This is the accepted manuscript made available via CHORUS. The article has been published as:

Physical theory of two-stage thermalization

Cheryne Jonay and Tianci Zhou

Phys. Rev. B **110**, L020306 — Published 22 July 2024

DOI: 10.1103/PhysRevB.110.L020306

A Physical Theory of Two-stage Thermalization

Cheryne Jonay[1] and Tianci Zhou[2,3]

[1] *Department of Physics, Stanford University, Stanford, CA 94305*
[2] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3] *Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA*

(Dated: July 8, 2024)

Thermalization time marks the scale at which the subsystem entanglement reaches its thermal value. It was recently shown to exhibit two distinct stages in a generic class of quantum circuits with decay rates r_1 and r_2 before and after thermalization. The two rates are observed in the Floquet dynamics of the state purity, entanglement spectra, and (local) correlation functions, both in clean and disordered systems. Entanglement membrane theory interprets r_1 as domain wall energy, and r_2 as either a geometric effect ("phantom eigenvalue") or a competition with an emergent magnon mode. In the case when the domain wall dominates, this offers a practical measure for entanglement growth via local correlations.

Thermalization in closed quantum systems has been the subject of extensive study due to its critical role in quantum statistical mechanics. In general, local subsystems are known to thermalize as they become maximally entangled with their surroundings. However, while the state at large may have equilibrated, dynamics at the microscopic scale continue past this thermalization time. A series of recent works have uncovered an intriguing phenomenon of two-stage thermalization in random quantum circuits[1–3]: the relaxation to equilibrium has two stages, characterized by different rates before and after the thermalization time. This opens up inquiries into the microscopic processes driving thermalization.

For a concrete setup of two-stage thermalization, let us consider a pure state evolved unitarily by a local quantum circuit (see e.g. insets of Fig. 1(a)). To hint at its statistical mechanical nature we denote the half-system purity as a partition function $Z(t)$. When thermalizing, $Z(t)$ approaches the equilibrium value $Z(\infty)$. Fig. 1(a) shows the convergence $\ln|Z(t)-Z(\infty)|$: The exponential decay towards saturation value is $e^{-r_1 t\ln(2)}$ before and $e^{-r_2 t\ln(2)}$ after the saturation time. Open and periodic boundary conditions in certain gate parameter regimes can yield different values r_2, see Fig. 1(b).

Crucially, the two-stage thermalization is a general feature of chaotic (meaning far from integrable) local systems. First, the phenomenon is not restricted to random quantum circuits but has been observed in clean translation invariant Floquet circuits [4]. Our theory suggests that analog Floquet simulations also work. Moreover, the observable is not confined to the subsystem purity. The entropy itself – here the entanglement entropy of the subsystem (including the von Neumman entropy) has this signature. The first stage corresponds to the linear growth of the entanglement before equilibrium, and the second stage corresponds to a size $2^{-r_2 t}$ fluctuation of entropy around its equilibrium values. The natural hypothesis is that this two-stage decay should be observable directly in the entanglement spectrum, which we confirm in the Supplemental Material [5] (see also Ref. [**?**] therein). We further observe two stages in the relaxation of out-of-time-ordered correlators, which characterize the scrambling of quantum information [6–8], and local correlation functions, which we propose has important implications for the experimental measure of entanglement.

Until now, the origin of two stages was attributed to the spectral properties of the (effective) non-Hermitian transfer matrix of the problem. When $r_1 > r_2$, r_1 lives in the gap of the transfer matrix and is regarded as a "phantom eigenvalue". Because of the exponentially large amplitude of the localized edge modes, it dominates the first stage decay. This phenomenon is known as the "non-Hermitian skin effect" and occurs in the effective dynamics of unitary evolution.

In this work, we lay out an intuitive physical theory for two-stage thermalization in its full generality. The two decay rates correspond to distinct "particle" excitations - domain walls and magnons, which compete to dominate the purity decay. The interplay between these excitations is also influenced by the system's geometric configuration, which dictates whether the decay rate accelerates or decelerates. This picture of emergent particle-like excitations is crucial in understanding how entanglement not only facilitates thermal equilibrium but also drives dynamics post-saturation.

