Separation of Out-Of-Time-Ordered Correlation and Entanglement

Aram W. Harrow,[1,*] Linghang Kong[1,†] Zi-Wen Liu[2,‡] Saeed Mehraban,[3,§] and Peter W. Shor[1,4,¶]

[1]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[2]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*

[3]*Institute for Quantum Information and Matter, Caltech, Pasadena, California 91125, USA*

[4]*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

The out-of-time-ordered correlation (OTOC) and entanglement are two physically motivated and widely used probes of the "scrambling" of quantum information, a phenomenon that has drawn great interest recently in quantum gravity and many-body physics. We argue that the corresponding notions of scrambling can be fundamentally different, by proving an asymptotic separation between the time scales of the saturation of OTOC and that of entanglement entropy in a random quantum-circuit model defined on graphs with a tight bottleneck, such as tree graphs. Our result counters the intuition that a random quantum circuit mixes in time proportional to the diameter of the underlying graph of interactions. It also provides a more rigorous justification for an argument in our previous work [Shor P.W., Scrambling time and causal structure of the photon sphere of a Schwarzschild black hole, arXiv:1807.04363 (2018)], that black holes may be slow information scramblers, which in turn relates to the black-hole information problem. The bounds we obtain for OTOC are interesting in their own right in that they generalize previous studies of OTOC on lattices to the geometries on graphs in a rigorous and general fashion.

I. INTRODUCTION AND OVERVIEW

The "scrambling" of quantum information is a phenomenon of fundamental importance, deeply connected to many important research topics in physics, such as black holes [1–4] and many-body chaos [5,6]. In recent years, a great amount of research effort has been devoted to the detection and characterization of scrambling. The so-called out-of-time-ordered correlation (OTOC) [7] is a commonly used measure of quantum chaos and scrambling. A variant based on commutators (also known as the OTO commutator) is given by

$$C(t) = \frac{1}{2}\langle [O_1(x,0), O_2(y,t)]^\dagger [O_1(x,0), O_2(y,t)]\rangle, \quad (1)$$

* aram@mit.edu

† linghang@mit.edu

‡ zliu1@perimeterinstitute.ca

§ mehraban@mit.edu

¶ shor@math.mit.edu

where $O_1(x,0)$ is an operator acting on site x, and $O_2(y,t)$ is a Heisenberg operator at time t that acts only on y at time 0, i.e., $O_2(y,t) = U^\dagger(t)O_2(y,0)U(t)$, where $U(t)$ is the unitary for the evolution from time 0 to t. The average is taken with respect to the thermal state at some temperature, which we take to be infinite in this work. Intuitively speaking, it characterizes parameters like sensitivity to initial conditions via the spread of local operators. Also notice that the scrambling phenomena exhibit a truly quantum nature—the state of the entire system remains pure during the unitary evolution (although it is effectively randomized), thus no information is really lost; the generation of global entanglement leads to the scrambling of initially localized quantum information, spreading and hiding it from observers that only have access to part of the system. This observation leads to another fundamental probe of a stronger form of scrambling, namely the entanglement between parts of the system [2,8–10] (which measures the equivalent effect as the tripartite information [9] in the case of unitary dynamics; see Ref. [10] for more detailed discussions).

To understand and characterize the dynamical behaviors of scrambling systems, several explicit models have been proposed and investigated, such as the Sachdev-Ye-Kitaev (SYK) model [11,12]. Another leading approach is the random quantum-circuit model (i.e., sequential applications of random local quantum gates), which is a

widely used model of chaotic dynamics that make minimal assumptions about the nature of the local degrees of freedom (see, e.g., Refs. [1,9,13–17]). They capture the key kinematic feature of chaos that the evolution appears to be random, and the locality of physical interactions. In these well-studied physical scrambling models, the saturation of OTOC and that of entanglement are expected to occur at a similar time scale [9,13,14]. More generally, one could consider the dynamics of many small quantum systems (say qubits) connected according to some graph [18,19], with random unitary gates being applied to each edge. Suppose that we apply gates in a random order such that on average each edge has one gate applied to it per unit time. A natural conjecture here, which would be compatible with all previous results, is that the scrambling time for a constant degree graph is proportional to its diameter, i.e., the maximum distance between any two vertices. This would correspond to information traveling through the graph at a linear velocity and is assumed implicitly in previous works. However, no proof exists, outside of the special case of Euclidean lattices in a fixed number of dimensions. Even for Euclidean lattices in more than one dimension, this result was only recently proven [20].

Our main results are the following. The first one (Theorems 1 and 2) shows that for arbitrary graphs with sufficiently low degree, the OTOC saturation time scales linearly in the graph diameter. Here by low degree, we mean $d^2 \geq z$ where d is the dimension of the quantum system and z is the degree of the graph. On the other hand, we use bounds on entanglement growth to show that the time needed to establish substantial entanglement between parts of the system scale at least as the number of vertices and thus could be longer than the OTOC saturation time, for graphs with bottlenecks (see Theorem 3). Such graphs include, e.g., binary trees, which we explicitly analyze in this paper, and discretizations of hyperbolic space around black holes, originally proposed by Ref. [21], which are expected to exhibit similar behaviors (as argued below). In other words, we establish an *asymptotic* separation between the time scales of OTOC and entanglement saturation. Recently, Refs. [22,23] studied scrambling on certain peculiar graphs via a Hamiltonian model, but the relations between OTOC and entanglement were not fully understood and the physical correspondences were not clear; Ref. [24] considered the relations between specific forms of OTOC and entanglement, but it focused on a system with a specific quantum optical Hamiltonian and the main message is about the connections rather than differences between OTOC and entanglement. Here we make key progress in the understanding of scrambling measures by rigorously demonstrating the fundamental difference between OTOC and entanglement in a general setting and studying the implications to physics.

Our results have the following major implications.

(i) *Scrambling in non-Euclidean geometries.* Existing work mostly studied scrambling on Euclidean lattices [13,14,25]. The general assumption is that after time t, a localized perturbation will affect everything within some ball of radius $v_{\text{butterfly}}t$, where $v_{\text{butterfly}}$ is known as the "butterfly velocity," which characterizes the speed of information spreading. However, this has not been proved and previous works gave only heuristic arguments for it that included uncontrolled approximations. For the random circuit models defined on general graphs, we find that if the local dimension is large relative to the graph degree then indeed there is a linear butterfly velocity. We also discuss apparent counterexamples, which suggest that linear butterfly velocity no longer holds for high-degree graphs. More specifically, we give a heuristic argument that for tree graphs with high degree, the scrambling time could be exponential in the diameter. The star graph studied in Ref. [23] is also an example where the scrambling time grows much faster than the diameter. On the other hand, for complete graphs [26] the scrambling time is sublinear in the diameter. We also show that by replacing edges in a graph by some gadget with high degree, the scrambling time could be arbitrarily small.

