

Free Probability in a Minimal Quantum Circuit Model

Felix Fritzsche^{1,*} and Pieter W. Claeys^{1,2}

¹*Max Planck Institute for the Physics of Complex Systems, 01087 Dresden, Germany*

²*School of Physics, Trinity College Dublin, Dublin 2, Ireland*

 (Received 24 June 2025; revised 24 April 2026; accepted 19 May 2026; published 5 August 2026)

Recent experimental and theoretical developments in many-body quantum systems motivate the study of their out-of-equilibrium properties through multitime correlation functions. We consider the dynamics of higher-order out-of-time-order correlators (OTOCs) in a minimal circuit model for quantum dynamics. This model mimics the dynamics of a structured subsystem locally coupled to a maximally random environment. We prove the exponential decay of all higher-order OTOCs and fully characterize the relevant timescales, showing how local operators approach free independence at late times. We show that the effects of the environment on the local subsystem can be captured in a higher-order influence matrix, which allows for a Markovian description of the dynamics provided an auxiliary degree of freedom is introduced. This degree of freedom directly yields a dynamical picture for the OTOCs in terms of free cumulants from free probability, consistent with recent predictions from the full eigenstate thermalization hypothesis. This approach and the relevant influence matrix are expected to be applicable in more general settings and present a first step to characterizing quantum memory in higher-order OTOCs.

DOI: [10.1103/6mpz-p85s](https://doi.org/10.1103/6mpz-p85s)

Subject Areas: Quantum Physics, Statistical Physics

I. INTRODUCTION

Recent decades have seen tremendous progress in our understanding of the thermalization of isolated quantum systems under their own unitary dynamics. The current paradigm through which we understand how local observables relax toward thermal equilibrium is known as the eigenstate thermalization hypothesis (ETH) [1–3]. Intuitively, ETH explains the relaxation of local subsystems toward thermal equilibrium by interpreting the complement of this subsystem as a large effective bath. This picture can be made precise within the influence-matrix approach to quantum dynamics [4–6]. Inspired by the Feynman-Vernon influence functional [7], the influence matrix at a given time captures the combined effect of the degrees of freedom in the effective bath on the local subsystem up to that time. In the case of maximally ergodic (discrete time) dynamics, the influence matrix represents perfectly dephasing boundary conditions, allowing for a Markovian description of the local subsystem evolution [5,6]. Away from the limit of a perfect Markovian bath, the influence-matrix approach represents a powerful numerical tool for capturing weak

memory effects and describing many-body quantum dynamics in a tractable way [8–28].

Nevertheless, the advent of novel experimental techniques, such as quantum simulation and (noisy, intermediate-scale) quantum computing platforms provide the opportunity for probing quantum dynamics with unprecedented precision and for asking more refined questions about thermalization. One example is the concept of deep thermalization, i.e., the dynamical emergence of quantum state designs in local subsystems under projective measurements on their environment [29–47]. A complementary approach asks about the relaxation of general multitime correlation functions, which characterize, for instance, the scrambling of quantum information in terms of out-of-time-order correlators (OTOCs) [48,49] and contribute to fluctuation-dissipation relations and transport phenomena in hydrodynamic theories beyond the linear-response regime [50–53]. OTOCs and their higher-order generalizations have recently been experimentally probed on Google’s quantum processor [54], where they were proposed as a path to quantum advantage, serving as additional motivation for theoretical studies that can guide classical approaches.

This move to more refined probes of thermalization was accompanied by corresponding extensions of ETH [55–63]. While two-point functions, appearing in linear response and dissipation-fluctuation relations, are described by ETH, the latter fails to correctly capture higher-order, multitime correlation functions. This failure can be traced back to the ETH ansatz neglecting correlations between eigenstates and matrix elements

*Contact author: fritzsche@pks.mpg.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

[64–67]. Reintroducing such correlations recently led to an extended, so-called full ETH ansatz, whose validity has been demonstrated in paradigmatic chaotic many-body quantum systems where it accurately describes the dynamics of OTOCs [60,63].

The discussion of such OTOCs and more general multi-time correlation functions is greatly simplified by translating the combinatorics of the full ETH ansatz into the combinatorial structure of free probability [57]. Free probability is a theory of noncommutative random variables and has found applications in random-matrix theory and operator algebras [68–73]. In the context of ETH, free probability allows for a decomposition of (multitime) correlation functions into elementary building blocks known as free cumulants [57,69,74]. The notion of freeness extends the notion of independence from commuting random variables to noncommuting random variables, and it is expected that in ergodic systems time-evolved observables become asymptotically free from static observables in the limit of large systems, i.e., in the thermodynamic or semiclassical limit. This emergence of freeness at late times has been linked to the emergence of unitary designs from physical time evolution [61] and was proposed as an indicator for quantum chaos [75–77].

On the level of correlation functions, asymptotic freeness at late times is characterized by the relaxation of generalized, higher-order, OTOCs toward a stationary value described by the correlations between free random variables. Analytical results on the dynamics of OTOCs, however, remain scarce. Notable exceptions include holographic and conformally invariant theories [78,79], minimal supersymmetric models [80], and the Sachdev-Ye-Kitaev model [81], as well as random quantum circuits [82–84] and dual-unitary circuits [85,86]. For dual-unitary circuits, partial analytical results on the dynamics of higher-order OTOCs were recently obtained [75]. To the best of our knowledge, no other analytical result on the dynamics of higher-order OTOCs exists, while a corresponding generalization of the influence-matrix approach even for perfectly Markovian baths is missing completely. Furthermore, it is not clear how the decomposition proposed by full ETH in terms of free cumulants, defined as appropriate summations over matrix elements, can be interpreted in a purely dynamical manner.