Finally, our theory also generates two concrete experimental protocols that are of great interest for NISQ devices. (i) Measuring entanglement growth through a local correlation function. (ii) Mitigating errors through the first-order transition between the domain wall and magnon in Fig. 1(b). (i) uses the fact that the two-stage decay also occurs in local correlation functions, and greatly simplifies the entanglement measurement protocol. (ii) takes advantage of the discontinuity to enlarge the signal.

Phenomenologies: We focus on the two-stage decay of the state purity and first review the quantitative phenomenologies observed in quantum circuits studied in Ref. [1]. The unitary gates in the circuits are nearest-neighbor and stacked either in a brickwall or staircase geometry (Fig. 1(a)) on a one-dimensional lattice of L qudits with local Hilbert space dimension q. The brickwall structure naturally arises in the Trotter limit and for modeling local interactions, which ensures a finite speed

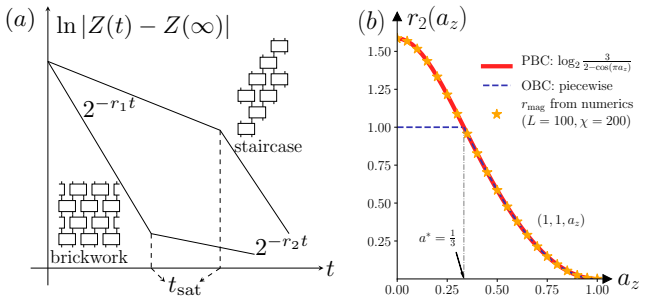


Figure 1. Two-stage thermalization. (a): Schematics of 2-stage decay for staircase and brickwall geometry. Both insets show circuits of 4 time steps. Generally $r_1 < r_2$ for staircase geometry and $r_1 \geq r_2$ for brickwork geometry. (b) For the dual unitary $(1, 1, a_z)$ circuits, r_2 depends on a_z and the boundary condition (periodic or open).

of information propagation. The staircase geometry has the advantage that dynamics can be traced backward along a single space-time path, which also reduces spatial cost in the quantum simulation if reset is possible [9]. We identify three specific scenarios $r_1 = r_2$, $r_1 < r_2$ and $r_1 > r_2$.

Scenario 0: $r_1 = r_2$ We first consider the paradigmatic example of random unitary circuits (RUC) in which the gates are independent random $\mathrm{U}(q^2)$ matrices. This choice allows us to access the typical behavior via disorder averaging(see e.g. [10–19] and a recent review [20]). We thus examine $\ln|\overline{Z(t)} - \overline{Z}(\infty)|$ where the over-line denotes averaging over Haar ensemble on $\mathrm{U}(q^2)$. In a brickwall geometry, there is a single decay rate given by

$$r_1 = r_2 = \frac{\ln \frac{q^2+1}{2q}}{\ln q}, \qquad \text{[brickwall, RUC]}. \tag{1}$$

Scenario 1: $r_1 < r_2$. For RUC in a staircase geometry, we obtain two decay rates after random averaging [2]

$$r_1 = \frac{1}{2}\frac{\ln \frac{q^2-q+1}{q}}{\ln q}, \quad r_2 = \frac{\ln \frac{q^2+1}{2q}}{\ln q}. \qquad \text{[S, RUC]} \tag{2}$$

This phenomenon is observed in more general (non-random) circuits with staircase geometry.

Scenario 2: $r_1 > r_2$ brickwall geometry can host different behaviors than scenario 0 if we allow more general gates than Haar random. Let us specialize to $q = 2$. An arbitrary 2-qubit gate can be parametrized through four single-qubit rotations u_i, $i = 1, 2, 3, 4$ and a symmetric 2-body interaction u_{sym} as