(ii) *Black-hole information scrambling.* Our results can be regarded as a more rigorous argument that fleshes out the idea of a recent paper by one of the authors [21], which concerns whether it is possible for the fast scrambling conjecture of black holes [2] to hold if one assumes that the causality structure of general relativity holds around a black hole, and if the medium by which the information is scrambled is Hawking radiation. In the model of Ref. [21], the space around the black hole is divided into cells, each of which contains a constant number of bits of Hawking radiation. It then gives arguments for why the Hawking radiation is not adequate for fast scrambling if the entanglement definition of scrambling is used. The cell structure around the black hole looks like a patch of a cellulation of hyperbolic geometry, where the cells on the event horizon are the boundary of this patch. The tree graph we consider captures a key feature of this geometry: the leaves lie on the event horizon, and the density of nodes decreases as one moves outwards radially. As the assumptions essentially suggest that information is processed via local interactions of the Hawking radiation, we may consider a random circuit defined on the underlying graph to be a toy model that captures key features of the black-hole scrambling process. Our mathematical results then indicate that the scrambling time scales given by entanglement and OTOC are fundamentally different. Another way to interpret our model is that the information "wavefront" could reach the farthest side rather quickly since there exists short paths, but it takes a longer time, which scales with the number of degrees of freedom, to establish truly global entanglement. This is consistent with recent holographic

calculations (see, e.g., Refs. [27,28]), which suggest that the entanglement entropy grows roughly linearly after a quench in chaotic systems.

We would also like to remark upon the task of recovering quantum information falling into the black hole from Hawking radiation, which is often referred to as Hayden-Preskill decoding [1]. This task plays central roles in recent studies of the black-hole information problem. Yoshida and Kitaev recently proposed an explicit protocol [29] whose decoding fidelity is at least the order of $1/d_A^2[1 - C(t)]$, where d_A is the Hilbert-space dimension of the input message. Here $C(t)$ takes the form of Eq. (1) and considers O_1 and O_2 averaged over all Pauli operators on the infalling system and Hawking radiation, respectively; see Secs. 2–4 of Ref. [29] for details. By simple calculations one can see that our results imply a possible time window in which the decoding could be achieved with high fidelity without substantial entanglement when the infalling quantum state is sufficiently small compared to the black hole. However, it appears that adding a small number of qubits to a Schwarzschild black hole can only be done by photons whose wavelength is comparable to the size of the black hole. It does not seem surprising that the information carried by such photons can be extracted by a black hole quickly; when the information is absorbed by the black hole, it is already spread out over the entire black hole, and so does not need to migrate from a localized region to a state where it is delocalized on the black hole.

(iii) *Inequivalence of convergence to 2-designs in different measures.* The speed of convergence of a random circuit to a 2-design (distributions that approximately agree with the Haar measure up to the first two moments, which have found many important applications as an efficient approximation to Haar randomness [30]) has been the subject of a large amount of research. In particular, Refs. [20,26,31–33] show that the time of convergence depends on the graph of interactions, and suggest that it should be proportional to the diameter. Note that 2-designs are very powerful measures of convergence, in the sense that a distribution being close to a 2-design implies that the distribution has mixed with respect to not only OTOC but also von Neumann and Rényi-2 entanglement entropies [9,34], and other important signatures of information scrambling such as decoupling [35]. Our work provides several examples where a random circuit approximates the OTOC but not the entanglement properties of a 2-design, and therefore implies that a strong approximation of 2-designs (in terms of, e.g., the frame operator [10]) may not be achieved in time proportional to the diameter.

II. MODELS AND NOTATION

Let G be a graph with V vertices and E edges. The model we study consists of a graph with a d-dimensional Hilbert space associated with each vertex of G. Each edge has Haar-random unitary gates applied to qudits on its endpoints according to a Poisson process with rate 1, meaning a Poisson distribution such that k unitaries are applied in time t with probability $t^k e^{-t}/k!$. The mixing times for OTOC and entanglement, $\tau_{\text{OTOC}}^{(x,y)}$ and $\tau_{\text{ent}}^{(A)}$, are defined as follows.

Definition 1. $\tau_{OTOC}^{(x,y)}$ *(respectively, $\tau_{ent}^{(A)}$) is defined to be the minimum amount of time needed for OTOC between vertices x and y (respectively, the entanglement entropy between A and the complement of A) to become at least a constant fraction of its equilibrium value. In this work we take the constant to be $1/(d^2 + 1)$.*

Here we expect that a qualitatively similar behavior will hold with $1/(d^2 + 1)$ replaced by any constant strictly between 0 and 1. We study how τ_{OTOC} and τ_{ent} scale with parameters such as local dimension, degree, and number of vertices.

We study the pair of (x,y) that has the largest $\tau_{\text{OTOC}}^{(x,y)}$, and the set A that has the largest $\tau_{\text{ent}}^{(A)}$, as they could best characterize OTOC and entanglement properties for G.

Instead of studying this model directly we may consider the process in which a random edge is picked every $1/E$ time units. This is because in our Poisson process model, each edge is equally likely to be picked. The number of unitaries applied within time t is of order Et (see Appendix A), so the two models above are equivalent up to a constant factor.

III. OTOC

To analyze the saturation time of OTOC, we formulate the process of operator spreading as a Markov chain. Consider an arbitrary Pauli operator $\sigma_{\vec{p}}$ acting on n d-dimensional qudits, $\vec{p} \in \{0, \ldots, d^2 - 1\}^n$, and apply some unitary U to it. We expand the resulting operator on Pauli basis and have

$$U^\dagger \sigma_{\vec{p}} U = \sum_{\vec{q}} \alpha_{\vec{q}} \sigma_{\vec{q}}, \quad \alpha_{\vec{q}} \equiv \frac{1}{d^n} \text{Tr}[U^\dagger \sigma_{\vec{p}} U \sigma_{\vec{q}}^\dagger].$$

The expected value of the cross term for $\alpha_{\vec{q}}$ averaged over the distribution of U is

$$\mathbb{E}_U \alpha_{\vec{q}} \alpha_{\vec{q}'}^* = \frac{1}{d^{2n}} \mathbb{E}_U \text{Tr}[U^{\dagger \otimes 2}(\sigma_{\vec{p}} \otimes \sigma_{\vec{p}}^\dagger) U^{\otimes 2}(\sigma_{\vec{q}}^\dagger \otimes \sigma_{\vec{q}'})].$$

According to the construction of random circuit, this is zero for $\vec{q} \neq \vec{q}'$ for U being the unitary in a single step. Therefore, in each step the values of $\overline{\alpha_{\vec{q}} \alpha_{\vec{q}}^*}$ undergo linear transformation, which we can interpret as a distribution because they are positive and sum to 1.

If we start from a Pauli operator located at a single vertex x, on each vertex all nonidentity Pauli operators will

have the same probability as long as x has been touched at least once in the process. So we care only if the operator on a vertex is identity (I) or nonidentity (N). And the norm of the time-evolved operator with a Pauli operator P on some vertex y would be proportional to the probability of having nonzero Pauli operator on that site, and the factor of proportionality is

$$\frac{1}{d^2 - 1} \sum_{i=1}^{d^2-1} \frac{1}{2d} \text{Tr}([\sigma_i, P]^\dagger [\sigma_i, P]) = \frac{d^2}{d^2 - 1},$$

which is just the commutator averaged over all nonidentity Pauli operators.