We here consider a minimal quantum-circuit model for the unitary dynamics of a local system coupled to an environment, motivated by a similar approach for deep thermalization [31]. While the local dynamics is structured, the bath is evolved using random unitary dynamics resulting in tractable yet nontrivial dynamics for local observables. This model can be interpreted as the circuit equivalent of Lindblad dynamics, with a separation of relaxation timescales between system and bath. Through an appropriate averaging over circuit realizations, we completely and rigorously characterize the resulting

dynamics of higher-order OTOCs, with the emergent structure reproducing predictions from full ETH and free probability. We demonstrate the emergence of free independence and explicitly characterize the relaxation dynamics in terms of free cumulants. The results on higher-order OTOC dynamics and the connections to full ETH and free probability are facilitated by introducing a higher-order influence matrix, which captures the effect of the bath on the system dynamics. The minimal structure exhibited by this influence matrix is directly linked to the dynamical emergence of free cumulants and can be expected to apply in more general models, up to possible non-Markovian corrections.

The main advances of this work are that we (i) fully characterize the dynamics of higher-order OTOCs in a minimal model of local dynamics, establishing asymptotic freeness and the dynamical emergence of free cumulants, (ii) show how this minimal model returns an influence matrix that extends the Markovian influence matrix for correlation functions to higher-order OTOCs, which can in turn be mapped to a Markovian process on the noncrossing partition lattice, and (iii) in this way, bridge full ETH and influence matrices, showing how the influence matrices for higher-order OTOCs are naturally expressed in terms of the noncrossing partitions underlying full ETH and free cumulants. Our study additionally highlights the different origins of full thermalization and deep thermalization. We expect this description and explicit structure of the influence matrix to be applicable beyond the specific setup of this work and emerge in more general dynamics in which a Markovian bath appears. The results from this work can be used as a starting point for more involved bath models and characterizations of quantum memory, in the same way that the Markovian limit can be used as the starting point for influence matrices on the level of correlation functions.

Before going into the details of our approach, we provide a synopsis of the main ideas and results in the following. We include an overview of the main relevant concepts from free probability and their connection with quantum dynamics before introducing our results. For the sake of readability, we keep this section informal and use simplified notation only. We refer the reader to the subsequent sections for proper definitions and proofs, and to Appendix A for a more formal introduction to free probability.

A. Setting

To model physical systems with spatially local interactions, we consider a minimal solvable circuit of just three sites, inspired by recent work on deep thermalization [31]. The left site A models a small local subsystem of a spatially extended many-body system, while the right site E models its environment acting as an effective bath. To incorporate the effect of spatial locality, the sites A and E are coupled via a small intermediate site C , which can be thought of as a local bottleneck preventing the direct flow of quantum

information between A and E . At every discrete time step, the sites A and C interact via a fixed two-site unitary U , which can in principle be time dependent, whereas the sites C and E interact via a Haar-random unitary V_t at time t . More precisely, the evolution operator over a single discrete time step reads

$$\mathcal{U}_t = (\mathbb{1}_A \otimes V_t)(U \otimes \mathbb{1}_E). \quad (1)$$

The Haar-random unitaries V_t at different time steps are taken independently from each other, and we denote the evolution operator up to time t by

$$\mathcal{U}(t) = \mathcal{U}_t \dots \mathcal{U}_2 \mathcal{U}_1 = \quad (2)$$

In this model, the thermodynamic limit is obtained by taking the bath size to infinity, thereby causing $D \rightarrow \infty$ with D the Hilbert-space dimension of the full system.

The main focus of this work is the correlation functions between local observables A and B acting nontrivially only in the subsystem A . We specifically focus on generalized OTOCs of arbitrary order k (k -OTOCs; see, e.g., Ref. [87]) with respect to the infinite-temperature state defined by

$$C^{(k)}(t) = \frac{1}{D} \text{Tr}([A(t)B]^k), \quad (3)$$

with $A(t) = \mathcal{U}(t)A\mathcal{U}(t)^\dagger$ the time-evolved observable. For $k = 1$, the above reduces to a two-point function, and for $k = 2$ it returns the usual OTOC.

B. Free probability

We here briefly review the relevant concepts from free probability, which will help to relate our results on the dynamics of k -OTOCs to this more general framework. A more formal summary can be found in Appendix A, while we refer to Refs. [70,71,88] for a general exposition and details of all introduced concepts and to Ref. [61] for the connection with quantum dynamics. Readers familiar with free probability can skip to Sec. IC.

Free probability is the study of noncommuting random variables and extends familiar notions of independence and cumulants to the noncommutative setting, leading to so-called free independence and free cumulants. Noncommutativity and randomness are two fundamental

properties of quantum mechanics and, as highlighted in the Introduction, in recent years free probability has gained significant attention in the context of quantum dynamics. Quantum-mechanical observables already correspond to (generally noncommuting) operators, replacing the commuting phase-space variables in a classical probabilistic description. Computing expectation values by integrating with respect to a probability measure is likewise replaced by computing expectation values within a given quantum state.

While two generic quantum-mechanical observables generally do not commute, this does not imply that these are freely independent. The utility of free probability in quantum dynamics rather follows from the following conjecture: Under ergodic quantum dynamics and in thermodynamically large systems, time-evolved observables are expected to become freely independent from static observables at late times. Furthermore, the way in which they approach this free independence can be systematically quantified using free cumulants.

Free independence of two variables a and b , or freeness in short, is defined from the vanishing of all alternating joint moments $\varphi(abab \dots ab) = 0$ provided that $\varphi(a) = \varphi(b) = 0$, where we denote the expectation value as $\varphi(\bullet)$ following the mathematical literature. The alternating property is crucial and highlights the importance of non-commutativity: For commuting independent variables $\varphi(abab) = \varphi(a^2)\varphi(b^2) \neq 0$, even if $\varphi(a) = \varphi(b) = 0$. This definition naturally relates to OTOCs, since Eq. (3) is exactly of this form when we identify $a = A(t)$, $b = B$, and fix the infinite-temperature expectation value $\varphi(\bullet) = \text{Tr}(\bullet)/D$. The vanishing of all k -OTOCs at time $t \rightarrow \infty$ is hence equivalent to the emergence of free independence between $A(t)$ and B , i.e., between time-evolved observables at late times and static observables, and is referred to as asymptotic freeness. Our minimal model is rigorously shown to exhibit asymptotic freeness under ergodic dynamics, where we additionally fully characterize the relaxation dynamics.