$$u = (u_1 \otimes u_2) u_{\mathrm{sym}} (u_3 \otimes u_4). \tag{3}$$

The symmetric part is generated by commutative operators $u_{\mathrm{sym}} = \exp\Big(-i\frac{\pi}{4}(\sum_{\alpha=x,y,z} a_\alpha \sigma^\alpha \sigma^\alpha)\Big)$. Here σ^α is the α-th Pauli matrix and $0 \leq a_\alpha \leq 1$. In this scenario, we average the four single-qubit unitaries u_i over $\mathrm{U}(2)$, hence the whole gate is indexed by (a_x, a_y, a_z). The choice $a_x = a_y = 1$ corresponds to a special class known as *dual unitary* gates, which means the gate is unitary viewed in both spatial and temporal directions (see e.g. [21, 22] and more details in a review [23]). We primarily focus on this one-parameter class of circuits abbreviated as $(1, 1, a_z)$. In the one-parameter family $(1, 1, a_z)$, $a_z = 0$ corresponds to an iSWAP gate; $a_z = 1$ a SWAP gate; the gate is not integrable elsewhere. It has been proven that the purity decays as $\mathrm{poly}(t)2^{-t}$ [24] on average for dual unitary circuits with high enough entangling power, while numerics suggest it holds more generally. Consequently, the first stage decay rate is $r_1 = 1$ for OBC and $r_1 = 2$ for PBC (the subregion A has two boundary points). The second stage decay rate r_2 is a function of a_z,

$$r_1 = 2, \quad r_2(a_z) = \frac{\ln \frac{3}{2-\cos(\pi a_z)}}{\ln 2}, \quad [(1, 1, a_z), \text{PBC}]. \tag{4}$$

We prove the $r_2(a_z)$ expression in the Supplemental Material [5], which was numerically computed and conjectured in [25]. For OBC, the expression remains the same for $a_z \geq \frac{1}{3}$ but caps to 1 for $0 \leq a_z \leq \frac{1}{3}$. See schematics in Fig. 1(b). The $r_1 > r_2$ phenomenon extends beyond the (a_x, a_y, a_z) circuits and applies even to Floquet circuits with space and time translation symmetries.

Effective magnet. The physical theory originates from analyzing random unitary circuits, where averaged entanglement dynamics map to the statistical mechanics of an effective magnet, which we review now [11, 16, 26–35]. The relevant partition function for the problem at hand is given by the purity of a time-evolved state restricted to subregion A,

$$Z_A(t) = \mathrm{Tr}\big(\rho_A(t)^2\big), \tag{5}$$

where $\rho_A(t) = \mathrm{Tr}_{\overline{A}}(|\psi(t)\rangle\langle\psi(t)|)$. This quantity contains two copies of the unitary circuit U and its conjugate U^*, combined as $U \otimes U^* \otimes U \otimes U^*$ acting on four copies of the L qudits. Since our sampling of random gates is independent in space and time, the overall random averaging of U reduces to the separate averaging over each individual gate. The local Haar average over the single-site unitary $\overline{u_1 \otimes u_1^* \otimes u_1 \otimes u_1^*}$ projects the 4 copies of the qudit Hilbert space into a two-dimensional subspace spanned by the states $\{|+\rangle, |-\rangle\}$ [36], which represent our effective spins. Physically these two states denote two different ways to pair a unitary and its conjugate,

$$|+\rangle := \overline{u_1 \otimes u_1^*} \otimes \overline{u_1 \otimes u_1^*}, \tag{6}$$

$$|-\rangle := \overline{u_1 \otimes \overline{u_1^* \otimes u_1} \otimes u_1^*} \tag{7}$$

The boundary spins are specified by how different copies of U and U^* are connected together in the partition function Eq. 5: For half-system purity, the top boundary is a domain wall state $|+\cdots+-\cdots-\rangle$ where region $\overline{A}$ (A) hosts $|+\rangle$ ($|-\rangle$) states; for an initial product state, the bottom boundary is free.

The local average $\overline{u \otimes u^* \otimes u \otimes u^*}$ is the local transfer matrix M that determines the update rules of adjacent spins. For RUCs, the update rules are

$$|{+}{+}\rangle \to |{+}{+}\rangle \tag{8}$$
$$|{+}{-}\rangle \to K\,|{+}{+}\rangle + K\,|{-}{-}\rangle\,, \tag{9}$$

and symmetric counterparts by exchanging $|+\rangle$ and $|-\rangle$ with $K = \frac{q}{q^2+1}$. Eq. (8) arises from the unitarity of the gates, while Eq. (9) is the microscopic theory of a propagating domain wall. At each time step, the domain wall can move to the left or right with equal probability, but the single-domain wall sector is preserved (Fig. 2(a)) – a special feature of Haar random circuits [37] (discussed in detail in the Supplemental Material [5], see also Ref [**?**] therein). The free energy of this random walk is the Rényi entropy. The average partition function is exact $\overline{Z(t)} = (2K)^t$ [27], and the decay rate coincides with r_2 in Eq. (2). For dual unitary circuits, the unitarity rules are the same as in Eq. (8). However, the domain wall (9) is not preserved, and can be transferred into more states