In summary, the object we study is the OTOC between Pauli operator on vertex y and time-evolved Pauli operator on vertex x after T steps of random circuits on graph G, which is equal to $d^2/(d^2 - 1)$ times the probability of having "N" on vertex y after T steps in the Markov chain M_0 defined below. The state space of M_0 is the set of all the configurations in which each vertex of G is assigned a label "N" or "I." The initial state of M_0 has "N" assigned to vertex x and "I" assigned to all other vertices. The update rule is that in each step a uniformly random edge is picked and the labels on the two corresponding vertices are updated. "II" remains "II", and otherwise they have a probability of $(d^2 - 1)/(d^4 - 1) = 1/(d^2 + 1)$ for becoming "IN" or "NI" each, and $(d^2 - 1)/(d^2 + 1)$ for becoming "NN" [26].

Now we prove an upper bound for the OTOC saturation time. Note that $O(\alpha)$ that appears here and in the following represents a quantity that scales asymptotically as α, i.e., a quantity between $c_1 \alpha$ and $c_2 \alpha$ for some constants c_1, c_2 such that $0 < c_1 < c_2$.

Theorem 1 (OTOC upper bound). *Let G be a graph with V vertices and E edges, and suppose the degree for each vertex at most d^2, where d is the Hilbert-space dimension for each vertex. Then for any pair of vertices x and y, $\tau_{OTOC}^{(x,y)} = O[D(x,y)]$ with high probability, where $D(x,y)$ is the distance between x and y. The probability of failure is exponentially small in $D(x,y)$. As a consequence the perfect binary tree has $\tau_{OTOC}^{(x,y)} = O(\ln V)$, where x and y are the farthest pair of vertices.*

Proof. The proof of this theorem follows from Lemmas 1 and 2 below. As explained earlier, the OTOC saturation time corresponds to the number of steps needed for M_0 to have constant probability of having a label "N" on y.

We first prove Lemma 1, which states that with probability $1 - e^{-O[D(x,y)]}$ the vertex y gets hit by a label "N" within order of $E \times D(x,y)$ steps. As shown in Appendix A, this needs order of $D(x,y)$ time units with high probability. Then we show in Lemma 2 that after this happens, the probability for having an "N" on y remains constant. ∎

Lemma 1. *Suppose that G is a graph with the degree for each vertex being at most d^2. For any pair of vertices x and y with distance $D(x,y)$, the expected number of steps for y to be labeled "N" is of order $E \times D(x,y)$ in M_0 starting from x. Besides, with high probability the vertex y gets labeled "N" in time of order $E \times D(x,y)$. The probability of failure is exponentially small in $D(x,y)$.*

To set up the proof, we first construct a Markov chain M, which has the same initial state as M_0, and in each step the update rule of M is applied, followed by changing all "N" into "I" except the one closest to vertex y. By a simple coupling argument the number of steps needed for y to get an "N" in M_0 is lower bounded by that in M. The distance between the vertex with label "N" and vertex y in Markov chain M_0 can be described by a biased random walk, from which we can obtain the desired bound.

Definition 2 (Markov chain M). *Markov chain M has the same state space and initial state as M_0. In each step the update rule for M_0 is applied, followed by setting all "N" labels into "I" except for the one closest to y (choose randomly if this is not unique).*

In this way the vertex with label "N" in M is always labeled "N" in M_0 in the most natural way of coupling M_0 to M, and therefore after any number of steps the probability that vertex is labeled "N" in M_0 is lower bounded by the corresponding probability in M.

The Bernstein inequality is needed for the proof of our theorem, which states that for independent zero-mean random variables $X_1, \ldots, X_n$ each with absolute value at most M,

$$\Pr\left[\sum_{i=1}^n X_i > t \right] \leq \exp\left(-\frac{(1/2)t^2}{\sum \mathbb{E}[X_i^2] + (1/3)Mt} \right).$$

This could be generalized to the case with nonzero mean. Suppose $Y_1, \ldots, Y_n$ has mean $\mu_1, \ldots, \mu_n$ and they satisfy $|Y_i - \mu_i| \leq M$ then by setting $X_i = Y_i - \mu_i$ we have

$$\Pr\left[\sum_{i=1}^n Y_i > t + \sum_{i=1}^n \mu_i \right]$$
$$\leq \exp\left(-\frac{(1/2)t^2}{\sum(\mathbb{E}[Y_i^2] + \mu_i^2) + (1/3)Mt} \right). \quad (2)$$

Now we give the proof of Lemma 1.

Proof. In this proof we first derive the transition rule of Markov chain M_0. We see that the distance between y and the vertex labeled "N" could only change by 1 in each step, the probability of decreasing is larger than that of increasing. Then by Bernstein in equality in Eq. (2) we can show

an upper bound for the number of steps needed for this distance to decrease to 0.

We would first restate the transition rule of M_0 here. For Markov chain M_0, an edge (u, v) is selected for each step. Suppose the labels on u and v are both "I," they remain unchanged. Otherwise the labels become "IN" or "NI" with probability $1/(d^2 + 1)$ each, and "NN" with probability $(d^2 - 1)/(d^2 + 1)$.

As mentioned, the upper bound for M defined in Definition 2 gives an upper bound for M_0. For Markov chain M_0, let v be the vertex with label "N." As long as $v \neq y$, there will be at least one neighbor u that is one step closer to y, and other neighbors are at most one step further from y due to the triangle inequality. If the edge (u, v) is selected, there is a chance of $d^2/(d^2 + 1)$ that u obtains label "N" according to the transition rule of M_0. In this case the distance between y and the closest "N" is one step shorter. If the edge between u and other neighbors is selected, there is a chance of $1/(d^2 + 1)$ that the label on u becomes "I," and the distance between y and the closest label "N" becomes one step longer. Let the degree of u be d_u, and the distance between the label "N" and vertex y will have probability of at least $(1/E)d^2/(d^2 + 1)$ to decrease by 1 and at most $(d_u - 1)/E[1/(d^2 + 1)]$ to increase by 1, where the probabilities depend on the specific vertex. Since the degree for any vertex is at most d^2, the time needed for the distance to drop from $D(x, y)$ to 0 is upper bounded by the time in the following biased random walk W. W has states $\{0, 1, \ldots, d_{max}\}$ where d_{max} is the maximum possible distance to y, and starting from vertex $D(x, y)$ it has a fixed probability of $d^2/[E(d^2 + 1)]$ for decreasing by 1 and $(d^2 - 1)/[E(d^2 + 1)]$ for increasing by 1.

Extension of this finite chain to an infinite one could increase only the hitting time of vertex 0, because the finiteness at vertex d_{max} prevents us from getting too far from vertex 0. The displacement of a random walk on an infinite chain (i.e., the difference of the final position and initial position) is the sum of displacement for each step, which has probability $d^2/[E(d^2 + 1)]$ of being -1 and $(d^2 - 1)/[E(d^2 + 1)]$ probability of being $+1$, and otherwise it is 0. The mean and variance for displacement at each step is

$$\mu_0 = -\frac{1}{E(d^2 + 1)}, \quad \sigma_0^2 = \frac{2d^2 + 1}{E(d^2 + 1)} - \mu_0^2.$$

The expected number of steps needed to reach vertex 0 in this random walk is $-D(x, y)/\mu_0 = \Theta[E \times D(x, y)]$. We can also use Eq. (2) to bound the probability that the total displacement of T steps is larger than $-D(x, y)$, where we set T to be twice the expected number of steps needed and $t = D(x, y)$. M can be set to be 2. The denominator in the exponent will be $T(\sigma_0^2 + 2\mu_0^2) + \frac{2}{3}t = \Theta[D(x, y)]$,

so Eq. (2) gives a probability of at most $e^{-\Theta[D(x,y)]}$ for not reaching vertex 0. ∎

Lemma 2. *After a label "N" reaches the target vertex y, the probability for having an "N" on y will remain order one.*

We again consider the modified chain, which keeps only one label "N" after each step. We show that vertex y has constant probability of having label "N" in the equilibrium distribution. This probability is monotonically nonincreasing as a function of the number of steps, so the probability is order one in any step.

