, each having $(n - 4)/2$ eigenvalues, e.g., 4 for the $n = 12$ example in Fig. 12]. In the TDL, though, not all eigenvalues contribute in the purity expansion [Eq. (33)]. It turns out that only those on the real axis contribute, that is, one with $\lambda_2 = \frac{1}{3}$ (orange point) and $(n - 6)/2$ with $\lambda_j = \frac{1}{4}$ (red point). The relative weight d_j/d_2 of relevant eigenvalues $\frac{1}{4}$ grows with n , which is how the faster entanglement rate $r_E = \ln 4$ emerges in the TDL.

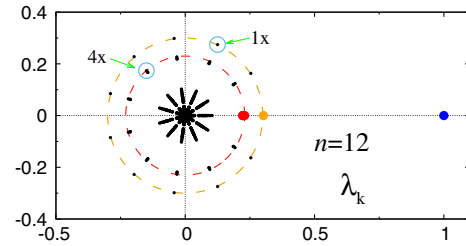


FIG. 12. Spectrum of M for S and PBC, $a_z = 0$ and $n = 12$. There are $n - 1$ eigenvalues with $|\lambda_j| \rightarrow \frac{1}{3}$ [black dots on the larger (orange) circle] and $(n - 1)(n - 4)/2$ with $|\lambda_j| \rightarrow \frac{1}{4}$ [smaller red circle; each visible point is composed of $(n - 4)/2$ close eigenvalues]. The leading contribution to $I(t)$ comes from one eigenvalue with $\lambda_k = \frac{1}{3}$ (orange point) and $(n - 6)/2$ eigenvalues with $\lambda_k = \frac{1}{4}$ (the red point represents three almost degenerate eigenvalues).

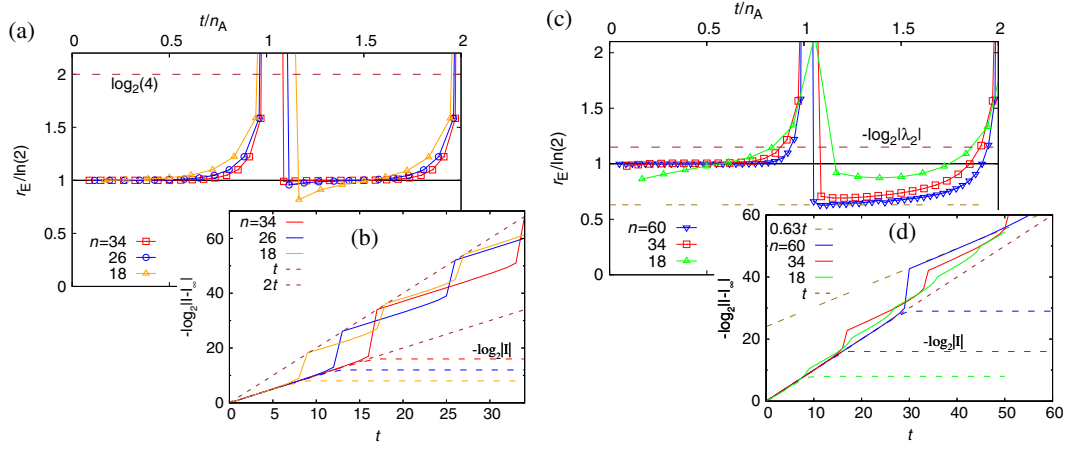


FIG. 13. Entanglement rate for the S configuration with OBC and $a_z = 0$ (a,b), and $a_z = 0.5$ (c,d). For $a_z = 0$, the rate is always $r_E = \ln 2$ but with discontinuous jumps in $I(t) - I_\infty$ at integer t/n_A [see panel (b)]. For $a_z = 0.5$, the rate is again $\ln 2$ for $t < t_c = n_A$; however, in the TDL, it discontinuously jumps to a smaller value [$r_E \approx 0.63 \ln 2$ at $a_z = 0.5$ in (c,d), and is not equal to $-\ln |\lambda_2|$].

Configuration S with OBC is even more interesting. In Fig. 13, we show numerical results for r_E [Eq. (32)]. The rate is $r_E = \ln 2$ and is therefore smaller than $\ln 4$ as suggested by the gap. For $a_z = 0$, such a rate persists for any finite time t/n_A ; however, looking at $I(t) - I_\infty$ [Fig. 13(b)], we note that there are discontinuous jumps at integer t/n_A , the size of which increases with n . At such times, purity $I(t)$ closes in on its asymptotic value I_∞ by a finite amount (that increases with n) in a single time step. For nonzero $a_z = 0.5$, the situation is similar, with the difference being that the rate at $t > t_c = n_A$ jumps to smaller values than $\ln 2$ [Fig. 13(c)]. In this case, though, the rate after t_c is not simply equal to $-\ln |\lambda_2|$ as in other cases studied.

The explanation of how one can get a slower rate $r_E = \ln 2$, despite all nontrivial eigenvalues of M being $|\lambda_2| \leq \frac{1}{4}$, is very interesting and is illustrated in Fig. 14. In frame (a), we can see that there are $n - 1$ eigenvalues with $|\lambda_j| = \frac{1}{4}$; however, only $n/2$ have nonzero overlap with the initial vector [red points in (b)]. The size $|c_j|$ of those coefficients

grows exponentially with n , and they come in conjugate pairs. The resulting expression for purity is therefore of the form $I(t) - I_\infty \approx \sum_{k=2}^{n/2+1} |d_k| \cdot |\lambda_k|^t \cos(\varphi_k t + \alpha_k)$, where φ_k is the phase of λ_k and α_k the phase of d_k . Coefficients d_k , Eq. (33), are such that the corresponding sum, including cosine terms, mimics exponential growth $\sim 2^t$ up to time $t = n/2$, when it discontinuously jumps to 0 in a single unit of time [frame (c)]. The end result is that their growth 2^t partially cancels $(\frac{1}{4})^t$ from the actual eigenvalues, resulting in a decay $(\frac{1}{2})^t$, as there would be a phantom eigenvalue of size $\frac{1}{2}$. Again, this is only possible because of the non-Hermitian M whose eigenvectors are not orthogonal. The size of discontinuous jumps in panel (c) grows exponentially with n and is responsible for the jumps that we observed in Fig. 13(b).

In Fig. 14(d), we see how this is reflected in $I(t)$: Discontinuous jumps are such that on times larger than $t_\infty = n/2$, one actually obtains an overall $\Delta S_2(t) \asymp t \ln 4$ as predicted from λ_2 ; however, in the TDL, the time t_∞ is pushed to infinity, and the physically relevant entanglement rate towards random-state entanglement is instead given by

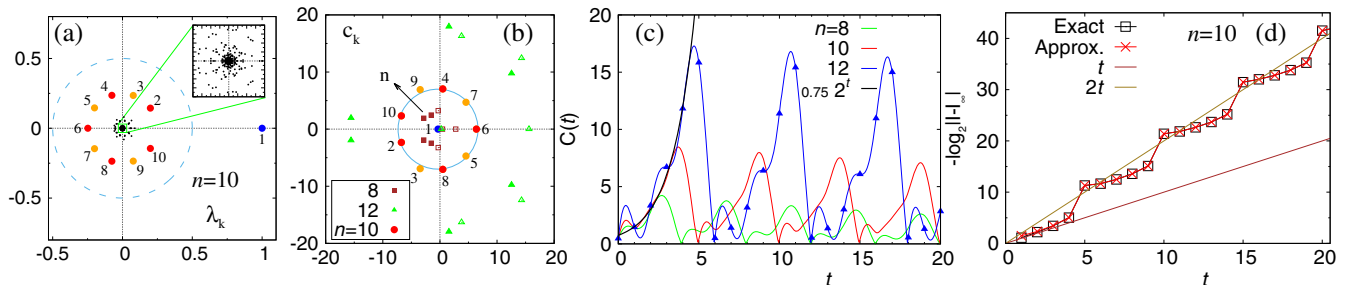


FIG. 14. Slower convergence than $-\ln |\lambda_2|$ for the S with OBC and the XY gate due to a phantom eigenvalue [dashed blue circle in (a)]. (a) Complex spectrum of M for $n = 10$ (even sector with 512 eigenvalues). (b) Expansion coefficients c_k [Eq. (30)]. Numbered points correspond to respective eigenvalues in panel (a); the five full red circles (2,4,6,8,10) are $n/2$ leading terms contributing to $I - I_\infty$. The size of $|c_k|$ grows exponentially with n . (c) Sum $C(t) = 4^t |\sum_{k=2}^{n/2+1} d_k \lambda_k^t|$ [Eq. (33)] of corresponding terms from panel (b) (blue triangles are at physical integer t). (d) Comparison of the exact purity and the approximation with $n/2$ terms from panel (c).

the phantom eigenvalue $\frac{1}{2}$. Note that even though the local entanglement rate is, for noninteger t/n_A , always $\ln 2$, because of discontinuities in $I(t)$ at integer t/n_A , the overall growth of $\Delta S_2(t)$ for $a_z = 0$ is given by $\ln 4$ and not the local rate [Fig. 14(d)]. For $a_z > a_c \approx 0.32$, when $|\lambda_2|$ moves, the local rate at $t > t_c = n_A$ also decreases; see Fig. 13(c). However, it is not equal to $-\ln |\lambda_2|$ (the rate decreases from $\ln 2$, whereas the eigenvalue decreases from $\ln 4$). Note that $r_E = \ln 2$ for all $a_z < 1$ regardless of $|\lambda_2|$ increasing with a_z (Fig. 4). This is all reflected in the phase diagram in Fig. 11(b).

We remark that one could, in principle, get a slower effective decay than $|\lambda_2|^t$ for special initial states that would have a very small overlap with the corresponding eigenvector; in the asymmetric simple exclusion process that exhibits a cutoff [67] such states are those with an extensive empty or occupied sections. However, this is not what happens in our case; we get such a phantom eigenvalue decay for the relevant initial state, i.e., generic bipartition (provided n_A is extensive). The origin of the slower decay is different in our case. It is the left and right eigenvectors corresponding to $|\lambda_j| \approx 1/4$ [red points in Fig. 14(a)] that must be special as they are the ones that mimic the exponential function [Fig. 14(c)].

We have seen that, in most of the circuits we studied, the relevant entanglement rate r_E is not given by the largest nontrivial eigenvalue of M . The ratio of $\ln |\lambda_2|$ between the PBC and the OBC (Table II) for a fixed W is not equal to 2 for the BW configuration or for the S ; in particular, it changes with the a_z of the XXZ gate. However, the correct rates r_E we identified in this section are always in the expected ratio of 2 : 1 for the two boundary conditions. Rates that are different than $-\ln |\lambda_2|$ are made possible because of the nonsymmetric nature of M (despite individual gates $M_{i,j}$ being symmetric matrices), which results in expansion coefficients growing with n . Another observation is that the ratio of r_E between the BW and S configurations for fixed W and boundary conditions is always 2 for the XXZ gates, which, however, is not the case for all 2-qubit gates [68].

It is worth noting that, in general, the exponential growth of coefficients, in our case, cannot be traced to degeneracies. For instance, taking a simple nonsymmetric 2×2 matrix

$$\begin{pmatrix} 1 & 1 - \epsilon \\ 0 & 1 - \epsilon \end{pmatrix}, \quad (34)$$

the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 1 - \epsilon$, with the associated right eigenvectors $\mathbf{x}_1 = (1, 0)$ and $\mathbf{x}_2 = (\epsilon - 1, \epsilon)$. In the limit $\epsilon \rightarrow 0$, the two eigenvalues coalesce (an “exceptional point”; for $\epsilon = 0$, one has a nondiagonalizable Jordan canonical form), and the right eigenvectors are almost collinear. Expanding $\mathbf{y} = (a, 1)$ over such a basis will result in large expansion coefficients of the order of the inverse eigenvalue separation that scales as $1/|\lambda_1 - \lambda_2| \sim (1/\epsilon)$,

$\mathbf{y} = [a - 1 + (1/\epsilon)]\mathbf{x}_1 + (1/\epsilon)\mathbf{x}_2$. While in some cases one indeed has $\sim n$ -fold degeneracy (S with PBC, Fig. 12), in others, one does not [S with OBC, Fig. 14(a)]. Thus, it does not seem like the phenomena we found can be simply ascribed to degeneracies.