1. Free cumulants

The approach to freeness can be quantified through free cumulants. Free cumulants κ_n satisfy the standard cumulative property $\kappa_n(a+b) = \kappa_n(a) + \kappa_n(b)$ whenever a and b are freely independent. The n th free cumulant $\kappa_n(a)$ can additionally be expressed in terms of the n th and lower moments of a , i.e., as a linear combination of $\varphi(a^m)$ with $m \leq n$.

Free cumulants are recursively defined through the combinatorics of noncrossing partitions, the fundamental objects in both free probability and full ETH [57]. A partition σ of $\{1, 2, \dots, k\}$ with blocks $V_1, V_2, \dots, V_{|\sigma|}$ is said to be noncrossing if, when the elements p_i, q_i and p_j, q_j belong to distinct blocks V_i and V_j , there are no “crossings” in which $p_i < p_j < q_i < q_j$. These noncrossing partitions can be represented graphically by

arranging the elements on the circle and connecting elements that are in the same block. For example, the partition (12)(34) is noncrossing and (13)(24) is crossing, as made explicit graphically:

$$(12)(34) = 4 \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ 2 \\ \diagdown \quad \diagup \\ 3 \end{array} 2, \quad (13)(24) = 4 \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ 2 \\ \diagdown \quad \diagup \\ 3 \end{array} 2. \quad (4)$$

The noncrossing partitions of four elements are listed below, where we drop the labels for ease of notation:

$$\begin{array}{c} \text{Diagram 1: A square with all four vertices connected to each other (K4).} \\ \text{Diagram 2: A square with two diagonals (K4 minus one edge).} \\ \text{Diagram 3: A square with one diagonal (K4 minus two edges).} \\ \text{Diagram 4: A square with no diagonals (K4 minus three edges).} \\ \text{Diagram 5: A square with two opposite vertices connected (K4 minus two edges).} \\ \text{Diagram 6: A square with one diagonal and one edge (K4 minus one edge).} \\ \text{Diagram 7: A square with two diagonals and one edge (K4).} \\ \text{Diagram 8: A square with two diagonals and two edges (K4).} \\ \text{Diagram 9: A square with two diagonals and three edges (K4).} \\ \text{Diagram 10: A square with two diagonals and four edges (K4).} \end{array} \quad (5)$$

At the bottom and top, we isolated two particularly important noncrossing partitions, where the elements are either all in separate blocks or all in the same block,

$$(1)(2)(3)(4) = 4 \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ 2 \\ \diagdown \quad \diagup \\ 3 \end{array} 2, \quad (1234) = 4 \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ 2 \\ \diagdown \quad \diagup \\ 3 \end{array} 2, \quad (6)$$

which we will refer to as the identity and the cyclic partition, respectively.

Free cumulants are implicitly defined from the relations

$$\varphi(a^n) = \sum_{\sigma \in \text{NC}(n)} \kappa_{\sigma}(a), \quad \kappa_{\sigma}(a) = \prod_{V \in \sigma} \kappa_{|V|}(a). \quad (7)$$

Here, $\kappa_{\sigma}(a)$ factorizes according to the blocks in the partition σ , with each block V returning the free cumulant $\kappa_{|V|}(a)$ set by the size $|V|$ of this block. For the first four moments, Eq. (7) returns

$$\varphi(a) = \kappa_1(a), \quad (8)$$

$$\varphi(a^2) = \kappa_2(a) + \kappa_1^2(a), \quad (9)$$

$$\varphi(a^3) = \kappa_3(a) + 3\kappa_2(a)\kappa_1(a) + \kappa_1^3(a), \quad (10)$$

$$\begin{aligned} \varphi(a^4) &= \kappa_4(a) + 2\kappa_2^2(a) + 6\kappa_2(a)\kappa_1^2(a) \\ &\quad + 4\kappa_3(a)\kappa_1(a) + \kappa_1^4(a), \end{aligned} \quad (11)$$

using that for $n = 1$ there is only the identity partition (1), for $n = 2$ there is the cyclic partition (12), and the identity partition (1)(2), etc. By inverting for κ_n , this leads to

$$\kappa_1(a) = \varphi(a), \quad (12)$$

$$\kappa_2(a) = \varphi(a^2) - \varphi(a)^2, \quad (13)$$

$$\kappa_3(a) = \varphi(a^3) - 3\varphi(a^2)\varphi(a) + 2\varphi(a)^3, \quad (14)$$

$$\begin{aligned} \kappa_4(a) &= \varphi(a^4) - 2\varphi(a^2)^2 - 4\varphi(a)\varphi(a^3) \\ &\quad + 10\varphi(a^2)\varphi(a)^2 - 5\varphi(a)^4. \end{aligned} \quad (15)$$

Standard cumulants can be defined in the same way, with the summation running over all partitions rather than just the noncrossing ones. The first three cumulants are identical to the standard cumulants, reflecting that for $n < 4$ all partitions of n elements are noncrossing, whereas for $n = 4$ the single crossing partition (13)(24) needs to be excluded. In Eq. (11), this change results in a prefactor 2 for the free cumulant term $\kappa_2^2(A)$ instead of 3 for the standard cumulant, leading to a corresponding change in Eq. (15).

It is more generally possible to define mixed free cumulants as

$$\varphi(a_1, a_2, \dots, a_n) = \sum_{\sigma \in \text{NC}(n)} \kappa_{\sigma}(a_1, a_2, \dots, a_n), \quad (16)$$

where we defined $\varphi(a_1, a_2, \dots, a_n) = \varphi(a_1 a_2 \dots a_n)$, and $\kappa_{\sigma}(a_1, a_2, \dots, a_n)$ factorizes according to the blocks of σ (for more details, we again refer to Appendix A). The free cumulants from Eq. (7) are obtained by setting $a_1 = \dots = a_n = a$. An alternative definition of free independence is then the vanishing of mixed free cumulants, i.e.,

$$\kappa_n(\dots, a, \dots, b, \dots) = 0, \quad (17)$$

whenever a and b are freely independent.