To give a full proof of this lemma, we need the following lemma.

Lemma 3. *Consider a reversible Markov chain M with transition matrix $P(x, y)$, $x, y \in \Omega$. M starts deterministically from state $x_0 \in \Omega$. If all the eigenvalues of P are non-negative, then the probability for x_0 will be monotonically nonincreasing as a function of the number of steps.*

Proof. Let $\pi(x)$ be the stationary distribution. The reversibility implies that $A(x, y) \equiv \sqrt{\pi(x)/\pi(y)}P(x, y)$ is a symmetric matrix, and therefore has orthonormal eigenvectors $f_k(x)$ with corresponding eigenvalues λ_k. The eigenvectors for $P(x, y)$ will then be $g_k(x) = f_k(x)/\sqrt{\pi(x)}$. Now we want to expand the initial distribution $p_0(x) = \delta_{x,x_0}$ in terms of $g_k(x)$,

$$p_0(x) = \sum_k \alpha_k g_k(x), \quad \alpha_k = \pi(x_0)g_k(x_0),$$

which can be verified using the orthogonality of f_k. After t steps, the probability distribution would be

$$p_t(x) = \sum_k \alpha_k \lambda_k^t g_k(x),$$

and the probability for state x_0 would be

$$p_t(x_0) = \sum_k \alpha_k \lambda_k^t g_k(x_0) = \sum_k \pi(x_0)g_k(x_0)^2 \lambda_k^t,$$

which is a monotonically nonincreasing function of t given λ_k are all non-negative. ∎

Now we can prove Lemma 2.

Proof. We again keep track only of the label "N" closest to y. When it is at y, there is a probability of $d_y/E[1/(d^2 + 1)]$ that an "I" is left on y and the closest "N" becomes one of the neighbors of y. Here, d_y

is the degree of y. Otherwise, suppose it is at a vertex u with degree d_u. There is at least one neighbor of u that is one step closer to y, and other neighbors are at most one step farther. This corresponds to a probability of $1/E[d^2/(d^2+1)]$ to move one step closer to y, and $(d_u - 1)/E[1/(d^2+1)]$ to move one step farther.

Then we need only to consider a random walk on the states $\{0, 1, \ldots, d_{\max}\}$, where $d_{\max}$ is the largest possible distance from y. At state 0 there is a chance of $d^2/[E(d^2+1)]$ moving to 1, and at other state k the probabilities for moving to $k+1$ and $k-1$ are $(d^2-1)/[E(d^2+1)]$ and $d^2/[E(d^2+1)]$ respectively. From Ref. [36] we can get the probability for state 0 in the stationary distribution, which is

$$\frac{1}{1 + \sum_{t=0}^{d_{\max}-1} \left[(d^2-1)/d^2 \right]^t} > \frac{1}{d^2+1},$$

which is $\Omega(1)$. Note that at large E, the probability for staying at the same state is larger than $\frac{1}{2}$, which means that all eigenvalues are positive. Also the random walk on a finite chain is reversible, so by Lemma 3, the probability for having "N" on t remain $\Omega(1)$. ∎

Theorem 1 states that the number of steps needed for OTOC saturation in a low-degree graph is at most of order $E \times D(x, y)$. However, we expect that in a graph with high degree, the number could be much larger. Some intuitions are given in Appendix C.

Besides this upper bound we also derive a lower bound for OTOC saturation time.

Theorem 2 (OTOC lower bound). *Let G be a graph with V vertices and E edges, and suppose the degree for each vertex is $O(1)$. Then for any pair of vertices x and y, $\tau_{OTOC}^{(x,y)}$ is at least of order $D(x, y)$ with high probability, where $D(x, y)$ is the distance between x and y. The probability of failure is exponentially small in $D(x, y)$.*

Proof. Consider a Markov chain M' on all configurations in which every vertex of G is assigned labels "N" or "I." Initially x is assigned "N" and other vertices are assigned "I." In each step a random edge is picked and the vertices it connects are changed to "N" if at least one of the vertices is previously assigned "N." Using a simple coupling argument the number of steps needed for y to be assigned "N" gives a lower bound for $\tau_{OTOC}^{(x,y)}$. To show a lower bound for the number of steps needed in M', we prove an even stronger statement that with probability at least $1 - e^{-\Theta(d)}$, all of the vertices with distance d to vertex x (denoted by S_d) have label "I" after $\Theta(E \times d)$ steps.

We start with a Chernoff-type bound for a sum of k independent identically distributed (IID) geometric variables. Let $X = X_1 + \cdots + X_k$ where X_i has probability of

$p(1-p)^{m-1}$ for value m, $m \geq 1$. For any $t > 0$, we have $\mathbb{E}e^{-tX_i} = p/[e^t - (1-p)]$, and therefore

$$\mathbb{E}e^{-tX} = \left(\frac{p}{e^t - (1-p)} \right)^k.$$

The mean for X is $\mu \equiv \mathbb{E}X = k/p$. By Markov inequality,

$$\begin{aligned}
\Pr[X \leq \lambda\mu] &= \Pr[e^{-tX} \geq e^{-t\lambda\mu}] \\
&\leq e^{t\lambda k/p} \mathbb{E}e^{-tX} \\
&= \exp\left[k \left(\ln \frac{p}{e^t - (1-p)} + \frac{t\lambda}{p} \right) \right],
\end{aligned}$$

for any $t > 0$. When $\lambda > p$, we take $t = \ln[(1-p)\lambda]/(\lambda - p)$ and get $\Pr[X < \lambda\mu] \leq \exp[-kf(p, \lambda)]$ where

$$f(p, \lambda) = -\ln \frac{\lambda - p}{1 - p} - \frac{\lambda}{p} \ln \frac{\lambda(1-p)}{\lambda - p}.$$

In Markov chain M', the number of steps needed for spreading labels "N" along a path with length d is a sum of d geometric variables with $p = 1/E$. Let C be the max degree of the vertices. From vertex x to vertices in S_d, there could be at most $(C-1)^d$ paths with length d, so by union bound the probability that any vertex in S_d gets label "N" after λdE steps is at most $(C-1)^d \exp[-df(p, \lambda)]$. Knowing that $f(p, \lambda) = -\log \lambda + \lambda - 1 + O(p)$ can be arbitrarily large for small constant λ, we can pick λ such that the probability becomes $e^{-\Theta(d)}$. By Appendix A, the λdE steps here correspond to $\Theta(\lambda d) = \Theta(d)$ time units. ∎

IV. ENTANGLEMENT

We now turn our attention to entanglement. Here we need only to consider the case where the evolution is unitary and the system is pure. As a standard entanglement measure, the entanglement entropy of a pure state $|\psi\rangle_{AB}$ is given by $E(|\psi\rangle) := S(\rho_A)$ where $\rho_A = \text{Tr}_B[|\psi\rangle\langle\psi|]$ and S is the von Neumann entropy. Notice the following simple, general fact.