The phantom eigenvalue phenomenon—and, more generally, of multistage relaxation in which the relaxation rate exhibits a sudden transition—has been identified in random circuits. Considering that random circuits are often used as models of chaotic systems, one naturally asks if the randomness explicitly present in space and time due to random 1-qubit unitaries is actually necessary. A detailed study of this interesting question goes beyond the present work; however, we have numerically checked (see Appendix F 2) that, in the TDL, one gets the same multistage thermalization even in a single circuit realization and without randomness in both space and time; i.e., one can use the same 1-qubit unitary on all qubits and at all time steps.

C. Generic gates

So far, we have focused on the fastest 2-qubit gates in the spectral equivalence class $p = 1$ and $p = n/2$, which turn out to be the XXZ-type gates [Eq. (4)]. Random circuits with XXZ gates are dual unitary; thus, one might naturally ask if the observed multistage thermalization and phantom eigenvalues are perhaps limited to (some) dual-unitary circuits. The answer is no. All the phenomena we discussed can occur in circuits that are not dual unitary, and this is what we briefly demonstrate in this section. A full treatment goes beyond the present paper, so we will only show two examples.

The first is the case of the BW configuration with PBC and a generic 2-qubit gate. We pick a completely anisotropic non-dual-unitary gate with canonical parameters $\mathbf{a} = (0.9, 0.8, 0.5)$. In Fig. 15, we can see that one has

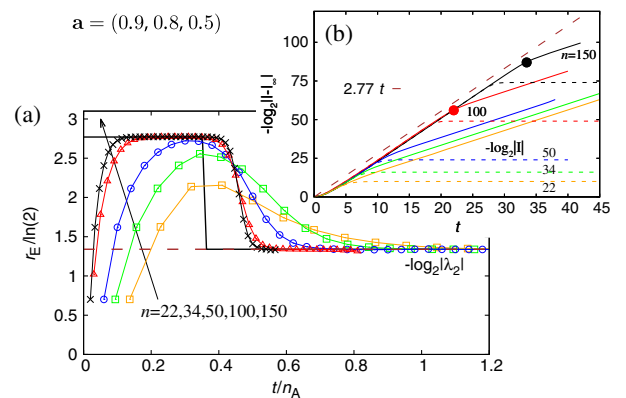


FIG. 15. Entanglement rate for the BW configuration with PBC and a 2-qubit gate with parameters $\mathbf{a} = (0.9, 0.8, 0.5)$. One has a multistage thermalization with rate $r_E / \ln 2 \approx 2.77$ that is nonmaximal (less than for XXZ gates) but still faster than the one calculated from $|\lambda_2|$ [dashed brown line at approximately 1.34 in (a)].

exactly the same phenomenology as for the XXZ gate (compare with Fig. 7). Namely, in the thermodynamic limit, the initial rate is larger than $-\ln|\lambda_2|$; therefore, one has a multistage thermalization. This local rate, though, is smaller than for the XXZ gates, where it is $r_E/\ln 2 = 4$, which is the reason why we conjecture that the XXZ gates are the only ones with the maximal entanglement production rate.

The second non-dual-unitary example that we show is a much-studied case of completely random 2-qubit gates distributed according to the Haar measure on $U(4)$. For previous results on the rates for different configurations, see Table I. We focus on the S configuration with OBC. Their dynamics can again be described by a Markovian matrix [23] with an elementary 4×4 two-site matrix [24]. Because of the spectral equivalence of all configurations with OBC, we also know that λ_2 for the S configuration that we study is the same as for the BW configuration. In the TDL, it has been calculated (Table I) that for the BW with OBC, one has $|\lambda_2| = (4/5)^2$. Therefore, if the rate is determined by λ_2 , one should have asymptotic decay $I(t) \asymp (4/5)^{2t}$, i.e., $r_E = 2 \ln(5/4)$. Numerical simulations in Fig. 16, however, show that this is not the case. Based on the data, we conjecture that, at any finite t , the purity for a half-half bipartition in the TDL is, in fact, exactly equal to

$$I(t) = \left(\frac{2}{3}\right)^t. \quad (35)$$

The decay is slower than one would predict from λ_2 ; therefore, one again deals with a phantom eigenvalue like in other staircase circuits that we studied. For the BW configuration $U(4)$ random circuit with open boundary conditions, see Ref. [69].

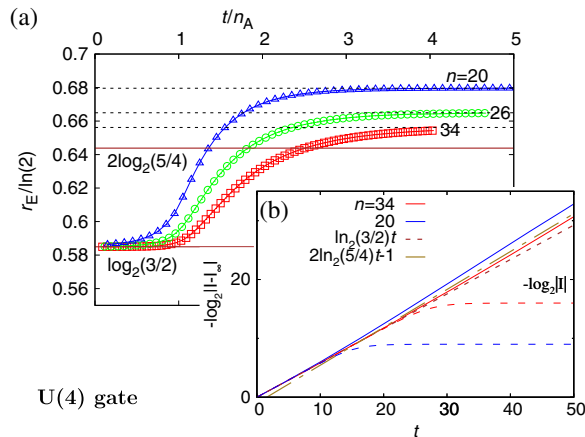


FIG. 16. Entanglement rate for the S configuration with OBC and Haar random 2-qubit gates from $U(4)$. In the TDL, one has $|\lambda_2| = (4/5)^2$; however, the entanglement rate is due to a phantom eigenvalue equal to $r_E = \ln(3/2)$. The three dashed lines in panel (a) show $r_E = -\log_2|\lambda_2|$, using the conjectured finite-size expression $|\lambda_2| = [\frac{4}{5} \cos(\pi/n)]^2$.

Because the Markovian matrix M essentially determines the evolution of squares of expansion coefficients like those of the density operator, one can expect that the phenomena identified in purity will also be present in other quantities that are quadratic in time-evolved coefficients. Such objects include the OTOC, for which our preliminary results show [19] the same effects.

D. Google's Sycamore processor

One of the largest experimental realizations of quantum computation is a recent work by Google AI Quantum performing a random-circuit-like computation on 53 qubits [3]. Here, we analyze the entanglement generation speed of the 2D qubit geometry that is used in the Sycamore quantum processor.

The Google experiment used a fixed 2-qubit Sycamore gate with canonical parameters $a_x = 1$, $a_y = 1$, $a_z = 1/6$, whereas for 1-qubit, the experiment randomly sampled from a set of square roots of Pauli matrices [3]. For the sake of simplicity, and to stay within the framework used in our paper, we take the 1-qubit gates to be random and uniformly distributed on $U(2)$. The geometry used in the Sycamore processor (and that we analyze) is schematically represented in Fig. 17. The protocols that we study are composed of commuting groups of 2-qubit gates denoted by four letters: A, B, C, and D (see Fig. 17). For instance, the protocol used in Google's supremacy experiment is ABCDCDAB. Because of the size limitation, we cannot simulate a circuit with $n = 53$ qubits, so we focus on the half-sized patch circuit with $n = 27$ used by Google. In addition, we show results for two other protocol steps (ABCD and ABCDCADB), which turn out to be optimal at smaller sizes.

In Table IV, we compare all three mentioned protocols, as well as different choices of a 2-qubit gate. In order to facilitate the comparison of entanglement generation rates between protocols with different numbers of gates, we compare the effective eigenvalue λ . The effective eigenvalue is equal to the true eigenvalue $|\lambda_2|$ renormalized to a

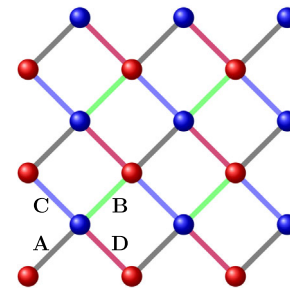


FIG. 17. Qubit connectivity used in Google's quantum processor Sycamore. Dots represent qubits, and lines are different 2-qubit gates. We focus on lattices of $3 \times m$ qubits (we show $m = 6$, giving $n = 18$). Gates labeled with the same letter (and color) are executed simultaneously (see text for details).

TABLE IV. Effective eigenvalue λ for different configurations, gates (XY, Sycamore, CNOT, and random), and system sizes n in a 2D geometry (Fig. 17). With random **a**, we denote the gate with randomly generated canonical parameters $a_x = 0.8501$, $a_y = 0.4628$, and $a_z = 0.1204$. The effective eigenvalue describes purity decay after the application of n 2-qubit gates.

Configuration	λ							
	XY			Sycamore			CNOT	Random a
	$n = 9$	$n = 18$	$n = 27$	$n = 9$	$n = 18$	$n = 27$	$n = 27$	$n = 27$
ABCDCDAB	0.319	0.161	0.141	0.336	0.188	0.169	0.548	0.427
ABCDCADB	0.192	0.128	0.143	0.236	0.160	0.176	0.547	0.428
ABCD	0.114	0.174	0.208	0.133	0.196	0.229	0.494	0.358

step with n gates; that is, in a protocol step with T local gates $M_{i,j}$, it is $\lambda = |\lambda_2|^{n/T}$. For example, protocol steps with eight letters have $T = (10n - 30)/3$, whereas the ABCD protocol has $T = (5n - 15)/3$. Similarly as in 1D protocols, fixing a configuration, the fastest gate is the XY gate. Comparing different configurations (for small n , we checked all possible permutations of eight-letter protocols), we find that while at smaller n there are configurations that have smaller λ than Google’s ABCDCDAB, at $n = 27$, Google’s choice seems to be the optimal one. We can see that the optimal eigenvalue $\lambda = 0.141$ for the XY gate is approximately 15% larger than for the BW PBC $|\lambda_2| = 1/9$ [Eq. (27)]; however, applying a BW protocol on the 2D geometry in Fig. 17 would necessitate additional long-range 2-qubit gates. A similar gain (approximately 10%) would also be obtained for the Sycamore gate. For non-optimal 2-qubit gates, like the CNOT, the optimal protocol can be different than ABCDCDAB even at $n = 27$; see Table IV.

In the 1D protocols studied in the rest of the paper, we have seen that the entanglement generation rate r_E is usually not given by the spectral gap. In Fig. 18, we show numerically calculated purity for the protocols discussed in this section. For even m , subsystem A for our bipartition is

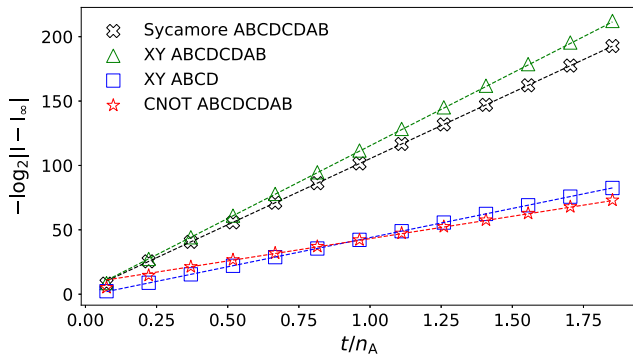


FIG. 18. Purity results for Google-like protocol (symbols, $n = 27$) and the asymptotic linear growth given by the leading eigenvalue $-\log_2 |\lambda_2|$ (dashed lines). The protocols examined suggest that the purity decay rate agrees with the value predicted from spectral analysis.

composed of $m/2$ bottom rows of qubits (see Fig. 17), while for odd m , we take the bottom $(m - 1)/2$ rows plus an additional leftmost qubit in the next row. We can see that, for all cases, the purity rate agrees with $-\log |\lambda_2|$. It is now evident that taking a CNOT gate instead of the XY or the Sycamore gate would result in a significantly slower entanglement generation. It is an important and nontrivial task to find the optimal configuration and choose the best gate for each topology at hand.