$$|{+}{-}\rangle \to h\,|{+}{+}\rangle + h\,|{-}{-}\rangle + b_+\,|{+}{-}\rangle + b_-\,|{-}{+}\rangle \tag{10}$$

where $h = (3-v)/9$, $b_\pm = (3 \pm 6u + 5v)/36$ and $u = \cos(\pi a_x) + \cos(\pi a_y) + \cos(\pi a_z)$, $v = \cos(\pi a_x)\cos(\pi a_y) + \cos(\pi a_y)\cos(\pi a_z) + \cos(\pi a_z)\cos(\pi a_x)$ [1]. For later reference, we denote the overall propagator $\widehat{\mathcal{M}}(t)$, which is an alternate product of the global transfer matrices at even and odd steps (the Supplemental Material [5]). The partition function is

$$Z(t) = \frac{1}{(q^2+q)^L} \sum_{s_i=\pm} \langle s_1, \cdots s_L | \widehat{\mathcal{M}}(t) | \cdots + - \cdots \rangle. \tag{11}$$

At scales much larger than the domain wall width, the membrane theory asserts that the free energy of the domain wall can be described by a macroscopic line tension function $\mathcal{E}(v)$, which only depends on the space-time (anti-)slope v. The partition function Eq. (11) asymptotically decays as $\exp(-\min_v \mathcal{E}(v) t \ln q)$. For RUCs, the average partition function has an exact line tension for a random walk:

$$\mathcal{E}(v) = \frac{\ln \frac{q^2+1}{q} + \frac{1+v}{2}\ln\frac{1+v}{2} + \frac{1-v}{2}\ln\frac{1-v}{2}}{\ln q}. \tag{12}$$

The minimum at $v = 0$ is a vertical random walk (Fig. 2(a)) and consistently gives r_2 in Eq. (2) of scenario 0.

Scenario 1: $r_1 < r_2$ S-geometry. The staircase geometry creates a tilted diamond region which constrains the domain wall's movement. As shown in Fig. 2, the bottom staircase boundary is tilted upward at 45°, allowing the domain wall to travel a shorter distance if it tilts towards this lower boundary. However, this comes at the price of a larger line tension $\mathcal{E}(v)$. Quantitatively, if the domain wall has an (anti-)slope v, the distance traveled (or time duration) is $t/(1+v)$. The free energy $F(v)$ is a trade-off between shorter path length and increased line tension:

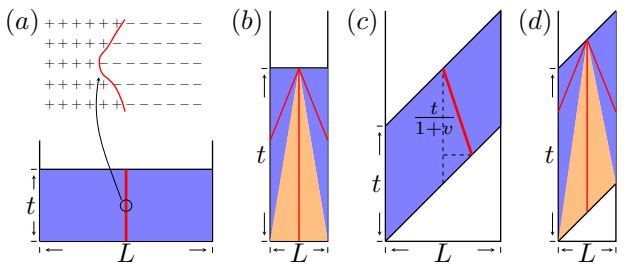


Figure 2. Domain wall configurations in brickwall ((a)(b)) and staircase ((c)(d)) geometries. (a) $t < t_{\rm sat}$, spin states at the lattice level (top) and domain wall random walk at the coarsed grained level (bottom). (b) $t > t_{\rm sat}$ for brickwall. (c) $t < t_{\rm sat}$, domain wall has shorter paths in the staircase. (d) Same as (b) for staircase.

$$F(v) = \min_v \frac{\mathcal{E}(v)}{1+v} \ln q. \tag{13}$$

For RUCs, the line tension is given in Eq. (12), and the minimum at $v^* = \frac{(q-1)^2}{q^2+1}$ reproduces $r_1 = \frac{\mathcal{E}(v^*)}{1+v^*} = \frac{1}{2}\frac{\ln\frac{q^2-q+1}{q}}{\ln q}$, which is consistent with Eq. (2). However, for $t \sim t_{\rm sat}$, the same trajectory exits at the spatial boundary (rather than the bottom boundary). For $t > t_{\rm sat}$, the partition function has almost saturated to a static value. Schematically (Fig. 2(d)), the saturation value comes from the contribution in which the domain wall hits the left and right boundaries. And the second stage decay comes from a subleading correction in which the domain wall continues to reach the bottom (yellow). The effect of the tilted bottom boundary is increasingly negligible compared to the bulk contribution. Consequently, the yellow part is dominated by an (almost) vertical domain wall $e^{-\mathcal{E}(0)t\ln q}$, which gives the decay rate $r_2 = \mathcal{E}(0)$. Taking the tilt angle to zero reduces the staircase to the brickwall geometry, which explains why the brickwall RUCs exhibit no geometric cross-over (scenario 0).