A major application of free probability is in calculating joint moments of freely independent operators whenever their individual expectation values do not vanish, i.e., $\varphi(a) \neq 0$ or $\varphi(b) \neq 0$, and expressing these in terms of free cumulants. As one example, free probability predicts that

$$\begin{aligned} \varphi(a_1 b_1 a_2 b_2 \dots a_k b_k) &= \sum_{\sigma \in \text{NC}(2k)} \kappa_{\sigma}(a_1, b_1, a_2, b_2, \dots, a_k, b_k) \\ &= \sum_{\sigma \in \text{NC}(k)} \kappa_{\sigma}(a_1, a_2, \dots, a_k) \varphi_{\sigma^*}(b_1, b_2, \dots, b_k), \end{aligned} \quad (18)$$

with σ^* another noncrossing partition, the so-called Kreweras complement of σ (see Appendix A).

Under the assumption of asymptotic freeness, at late times and in the thermodynamic limit the k -OTOC (3) admits the corresponding decomposition

$$\lim_{t \rightarrow \infty} C^{(k)}(t) = \sum_{\sigma \in \text{NC}(k)} \kappa_{\sigma}(A) \varphi_{\sigma^*}(B), \quad (19)$$

which is a limiting case of the more general expression of asymptotic freeness

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{D} \text{Tr}[A_1(t) B_1 A_2(t) B_2 \dots A_k(t) B_k] \\ = \sum_{\sigma \in \text{NC}(k)} \kappa_{\sigma}(A_1, A_2, \dots, A_k) \varphi_{\sigma^*}(B_1, B_2, \dots, B_k). \end{aligned} \quad (20)$$

One of the main results of this work is to establish this identity and the corresponding asymptotic freeness for the introduced minimal model. This stationary value reflects as much factorization of correlations and consequently as much independence between A and B as is allowed by free probability.

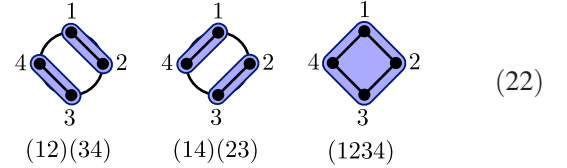
A key signature of quantum ergodicity is the emergence of random-matrix behavior in structured Hamiltonians or unitary evolution operators [3]. Asymptotic freeness here refines the expectation that at late times the unitary evolution operator $\mathcal{U}(t)$ resembles a random matrix. A standard example of freely independent variables is given by $D \times D$ matrices A and Haar-random rotated matrices $\mathcal{U}_{\text{Haar}} B \mathcal{U}_{\text{Haar}}^{\dagger}$, with $\mathcal{U}_{\text{Haar}}$ a Haar-random unitary matrix, in the limit $D \rightarrow \infty$. The unrotated matrices are freely independent from the rotated matrices, for generic choices of matrices A and B , as originally established in Ref. [89]. In the setting of quantum dynamics, the limit $D \rightarrow \infty$ effectively corresponds to the thermodynamic limit and late times, and asymptotic freeness implies that the unitary evolution operator $\mathcal{U}(t)$ becomes indistinguishable from a Haar-random unitary matrix $\mathcal{U}_{\text{Haar}}$ on the level of out-of-time-ordered correlation functions at late times [61]. While various signatures of random-matrix behavior have been analytically established for a wide range of models [32,83,90–93], it has not yet been shown how these lead to the emergence of free cumulants with nontrivial dynamics. Further applications of free probability include free convolution and the free central limit theorem, which can in turn be applied to quantum dynamics to identify further signatures of asymptotic freeness and corresponding universal random-matrix behavior.

However, both full ETH and our results build on free probability to go beyond purely establishing the steady-state behavior, reflecting that the unitary evolution operator remains distinct from a purely random matrix. The OTOC at finite times more generally allows a decomposition in terms of dynamical free cumulants,

$$C^{(k)}(t) = \sum_{\sigma \in \text{NC}(2k)} \kappa_{\sigma}(A, B, \dots, A, B; t), \quad (21)$$

where the terms in the rhs now factorize into time-dependent free cumulants according to the blocks of σ .

For concreteness, we consider the standard example where $k = 2$ and $\varphi(A) = \kappa_1(A) = 0 = \varphi(B)$. In this case, all the partitions which contain a singleton, i.e., a block of size one, give rise to a factor $\varphi(a) = 0$ and do not contribute. Consequently, one needs to consider only the partitions



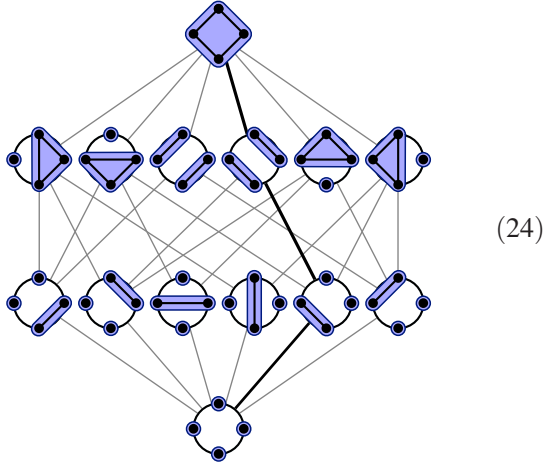
leading to the decomposition

$$\begin{aligned} C^{(2)}(t) &= \kappa_{(12)(34)}(A, B, A, B; t) + \kappa_{(14)(23)}(A, B, A, B; t) \\ &\quad + \kappa_{(1234)}(A, B, A, B; t) \\ &= \kappa_2(A, B; t) \kappa_2(A, B; t) + \kappa_2(A, B; t) \kappa_2(A, B; t) \\ &\quad + \kappa_4(A, B, A, B; t) \\ &= 2[\kappa_2(A, B; t)]^2 + \kappa_4(A, B, A, B; t). \end{aligned} \quad (23)$$

The decay of these mixed free cumulants, mixing both static and dynamic observables, then quantifies the approach to asymptotic freeness. The vanishing of free cumulants at increasingly higher orders encodes different orders of freeness (similar to the formation of approximate designs in deep thermalization) [61], and the corresponding timescales at which higher-order OTOCs and free cumulants decay have been a subject of recent study [54,59]. Within full ETH, these dynamical free cumulants are identified with specific time-resolved correlations between matrix elements of the operators A and B [60]. These correlations can be motivated by assuming that the unitary evolution operator resembles a random unitary matrix within each microcanonical energy shell only, rather than the full unitary resembling a random unitary matrix, and the time dependence of these free cumulants encodes correlations between different microcanonical energy shells [94]. We note, however, that this interpretation is lost here, since the minimal model does not support a meaningful notion of eigenstates. Remarkably, the dynamics still supports a decomposition in time-dependent free cumulants.