Lemma 4. *Let U_{AB} be a unitary operator acting on two d-dimensional systems AB. Then for any $|\psi\rangle_{AA'BB'}$ with ancilla systems A', B',*

$$E[(U_{AB} \otimes \text{id}_{A'B'})|\psi\rangle_{AA'BB'}] - E(|\psi\rangle_{AA'BB'}) \leq 2\log d.$$

Proof. Adapted from the proof of Lemma 1 of Ref. [37]. Suppose Alice holds AA' and Bob holds BB'. In addition, they share two copies of the maximally entangled state $|\Phi_d\rangle = (1/\sqrt{d}) \sum_{i=1}^{d} |i\rangle|i\rangle$, $E(|\Phi_d\rangle) = \log d$. Consider the following double teleportation protocol. Alice consumes

a $|\Phi_d\rangle$ and classical communication to teleport A to Bob, who performs U locally and then consumes a $|\Phi_d\rangle$ and classical communication to teleport system A back to Alice. The protocol is LOCC, under which the entanglement entropy between Alice and Bob is monotonically nonincreasing. Therefore, by the additivity of S (and thus E) on tensor products,

$$E(|\psi\rangle_{AA'BB'}) + 2E(|\Phi_d\rangle) \geq E[(U_{AB} \otimes \mathrm{id}_{A'B'})|\psi\rangle_{AA'BB'}],$$

and so the claimed bound follows. ∎

Note that the proof also applies to, e.g., the Rényi-2 entropy, which is a variant of the entanglement entropy that can be more easily measured in experiments [38,39].

By Lemma 4, the entanglement entropy between the two trees increases by at most $2\log d$ when the random unitary is acted across the middle edge. This edge has only a probability of $1/E \sim 1/V$ to be selected in each step. So in order to reach the maximum entropy of order $V\log d$, we need at least an order of V^2 steps or equivalently order V time. This is much larger than the OTOC time of order $\log V$.

From Lemma 4 we obtain the following result for a general graph.

Theorem 3 (Entanglement lower bound). *For a general graph G with vertices partitioned into sets A and B, the expected entanglement saturation time is at least of order $\min\{|A|, |B|\}/[C(A,B)]$, where $C(A,B)$ is the number of edges with one endpoint in A and one in B.*

Since approximate 2-designs exhibit nearly maximal entanglement [9,34], this result also provides lower bounds on the time it takes for the random circuit to converge to 2-designs.

V. OTOC VERSUS ENTANGLEMENT IN DIFFERENT MODELS

According to our results, a simple graph that can give a separation of OTOC and entanglement saturation times is a perfect binary tree, as depicted in Fig. 1. Here we consider the OTOC of operators located on the pair of farthest vertices in the graph (a leaf vertex of left subtree and a leaf vertex of the right subtree), and the entanglement entropy between left and right subtrees. The cell structure roughly equivalent to the hyperbolic geometry in three dimensions, or indeed any constant number of dimensions, exhibits such a separation as well. These graphs are regarded as toy models of quantum-information scrambling around black holes [21], as motivated in the introduction; see also Appendix B.

The behaviors of OTOC and entanglement on some other graphs are also studied. These include the following: (i) the "dumbbell graph" consisting of two complete
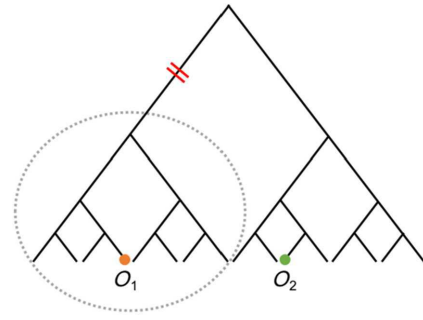

FIG. 1. Illustration of the binary-tree model of depth 4. We consider OTOC between local operators originally acting on two farthest vertices (a leaf of the left and right subtrees, respectively, for example, O_1 and O_2 in the diagram), and entanglement between the left subtree (dashed circle) and the rest of the graph (the cut shown by the red double line).

graphs connected by a bottleneck edge. A careful analysis could still show a separation between OTOC and entanglement; (ii) high-degree graphs. We demonstrate two examples in which the OTOC saturation time can be much longer or shorter than the bound we have on low-degree graphs. To be more precise, for a tree with high degree, the OTOC time could be exponential in the diameter of the graph. Another example is the star graph, which has constant diameter but the OTOC time is logarithmic in the number of vertices [23]. On the other hand, a complete graph has vanishing OTOC time when the graph size grows large, while the diameter remains constant. Another example is that by replacing edges in a graph by gadgets with arbitrarily large degree, the OTOC saturation time could be made arbitrarily small. See Sec. VI and Appendix C for more details. Also note that there is no separation on Euclidean lattices. These results are summarized in Table I.

Other than the Poisson process on each edge, different orders of choosing the edges are also studied below in Sec. VI.

TABLE I. Comparison of the time scales of OTOC and entanglement saturation in various graphs. n is the total number of vertices in the graph.

Models and graphs	OTOC	Entanglement
Euclidean lattices in D dimensions	$n^{1/D}$	$n^{1/D}$
Hyperbolic space with $D = 3$	$\log n$	$\sqrt{n}$[a]
Binary tree	$\log n$	n
Tree with degree $z \gg d^2$	$n^{1-(\log d^2/\log z)}$	n/z
Dumbbell graph	$\log n/n$	n

[a]See Sec. 4 of Ref. [21].

VI. VARIANTS OF OUR MODEL

As mentioned in the main text, one may be concerned that the slow entanglement saturation is due to the low chance of picking the cut edge in each step. Here we first consider a naturally modified model where the probability of picking the cut edge is much higher, albeit the OTOC and entanglement times remain separated. In this model we still apply Haar-random unitaries to vertices connected by randomly chosen edges. Instead of picking a random edge from all edges, in each step we pick a random edge from the left subtree (L), and then the cut edge (C), and finally a random edge from the rest of the edges (R). In this *LCR* model, the analysis for OTOC does not change by much and still has the $O(V \log V)$ upper bound. We argue that the entanglement saturation time is still $\Omega(V^2)$. Consider the sequence of picking edges $L_{i_1} C R_{j_1} L_{i_2} C R_{j_2} \ldots$ The probability that the random edge on L or R is connected to the cut edge is $\Theta(1/V)$. So the probability that neither R_{j_k} nor $L_{i_{k+1}}$ is connected to C is $1 - \Theta(1/V)$, in which case $R_{j_k} L_{i_{k+1}}$ commute with the neighboring C, and the chain reads $\ldots L_{i_k} C C R_{j_k} L_{i_{k+1}} R_{j_{k+1}} \ldots$, where the two C gates reduce to one effective C gate. The probability that such commutation does not happen is $\Theta(1/V)$, which means that the expected number of effective C gates in constant time steps is $\Theta(1/V)$. So the expected saturation time is $\Omega(V^2)$.