Scenario 2: $r_1 > r_2$ The Haar gates in the brickwall structure do not exhibit two-stage thermalization, but the dual-unitary $(1, 1, a_z)$ gates can. The new ingredient is a "magnon" mode. While the local update rules for Haar random circuits in Eq. (9) can only move domain walls, the ones for dual-unitary circuits in Eq. (10) permit swap processes $|{+}{-}\rangle \to b_-\,|{-}{+}\rangle$. This can create new pairs of domain walls; such a pair forms a magnon when bound. Before $t_{\rm sat}$, pair creation means that a magnon, once formed, can only coexist with a (dressed) domain wall. Since the free energy of a magnon on top of a domain wall is always higher than a single domain wall, the single domain wall mode dominates, and $r_1 = 1$ (2) for OBC (PBC) [24]. However, once the domain wall exits through the boundary after $t_{\rm sat}$, a standalone magnon can exist and compete with the domain wall. We believe

that the transition from a domain wall to a magnon mode creates the second stage with a smaller rate $r_2 < r_1$.

To confirm that magnon gives the rate r_2, we isolate the magnon contribution and compute a magnon (sub)partition function (for a detailed derivation, see the Supplemental Material [5] and Ref. [**?**] therein),

$$Z_{\mathrm{mag}}(x,t) = \langle \cdots +^* -^* +^* \cdots +^* |\widehat{\mathcal{M}}(t)| - + \cdots + \rangle. \quad (14)$$

The initial state $|- + + \cdots +\rangle$ (top boundary) is a domain wall on the left boundary, which creates a magnon. The final states are constructed from the dual basis $\{|+^*\rangle, |-^*\rangle\}$ (note $|+\rangle$ and $|-\rangle$ are not orthonormal), which satisfies $\langle i^*|j\rangle = \delta_{ij}$, $i,j \in \pm$. These dual states pin the magnon at spatial positions $x = 1, 2, \cdots, L$. The partition function (14) contains all trajectories that start with a magnon at the left spatial boundary and end with a magnon at x, with possible branching and broadening corrections of the magnon mode in between. We recursively solve for these corrections from the numerical data of $Z_{\mathrm{mag}}(x,t)$ and resum them to obtain the asymptotic decay rate r_{mag} defined by $\sum_x Z_{\mathrm{mag}}(x,t) \sim \exp(-r_{\mathrm{mag}} t \ln q)$. Fig. 1(b) displays the numerical results for r_{mag} in random $(1,1,a_z)$ circuits, confirming the predictions of Eq. (4). For PBC, $r_1 = 2$ always exceeds r_{mag}, meaning that the magnon dominates in the second stage. For OBC, however, r_1 surpasses r_{mag} only when $a_z \geq 1/3$. These findings explain the physical mechanism of r_2, and clarify how r_2 behaves under different boundary conditions.

Exact r_{mag} from correlation function. We can derive the analytic curve in Fig. 1(b) exactly via dual unitarity, which confines local correlations to propagate strictly *on* the light cone [38]. From the heatmap of $Z_{\mathrm{mag}}(x,t)$ (Fig. 3(a)), the value at $x = t$ dominates for each t and magnon thus also mainly propagates on the $x = t$ ray. Instead of relying on $Z_{\mathrm{t,t}}$, we introduce a modified partition function

$$\tilde{Z}_{\mathrm{mag}}(x,t) = \langle + \cdots + - + \cdots + |\widehat{\mathcal{M}}(t)| - + \cdots + \rangle. \quad (15)$$