V. DISCUSSION

Studying the purity entanglement generation speed in quantum circuits with random single-qubit unitaries and a given nearest-neighbor 2-qubit gate, we have identified circuits with the fastest entanglement production. They are the fastest possible and, by a significant factor, faster than previously studied cases, for instance, circuits with CNOT or random $U(4)$ gates. These fastest scramblers are circuits with a brick-wall pattern of applied gates, with either open or periodic boundary conditions, and the XXZ 2-qubit gate, a special case of which is the XY gate.

Such random circuits should be of interest in experiments where they can reduce the running time and therefore increase the fidelity. These are simple examples of “fastest” chaotic many-body systems; see also Ref. [10] for a Floquet system with such a property. Interestingly, relaxation towards the asymptotic random-state entanglement in optimal protocols, as well as others, does not proceed with a single relaxation rate but rather exhibits a phase transition: Relaxation (or, equivalently, the entanglement generation rate) is faster until the thermalization time (thermalization time is the time when the entanglement closes in on its volume-law saturation value of $\sim n_A$) but then discontinuously transitions into a smaller rate, which happens because the initial relaxation rate is not given by the transfer matrix gap, as one would naively expect. On a mathematical level, it arises because of the nonsymmetric Markovian transfer matrix, even though the action of each individual gate is described by a symmetric matrix. The discrepancy between the spectral gap prediction and the convergence rate happens because the expansion coefficients in the spectral

decomposition can get exponentially large in the size of a many-body system. As a consequence, a subleading eigenvalue, even if it is gapped away from the leading one, can give the dominant relaxation rate. Even stranger is the observation that, for the same non-Hermiticity reasons, such exponentially large terms can result in relaxation with a rate that is smaller than the one given by the leading eigenvalue. It is as if one has an additional “phantom” eigenvalue in the spectrum that is larger than any actual eigenvalue.

We demonstrated that the same phenomenology also occurs for nonextremal fixed 2-qubit gates (non-dual-unitary), as well as for the case of Haar-random 2-qubit gates. For the latter gate several conjectured exact results about the largest eigenvalue and the purity decay are left as open problems. We also show that spatial as well as temporal randomness of on-site unitaries is not necessary, which suggests that the same phenomenology of multi-stage relaxation should also arise in many other situations, including nonrandom systems. While we discussed the average entanglement, for large system sizes, explicit averaging over different circuit realizations is not necessary—a single circuit realization will already show the phenomenon.

Therefore, the phenomena seem to be rather general, a common point being non-Hermitian many-body matrices. One may wonder in how many other situations with non-Hermitian operators such physics can arise (e.g., dissipative systems, scattering problems, etc.). A promising approach would be to find a continuous-time system with the same phenomenology. Alternatively, starting with a Hamiltonian with noise (instead of with a Floquet system—a quantum circuit—with random unitaries), one can arrive at the Lindblad master equation after averaging over noise. In some cases or limits, the resulting [70] non-Hermitian generator is equal to some well-studied models, simplifying the insight. It has been shown [70] that some features in such a setting are the same as in random circuits.

We stress that one needs sufficiently large systems to observe the effects. We are therefore working with genuine many-body physics. In some cases, convergence with n is, namely, quite slow, and one needs good numerical methods to access the required sizes.

How other quantities, besides purity, behave in such random circuits is also an interesting open problem, which includes other Rényi entropies as well as operators, correlation functions, etc. For instance, von Neumann entropy shows multistage thermalization. Preliminary results show [19] that OTOCs also exhibit effects of a phantom eigenvalue. Because the extremal random circuits belong to a class of dual-unitary circuits, non-Hermitian effects that we identify could be at play in such situations and perhaps, even more generally, in some chaotic Floquet systems [10,47,71].

Do Markov chains describing random circuits exhibit a cutoff phenomenon known to occur in some Markovian chains [72,73], and which has been speculated [38] to actually occur in certain random circuits; if yes, are

phenomena we observe in any way related to it? In the cutoff phenomenon, relaxation towards the invariant state as measured by the total variational distance is sudden—up to the cutoff time, there can be “memory” of the initial state and no relaxation, which is then suddenly “forgotten” at the cutoff (mixing) time. Based on purity decay, which is always exponential but with a “wrong” rate that is different than the inverse gap by a factor, it would be tempting to conclude that there is no cutoff. Namely, a necessary condition for a cutoff [74] is that the mixing time is parametrically larger than the inverse gap (their ratio should diverge in the TDL). However, to probe the cutoff, one should study the full measure relaxation as quantified by the total variational distance, which is, in particular, concerned with a worst-case relaxation scenario. On the other hand, we studied initial states that correspond to a valid purity vector (purities for different bipartitions cannot be arbitrary) and relevant bipartitions with extensive n_A .

While we explained the behavior in terms of non-orthogonality and the way expansion coefficients behave, the underlying physics is still not analytically understood. When and why do such phase transitions in the rate occur? Are they associated with some change in the entanglement properties of the underlying state, like, e.g., their Schmidt spectrum being different than that of random states [50]; such features have, for instance, been identified in either appropriate random ensembles [75] or under random evolutions [76,77]. It has been observed that in random circuits, fluctuations qualitatively change around thermalization time [78].

We focused on qubits and the case of nearest-neighbor protocols; however, other topologies are also of interest, including systems in higher dimensions as well as systems with larger local Hilbert space (qudits). For instance, for the all-to-all coupling, we have calculated the exact asymptotic expression for the spectral gap for any gate (Appendix G); however, we did not explore the entanglement generation.

ACKNOWLEDGMENTS

We would like to thank A. Nahum, S. Gopalakrishnan, A. G. Green, and A. Lamacraft for discussions. Support from Grants No. J1-1698 and No. P1-0402 from the Slovenian Research Agency is acknowledged.

APPENDIX A: BRIEF DERIVATION OF TRANSFER MATRIX

In this section, we present a short derivation of Eq. (11); the original derivation with a more detailed explanation can be found in Ref. [27].

In order to sketch the derivation of the Markov chain description, we define a new set of operators on a doubled Hilbert space. These operators will be used to obtain a compact expression for purity for every possible bipartition of our system. Averaging over all independent one-site

Haar random gates results in the Markov chain description in Eq. (11).

Let us write down the purity of the state Ψ for every possible bipartition of our set of qubits. We encode a bipartition in a vector $\mathbf{s} = (s_1, s_2, \dots, s_n)$, where $s_i \in \{\uparrow, \downarrow\}$, depending on whether the i th qubit is in the subsystem A ($s_i = \uparrow$) or B ($s_i = \downarrow$) of a given bipartition. Note that Ψ is a physical state of a chain of n qubits on a Hilbert space $\mathcal{H} = \mathbb{C}_1^2 \otimes \mathbb{C}_2^2 \otimes \dots \otimes \mathbb{C}_n^2$. In order to write the purity, we define two local operators χ_{\downarrow_i} and χ_{\uparrow_i} acting nontrivially on a site i of the duplicated system $\mathcal{H}^{\otimes 2}$, namely,

$$\chi_{s_i} = \begin{cases} \mathbb{1}_i = \sum_{\alpha, \beta=0}^1 |\alpha\beta\rangle_i \langle \alpha\beta|_i & s_i = \downarrow \\ \text{SWAP}_i = \sum_{\alpha, \beta=0}^1 |\beta\alpha\rangle_i \langle \alpha\beta|_i & s_i = \uparrow, \end{cases} \quad (\text{A1})$$

where $|\alpha\rangle, |\beta\rangle$ are basis vectors $\{|0\rangle, |1\rangle\}$ of the local Hilbert space \mathbb{C}_i^2 . The index s_i labels the local part of the bipartition, and the operator χ_{s_i} thus depends on the choice of a bipartition. Defining χ_s as the tensor product of all local operators χ_{s_i} , we can neatly express the purity of a state Ψ for a given bipartition as

$$I_s(\Psi) = \text{tr}[\chi_s(|\Psi\rangle\langle\Psi|)^{\otimes 2}], \quad (\text{A2})$$

where the trace is evaluated on the duplicated system $\mathcal{H}^{\otimes 2}$.

We are interested in the average evolution of purity under the application of one elementary step $U_{i,j} = W_{i,j} V_i V_j$. Averaging over one-site Haar random unitaries V_i and V_j [27], one gets the average purity of the state $U_{i,j}\Psi$,

$$\mathbb{E}_{V_i, V_j \in \text{Haar}} [I_s(U_{i,j}\Psi)] = \sum_{\mathbf{t}, \mathbf{t}'} (W_{i,j})'_{\mathbf{s}, \mathbf{t}'} O_{\mathbf{t}', \mathbf{t}} I_{\mathbf{t}}(\Psi), \quad (\text{A3})$$

where scalars $(W_{i,j})'_{\mathbf{s}, \mathbf{t}'}$ are

$$(W_{i,j})'_{\mathbf{s}, \mathbf{t}'} = \text{tr}[\chi_s W_{i,j}^{\otimes 2} \chi_{\mathbf{t}'}^{\dagger} W_{i,j}^{\otimes 2}] \quad (\text{A4})$$

while scalars $O_{\mathbf{t}', \mathbf{t}}$ are matrix elements of the operator $O = O_1 \otimes O_2 \otimes \dots \otimes O_n$ defined on \mathcal{H} , with

$$O_i = \frac{1}{3} |\uparrow\rangle\langle\uparrow| - \frac{1}{6} |\uparrow\rangle\langle\downarrow| - \frac{1}{6} |\downarrow\rangle\langle\uparrow| + \frac{1}{3} |\downarrow\rangle\langle\downarrow|. \quad (\text{A5})$$

Summing over \mathbf{t}' in Eq. (A3) and defining matrix elements of $M'_{i,j}$ as $[M'_{i,j}]_{\mathbf{s}, \mathbf{t}} = \sum_{\mathbf{t}'} (W_{i,j})'_{\mathbf{s}, \mathbf{t}'} O_{\mathbf{t}', \mathbf{t}}$, we get a simple Markovian mapping of average purities after one gate, as written in Eq. (11).

APPENDIX B: CALCULATION OF STEADY-STATE PURITY

In this Appendix, we derive the eigenvector Φ_∞ of M with eigenvalue $\lambda = 1$. The steady state of M will be of great importance in numerical gap calculations (see

Appendix E). With the help of Φ_∞ , we are able to calculate the purity of random states [Eq. (6)].