Before continuing to our results, we introduce two final notions that will be important throughout. The first is that the set of noncrossing partitions supports a partial ordering: For two noncrossing partitions ν and σ we say that $\nu \subseteq \sigma$ if and only if all blocks of ν are contained in the blocks of σ . This ordering can be illustrated in the so-called noncrossing

partition lattice, where $\nu \subset \sigma$ if ν lies in a row below σ and these are connected by lines. For example, for $k = 4$,



with the identity partition at the bottom and the cyclic partition at the top. With this partial ordering, it is possible to define geodesics as ordered sequences from the identity to the cyclic partition. One such geodesic is illustrated in Eq. (24) by a bold line corresponding to the sequence

$$(1)(2)(3)(4) \subset (1)(2)(34) \subset (12)(34) \subset (1234). \quad (25)$$

When writing $\nu \subseteq \sigma$, we will always implicitly assume that both σ and ν are noncrossing.

With this notation, it is possible to formally invert the definition of the free cumulants to write

$$\kappa_\sigma(a) = \sum_{\nu \subseteq \sigma} \varphi_\nu(a) \mu(\nu, \sigma), \quad (26)$$

with $\mu(\nu, \sigma)$ the so-called Möbius function on the noncrossing partition lattice. Specific values of the Möbius function can be read off from Eq. (15). Here we also introduced generalized moments $\varphi_\nu(a) \equiv \prod_{V \in \nu} \varphi_{|V|}(a) \equiv \prod_{V \in \nu} \varphi(a^{|V|})$, for which the definition of free cumulants generalizes to

$$\varphi_\nu(a) = \sum_{\sigma \subseteq \nu} \kappa_\sigma(a). \quad (27)$$

The Möbius function explicitly expresses free cumulants in terms of the moments and will be discussed in detail throughout the text. The relation between moments and free cumulants, Eqs. (26) and (27), is a key result from free probability and will naturally appear throughout.

C. Main results

1. Emergence of the noncrossing partition lattice

As a first step in characterizing the k -OTOC dynamics, we reveal the emergence of noncrossing partitions as an

organizing principle of OTOC dynamics in Sec. II. In the thermodynamic limit and after averaging over the Haar-random unitaries, we find that the k -OTOC can be expressed as a summation over different terms indexed by paths on the noncrossing partition lattice.

More specifically, we find that the k -OTOC can be expressed as a sum over terms indexed by all sequences of $2t$ nondecreasing partitions on a geodesic, so-called multichains. In this expansion of the averaged k -OTOC (3), each multichain $\sigma_1 \subseteq \nu_1 \subseteq \dots \subseteq \sigma_t \subseteq \nu_t$ defines a sequence of transfer matrices $\mathcal{M}_{\nu\sigma}$,

$$\mathcal{M}_{\nu_1\sigma_1} \dots \mathcal{M}_{\nu_{2t}\sigma_{2t}} \mathcal{M}_{\nu_{2t}\sigma_{2t}}. \quad (28)$$

This product of transfer matrices acts on k replicas of the subsystem A , with the noncrossing partitions encoding interactions between the replicas. The corresponding contribution to the k -OTOC is obtained by contracting this sequence of transfer matrices with k copies of the initial and final operators A and B , respectively. The full k -OTOC (3) corresponds to a weighted sum over all these contributions, with weight expressed in terms of the Möbius functions and the dimensionality of the different Hilbert spaces.

2. Timescale for asymptotic freeness

Using this expansion into multichains, in Sec. III we establish late-time asymptotic freeness as characterized for traceless observables by the vanishing of all k -OTOCs in the thermodynamic limit $\lim_{t \rightarrow \infty} \lim_{D \rightarrow \infty} C^{(k)}(t) = 0$ [95]. We achieve this by first showing that the contribution from each multichain exponentially decays in time. We identify the multichains leading to the slowest decay, which dominate the dynamics for all but the earliest times.

In the simplest case $k = 1$, there is only a single partition, the identity partition, and consequently only a single multichain. In this case, all dynamics are described by the unique transfer matrix associated with the identity partition. This transfer matrix is a unital single-particle quantum channel $\mathcal{M}$ acting on A . The subleading eigenvalue $\lambda < 1$ of the quantum channel dominates the late-time dynamics as

$$C^{(k=1)}(t) \propto \lambda^t. \quad (29)$$

For arbitrary $k > 1$, the combinatorics of noncrossing partitions guarantees that each of the transfer matrices $\mathcal{M}_{\nu\sigma}$ contains at least two copies of the channel $\mathcal{M}$ that can be factorized out. It is always possible to find multichains in which no additional factoring can be performed, and the decay is dominated by these two copies, and those multichains give rise to the slowest-decaying contribution to the expansion of the k -OTOC. They dominate the k -OTOC dynamics, yielding

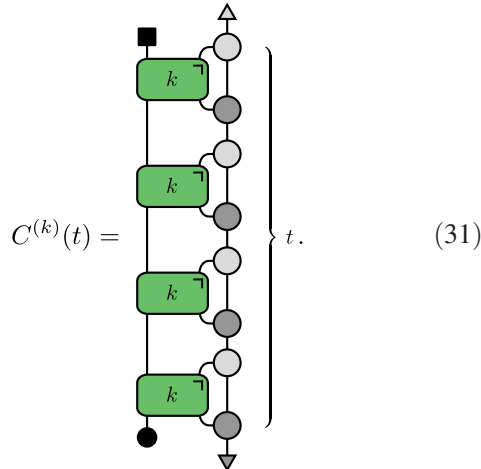
$$C^{(k \geq 2)}(t) \propto \lambda^{2t}; \quad (30)$$

i.e., all higher k -OTOCs decay exponentially with the same decay rate. More precisely, we obtain rigorous bounds on the dynamics of the k -OTOCs, which scale as the right-hand side of Eq. (30) up to polynomial corrections. This result demonstrates, in particular, that for $k > 2$ the k -OTOC cannot be decomposed as a product of two-point functions, as would be expected for free dynamics or Gaussian systems. A notable exception to the above behavior occurs when the subsystem A hosts a single qubit. In this case, the assumptions on the local observables imply further algebraic relations causing the k -OTOC to decay as $C^{(k)}(t) \propto \lambda^{kt}$.