We can also consider two complete graphs connected by a cut edge. Let V be the number of vertices, and the number of edges is $E = \Theta(V^2)$. We still have a similar Markov chain describing the evolution of OTOC. In the *LCR* model for this graph, by Ref. [26] the complete graph on the left reaches equilibrium in $O(\log V)$ time. Then after a constant number of steps there will be a label "N" propagated to the right, and another $O(\log V)$ steps will be needed for OTOC to equilibrate. On the other hand, $\Omega(V)$ steps will be needed for entanglement. If a random edge is picked in each step, we expect the OTOC time scales to stay the same. In this case, though, the entanglement takes time $\Omega(E) = \Omega(V^2)$ to saturate.

Besides the two ways of picking random edges, we can also pick a random matching in each step (assuming the number of vertices on each side is even), followed by picking the cut edge. The entanglement entropy takes $\Omega(V)$ steps to saturate. For OTOC, the number of edges connecting a vertex with label "N" and a vertex with label "I" in a random matching should be proportional to the total number of "N," so we expect the number of vertices with "N" to grow exponentially. Therefore, it would take $O(\log V)$ steps for an "N" to go across the middle edge, and a total number of $O(\log V)$ steps for OTOC to saturate.

VII. CONCLUSION

Random quantum circuits have widespread applications in quantum information, and are also very important models of scrambling and chaotic quantum systems in theoretical physics. There are several ways to characterize scrambling and randomness in quantum processes, among which the OTOC and entanglement are two important types of measures. This work aims to understand whether they are equivalent to each other as the signature of scrambling. To this end, we carefully analyze local random quantum circuits defined on, e.g., a binary tree, which exhibit the property that OTOC mixes rather fast since the light cone can quickly reach the far end (time of order $\ln V$), while it takes a much longer time for entanglement between the left and right subtrees to grow (time at least of order V). We furthermore generalize the result to any bounded-degree graph with a tight bottleneck. That is, the generation of entanglement is slow, even if the graph has small diameter. Our result indicates that unitary t-designs can be much more expensive than we thought: they require a random quantum circuit to have depth much larger than the diameter of the underlying graph. This result provides a more rigorous evidence for arguments made in Ref. [21]: if we consider the model discussed in Refs. [2,21], then the scrambling of quantum information as seen by strong measures such as entanglement or decoupling can be much slower than we thought before. It would be interesting to explore further implications of this phenomenon, and more generally, quantum information processing, to the black hole information problem, many-body physics, and beyond.

APPENDIX A: POISSON CLOCK

Here we argue that the Poisson clock model and the simplified model in which a random edge is picked every $1/E$ time units are essentially equivalent, by the following fact.

Proposition 1. *Within time t, the number of unitaries applied in the Poisson clock model is between $\frac{1}{2}Et$ and $\frac{3}{2}Et$ with probability $1 - e^{-\Omega(Et)}$.*

Proof. The number of unitaries applied to each edge is a Poisson distribution with mean t, so the total number of unitaries λ is a Poisson distribution with mean Et. By Ref. [40],

$$\Pr\left[|\lambda - Et| \geq \frac{1}{2}Et\right] \leq 2\exp\left\{-\frac{[(1/2)Et]^2}{2[Et + (1/2)Et]}\right\}$$
$$= e^{-\Theta(Et)}.$$

∎

APPENDIX B: CELLULATION OF HYPERBOLIC GEOMETRY

Here we include the figure from Ref. [21] that roughly depicts the cellulation of hyperbolic geometry (see Fig. 2), representing the causality structure of a black hole in Schwarzschild coordinates.

The line element in this geometry is

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2$$
$$+ r^2(d\theta^2 + \sin^2\theta \, d\phi^2).$$

The black hole has radius $2M$, the photon sphere has radius $3M$, and we are interested in the region $2M \leq r \leq 3M$. The total number of cells N is $O(M^2)$. A great circle of radius $r = 2M + h$ traverses approximately $\sqrt{M/h}$ cells while a line from radius $2M + h$ to radius $3M$ traverses approximately $\log(M/h)$ cells. Thus the graph diameter is $O(\log N)$.
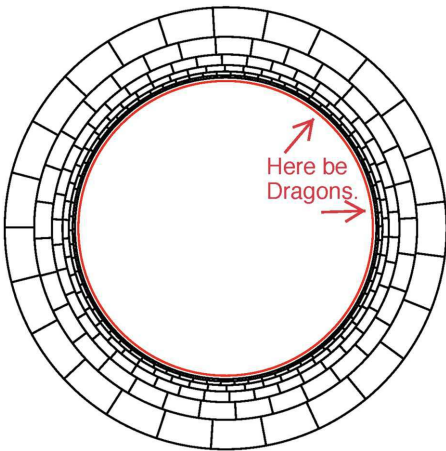

FIG. 2. The cell structure roughly equivalent to the hyperbolic geometry, depicted in two dimensions. In [21], this represents the black-hole cell structure in Schwarzschild coordinates, where each cell carries one qubit of Hawking-Unruh radiation.

APPENDIX C: OTOC FOR THE HIGH-DEGREE CASE

1. Examples with long OTOC time

In this section we describe two examples of a high-degree graph where the OTOC time is asymptotically larger than the graph diameter. One example is the star graph (see Fig. 3) where this phenomenon was previously observed by Lucas [23]. A second example is a binary tree with high but constant degree. We summarize the parameters of these two examples in Table II.

In this section we sketch proofs of both of the claimed OTOC times. First we give a simplified explanation of the $O(\log n)$ OTOC time for the star graph. (A more precise but also more complicated argument was given in Ref. [23].) In the star graph there are $n - 1$ vertices each connected to a central vertex.

Consider the OTOC between Paulis on two of the noncentral vertices, say 1 and 2. We begin with a single N on vertex 1 and I's everywhere else. Vertices 0 and 1 interact after expected time 1, and when they interact there is a 4/5 chance that vertex 0 ends up with an N. Thus the expected time for vertex 0 to turn to N is 5/4 and after this happens the expected number of noncentral N's is 3/4.

After this, vertex 0 rapidly interacts with the other $n - 1$ vertices. Each time there is a 3/5 chance that the other vertex will turn from I to N, a 1/5 chance that the other vertex will remain I (or change from N to I, but consider for simplicity the early stages when most vertices are I), and a 1/5 chance that the other vertex will become N and vertex 0 will revert to I. The number of N's created here before vertex 0 turns back to I is given by a geometric random variable with expectation 3, and the time this takes has expectation approximately equal to $4/n$. For simplicity, we ignore the fluctuations and assume that the process creates exactly three N's and takes time exactly $4/n$.

Now there (on expectation) 3.75 N's in noncentral vertices and I's elsewhere. Again there is a long wait until the central vertex turns back to N, but now the expected wait is only $\frac{5}{4} \times 1/(3 + \frac{3}{4})$. (We pretend, crudely, that there are exactly 3.75 N's in noncentral vertices.) Again (on average) three more noncentral vertices turn to N and this again takes expected time $4/n$.
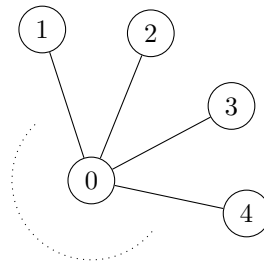

FIG. 3. Star graph.

TABLE II. Two examples with OTOC time asymptotically larger than the graph diameter.