Keeping in mind that we must consider only even-parity vectors in the basis of M , we can try to calculate the purity to which we will converge after infinite time. To do so, let us first derive the even-parity eigenvector Φ_∞ with eigenvalue 1. The eigenvector Φ_∞ can be constructed by demanding that, for every $i \in \{1, \dots, n\}$, we have $M_{i,i+1} \Phi_\infty = \Phi_\infty$. From Eq. (19), we see that the coefficients before basis vectors with the same number of up spins must be equal, and that the coefficients before vectors with k up spins must be 3 times greater than those with $(k-2)$ up spins. The eigenvector Φ_∞ is thus

$$\Phi_\infty = \sum_{\mathbf{e}_k \in \{\uparrow, \downarrow\}^{\otimes n}} 3^{n-w} \mathbf{e}_k, \quad (\text{B1})$$

where the sum runs over all basis vectors \mathbf{e}_k of M with an even number of down spins and where $w = w(\mathbf{s})$ is the Hamming weight of the n -bit string \mathbf{e}_k , i.e., the number of digits equal to \downarrow in \mathbf{e}_k . For instance, when $n = 4$, the vector Φ_∞ takes the form

$$\begin{aligned} \Phi_\infty = & 9|\uparrow\uparrow\uparrow\uparrow\rangle \\ & + 3(|\uparrow\uparrow\downarrow\downarrow\rangle + |\uparrow\downarrow\uparrow\downarrow\rangle + |\downarrow\uparrow\uparrow\downarrow\rangle) \\ & + 3(|\uparrow\downarrow\downarrow\uparrow\rangle + |\downarrow\uparrow\downarrow\uparrow\rangle + |\downarrow\downarrow\uparrow\uparrow\rangle) \\ & + 1|\downarrow\downarrow\downarrow\downarrow\rangle. \end{aligned} \quad (\text{B2})$$

We also present an alternative construction of the eigenvector Φ_∞ , which will be especially useful in the derivation of I_∞ that will follow. The new way to obtain Φ_∞ is by projecting the vector

$$\phi = \bigotimes_{i=1}^n (\sqrt{3}, 1) \quad (\text{B3})$$

onto the subspace of even parity, that is,

$$\Phi_\infty = \frac{1}{2} (\mathbb{1} + Z) \phi = \frac{1}{2} \left(\mathbb{1} + \prod_i \sigma_i^z \right) \phi. \quad (\text{B4})$$

After infinite time, we converge to the eigenvector Φ_∞ . If we want to read the purities from the coefficients of our final vector, we must convert Φ_∞ back to the basis of M' ; i.e., we must obtain the vector $\Phi'_\infty = A \Phi_\infty$. We get

$$\begin{aligned} \Phi'_\infty &= A \frac{1}{2} (\mathbb{1} + Z) \phi = \prod_i A_i \frac{1}{2} \left(\mathbb{1} + \prod_i \sigma_i^z \right) \phi \\ &= \left(\frac{1}{\sqrt{3}} \right)^n \left[\bigotimes_{i=1}^n (1, 2) + \bigotimes_{i=1}^n (2, 1) \right]. \end{aligned} \quad (\text{B5})$$

The first term on the right-hand side of Eq. (B5) will be, up to the normalization factor $(1/\sqrt{3})^n$, equal to 2^n , and the

second term will be 2^{n_B} . We normalize the coefficients of Φ'_∞ by demanding $I_\infty(\infty) = 1$, so

$$I_\infty = \frac{2^{n_A} + 2^{n_B}}{1 + 2^n}, \quad (\text{B6})$$

which agrees with Eq. (6).

APPENDIX C: MATRIX PRODUCE STATE FORM OF STEADY STATE

The steady state in the even sector $|\Phi'_\infty\rangle$ [Eq. (20)] is an (unnormalized) eigenvector of M' with eigenvalue 1. It has Schmidt rank 2 for any bipartition and can therefore be written in the MPS ansatz with matrices of size 2.

Let us define two non-normalized orthogonal vectors $|x_+^{(p)}\rangle$ and $|x_-^{(p)}\rangle$ on p qubits,

$$|x_\pm^{(p)}\rangle = \sum_{\mathbf{s} \in \{0,1\}^{\otimes p}} \frac{2^{p-w} \pm 2^w}{2^p \pm 1} |\mathbf{s}\rangle, \quad (\text{C1})$$

where $w = w(\mathbf{s})$ is the Hamming weight of a p -bit string \mathbf{s} , i.e., the number of s_j equal to \downarrow in the string \mathbf{s} . The norm of those vectors is

$$|x_\pm^{(p)}|^2 = \frac{2(5^p \pm 4^p)}{(2^p \pm 1)^2}. \quad (\text{C2})$$

Observe that $|\Phi'_\infty\rangle = |x_+^{(n)}\rangle$.

The Schmidt decomposition of an n -qubit $|x_+^{(n)}\rangle$ for a bipartition into first r (subsystem A) plus last $n-r$ qubits (subsystem B) is

$$|x_+^{(n)}\rangle = \mu_+(n, r) \frac{|x_+^{(r)}\rangle |x_+^{(n-r)}\rangle}{|x_+^{(r)}| |x_+^{(n-r)}|} + \mu_-(n, r) \frac{|x_-^{(r)}\rangle |x_-^{(n-r)}\rangle}{|x_-^{(r)}| |x_-^{(n-r)}|}, \quad (\text{C3})$$

where the two Schmidt coefficients are

$$\mu_\pm(n, r) = \sqrt{\frac{(5^r \pm 4^r)(5^n 4^r \pm 5^r 4^n)}{5^r}} \frac{1}{2^r(2^n + 1)}. \quad (\text{C4})$$

The Schmidt decomposition of the second vector is, on the other hand,

$$|x_-^{(n)}\rangle = \nu_+(n, r) \frac{|x_+^{(r)}\rangle |x_-^{(n-r)}\rangle}{|x_+^{(r)}| |x_-^{(n-r)}|} + \nu_-(n, r) \frac{|x_-^{(r)}\rangle |x_+^{(n-r)}\rangle}{|x_-^{(r)}| |x_+^{(n-r)}|}, \quad (\text{C5})$$

with Schmidt coefficients $\nu_\pm(n, r)$,

$$\nu_\pm(n, r) = \sqrt{\frac{(5^r \pm 4^r)(5^n 4^r \mp 5^r 4^n)}{5^r}} \frac{1}{2^r(2^n - 1)}. \quad (\text{C6})$$

This recursive two-state structure of the Schmidt decompositions guarantees a MPS representation with matrices of size 2. More explicitly, writing

$$I_\infty(\mathbf{s}) = \langle y | A_1^{(s_1)} A_2^{(s_2)} \cdots A_n^{(s_n)} | y \rangle \quad (\text{C7})$$

where $y = (1, 0)$, the matrices are

$$A_1^{(\uparrow)} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mu_+(n, 1) & -\mu_-(n, 1) \\ 0 & 0 \end{pmatrix},$$

$$A_1^{(\downarrow)} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mu_+(n, 1) & \mu_-(n, 1) \\ 0 & 0 \end{pmatrix} \quad (\text{C8})$$

on the first site,

$$A_n^{(\uparrow)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix},$$

$$A_n^{(\downarrow)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad (\text{C9})$$

on the last, and

$$A_r^{(\uparrow)} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{\mu_+(n-r, 1)}{|x_+^{(n-r)}|} & -\frac{\mu_-(n-r, 1)}{|x_+^{(n-r)}|} \\ -\frac{\nu_-(n-r, 1)}{|x_-^{(n-r)}|} & \frac{\nu_+(n-r, 1)}{|x_-^{(n-r)}|} \end{pmatrix},$$

$$A_r^{(\downarrow)} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{\mu_+(n-r, 1)}{|x_+^{(n-r)}|} & \frac{\mu_-(n-r, 1)}{|x_+^{(n-r)}|} \\ \frac{\nu_-(n-r, 1)}{|x_-^{(n-r)}|} & \frac{\nu_+(n-r, 1)}{|x_-^{(n-r)}|} \end{pmatrix} \quad (\text{C10})$$

on sites $r = 2, \dots, n-1$.

APPENDIX D: SPECTRUM ANALYSIS

Let us first note an elementary fact about the matrices that we will repeatedly use. Let A and B be two finite-dimensional square matrices. Then, the product AB has the same spectrum as BA , $AB \simeq BA$. One way to see this is to write down the characteristic polynomial $p(\lambda)$ whose expansion coefficients can all be expressed in terms of invariants, for instance, the traces $\text{tr}(AB)^r$. Because $\text{tr}(AB)^r = \text{tr}(BA)^r$, the above equivalence immediately follows. A corollary of $AB \simeq BA$ is also the spectral equivalence under cyclic permutations that we shall use, e.g., $ABC \simeq CAB$. The other property that we need is the spectral equivalence under transposition, $A^T \simeq A$.

Physical initial states always have positive (even) parity Z ; therefore, only even eigenvalues of M matter for the purity decay. Because all elementary gates $M_{i,j}$ preserve the parity, the parity of eigenvectors for spectrally equivalent configurations $A \simeq B$ corresponding to the same eigenvalue λ is the same. More explicitly, for cyclic permutations, if one has a nonzero λ and the associated eigenvector \mathbf{x} of

ABC , then $CABC\mathbf{x} = C\lambda\mathbf{x} = CAB\mathbf{y} = \lambda\mathbf{y}$, where $\mathbf{y} = C\mathbf{x}$. Because all matrices conserve parity, the parity of \mathbf{x} is the same as that of \mathbf{y} . The same holds for the transposition; parities of eigenvectors \mathbf{x} and \mathbf{y} with eigenvalue λ corresponding to A and A^T , respectively, is the same (in a basis with good parity, A and A^T have a block structure). In the two theorems we are going to prove, we therefore do not have to keep track of the parity—an eigenvalue with good parity will have the same parity in all members of the equivalence class.

1. Proof of Theorem 1

Proof.—Let us begin with an arbitrary permutation of $n - 1$ gates M (an example for $n = 4$ is $M = M_{3,4}M_{1,2}M_{2,3}$). Our first step is to put the matrix $M_{1,2}$ at the rightmost position in the product (the first gate) using cyclic permutations. We thus have $M \simeq AM_{1,2} = AS_1^{(1)}$, where A is a product of matrices that do not contain $M_{1,2}$. We next increase the length of $S_1^{(1)}$ by one using the following step.

Suppose we have

$$RM_{i,i+1}AS_1^{(i-1)}, \quad (\text{D1})$$

where R and A are a product of gates not including $M_{i,i+1}$ or any of the gates in $S_1^{(i-1)} = M_{i-1,i} \cdots M_{1,2}$. Note that the only gate with which the gates in $S_1^{(i-1)}$ do not commute is $M_{i,i+1}$; therefore, one always has $[S_1^{(i-1)}, A] = 0$. Thus, $RM_{i,i+1}AS_1^{(i-1)} \simeq RM_{i,i+1}S_1^{(i-1)}A \simeq ARM_{i,i+1}S_1^{(i-1)} = ARS_1^{(i)}$. We have increased the length of $S_1^{(i-1)}$ by one. Iterating this step, we end up with $M \simeq S_1^{(n-1)}$. This concludes the proof. ■

2. Proof of Theorem 2

Proof.—In the first step, we bring, by cyclic permutations, the gate $M_{1,2}$ to the rightmost position. Then, in step 1 (described below), we try to construct $B_1^{(3)}$ ($B_1^{(3)}$ is the first nontrivial brick-wall because $B_1^{(1)}$ is equal to $S_1^{(1)} = M_{1,2}$). Step 1 will either succeed in constructing $B_1^{(3)}$, or we will end up in the canonical form with $p = 1$ (see Fig. 6).

Step 1.—First, we bring $M_{2,3}$ to the position of a second gate. Writing $RM_{2,3}AM_{1,2}$, where R and A are arbitrary products excluding $M_{1,2}$ and $M_{2,3}$, we have two possibilities ($M_{2,3}$ does not commute with only two gates, $M_{1,2}$ and $M_{3,4}$; for $n = 3$, when $M_{3,4}$ is, in fact, equal to $M_{3,1}$, the presented argument still works):

- If $M_{3,4} \notin A$, so that $[M_{2,3}, A] = 0$, we have $RM_{2,3}AM_{1,2} \simeq RAM_{2,3}M_{1,2}$.
- If $M_{3,4} \in A$, we use $[M_{2,3}, R] = 0$ and cyclic permutations to get $RM_{2,3}AM_{1,2} \simeq M_{1,2}M_{2,3}RA$. Using the

fact that all $M_{i,i+1}$ are real and symmetric, and that M^T and M have the same spectra, we write $M_{1,2}M_{2,3}RA \simeq A^T R^T M_{2,3} M_{1,2}$.