In general, we establish late-time asymptotic freeness between local observables within this solvable model whenever the choice of unitary gates U results in ergodic dynamics for the two-point correlation functions. When choosing U to be dual unitary [96], i.e., also unitary when switching the roles of space and time, the dynamics is shown to be maximally ergodic corresponding to $\lambda = 0$, and we explicitly characterize the stability of this maximally ergodic point.

3. Influence matrix and Markov process

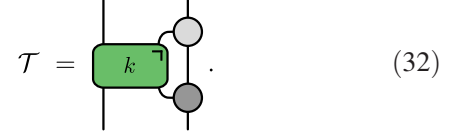
In Sec. IV, we provide an alternative point of view on k -OTOC dynamics in terms of a Markov process on the noncrossing lattice. This interpretation allows for both an efficient numerical evaluation of the k -OTOC and for an analytical characterization of the steady state and the transient dynamics. We obtain this picture by first introducing a convenient way to track the combinatorial structure of multichains in terms of a temporal matrix product state in a replica picture, whose bond degree of freedom corresponds to noncrossing partitions. This matrix product state provides a multireplica generalization of the usual influence matrix, extending this approach from time-ordered to out-of-time-ordered correlators. With this influence matrix, the averaged k -OTOC can be expressed as a tensor network of the form



$$C^{(k)}(t) = \text{[Diagram]} \quad (31)$$

Compared with Eq. (2), the two-site tensors on the left (rectangles) now indicate k replicas of the two-site unitary U , and the tensors on the right (circles) define the matrix-product-state representation of the influence matrix. Its bond dimension is constant in time and equals the number of noncrossing partitions of k elements.

This tensor network can be efficiently contracted by identifying a transfer operator $\mathcal{T}$ (generator) in the temporal direction



$$\mathcal{T} = \text{[Diagram]} \quad (32)$$

Its left legs act on k copies of the subsystem A , while the right legs act on the noncrossing degrees of freedom. The full eigenspectrum of this generator can be analytically obtained, and we explicitly construct the eigenstates corresponding to the leading and the dominant subleading eigenvalues. The leading eigenvalue 1 encodes the stationary state (20), whereas the subleading eigenvalue equals λ and encodes the approach toward equilibrium (21).

Concretely, the generalized moments appear as the leading right eigenvectors of $\mathcal{T}$, with free cumulants as the corresponding biorthogonal left eigenvectors. The stationary value, Eq. (20), is the projection onto the leading eigenspace expressed in terms of the above eigenvectors. The eigenstates corresponding to the dominant subleading eigenvalues governing the decay to equilibrium are shown to exhibit the combinatorics of mixed free cumulants, mixing a single static observable with dynamic observables (or vice versa), leading to the dynamical decomposition (21). For $k > 1$, the boundary conditions at initial and final time set by traceless A and B do not overlap with the leading and subleading eigenspaces, giving rise to the next-to-subleading eigenvalue λ^2 dominating the decay as $C^{(k)}(t) \sim \lambda^{2t}$.

To summarize, we recover the full structure predicted by free probability from the spectral properties of the Markov process, including the decomposition of k -OTOCs into dynamic free cumulants and their decay to the expected steady-state value. This expansion corresponds exactly to the prediction from full ETH [55,57,60,63]. These results are argued to remain valid when adding additional spatial structure to either the subsystem or effective bath. This stability of the late-time behavior indicates that the minimal circuit model captures the essential dynamics of multitime correlation functions in strongly ergodic systems.

II. SOLVABLE CIRCUIT MODEL AND k -OTOCs

We here present a more detailed introduction of the model and higher-order OTOCs, fixing notation along the way. The replica picture is introduced as a convenient

representation for the k -OTOCs dynamics, and we show how averaging over the bath dynamics returns an expansion for the k -OTOCs in terms of paths on the noncrossing partition lattice.

A. Time evolution

The minimal circuit model considered here was introduced as a minimal solvable model for deep thermalization in Ref. [31]. The model is built from a lattice of just three sites, modeling a local subsystem A of a many-body system and its complement E , with intermediate site C mediating the local coupling between A and E . The local Hilbert spaces associated with the three sites are $\mathcal{H}_A = \mathbb{C}^{d_A}$, $\mathcal{H}_C = \mathbb{C}^{d_C}$, as well as $\mathcal{H}_E = \mathbb{C}^{d_E}$ with dimensions $d_A, d_C \ll d_E$. The full Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_C \otimes \mathcal{H}_E$ has dimension $D = d_A d_C d_E$. We keep d_A and d_C fixed, while allowing $d_E \rightarrow \infty$ and hence $D \rightarrow \infty$ to model the thermodynamic limit. We denote the canonical (computational) basis in $\mathcal{H}_x$ by $|i\rangle_x$ for $i = 1, \dots, d_x$ and $x = A, C, E$.