and seek for $\tilde{Z}_{\mathrm{mag}}(t,t)$. Contrary to $Z_{\mathrm{mag}}(x,t)$ in Eq.(14), where the magnon was strictly pinned to position x, the partition function in Eq. (15) does not entirely eliminate but penalize other configurations with exponential cost. Thus it allows small $\mathcal{O}(1)$ fluctuations of the magnon final position around x and crucially does not affect the asymptotic decay rate r_{mag}. Moreover, dual-unitarity reduces $\tilde{Z}_{\mathrm{mag}}(t,t)$ to repeated applications of a quantum channel acting on the $|-\rangle$ state, as discussed in the Supplemental Material [5] and Ref. [38]. After averaging over single sites, the quantum channel is simplified and has only two eigenvalues: $\lambda_+ = 1$, which gives the dissipationless propagation of the $|+\rangle$ state, and $\lambda_- = 2^{-r_{\mathrm{mag}}}$, where r_{mag} is the analytic expression in Eq. (4).

Generic systems Importantly, this theory can be applied beyond random averaging to generic chaotic systems with either or both space and time translation invariance. As shown in Ref. [28], going away from random averaging leads to bubble corrections that widen the domain walls in Fig. 2(a) to have $O(1)$ width. The line tension function still exists after the renormalization by the bubbles [28]. Similar corrections can occur for the magnon (see the Supplemental Material [5]), so long as the bound state has lower energy than two separate domain walls. Otherwise, the magnon can dissolve through a binding transition, as discussed in Ref. [37].

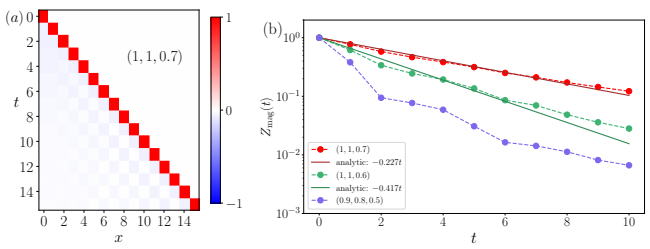


Figure 3. Magnon partition function and magnon decay rate. (a) heatmap of $Z_{\mathrm{mag}}(x,t)/\max_x|Z_{\mathrm{mag}}(x,t)|$. The magnon mostly travels on the light cone. (b) Magnon decay rate for clean dual unitary circuits $(1,1,0.7)$, $(1,1,0.6)$ and clean non-dual unitary Floquet circuit $(0.9, 0.8, 0.5)$.

In Fig. 3, we numerically resumed the corrections to the magnon in a Floquet dual unitary circuit. It has parameters $(a_x, a_y, a_z) = (1,1,0.5)$ and a fixed single site unitary $u_1 = u_2 = u_3 = u_4 = e^{i(\sin(\phi)\sigma^x + \cos(\phi)\sigma^z)}$ with $\phi = 0.6$. As in the average case, dual unitarity allows us to analytically solve r_{mag} from the eigenvalues of the quantum channel. The resummed value of r_{mag} even in systems of $L = 12$ sites converges to the analytic prediction from the channel calculation above, as shown in the inset of Fig. 3.

Discussion Our physical theory of the emergent domain wall and magnon has quantitatively explained all the observed finite system phenomenologies of r_1 and r_2: $r_1 < r_2$ is a geometric effect, while $r_2 < r_1$ is a magnon mode winning over the domain wall. Numerical evidence suggests that our theory works beyond random averaging and applies to generic (time-periodic) chaotic systems.

Interestingly, there is a similar two-stage thermalization for a standalone magnon. Instead of starting from a domain wall, as in the case of the purity, we consider the square of a local correlation function, which after averaging over the operator creates a magnon as the initial state. If the magnon is at the center, r_1 will be the magnon decay rate as the domain wall cannot exist alone before t_{sat}. For $t > t_{\mathrm{sat}}$, r_2 is once more determined by the competition between the domain wall and magnon. This is a *reverse* transition from the magnon rates to whichever wins.

If the domain wall decay rate is smaller, such a reverse transition can be used to measure the domain wall rate $\mathcal{E}(0)$ from r_2. This is the case for $(1,1,a_z)$ circuits with $a_z < 1/3$ after random averaging. An improvement of the protocol is to place the operator at the spatial boundary, so the magnon is instantaneously in a position

to compete with the domain wall. In the Supplemental Material [5], we compute the boundary correlation function for dual unitary circuits $(1, 1, a_z)$. The decay rate is 1 when $a_z \lesssim \frac{1}{3}$, which is the domain wall rate. When $a_z \gtrsim \frac{1}{3}$ the decay rate is close to the ones obtained from the gap of the quantum channel, indicating it as the magnon rate. Our theory thus also works for the local correlation function. It provides a practical scheme to measure entanglement, a highly non-local quantity, from the decay of local observables.