Graph	Diameter	OTOC time	Source
n qubits in star graph	2	$O(\log n)$	[23]
Tree with degree $z \gg d^2$	$O(\log n/\log z)$	$\tilde{O}\left(n^{1-[\ln(d^2)/\ln z]}\right)$	This section

The time for $O(n)$ of the noncentral vertices to turn to N is thus dominated by the waits for the central vertex to turn back to N, which add up to

$$\frac{5}{4}\left[1 + \frac{1}{1\times 3 + (3/4)} + \frac{1}{2\times 3 + (3/4)} + \cdots + \frac{1}{O(n)\times 3 + (3/4)}\right], \quad (C1)$$

and is equal to $O(\log n)$.

This is not quite rigorous as we replace random variables with their expectations in several places, but otherwise is strong evidence for the OTOC time of $O(\log n)$ for the star graph. A more rigorous argument is given in Sec. 5 of Ref. [23]. There it is also observed that dynamics based on random Hamiltonians give an asymptotically different OTOC time.

We can obtain a stronger separation between diameter and OTOC time by considering a tree. Here we need the degree z only to be a sufficiently large constant and can obtain an OTOC time of $n^{1-\delta}$ with $\delta \to 0$ as $z \to \infty$. However, the diameter is also asymptotically growing as $O(\log n)$, unlike the star graph where the diameter is only 2.

Suppose that degree $z \gg d^2 \gg 1$. Consider a tree of depth H, so that there are $(z-1)^h$ vertices at level h for $h = 0, \ldots, H$. The total number of vertices is thus $n = 1 + (z-1) + (z-1)^2 + \cdots + (z-1)^H = z^{H+1} - 1$. Each vertex with $h > 0$ has one parent and each vertex with $h < H$ has $z - 1$ children.

Consider an N at vertex v, with height h. Suppose that $x(z-1)$ of its children are in state N and $(1-x)(z-1)$ are in state I for some $x \in [0,1]$. We approximate the dynamics as follows.

(a) The parent of v becomes N at rate $1 - 1/(d^2+1) \approx 1$, regardless of its current state.

(b) A child of v turns into N at a rate $(z-1)[1 - 1/(d^2+1)] \approx z$, again regardless of its current state. This means that dx/dt has an expected contribution of $(1-x)$.

(c) A child of v currently in state N turns to I, at rate $x(z-1)/(d^2+1) \approx xz/d^2$, yielding a contribution to dx/dt of $-x/d^2$.

(d) v itself turns to I at a rate of $(z-1)/d^2 \approx z/d^2$. (At the same time, one of v's children turns to N, but we can neglect this.)

Overall we find that x evolves according to

$$\frac{dx}{dt} \approx 1 - (1 + 1/d^2)x, \quad (C2)$$

which asymptotically approaches $(1 + 1/d^2)^{-1} \approx 1 - 1/d^2$. However, v will turn to I before this happens, typically after time d^2/z. During this time, the probability of v's parent turning to N is $\approx z/d^2$.

Once v turns to I, its children can turn it back to N. This happens at rate approximately equal to xd.

First, let us analyze more carefully the dynamics of x. If $h \le H - 2$ then the children of v will themselves undergo the same process. In particular, they will turn to I at rate z/d^2. This means Eq. (C2) is modified to become

$$\frac{dx}{dt} \approx 1 - (1 + 1/d^2 + z/d^2)x \approx 1 - \frac{z}{d^2}x, \quad (C3)$$

Now x will saturate at $x = d^2/z$.

Putting this together, the N's will "fall" to lower levels more quickly than they can climb or be replaced.

By the time they have fallen to the base, the N's correspond to the leaves of a subtree S with branching factor d^2. Then they begin slowly climbing back up.

Let us start at level H. Fix a vertex $v \in S$ at height $H - 1$. After time $1/d^2$, v will turn to N. Then it will fall back down again after time d^2/z, during which it will have created another d^2 N's at level H. Now there are $2d^2$ N's below v. This means it will take time $1/2d^2$ to turn v back to N. Again it falls down after time d^2/z, creating another d^2 N's. This process continues z/d^2 times, for a total time of

$$\frac{1}{d^2}\left(1 + \frac{1}{2} + \cdots + \frac{1}{z/d^2}\right) + \frac{z}{d^2} \times \frac{d^2}{z} \approx 1 + \frac{\ln(z)}{d^2}.$$

This happens in parallel for each height-$(H-1)$ vertex in S. These vertices no longer fall because most of their children are (and remain) equal to N.

Now consider a height-$(H-2)$ vertex in S, again called v. By the previous arguments, it has d^2 children that remain mostly permanently set to N. So v will become N after time $1/d^2$. However, after time d^2/z it will fall down to level $H - 1$, creating d^2 N's at level $H - 1$ in the process; call this set A. These in turn will fall to level H where they create d^4 N's. Again it takes time $1 + \ln(z)/d^2$ to cause the vertices in A to turn permanently into N's. After this it takes time $1/2d^2$ to again turn v to N, etc. There are again z/d^2 rounds, and focusing only on the $1 + \ln(z)/d^2$

contribution to the time for one round, we obtain a total time of

$$\frac{z}{d^2}[1 + \ln(z)/d^2].$$

Now we move on to the height-$(H - 3)$ part of S. Each vertex there again turns to N and then falls to the bottom of the tree z/d^2 times. But now climbing back up takes time

$$\approx \left(\frac{z}{d^2}\right)^2 [1 + \ln(z)/d^2].$$

We conclude that the entire OTOC time is

$$\approx \left(\frac{z}{d^2}\right)^{H-1} \left(1 + \frac{\ln(z)}{d^2}\right). \tag{C4}$$

Using that $H + 1 = \ln(n + 1)/\ln(z)$, the OTOC time is approximately

$$\left(\frac{d^2}{z}\right)^2 \left(\frac{z}{d^2}\right)^{H+1} = \left(\frac{d^2}{z}\right)^2 (n + 1)^{\ln(z/d^2)/\ln(z)}$$
$$= \left(\frac{d^2}{z}\right)^2 (n + 1)^{1 - [\ln(d^2)/\ln(z)]}.$$

Thus by taking $z = d^{1/\delta}$ we obtain OTOC scaling as $n^{1-\delta}$, which is much larger than the diameter of $O(H) = O[\ln(n)/\ln(z)] = O[\delta \ln(n)/\ln(d)]$.

2. Examples with short OTOC time

For a complete graph with V vertices, $\Theta(V \log V)$ edges should be picked in order for the OTOC to be $\Theta(1)$, according to Ref. [26]. By Proposition 1, the OTOC saturation time is $\Theta[\log(V)/V]$. In contrast, the distance between any pair of vertices is 1.

Another example is that the OTOC saturation time on any graph can be scaled down by a factor of $\Theta(1/k)$ if we replace each edge by k-fold parallel edges. Alternatively we can replace each edge (u, v) by $2k$ edge (u, z_i) and (z_i, v) for $1 \leq i \leq k$, where $\{z_i\}_{i=1}^{k}$ are newly added vertices (there are $2kE$ new vertices in total, k for each edge). Since k could be arbitrary, this violates the lower bound we obtain at finite degree.