Having a form $RM_{3,4}AM_{2,3}M_{1,2}$, we now try to bring $M_{3,4}$ to the rightmost position. There are, again, two possibilities ($n = 4$ does not influence the argument):

- If $M_{4,5} \notin A$, we have $RM_{3,4}AM_{2,3}M_{1,2} \simeq RAM_{3,4}M_{2,3}M_{1,2} = RAS_1^{(3)}$.
- If $M_{4,5} \in A$, we use $[M_{3,4}, R]$ and a cyclic permutation to write $RM_{3,4}AM_{2,3}M_{1,2} \simeq RAM_{2,3}M_{1,2}M_{3,4} = RAB_1^{(3)}$.

At this point, we therefore either have $RAB_1^{(3)}$ and have succeeded (finished step 1), or we have $RAS_1^{(3)}$ and continue by trying to include $M_{4,5}$ into $S_1^{(3)}$.

The procedure is, by now, familiar; writing $RM_{4,5}AS_1^{(3)}$, we have either $M_{5,6} \in A$, in which case $RM_{4,5}AS_1^{(3)} \simeq RAS_1^{(4)}$, or $M_{5,6} \notin A$, in which case $RM_{4,5}AS_1^{(3)} \simeq RAS_1^{(4)}$, or $M_{5,6} \notin A$, in which case $RM_{4,5}AS_1^{(3)} \simeq RAS_1^{(3)}M_{4,5} \simeq RAM_{3,4}M_{4,5}M_{2,3}M_{1,2} \simeq RAM_{4,5}B_1^{(3)}$, where, in the last equivalence, we used that $M_{3,4}$ commutes with all the gates in A and R . Therefore, we have either increased the number of gates in $S_1^{(3)}$ by one or obtained the desired $B_1^{(3)}$.

We are now in a position to write the generic iteration step: In $RM_{i+1,i+2}AS_1^{(i)}$, one has either (i) $M_{i+2,i+3} \notin A$ or (ii) $M_{i+2,i+3} \in A$. In case (i), we immediately get $\simeq RAS_1^{(i+1)}$. In case (ii), we have $RM_{i+1,i+2}AS_1^{(i)} \simeq RAS_1^{(i)}M_{i+1,i+2} \simeq RAM_{i,i+1}M_{i+1,i+2}M_{i-1,i} \cdots M_{3,4}M_{2,3}M_{1,2}$. Using the fact that gates $M_{3,4}$ to $M_{i,i+1}$ commute with all gates in R and A , we can use cyclic permutations to first

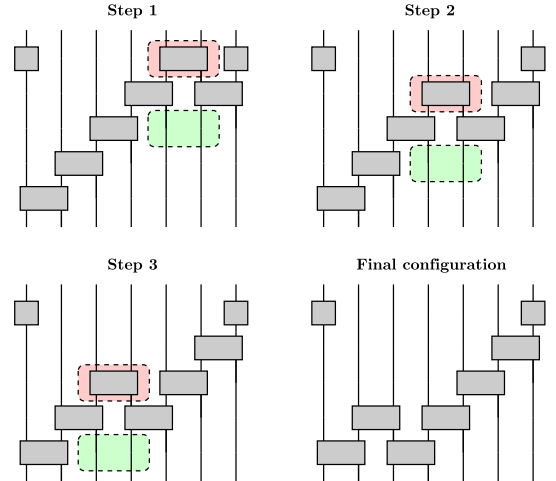


FIG. 19. Example of a transformation from the configuration $M = RS_1^{(i)}M_{i,i+1}$ to the configuration $M = RB_1^{(3)}$. Using commutation relations and cyclic permutations, we must transform the operator in the red box to the position denoted by the green box.

bring $M_{i,i+1}$ to the rightmost position, then $M_{i-1,i}$, and so on, until we bring $M_{3,4}$ to the first position, resulting in $B_1^{(3)}$. An example of such a transformation can be seen in Fig. 19.

We see that repeating this procedure eventually either produces $B_1^{(3)}$ at the position of the first three gates, or we end up with $S_1^{(n)}$, i.e., the canonical configuration with $p = 1$. If we have the situation $RB_1^{(3)}$, we continue with step 2, which describes a generic step starting with $RB_1^{(i)}$. It either increases the number of brick-wall gates, bringing us to $AB_1^{(i+2)}$, or it will end up in $S_{i+1}^{(n-i)} B_1^{(i)}$.

Step 2.—Starting with $RB_1^{(i)}$ (with odd i), we see that $B_1^{(i)}$, which acts on sites $1, \dots, i+1$, does not commute only with two gates, $M_{n,1}$ and $M_{i+1,i+2}$. It behaves in exactly the same way as gate $M_{1,2}$ in step 1 because none of the gates in R acts on the inner qubits $2, \dots, i$ of the block $B_1^{(i)}$. Following exactly the same steps as in step 1, we try to add two gates to $B_1^{(i)}$ in order to obtain $M_{i+1,i+2} B_1^{(i)} M_{i+2,i+3} = B_1^{(i+2)}$. In step 1(b), we used that $M_{1,2}^T = M_{1,2}$, which is not true for $B_1^{(i)}$; however, the transpose of $B_1^{(i)}$ can always be transformed into $B_1^{(i)}$ by using cyclic permutations and the fact that all the “inner” gates in $B_1^{(i)}$ acting on qubits $2, \dots, i$ commute with all the gates in R . The procedure analogous to the one in step 1 will therefore either result in $RB_1^{(i)} \simeq AB_1^{(i+2)}$ or in $RB_1^{(i)} \simeq S_{i+1}^{(n-i)} B_1^{(i)}$, which is the canonical configuration with $p = (i+1)/2$. This concludes the proof. ■

3. Proof of eigenvalue 1/9

We can also prove that $\lambda = 1/9$ is an eigenvalue of M for the brick-wall configuration with PBC and the XY gate.

Proof.—Let us, for simplicity, focus on even n . Taking PBC BW $M = B_1^{(n)}$ with the XY gate, we can explicitly construct an eigenvector \mathbf{v} of M with eigenvalue $1/9$. The ansatz for the eigenvector is

$$\mathbf{v} = \sum_{i=1}^{n/2} T^{2i} (\mathbf{v}_\alpha \otimes \mathbf{v}_\alpha \otimes \dots \otimes \mathbf{v}_\alpha), \quad (\text{D2})$$

where $\mathbf{v}_\alpha = \mathbf{v}_1 + \alpha \mathbf{v}_2$, where $\mathbf{v}_{1,2,3}$ are the eigenvectors of a 2-qubit $M_{i,i+1}$ [Eq. (19)] and T is the translation operator by one site on a circle, e.g., $T^2 \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} = \mathbf{c} \otimes \mathbf{a} \otimes \mathbf{b}$. Let us write the brick-wall step M as $M = M_e M_o$, where M_o is the first layer (half-step) of the BW; i.e., it contains all $M_{i,i+1}$ with odd i , and M_e is the second layer ($M_{i,i+1}$ with even i). For an arbitrary α , the ansatz vector \mathbf{v} is already the eigenvector of M_o with $\lambda = -1/3$. If we want it to be the eigenvector of the whole M , we must fix the constant α so that it is also the eigenvector of M_e . Let us focus on a pair of

qubits $(i, i+1)$ with even i . We can expand \mathbf{v} as a linear combination of $n4^{(n/2)-1}$ basis vectors \mathbf{e}_k [we choose the same basis as in Eq (16)]. Each basis vector can be further decomposed as $\mathbf{e}_k = \mathbf{w}_{1,i-1}^k \otimes \mathbf{v}_{i,i+1}^k \otimes \mathbf{w}_{i+2,n}^k$, where the subscripts denote the range of qubits that these decomposition vectors describe. For every $\mathbf{v}_{i,i+1}^k$, we want either $M_{i,i+1} \mathbf{v}_{i,i+1}^k = \mathbf{v}_{i,i+1}^k$ or $M_{i,i+1} \mathbf{v}_{i,i+1}^k = -\frac{1}{3} \mathbf{v}_{i,i+1}^k$, so $\mathbf{v}_{i,i+1}^k$ must be either $(3,0,0,1)$ or $(0, \pm 1, 1, 0)$. The vectors $\mathbf{v}_{i,i+1}^k$ depend on the neighboring pairs $(i-1, i)$ and $(i+2)$, which can be $\mathbf{v}_\alpha \otimes \mathbf{v}_\alpha$, $\mathbf{v}_\alpha \otimes \mathbf{v}_3$, or $\mathbf{v}_3 \otimes \mathbf{v}_\alpha$ [Eq. (D2)]. We have 32 possible combinations on qubits $(i, i+1)$, but for now, let us focus on four instances: $(9, 0, 0, \alpha^2)$, $(\alpha^2, 0, 0, 1)$, $(0, 3, \alpha^2, 0)$, and $(0, \alpha^2, 3, 0)$. We obtain the desired vector if we fix $\alpha = \sqrt{3}$. For the other 28 possibilities, our demands are automatically fulfilled without the specification of α ; one can check this by writing down all 32 possibilities that come from terms $\mathbf{v}_\alpha \otimes \mathbf{v}_\alpha$, $\mathbf{v}_\alpha \otimes \mathbf{v}_3$, and $\mathbf{v}_3 \otimes \mathbf{v}_\alpha$. For small systems ($n = 4, 6, 8$), we numerically checked the correctness of our guess. The eigenvector \mathbf{v} does not have a good parity and can be decomposed as $\mathbf{v} = \mathbf{v}_e + \mathbf{v}_o$, into even- and odd-parity parts, both of which are, because of parity conservation of M , still eigenvectors. This concludes the proof. ■

4. Numerical data for λ_2

Here, we present the numerical results used to determine the fastest asymptotic scrambler for PBC and OBC protocols. To calculate λ_2 , we use either exact diagonalization or the power method as described in Appendix E.

Figure 20 shows color-coded plots of $|\lambda_2|$ for the S configuration with OBC for different values of all three

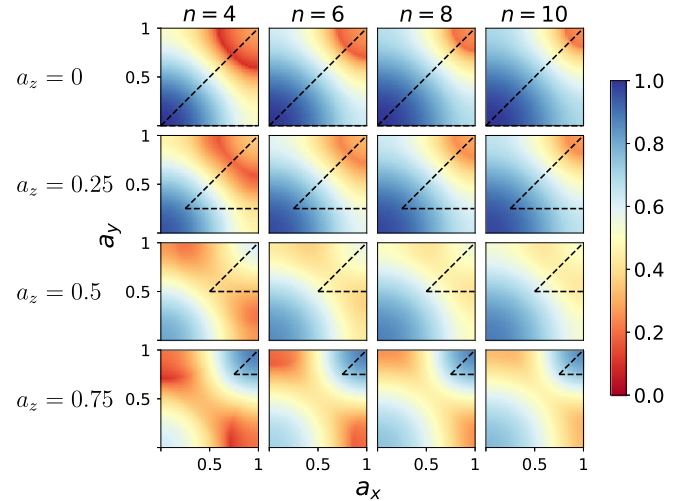


FIG. 20. Second-largest eigenvalue $|\lambda_2|$ of M for the OBC S protocol and different canonical gate parameters \mathbf{a} [Eq. (3)]. The columns represent a fixed number of qubits, and the rows are for different a_z . The dotted triangles indicate the relevant set of irreducible parameters a_x, a_y .