We note that our statistical mechanical model can be applied to Rényi entropies with index $n \geq 2$, for which the spin lives in the the permutation group S_n. This amounts to vastly more modes, with more types of domain walls, magnons and other bound states. The first stage is still set by a domain wall whose rate can be determined through the line tension method, however, in the second stage, the rate will be given by the competition between all modes. We don't have a good understanding of which mode 'wins' in this scenario. Further, we don't fully understand how to relate the decay of the these modes to the decay of individual eigenvalues in the entanglement spectra.The first step would be to understand S_∞, where only the largest entanglement eigenvalue is concerned.

Finally we note that the point $a_z = a^*$ marks a first order phase transition. The first order derivative of $r(a_z)$ has a discontinuous jump from 0 at $a_z = a^*$, which can be a sensitive signal in experiments. Mathematically, the derivative is given by the ratio $\frac{\partial_{a_z}\langle\psi(t)|Z_0|\psi(t)\rangle}{\langle\psi(t)|Z_0|\psi(t)\rangle}$. Through the parameter-shift rule [39], the derivative term on the numerator can be measured as the difference of two correlators using almost identical circuits with one parameter shifted. The discontinuity in $\partial_{a_z} r(a_z)$ can thus be detected by measuring correlators on order $\mathcal{O}(tL)$ parameter-shifted circuits. We anticipate applications in error detection and mitigation on modern quantum simulation platforms and leave the practical implementation and study of noise to future works.

Acknowledgements. C. J. thanks Tibor Rakovszky for his guidance on numerically implementing matrix product states. T.Z. thanks E. Barnes for suggestions about the manuscript. We acknowledge S. Gopalakrishnan for the discussion leading us to examine the entanglement spectrum. T. Z. was supported as a post-doctoral researcher from NTT Research Award No. AGMT DTD 9.24.20. and the Massachusetts Institute of Technology. This work was supported by the US Department of Energy, Office of Science, Basic Energy Sciences, under Early Career Award Nos. DE-SC0021111 (C.J, under Prof. Vedika Khemani's grant). We acknowledge the accommodation of the KITP program "Quantum Many-Body Dynamics and Noisy Intermediate-Scale Quantum Systems" and the Simons Center for Geometry and Physics program "Fluctuations, Entanglements, and Chaos: Exact Results" in which parts of the work took place.

[1] J. Bensa and M. Žnidarič, Fastest local entanglement scrambler, multistage thermalization, and a non-hermitian phantom, Phys. Rev. X **11**, 031019 (2021).
[2] M. Žnidarič, Solvable non-hermitian skin effect in many-body unitary dynamics, Phys. Rev. Research **4**, 033041 (2022).
[3] M. Znidaric, Phantom relaxation rate due to Jordan non-Hermitian skin effect and magic sums, (2023), 10.48550/arXiv.2306.07876.
[4] M. Znidaric, Two-step relaxation in local many-body floquet systems, (2023).
[5] See Supplemental Material for an overview of two-stage decay for OTOCs, an exact derivation of the second stage decay in Eq. (4), explanations of the different partition functions Eq. (11), Eq. (14), Eq. (15), additional details on the mapping to the stat mech model, the perturbative resummation method to numerically extract r_2, numerical techniques, and analysis of the entanglement spectrum.
[6] A. I. Larkin and Y. N. Ovchinnikov, Quasiclassical Method in the Theory of Superconductivity, Soviet Journal of Experimental and Theoretical Physics **28**, 1200 (1969).
[7] A. Kitaev, (2015), talks at KITP, April 7, 2015 and May 27, 2015.
[8] J. Maldacena, S. H. Shenker, and D. Stanford, A bound on chaos, Journal of High Energy Physics **2016**, 106 (2016).
[9] G. Anikeeva, I. H. Kim, and P. Hayden, Recycling qubits in near-term quantum computers, Phys. Rev. A **103**, 042613 (2021).
[10] A. Nahum, S. Vijay, and J. Haah, Operator spreading in random unitary circuits, Phys. Rev. X **8**, 021014 (2018).
[11] 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).
[12] M. Žnidarič, Exact convergence times for generation of random bipartite entanglement, Phys. Rev. A **78**, 032324 (2008).
[13] A. W. Harrow and R. A. Low, Random quantum circuits are approximate 2-designs, Commun. Math. Phys. **291**, 257 (2009).
[14] 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).
[15] B. Skinner, J. Ruhman, and A. Nahum, Measurement-induced phase transitions in the dynamics of entanglement, Phys. Rev. X **9**, 031009 (2019).
[16] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, Quantum entanglement growth under random unitary dynamics, Phys. Rev. X **7**, 031016 (2017).
[17] Y. Li, X. Chen, and M. P. A. Fisher, Measurement-driven entanglement transition in hybrid quantum cir-