---

[1] Patrick Hayden and John Preskill, Black holes as mirrors: Quantum information in random subsystems, J. High Energy Phys. **2007**, 120 (2007).

[2] Yasuhiro Sekino and Leonard Susskind, Fast scramblers, J. High Energy Phys. **2008**, 065 (2008).

[3] Juan Maldacena, Stephen H. Shenker, and Douglas Stanford, A bound on chaos, J. High Energy Phys. **2016**, 106 (2016).

[4] Stephen H. Shenker and Douglas Stanford, Black holes and the butterfly effect, J. High Energy Phys. **2014**, 67 (2014).

[5] Rahul Nandkishore and David A. Huse, Many-body localization and thermalization in quantum statistical mechanics, Annu. Rev. Condens. Matter Phys. **6**, 15 (2015).

[6] Arijeet Pal and David A. Huse, Many-body localization phase transition, Phys. Rev. B **82**, 174411 (2010).

[7] A. I. Larkin and Yu N. Ovchinnikov, Quasiclassical method in the theory of superconductivity, Sov. Phys. JETP **28**, 1200 (1969).

[8] Nima Lashkari, Douglas Stanford, Matthew Hastings, Tobias Osborne, and Patrick Hayden, Towards the fast scrambling conjecture, J. High Energy Phys. **2013**, 22 (2013).

[9] Pavan Hosur, Xiao-Liang Qi, Daniel A. Roberts, and Beni Yoshida, Chaos in quantum channels, J. High Energy Phys. **2016**, 4 (2016).

[10] Zi-Wen Liu, Seth Lloyd, Elton Zhu, and Huangjun Zhu, Entanglement, quantum randomness, and complexity beyond scrambling, J. High Energy Phys. **2018**, 41 (2018).

[11] Subir Sachdev and Jinwu Ye, Gapless Spin-Fluid Ground State in a Random Quantum Heisenberg Magnet, Phys. Rev. Lett. **70**, 3339 (1993).

[12] Alexei Kitaev, http://online.kitp.ucsb.edu/online/entangled15/kitaev/, http://online.kitp.ucsb.edu/online/entangled15/kitaev2/ (2015).

[13] Adam Nahum, Sagar Vijay, and Jeongwan Haah, Operator Spreading in Random Unitary Circuits, Phys. Rev. X **8**, 021014 (2018).

[14] Adam Nahum, Jonathan Ruhman, Sagar Vijay, and Jeongwan Haah, Quantum Entanglement Growth under Random Unitary Dynamics, Phys. Rev. X **7**, 031016 (2017).

[15] Vedika Khemani, Ashvin Vishwanath, and David A. Huse, Operator Spreading and the Emergence of Dissipative Hydrodynamics under Unitary Evolution with Conservation Laws, Phys. Rev. X **8**, 031057 (2018).

[16] C. W. von Keyserlingk, Tibor Rakovszky, Frank Pollmann, and S. L. Sondhi, Operator Hydrodynamics, fdicke OTOCs, and Entanglement Growth in Systems Without Conservation Laws, Phys. Rev. X **8**, 021013 (2018).

[17] Lorenzo Piroli, Christoph Sünderhauf, and Xiao-Liang Qi, A random unitary circuit model for black hole evaporation, J. High Energy Phys. **4**, 63 (2020).

[18] Matthew B. Hastings and Tohru Koma, Spectral gap and exponential decay of correlations, Commun. Math. Phys. **265**, 781 (2006).

[19] Bruno Nachtergaele and Robert Sims, Lieb-robinson bounds and the exponential clustering theorem, Commun. Math. Phys. **265**, 119 (2006).

[20] Aram Harrow and Saeed Mehraban, Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates, ArXiv:1809.06957 (2018).

[21] Peter W. Shor, Scrambling time and causal structure of the photon sphere of a Schwarzschild black hole, ArXiv:1807.04363 (2018).

[22] Gregory Bentsen, Yingfei Gu, and Andrew Lucas, Fast scrambling on sparse graphs, Proc. Natl. Acad. Sci. 116, 201811033 (2019).

[23] Andrew Lucas, Quantum many-body dynamics on the star graph, ArXiv:1903.01468 (2019).

[24] R. J. Lewis-Swan, Arghavan Safavi-Naini, John J. Bollinger, and Ana M. Rey, Unifying scrambling, thermalization and entanglement through measurement of

fidelity out-of-time-order correlators in the Dicke model, Nat. Commun. **10**, 1 (2019).

[25] Cheryne Jonay, David A. Huse, and Adam Nahum, Coarse-grained dynamics of operator and state entanglement, ArXiv:1803.00089 (2018).

[26] Aram W. Harrow and Richard A. Low, Random quantum circuits are approximate 2-designs, Commun. Math. Phys. **291**, 257 (2009).

[27] Hong Liu and S. Josephine Suh, Entanglement Tsunami: Universal Scaling in Holographic Thermalization, Phys. Rev. Lett. **112**, 011601 (2014).

[28] Josiah Couch, Stefan Eccles, Phuc Nguyen, Brian Swingle, and Shenglong Xu, The Speed of Quantum Information Spreading in Chaotic Systems, ArXiv:1908.06993 (2019).

[29] Beni Yoshida and Alexei Kitaev, Efficient decoding for the Hayden-Preskill protocol, ArXiv:1710.03363 (2017).

[30] Richard A. Low, Pseudo-randomness and learning in quantum computation, ArXiv:1006.5227 (2010).

[31] Fernando GSL Brandao, Aram W. Harrow, and Michał Horodecki, Local random quantum circuits are approximate polynomial-designs, Commun. Math. Phys. **346**, 397 (2016).

[32] Winton Brown and Omar Fawzi, Scrambling speed of random quantum circuits, ArXiv:1210.6644 (2012).

[33] Winton Brown and Omar Fawzi, Decoupling with random quantum circuits, Commun. Math. Phys. **340**, 867 (2015).

[34] Zi-Wen Liu, Seth Lloyd, Elton Yechao Zhu, and Huangjun Zhu, Generalized Entanglement Entropies of Quantum Designs, Phys. Rev. Lett. **120**, 130502 (2018).

[35] Oleg Szehr, Frédéic Dupuis, Marco Tomamichel, and Renato Renner, Decoupling with unitary approximate two-designs, New J. Phys. **15**, 053022 (2013).

[36] David A. Levin and Yuval Peres, *Markov Chains and Mixing Times* (American Mathematical Soc., Providence, Rhode Island, 2017), Vol. 107

[37] Charles H. Bennett, Aram Wettroth Harrow, Debbie W. Leung, and John A. Smolin, On the capacities of bipartite hamiltonians and unitary gates, IEEE Trans. Inf. Theory **49**, 1895 (2003).

[38] Rajibul Islam, Ruichao Ma, Philipp M. Preiss, M. Eric Tai, Alexander Lukin, Matthew Rispoli, and Markus Greiner, Measuring entanglement entropy in a quantum many-body system, Nature **528**, 77 (2015).

[39] Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier, Ben P. Lanyon, Peter Zoller, Rainer Blatt, and Christian F. Roos, Probing entanglement entropy via randomized measurements, ArXiv:1806.05747 (2018).

[40] David Pollard, Mini empirical, *Manuscript*, http://www.stat.yale.edu/pollard/Books/Mini (2015).