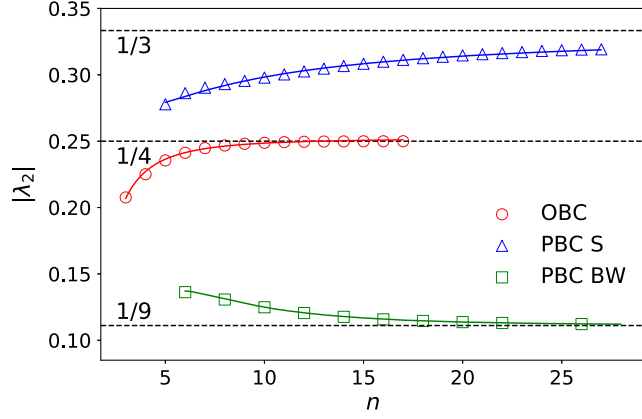


FIG. 21. Eigenvalue $|\lambda_2|$ for OBC, and PBC with the S configuration (the equivalence class $p = 1$) and BW configuration ($p = n/2$) for the 2-qubit XY gate. The extrapolated values in the TDL (lines) are obtained from the fits (full curves) to numerical data (symbols). The fits are $|\lambda_2(n)| = 0.252 + 0.008/n - 0.43/n^2$ (OBC), $|\lambda_2(n)| = 0.334 - 0.44/n + 0.83/n^2$ (PBC, $p = 1$), and $|\lambda_2(n)| = 0.114 - 0.20/n + 4.7/n^2 - 16/n^3$ (PBC, $p = n/2$).

canonical parameters a_j . Looking for the smallest $|\lambda_2|$ and for how things change with increasing n , a promising candidate for the smallest $|\lambda_2|$ is $\mathbf{a} = (1, 1, 0)$, i.e., the XY gate. Studying more closely how $|\lambda_2|$ depends on n for this XY gate, Fig. 21, we can conjecture that, in the TDL, one has $|\lambda_2| = 1/4$ for any protocol with OBC (remember that, for OBC, the spectrum does not depend on the configuration).

Figure 22 shows the values of $|\lambda_2|$ for different 2-qubit gates (parameters a_x, a_y, a_z) with PBC and different configurations (parameter p) for fixed $n = 10$. From Fig. 22, we learn that $|\lambda_2|$ monotonically decreases as we increase p . Moreover, the fastest entanglement generation comes from the region $a_z = 0$. Figures 23 and 24 show the dependence of $|\lambda_2|$ on $a_x, a_y, a_z = 0$ and n for $p = 1$

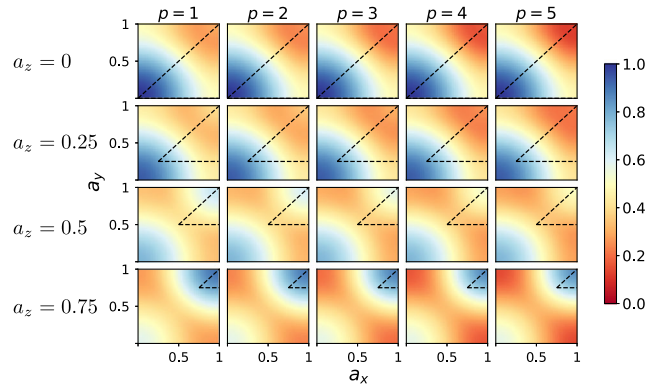


FIG. 22. $|\lambda_2|$ for the PBC case at $n = 10$. The columns represent a fixed configuration, and the rows represent a fixed value for the parameter a_z .

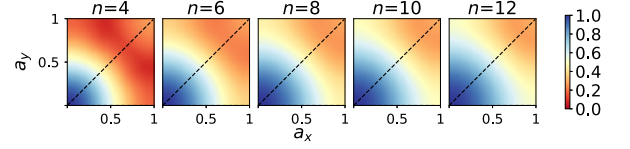


FIG. 23. $|\lambda_2|$ of configurations $p = 1$ for different values of parameters a_x, a_y at $a_z = 0$.

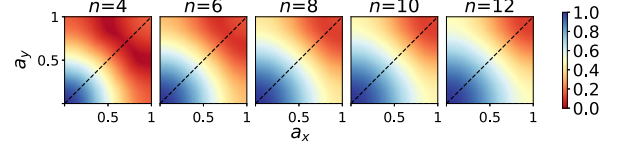


FIG. 24. $|\lambda_2|$ of configurations $p = \lfloor n/2 \rfloor$ for different values of parameters a_x, a_y at $a_z = 0$.

(slowest scrambler) and $p = \lfloor n/2 \rfloor$ (fastest scrambler), respectively. Contrary to OBC protocols, where the fastest asymptotic scrambler is obtained for all $a_x = a_y = 1, a_z \in [0, a_c \approx 0.32]$, the smallest $|\lambda_2|$ for PBC protocols comes from a single choice of 2-qubit gates, namely, $a_x = a_y = 1, a_z = 0$ (XY gates).

APPENDIX E: NUMERICAL CALCULATION OF SPECTRAL GAPS FOR LARGE n

Whenever exact numerical diagonalization is too demanding in terms of memory, we calculate the dominant eigenvalue $|\lambda_2|$ using the power method. Briefly, the power method consists of iterations of the form

$$\mathbf{u}_{i+1} = M\mathbf{w}_i, \quad (\text{E1})$$

$$\mathbf{w}_{i+1} = \mathbf{u}_{i+1}/\|\mathbf{u}_{i+1}\|, \quad (\text{E2})$$

where M is the matrix for which we want $|\lambda_2|$ and \mathbf{w}_k is the vector we iterate. For almost all initial vectors \mathbf{w}_0 , the power method converges to the vector with the greatest eigenvalue in absolute value.

In our case, the matrix M has one even-parity eigenvector Φ_∞ with eigenvalue equal to 1 (for details, see Appendix B). If we wish to get $|\lambda_2|$ from our power iteration, we must subtract the component corresponding to Φ_∞ from \mathbf{w}_i at every step i . We iterate

$$\mathbf{u}_{i+1} = M\mathbf{w}_i - \Phi_\infty \langle \Phi_\infty, \mathbf{w}_i \rangle, \quad (\text{E3})$$

$$\mathbf{w}_{i+1} = \mathbf{u}_{i+1}/\|\mathbf{u}_{i+1}\|, \quad (\text{E4})$$

where $\langle *, * \rangle$ is the standard inner product.

If $|\lambda_2|$ is nondegenerate, the power iteration converges to the corresponding eigenvector. In general, $\lambda_2 \in \mathbb{C}$. Because the product M is real, if $\lambda_2 \in \mathbb{C}$ there is also a $\bar{\lambda}_2$ in the spectrum of M . Suppose we want to extract the value $|\lambda_2|$

from our power iteration. We begin with a vector \mathbf{w}_0 and denote with \mathbf{w}_∞ the eigenvector of M with eigenvalue λ_2 . After $k \gg 1$ iterations, we can write

$$M^k \mathbf{w}_0 = \lambda_2^k c \mathbf{w}_\infty + \bar{\lambda}_2^k \bar{c} \bar{\mathbf{w}}_\infty, \quad (\text{E5})$$

where $c = \langle \mathbf{w}_\infty, \mathbf{w}_0 \rangle$. In Eq. (E5), we use the fact that, at every step, we subtract the component resulting from the eigenvector with $\lambda = 1$; we can, equivalently, start with a vector orthogonal to Φ_∞ : $\langle \mathbf{w}_0, \Phi_\infty \rangle = 0$.

We are interested only in the value $|\lambda_2|$, so we can compute the norm of $M^k \mathbf{w}_0$ and compare it with $M^{k+1} \mathbf{w}_0$. Writing $\lambda_2 = |\lambda_2| e^{i\phi}$, $c = |c| e^{i\psi_c}$, and $\langle \mathbf{w}_\infty, \bar{\mathbf{w}}_\infty \rangle = V e^{i\psi_v}$, we get

$$\|M^k \mathbf{w}_0\| = 2|\lambda_2|^{2k} |c|^2 [1 + V \cos(\tilde{k}\phi)]. \quad (\text{E6})$$

For the sake of simplicity, we define $\tilde{k} = k + (\psi_c/2\psi) + (\psi_v/2\psi) = k + \text{const}$.

An arbitrary quotient of the $(k+1)$ th and k th steps can be written as

$$\frac{\|M^{k+1} \mathbf{w}_0\|}{\|M^k \mathbf{w}_0\|} = |\lambda_2|^2 \frac{1 + V \cos((\tilde{k}+1)\phi)}{1 + V \cos(\tilde{k}\phi)} = |\lambda_2|^2 \frac{v_{k+1}}{v_k}, \quad (\text{E7})$$

where we use the notation $v_k = 1 + V \cos(\tilde{k}\phi)$. Now, suppose $\phi = (l/m)2\pi$ for integers l, m , and then $v_{k+m} = v_k$ for every k . Taking the geometric mean of norms $\|M^k \mathbf{w}_0\|$ of m subsequent iterations, we get

$$\sqrt[m]{\frac{v_{k+1}}{v_k} \frac{v_{k+2}}{v_{k+1}} \cdots \frac{v_{k+m}}{v_{k+m-1}}} = |\lambda_2|^2. \quad (\text{E8})$$

We find that, for most cases, $\phi \approx (l/m)2\pi$ holds; hence, a good estimation of $|\lambda_2|$ is possible.

APPENDIX F: FURTHER ENTANGLEMENT DATA

1. von Neumann entropy

In Fig. 25, we compare the von Neumann entropy with the logarithm of the average purity that we studied in the rest of the paper. As we can see, von Neumann entropy behaves rather similarly to purity or, more precisely, as $-\log \langle I(t) \rangle$. In particular, it also exhibits a phase transition in the local rate at $t \approx t_\infty$ ($t_\infty \approx 8$ in the figure).

2. Fluctuations and randomness

For sufficiently large times, the state reached under random-circuit evolution is close to a random state. Because of the measure concentration in a large Hilbert space, one can obtain good self-averaging properties even for a single random-circuit realization. In Fig. 26, we numerically check such a self-averaging property for the S

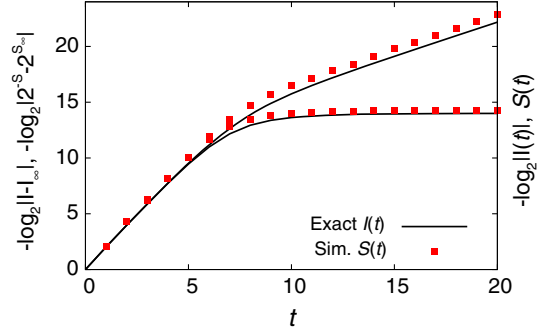


FIG. 25. Exact average results for purity (black curves) are compared with the von Neumann entropy $S(t)$ (red squares). In the upper data, the thermal value is subtracted (left label), while for the lower two saturating sets, it is not (right label). We use $n = 30$ and the S configuration with PBC and the XXZ gate with $a_z = 0.5$ [the same parameters as in Figs. 10(c) and 10(d)] for which the rate is $r_E = 2 \ln 2$.

configuration with PBC and two different system sizes in order to get insight into how fluctuations behave with n . First, we notice that a single realization with random Haar independent identically distributed 1-qubit unitaries

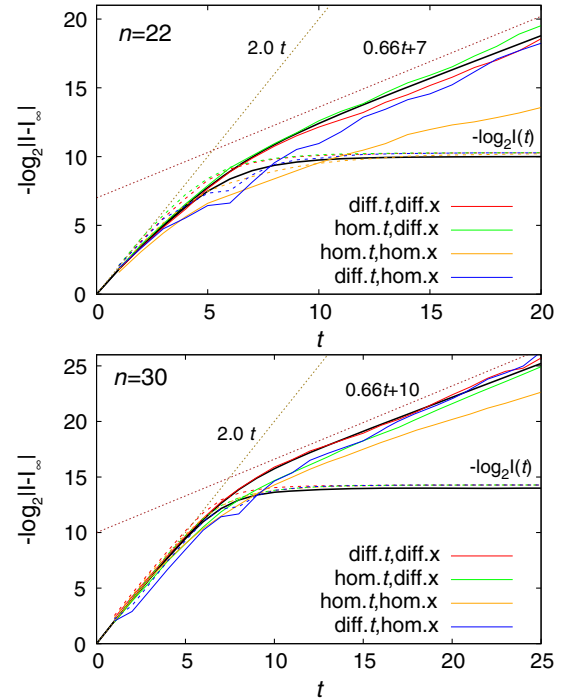


FIG. 26. Single random-circuit realization results for a single random-product initial state (four colored curves with labels) and the average purity (smooth black solid curves). We use the S configuration with PBC and $a_z = 0.5$ [the same parameters as in Figs. 10(c) and 10(d)] and system size $n = 22$ (top) and $n = 30$ (bottom). Labels for the colored curves denote whether 1-qubit Haar unitaries are different on each site (“diff.x”) or the same (“hom.x”), as well as whether they are the same at every time step (“hom.t”) or different at every circuit layer (“diff.t”).

in both space and time (red curves labeled by “diff.t,diff.x”) is, for large n , almost on top of the average purity. Therefore, while an explicit averaging over single-site Haar random unitaries simplifies the analytical treatment (and results in a Markovian process), it is not necessary for the observed phenomena.