Local interactions between A and C are modeled by a fixed unitary two-qudit gate U of dimension $d_A d_C \times d_A d_C$, and the interactions between C and E by independently drawn Haar-random $d_C d_E \times d_C d_E$ unitaries V_t at each time step t . Graphically, we denote the unitary gates as

$$U = \begin{array}{|c|} \hline \text{[Green box with } U \text{]} \\ \hline \end{array} \begin{array}{|c|} \hline \text{[Green box with } U \text{]} \\ \hline \end{array}, \quad V_t = \begin{array}{|c|} \hline \text{[Blue box with } V_t \text{]} \\ \hline \end{array} \begin{array}{|c|} \hline \text{[Blue box with } V_t \text{]} \\ \hline \end{array}. \quad (33)$$

Here the labels indicate the relevant Hilbert space. Evolution for the t th time step is then given by the minimal two-layer brickwork circuit $\mathcal{U}_t$ from Eq. (1). For a fixed realization of the circuit corresponding to a fixed sequence of independent Haar-random unitaries $\{V_1, \dots, V_t\}$, the evolution operator $\mathcal{U}(t)$ is given in terms of Eq. (2) as

$$\mathcal{U}(t) = \begin{array}{|c|} \hline \text{[Blue box with } V_t \text{]} \\ \hline \text{[Green box with } U \text{]} \\ \hline \vdots \\ \hline \text{[Green box with } U \text{]} \\ \hline \text{[Blue box with } V_1 \text{]} \\ \hline \text{[Green box with } U \text{]} \\ \hline \end{array}. \quad (34)$$

The resulting model for time evolution is not homogeneous in time and should be thought of as a nonperiodically driven system. It is not a Floquet system, and hence there is no notion of eigenstates, rendering concepts from ETH and full ETH ill-defined. The dynamical picture in terms of free cumulants put forward by full ETH, however, persists for the present model.

B. Higher-order OTOCs

We consider generalized OTOCs of order k defined by Eq. (3) for local observables A and B acting nontrivially only on the subsystem A . These observables are determined by Hermitian observables a and b acting on $\mathcal{H}_A$ which are embedded into the full system as $A = a \otimes \mathbb{1}_C \otimes \mathbb{1}_E$ and $B = b \otimes \mathbb{1}_C \otimes \mathbb{1}_E$, respectively. Unless stated otherwise, we assume the observables to be traceless, $\text{Tr}(a) = \text{Tr}(b) = 0$. We define evolution in time as

$$A(t) = \mathcal{U}(t) A \mathcal{U}(t)^\dagger. \quad (35)$$

Note that this dynamics corresponds to that of density matrices rather than observables in the Heisenberg picture, which will help to simplify the following expressions.

We will also consider more general k -OTOCs defined between k time-evolved observables $A_1(t), \dots, A_k(t)$ and k static observables $B_1, \dots, B_k$ built from k -tuples of local observables $a = (a_1, \dots, a_k)$ and $(b_1, \dots, b_k)$, respectively. The corresponding k -OTOC reads

$$c_{ab}^{(k)}(t; d_E) = \frac{1}{D} \text{Tr}[A_1(t) B_1 A_2(t) B_2 \dots A_k(t) B_k]. \quad (36)$$

While we dropped the dependence on the dimension of the environment in $\mathcal{U}(t)$ as well as in A_i, B_i , we explicitly include it in the single-realization k -OTOC. Originating from a random time evolution, the k -OTOC is a random variable itself, and we are generally interested in its expectation value when averaged over the independent Haar-random unitaries V_t in the thermodynamic limit $d_E \rightarrow \infty$. We denote this limit as

$$C_{ab}^{(k)}(t) = \lim_{d_E \rightarrow \infty} \mathbb{E}[c_{ab}^{(k)}(t; d_E)]. \quad (37)$$

We emphasize that we take the thermodynamic limit at an arbitrary but fixed time and only afterward study the late-time behavior. This order of limits, thermodynamic limit first and late time second, is crucial as for finite systems the k -OTOCs for both a single realization and on average are expected to exhibit residual finite-size fluctuations even after relaxation. Note also that we take the limit of the finite-size OTOCs, i.e., a limit of ordinary real or complex numbers, rather than explicitly constructing observables for the infinite system and computing OTOCs between them. This former type of limit is the usual limit considered in free probability theory and gives rise to a convergence in distribution. In Appendix B, we additionally prove the stronger almost sure convergence, ensuring that the above average describes typical circuit realizations as well.

C. Replica picture

A convenient choice for representing the k -OTOCs diagrammatically and for analytical considerations is the so-called folded or replica picture. Graphically, the folded

picture is obtained by folding all copies (replicas) of $\mathcal{U}(t)$ and its inverse on top of each other in an alternating fashion. This folding gives rise to a single evolution operator

$$\mathcal{U}^{(k)}(t) = (\mathcal{U}(t) \otimes \mathcal{U}(t)^*)^{\otimes k}, \quad (38)$$

with $*$ denoting complex conjugation, acting on k replicas, each of which consists of a forward and a backward time sheet. This k -replica evolution operator is built from the k -folded gate

$$U^{(k)} = \frac{1}{d_C^k} (U \otimes U^*)^{\otimes k}, \quad (39)$$

or, graphically,

$$k = \frac{1}{d_C^k} U, \quad (40)$$

here illustrated for $k = 3$. Note that we include an unconventional prefactor d_C^k in the folded gate in order to simplify all following graphical expressions. The green (red) gates denote U (U^*). A folded version of V_i can be similarly defined.

In this representation, the k -OTOC can be obtained as a matrix element of $\mathcal{U}^{(k)}(t)$ with respect to an initial and a final state on $\mathcal{H}^{\otimes 2k}$, which encode both the observables A and B as well as how the replicas are connected. Graphically [97],

$$c_{ab}^{(k)}(t; d_E) \propto$$

(41)

The boundary conditions, here represented by squares and circles, can be expressed in terms of permutation states. We denote the product basis on $\mathcal{H}_x^{\otimes 2k}$ by $|i_1, i'_1, \dots, i_k, i'_k\rangle_x$ for $x = A, C, E$ and introduce the unnormalized permutation state $|\sigma\rangle$ for a permutation $\sigma \in S_k$ of k elements via

$$(i_1, i'_1, \dots, i_k, i'_k | \sigma) = \delta_{i_{\sigma(1)}, i'_1} \dots \delta_{i_{\sigma(k)}, i'_k}. \quad (42)$$

Graphically, such a permutation state connects the backward sheet of the i th replica with the forward sheet of the $\sigma(i)$ th replica, where, e.g., for $\sigma = (12)(3)$,

$$|\sigma = (12)(3)\rangle = \textcircled{\sigma} = \text{diagram with a circle containing } \sigma \text{ and a line entering from the top, branching into two lines exiting to the right.} \quad (43)$$

Note that we denote permutations using cycle notation. We write the identity permutation as $\circ = (1)(2)\dots(k)$ and the cyclic permutation as $\square = (123\dots k)$, with corresponding states of the form

$$|\circ\rangle = \text{circle with a vertical line} = \text{diagonal lines}, \quad |\square\rangle = \text{square with a vertical line} = \text{diagonal lines}, \quad (44)$$

again illustrated for $k = 3$ for concreteness. Permutation states are neither normalized nor orthogonal, since for $\sigma, \nu \in S_k$ one has

$$_x(\sigma|\nu)_x = d_x^{|\sigma^{-1}\nu|}, \quad (45)$$

with $|\pi|$ the number of cycles of a permutation $\pi \in S_k$.