cuits, Phys. Rev. B **100**, 134306 (2019).
[18] A. Chan, A. De Luca, and J. T. Chalker, Spectral statistics in spatially extended chaotic quantum many-body systems, Phys. Rev. Lett. **121**, 060601 (2018).
[19] A. Chan, R. M. Nandkishore, M. Pretko, and G. Smith, Unitary-projective entanglement dynamics, Phys. Rev. B **99**, 224307 (2019).
[20] M. P. A. Fisher, V. Khemani, A. Nahum, and S. Vijay, Random Quantum Circuits, (2022), arXiv:2207.14280 [cond-mat, physics:quant-ph].
[21] B. Bertini, P. Kos, and T. Prosen, Exact correlation functions for dual-unitary lattice models in $1+1$ dimensions, Phys. Rev. Lett. **123**, 210601 (2019).
[22] 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).
[23] T. Prosen, Many-body quantum chaos and dual-unitarity round-a-face, Chaos: An Interdisciplinary Journal of Nonlinear Science **31**, 093101 (2021).
[24] A. Foligno, T. Zhou, and B. Bertini, Entanglement, , 125 (2011).
[25] J. Bensa and M. Žnidarič, Two-step phantom relaxation of out-of-time-ordered correlations in random circuits, Phys. Rev. Research **4**, 013228 (2022).
[26] 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).
[27] T. Zhou and A. Nahum, Emergent statistical mechanics of entanglement in random unitary circuits, Phys. Rev. B **99** (2019), 10.1103/physrevb.99.174205.
[28] T. Zhou and A. Nahum, Entanglement membrane in chaotic many-body systems, Phys. Rev. X **10**, 031066 (2020).
[29] 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** (2018), 10.1103/physrevx.8.031057.
[30] R. Vasseur, A. C. Potter, Y.-Z. You, and A. W. W. Ludwig, Entanglement transitions from holographic random tensor networks, Phys. Rev. B **100** (2019), 10.1103/physrevb.100.134203.
[31] C.-M. Jian, Y.-Z. You, R. Vasseur, and A. W. W. Ludwig, Measurement-induced criticality in random quantum circuits, Phys. Rev. B **101** (2020), 10.1103/physrevb.101.104302.
[32] Y. Bao, S. Choi, and E. Altman, Theory of the phase transition in random unitary circuits with measurements, Phys. Rev. B **101** (2020), 10.1103/physrevb.101.104301.
[33] N. Hunter-Jones, Unitary designs from statistical mechanics in random quantum circuits, (2019).
[34] H. Liu and S. Vardhan, Entanglement entropies of equilibrated pure states in quantum many-body systems and gravity, PRX Quantum **2** (2021), 10.1103/prxquantum.2.010344.
[35] M. P. Fisher, V. Khemani, A. Nahum, and S. Vijay, Random quantum circuits, Annu. Rev. Condens. Matter Phys. **14**, 335 (2023).
[36] B. Collins and P. Śniady, Integration with respect to the haar measure on unitary, orthogonal and symplectic group, Commun. Math. Phys. **264**, 773 (2006).
[37] A. Nahum, S. Roy, S. Vijay, and T. Zhou, Real-time correlators in chaotic quantum many-body systems, Phys. Rev. B **106**, 224310 (2022).
[38] L. Piroli, B. Bertini, J. I. Cirac, and T. Prosen, Exact dynamics in dual-unitary quantum circuits, Phys. Rev. B **101**, 094304 (2020).
[39] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Physical Review A **99**, 032331 (2019).