We also check how randomness in one-site unitaries influences our results. To that end, we compare our canonical case where one-site unitaries at each site and time step are independent (“diff.t,diff.x”) with a situation where unitaries are the same at every site (“hom.x”) and/or the same at every time step (“hom.t”). Interestingly, we see that if we use the same random 1-qubit unitary at every site, as well as at every time (“hom.x,hom.t”)—i.e., the whole single realization of a circuit uses only one Haar 1-qubit unitary—one will get the same behavior in the thermodynamic limit. Based on the data, we can conjecture that, in the TDL, explicit randomness is not necessary neither in space nor in time. Comparing fluctuations between the four cases shown, they are expectedly the smallest for 1-qubit unitaries that are independently random in space and time, followed by the case of random in space and the same in time (every circuit layer uses the same 1-qubit unitaries), and then the case with no spatial randomness but with new unitaries at every time. Lastly, the largest fluctuations are observed for the circuits that are homogeneous in space and time.

APPENDIX G: RANDOM ALL-TO-ALL COUPLING

Here, we compute the spectral gap of the transfer matrix describing random quantum circuits, where, on every step, we randomly choose two qubits i, j on which we apply a 2-qubit gate. Unlike in the main text, we therefore allow coupling between an arbitrary pair of qubits (not just NN). The average step can be written as

$$\bar{M} = \frac{2}{n(n-1)} \sum_{i,j=1;i < j}^n M_{i,j}, \quad (\text{G1})$$

where $M_{i,j}$ is the familiar matrix from Eq. (16). We follow the procedure from Ref. [24], where the gap has been calculated for Clifford gates. Using Pauli notation (17) and rewriting everything in terms of total spin operators $S_\alpha = 1/2 \sum_{i=1}^n \sigma_i^\alpha$, we get

$$\begin{aligned} \bar{M} &= \frac{2h}{n} S_z + \frac{4}{n(n-1)} (J_x S_x^2 + J_y S_y^2 + J_z S_z^2) \\ &+ \left(d - \frac{J_x + J_y + J_z}{n-1} \right) \mathbb{1}. \end{aligned} \quad (\text{G2})$$

This represents the Lipkin-Meshkov-Glick (LMG) model [79], for which one can calculate the spectral gap by taking the semiclassical limit $n \gg 1$ (i.e., spin size $S \rightarrow \infty$) and replacing spin operators with classical spins

$$S_x = S \cos \phi \sqrt{1 - \mu^2}, \quad (\text{G3})$$

$$S_y = S \sin \phi \sqrt{1 - \mu^2}, \quad (\text{G4})$$

$$S_z = S\mu. \quad (\text{G5})$$

Expanding \bar{M} around the energy maximum and quantizing the resulting harmonic oscillator, we get the spectral gap

$$1 - \lambda_2 = \frac{3h}{n} + \mathcal{O}(1/n^2), \quad (\text{G6})$$

where $h = \frac{1}{9}(3 - v)$, $v = \cos(\pi a_x) \cos(\pi a_y) + \cos(\pi a_x) \times \cos(\pi a_z) + \cos(\pi a_y) \cos(\pi a_z)$. The gap is maximal and equal to $1 - \lambda_2 = (4/3n)$ at $a_x = 1, a_z = 0$, and an arbitrary a_y , which includes both XY ($a_y = 1$) and CNOT ($a_y = 0$). Those gates are therefore the fastest scramblers for the all-to-all coupling (numerically identified in Ref. [38]). For the Clifford XY and CNOT gates, the gap has already been calculated in Ref. [24]; Eq. (G6), though, extends it to any gate. Note that the normalization in Eq. (G1) means that \bar{M} represents the average action of a single 2-qubit gate. If we measure time in such units that $T = n$ gates are applied per unit of time, we have $I(t) - I_\infty \asymp (1 - 3h/n)^{tn} \rightarrow \exp(-3ht)$, giving the purity rate $r_E = 3h$, which is equal to $r_E = \frac{4}{3}$ for optimal gates.

-
- [1] B. Zheng, X. Chen, D. L. Zhou, and X. G. Wen, *Quantum Information Meets Quantum Matter* (Springer, New York, 2019).
 - [2] J. Emerson, Y. S. Weinstein, M. Saraceno, S. Lloyd, and D. G. Cory, *Pseudo-Random Unitary Operators for Quantum Information Processing*, *Science* **302**, 2098 (2003).
 - [3] F. Arute, K. Arya, R. Babbush *et al.*, *Quantum Supremacy Using a Programmable Superconducting Processor*, *Nature (London)* **574**, 505 (2019).
 - [4] C. W. von Keyserlingk, T. Rakovszky, F. Pollmann, and S. L. Sondhi, *Operator Hydrodynamics, OTOCs, and Entanglement Growth in Systems without Conservation Laws*, *Phys. Rev. X* **8**, 021013 (2018).
 - [5] A. Nahum, S. Vijay, and J. Haah, *Operator Spreading in Random Unitary Circuits*, *Phys. Rev. X* **8**, 021014 (2018).
 - [6] V. Khemani, A. Vishwanath, and D. A. Huse, *Operator Spreading and the Emergence of Dissipative Hydrodynamics under Unitary Evolution with Conservation Laws*, *Phys. Rev. X* **8**, 031057 (2018).
 - [7] N. Hunter-Jones, *Operator Growth in Random Quantum Circuits with Symmetry*, [arXiv:1812.08219](https://arxiv.org/abs/1812.08219).
 - [8] A. Chan, A. De Luca, and J. T. Chalker, *Solution of a Minimal Model for Many-Body Quantum Chaos*, *Phys. Rev. X* **8**, 041019 (2018).
 - [9] P. Calabrese and J. L. Cardy, *Evolution of Entanglement Entropy in One-Dimensional Systems*, *J. Stat. Mech.* (2005) P04010.

- [10] B. Bertini, P. Kos, and T. Prosen, *Entanglement Spreading in a Minimal Model of Maximal Many-Body Quantum Chaos*, *Phys. Rev. X* **9**, 021033 (2019).
- [11] P. Hayden and J. Preskill, *Black Holes as Mirrors: Quantum Information in Random Subsystems*, *J. High Energy Phys.* **09** (2007) 120.
- [12] Y. Sekino and L. Susskind, *Fast Scramblers*, *J. High Energy Phys.* **10** (2008) 065.
- [13] L. Piroli, C. Sünderhauf, and X.-L. Qi, *A Random Unitary Circuit Model for Black Hole Evaporation*, *J. High Energy Phys.* **04** (2020) 063.
- [14] H. Liu and S. J. Suh, *Entanglement Tsunami: Universal Scaling in Holographic Thermalization*, *Phys. Rev. Lett.* **112**, 011601 (2014).
- [15] H. Casini, H. Liub, and M. Mezei, *Spread of Entanglement and Causality*, *J. High Energy Phys.* **07** (2016) 077.
- [16] M. Mezei and D. Stanford, *On Entanglement Spreading in Chaotic Systems*, *J. High Energy Phys.* **05**(2017) 065.
- [17] W. Brown and O. Fawzi, *Decoupling with Random Quantum Circuits*, *Commun. Math. Phys.* **340**, 867 (2015).
- [18] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, *From Quantum Chaos and Eigenstate Thermalization to Statistical Mechanics and Thermodynamics*, *Adv. Phys.* **65**, 239 (2016).
- [19] J. Bensa and M. Žnidarič (to be published).
- [20] T. Mori and T. Shirai, *Resolving a Discrepancy between Liouvillian Gap and Relaxation Time in Boundary-Dissipated Quantum Many-Body Systems*, *Phys. Rev. Lett.* **125**, 230604 (2020).
- [21] M. Žnidarič, *Relaxation Times of Dissipative Many-Body Quantum Systems*, *Phys. Rev. E* **92**, 042143 (2015).
- [22] J. Emerson, E. Livine, and S. Lloyd, *Convergence Conditions for Random Quantum Circuits*, *Phys. Rev. A* **72**, 060302(R) (2005).
- [23] R. Oliveira, O. C. O. Dahlsten, and M. B. Plenio, *Generic Entanglement Can Be Generated Efficiently*, *Phys. Rev. Lett.* **98**, 130502 (2007); O. C. O. Dahlsten, R. Oliveira, and M. B. Plenio, *The Emergence of Typical Entanglement in Two-Party Random Processes*, *J. Phys. A* **40**, 8081 (2007).
- [24] M. Žnidarič, *Exact Convergence Times for Generation of Random Bipartite Entanglement*, *Phys. Rev. A* **78**, 032324 (2008).
- [25] W. G. Brown and L. Viola, *Convergence Rates for Arbitrary Statistical Moments of Random Quantum Circuits*, *Phys. Rev. Lett.* **104**, 250501 (2010).
- [26] P. Ćwikliński, M. Horodecki, M. Mozrymas, Ł. Pankowski, and M. Studziński, *Local Random Quantum Circuits Are Approximate Polynomial-Designs: Numerical Results*, *J. Phys. A* **46**, 305301 (2013).
- [27] W.-T. Kuo, A. A. Akhtar, D. P. Arovas, and Y.-Z. You, *Markovian Entanglement Dynamics under Locally Scrambled Quantum Evolution*, *Phys. Rev. B* **101**, 224202 (2020).
- [28] R. Belyansky, P. Bienias, Y. A. Kharkov, A. V. Gorshkov, and B. Swingle, *Minimal Model for Fast Scrambling*, *Phys. Rev. Lett.* **125**, 130601 (2020).
- [29] D. Gross, K. Audenaert, and J. Eisert, *Evenly Distributed Unitaries: On the Structure of Unitary Designs*, *J. Math. Phys. (N.Y.)* **48**, 052104 (2007).
- [30] A. W. Harrow and R. Low, *Random Quantum Circuits Are Approximate 2-Designs*, *Commun. Math. Phys.* **291**, 257 (2009).
- [31] F. G. S. L. Brandao, A. W. Harrow, and M. Horodecki, *Local Random Quantum Circuits Are Approximate Polynomial Designs*, *Commun. Math. Phys.* **346**, 397 (2016).
- [32] F. G. S. L. Brandao, A. W. Harrow, and M. Horodecki, *Efficient Quantum Pseudorandomness*, *Phys. Rev. Lett.* **116**, 170502 (2016).
- [33] J. Haferkamp and N. Hunter-Jones, *Improved Spectral Gaps for Random Quantum Circuits: Large Local Dimensions and All-to-All Interactions*, [arXiv:2012.05259](https://arxiv.org/abs/2012.05259).
- [34] A. Hamma, S. Santra, and P. Zanardi, *Quantum Entanglement in Random Physical States*, *Phys. Rev. Lett.* **109**, 040502 (2012).
- [35] T. Zhou and A. Nahum, *Emergent Statistical Mechanics of Entanglement in Random Unitary Circuits*, *Phys. Rev. B* **99**, 174205 (2019).
- [36] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, *Quantum Entanglement Growth under Random Unitary Dynamics*, *Phys. Rev. X* **7**, 031016 (2017).
- [37] L. Arnaud and D. Braun, *Efficiency of Producing Random Unitary Matrices with Quantum Circuits*, *Phys. Rev. A* **78**, 062329 (2008).
- [38] M. Žnidarič, *Optimal Two-Qubit Gate for Generation of Random Bipartite Entanglement*, *Phys. Rev. A* **76**, 012318 (2007).
- [39] There is an interesting observation about r_E (or, equivalently, v_E) between the U(4) gate in the BW and random NN configurations. In the random NN protocol the Markovian matrix is equal to a sum of two-site matrices, whereas it is equal to a product of the same two-site matrices in the BW case. Based on exact diagonalization results, it was guessed [40] that the largest nontrivial eigenvalue for the BW case is $\lambda_2 = [\frac{4}{5}\cos(\pi/n)]^2$ for the OBC and $\lambda_2 = (\frac{4}{5}\cos\frac{\pi}{n})^4$ for the PBC (expressions are exact for any n and agree with the asymptotic results for r_E in Table I). For the random NN protocol, on the other hand, one has [24] $\lambda_2 = \{1 - \frac{1}{n-1}[1 - \frac{4}{5}\cos(\pi/n)]\}^{n-1}$ for the OBC and $\lambda_2 = \{1 - \frac{2}{n}[1 - \frac{4}{5}\cos(\pi/n)]\}$ for the PBC.
- [40] M. Žnidarič, *Entanglement Growth in Diffusive Systems*, *Commun. Phys.* **3**, 100 (2020).
- [41] S. A. Rather, S. Aravinda, and A. Lakshminarayan, *Creating Ensembles of Dual Unitary and Maximally Entangling Quantum Evolutions*, *Phys. Rev. Lett.* **125**, 070501 (2020).
- [42] D. Goyeneche, D. Alsina, J. I. Latorre, A. Riera, and K. Życzkowski, *Absolutely Maximally Entangled States, Combinatorial Designs, and Multiunitary Matrices*, *Phys. Rev. A* **92**, 032316 (2015).
- [43] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, *Holographic Quantum Error-Correcting Codes: Toy Models for the Bulk/Boundary Correspondence*, *J. High Energy Phys.* **06**, 149 (2015).
- [44] L. Piroli, B. Bertini, J. I. Cirac, and T. Prosen, *Exact Dynamics in Dual-Unitary Quantum Circuits*, *Phys. Rev. B* **101**, 094304 (2020).
- [45] S. Gopalakrishnan and A. Lamacraft, *Unitary Circuits of Finite Depth and Infinite Width from Quantum Channels*, *Phys. Rev. B* **100**, 064309 (2019).

- [46] L. Piroli and B. Bertini, *Scrambling in Random Unitary Circuits: Exact Results*, *Phys. Rev. B* **102**, 064305 (2020).
- [47] P. W. Claeys and A. Lamacraft, *Maximum Velocity Quantum Circuits*, *Phys. Rev. Research* **2**, 033032 (2020).
- [48] Another choice in taking the thermodynamic limit would be to hold n_A fixed and let $n_B \rightarrow \infty$. In such a case, one would probe the local rather than the global thermalization and would have strong kinematic effects. Namely, for a typical state in a large Hilbert space, the reduced density operator ρ_A will be very close to an identity [49]. For instance, tracing a random state over $n_B \gg n_A$ results in the spectrum of ρ_A whose relative deviations from a flat one are negligible, scaling as $\sim q^{-(n_B - n_A)/2}$ [50]. Such a measure concentration can swamp out any dynamical effects (e.g., integrable vs chaotic evolution).
- [49] V. M. Kendon, K. Życzkowski, and W. J. Munro, *Bounds on Entanglement in Qudit Subsystems*, *Phys. Rev. A* **66**, 062310 (2002).
- [50] M. Žnidarič, *Entanglement of Random Vectors*, *J. Phys. A* **40**, F105 (2007).
- [51] Often it is incorrectly stated that a one 2-qubit gate can increase entanglement of a separable state by $\ln q$ (i.e., by 1 ebit). That is not true—one 2-qudit gate acting on a bipartite separable state can produce maximal entanglement of $2 \ln q$ [52]. In a special case of acting on a separable state of 2-qubits, the maximum is indeed 1 ebit; however, if one has $2 + 2$ (or more qubits), the maximum is 2 ebits. Therefore, there are three possible sources of confusion (sometimes misplaced in the literature): (i) a factor of 2 in the number of cuts between the OBC (1 boundary link connecting subsystems A and B) and PBC (2 boundary links), (ii) how time is counted—one full BW is counted as $t = 1$ (i.e., n gates per unit of time, which is the unit we use), or as $\tau = 2$ ($n/2$ gates per $\tau = 1$), and (iii) one 2-qudit gate can increase entanglement of a bipartite separable state maximally by $2 \ln q$.
- [52] C. H. Bennett, A. W. Harrow, D. W. Leung, and J. A. Smolin, *On the Capacities of Bipartite Hamiltonians and Unitary Gates*, *IEEE Trans. Inf. Theory* **49**, 1895 (2003).
- [53] B. Bertini, P. Kos, and T. Prosen, *Operator Entanglement in Local Quantum Circuits I: Chaotic Dual-Unitary Circuits*, *SciPost Phys.* **8**, 067 (2020).
- [54] N. Khaneja, R. Brockett, and S. J. Glaser, *Time Optimal Control in Spin Systems*, *Phys. Rev. A* **63**, 032308 (2001).
- [55] B. Kraus and J. I. Cirac, *Optimal Creation of Entanglement Using a Two-Qubit Gate*, *Phys. Rev. A* **63**, 062309 (2001).
- [56] M. Blaauboer and R. L. de Visser, *An Analytical Decomposition Protocol for Optimal Implementation of Two-Qubit Entangling Gates*, *J. Phys. A* **41**, 395307 (2008).
- [57] E. Lubkin, *Entropy of an n -System from Its Correlation with a k -Reservoir*, *J. Math. Phys. (N.Y.)* **19**, 1028 (1978).
- [58] P. Zanardi, C. Zalka, and L. Faoro, *Entangling Power of Quantum Evolutions*, *Phys. Rev. A* **62**, 030301(R) (2000).
- [59] Y. Makhlin, *Nonlocal Properties of Two-Qubit Gates and Mixed States, and the Optimization of Quantum Computations*, *Quantum Inf. Process.* **1**, 243 (2002).
- [60] J. Zhang, J. Vala, S. Sastry, and K. B. Whaley, *Geometric Theory of Nonlocal Two-Qubit Operations*, *Phys. Rev. A* **67**, 042313 (2003).
- [61] The bipartite entangling power $e(W)$ is defined [58] as $e(W) = 1 - \langle I(W|\psi_A)\psi_B \rangle$, where the averaging is over the Haar measure on $U(2^{n_A}) \otimes U(2^{n_B})$ of initial bipartite product states; see, e.g., Ref. [62] and references therein for studies of $e(U)$. However, in random circuits, one is typically interested in fully factorized initial product states [$U(2) \otimes U(2) \otimes \dots \otimes U(2)$] and not bipartite product initial states. While one can generalize the entangling power with respect to multipartite product initial states, e.g., Ref. [63], in random circuits, things are further complicated by having random one-site unitaries at each step of a protocol (not just at the beginning) and the fact that the unitary propagator for the whole circuit depends, in a nontrivial way, on the number of layers (time) and the 2-qubit gate W . While the entangling power has been of use in small systems, e.g., 2 qubits, where exact results are simple to obtain, it is not clear how to make use of it (go beyond the Markovian techniques that we use) in the TDL and a multipartite setting.
- [62] B. Jonnadula, P. Mandayam, K. Życzkowski, and A. Lakshminarayan, *Entanglement Measures of Bipartite Quantum Gates and Their Thermalization under Arbitrary Interaction Strength*, *Phys. Rev. Research* **2**, 043126 (2020).
- [63] A. J. Scott, *Multipartite Entanglement, Quantum-Error-Correcting Codes, and Entangling Power of Quantum Evolutions*, *Phys. Rev. A* **69**, 052330 (2004).
- [64] S.-H. Lin, R. Dilip, A. G. Green, A. Smith, and F. Pollmann, *Real- and Imaginary-Time Evolution with Compressed Quantum Circuits*, *PRX Quantum* **2**, 010342 (2021).
- [65] S. Gopalakrishnan and M. J. Gullans, *Entanglement and Purification Transitions in Non-Hermitian Quantum Mechanics*, *Phys. Rev. Lett.* **126**, 170503 (2021).
- [66] G. Vidal, *Efficient Classical Simulation of Slightly Entangled Quantum Computations*, *Phys. Rev. Lett.* **91**, 147902 (2003).
- [67] C. Labbe and H. Lacoïn, *Cutoff Phenomenon for the Asymmetric Simple Exclusion Process and the Biased Card Shuffling*, *Ann. Probab.* **47**, 1541 (2019).
- [68] Stacking odd- n consecutive S configurations, we can write it in terms of a BW configuration as $[S_1^{n-1}]^n = L[B_1^{n-1}]^{(n+3)/2}U$, where L is a “lower triangle” block of gates and U an “upper triangle” of gates, each containing $(n-1)(n-3)/2$ gates; in the TDL, disregarding L and U would imply $\lambda_2(S) = \sqrt{\lambda_2(\text{BW})}$.
- [69] We conjecture that in the TDL and for Haar-random $U(4)$ gates in the BW configuration with OBC and half-half symmetric bipartition purity at any finite fixed t is equal to $I(t) = (16/25)^t$. This agrees with the asymptotic decay $(4/5)^{2t}$ obtained in Refs. [4,5] [the explicit finite- n expression in Eq. (25) of Ref. [4], however, is not equal to this result because one has to correctly account for boundary conditions].
- [70] D. A. Rowlands and A. Lamacraft, *Noisy Coupled Qubits: Operator Spreading and the Fredrickson-Andersen Model*, *Phys. Rev. B* **98**, 195125 (2018).
- [71] T. Zhou and A. Nahum, *Entanglement Membrane in Chaotic Many-Body Systems*, *Phys. Rev. X* **10**, 031066 (2020).
- [72] P. Diaconis, *The Cutoff Phenomenon in Finite Markov Chains*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1659 (1996).

- [73] M.J. Kastoryano, D. Reeb, and M.M. Wolf, *A Cutoff Phenomenon for Quantum Markov Chains*, *J. Phys. A* **45**, 075307 (2012).
- [74] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Time* (AMS, 2017).
- [75] P. Facchi, U. Marzolino, G. Parisi, S. Pascazio, and A. Scardicchio, *Phase Transitions of Bipartite Entanglement*, *Phys. Rev. Lett.* **101**, 050502 (2008).
- [76] Vinayak and M. Žnidarič, *Subsystem Dynamics under Random Hamiltonian Evolution*, *J. Phys. A* **45**, 125204 (2012).
- [77] P.-Y. Chang, X. Chen, S. Gopalakrishnan, and J.H. Pixley, *Evolution of Entanglement Spectra under Generic Quantum Dynamics*, *Phys. Rev. Lett.* **123**, 190602 (2019).
- [78] J. Cotler, N. Hunter-Jones, and D. Ranard, *Fluctuations of Subsystem Entropies at Late Times*, [arXiv:2010.11922](https://arxiv.org/abs/2010.11922).
- [79] H. J. Lipkin, N. Meshkov, and A. J. Glick, *Validity of Many-Body Approximation Methods for a Solvable Model: Exact Solutions and Perturbation Theory*, *Nucl. Phys.* **62**, 188 (1965).