The relevant boundary vectors for the OTOC in Eq. (41) are obtained by dressing these permutation states with operators as

$$| \bullet \rangle = | a_\sigma \rangle_A = (a_1 \otimes \mathbb{1}_A \otimes \cdots \otimes a_k \otimes \mathbb{1}_a) | \sigma \rangle_A, \quad (46)$$

$$\blacksquare =_A (b_\sigma | =_A (\square | (\mathbb{1}_A \otimes b_1 \otimes \cdots \otimes \mathbb{1}_A \otimes b_k), \quad (47)$$

and similarly for arbitrary permutations σ . In terms of matrix elements, these read

$$_A(i_1, i'_1, \dots, i_k, i'_k | a_\circ)_A = (a_1)_{i_1 i'_1} (a_2)_{i_2 i'_2} \dots (a_k)_{i_k i'_k}, \quad (48)$$

$$_A(b_{\square}|i_1, i'_1, \dots, i_k, i'_k)_A = (b_1)_{i'_1 i_2} (b_2)_{i'_2 i_3} \dots (b_k)_{i'_k i_1}, \quad (49)$$

with $a_{ij} =_A \langle i|a|j\rangle_A$ and $b_{ij} =_A \langle i|b|j\rangle_A$.

D. Averaged k -OTOC

The averaging over unitary gates V_i to obtain the averaged k -OTOC in the thermodynamic limit [Eq. (37)] can now be performed using the Weingarten calculus [98,99]. This calculation is detailed in Appendix B, and we here first state the final result and then sketch the derivation. While this averaging involves standard methods, care needs to be taken due to the nontrivial interplay between the overlaps between the permutations themselves

with the boundary conditions corresponding to the identity and the cyclic permutation [cf. Eq. (45)]. The final expression is given by

$$C_{ab}^{(k)}(t) = d_A^{-1} d_C^{k-1} \sum_{\substack{\sigma_1, \nu_1 \in \text{NC}(k) \\ \dots \subseteq \sigma_i \subseteq \nu_i \subseteq \sigma_{i+1} \subseteq \dots}} \left(\prod_{i=1}^{t-1} W_{\nu_i \sigma_{i+1}} \right) \times (b_{\square} | \mathcal{M}_{\nu_1 \sigma_1} \dots \mathcal{M}_{\nu_{2t} \sigma_{2t}} \mathcal{M}_{\nu_1 \sigma_1} | a_{\circ}). \quad (50)$$

Each term in this summation is indexed by a set of permutations $\sigma_1, \nu_1, \dots, \sigma_t, \nu_t$. These permutations are constrained to be noncrossing. We can identify permutations with partitions by choosing the blocks of the partition as the cycles of the permutation, and we will freely switch between these interpretations. As mentioned above, noncrossing partitions support a partial ordering, where $\nu \subseteq \sigma$ if all cycles of ν are contained in the cycles of σ . In Eq. (50), the permutations are further restricted to so-called multichains on the noncrossing partition lattice, i.e., geodesics connecting the identity with the cyclic permutation as

$$\circ = \sigma_1 \subseteq \nu_1 \subseteq \sigma_2 \subseteq \nu_2 \dots \subseteq \sigma_t \subseteq \nu_t = \square. \quad (51)$$

The factors $W_{\nu\sigma}$ originate from the Weingarten calculus as

$$W_{\nu\sigma} = d_C^{-k+|\nu^{-1}\sigma|} \mu(\nu, \sigma), \quad (52)$$

where $\mu(\nu, \sigma)$ is shorthand notation for

$$\mu(\nu, \sigma) = \prod_{V \in \nu^{-1}\sigma} (-1)^{|V|-1} C_{|V|-1}, \quad (53)$$

with V the different cycles of $\nu^{-1}\sigma$. Crucially, this function is a Möbius function for the noncrossing partition lattice, satisfying

$$\sum_{\sigma \subseteq \rho \subseteq \nu} \mu(\nu, \rho) = \delta_{\nu\sigma}, \quad (54)$$

which can also be taken as its defining property. From the above definition, it also directly follows that the Möbius function encodes the transformation from moments to free cumulants; cf. Eq. (26). This connection will be directly responsible for the appearance of free cumulants in the OTOC dynamics.

The remaining terms $\mathcal{M}_{\nu\sigma}$ are nonexpanding linear maps acting on the space of k replicas [as in Eqs. (46) and (47)] defined as

$$\mathcal{M}_{\nu\sigma} = \text{diagram of a box labeled } k \text{ with } \nu \text{ and } \sigma \text{ inputs/outputs}. \quad (55)$$

Written out explicitly, these maps are given by

$$\mathcal{M}_{\nu\sigma} = d_C^{-k} (1_A^{\otimes 2k} \otimes_C |\nu\rangle)(U \otimes U^*)^{\otimes k} (1_A^{\otimes 2k} \otimes |\sigma\rangle_C). \quad (56)$$

We now sketch how this averaging is performed, highlighting the appearance of noncrossing permutations and multichains. The Haar average over $V^{(k)}$ projects onto the subspace spanned by all permutation states $|\sigma\rangle_{CE}$ on sites C and E as

$$\mathbb{E}[V_i^{(k)}] = \sum_{\nu, \sigma \in S_k} W(\nu, \sigma; d) \begin{array}{c} \text{diagram of } V_i^{(k)} \end{array}, \quad (57)$$

with W the Weingarten function [100]. For large dimension d_E , the Weingarten functions can be asymptotically expressed as

$$W(\nu, \sigma; d = d_E d_C) = (d_E d_C)^{-2k+|\nu^{-1}\sigma|} \mu(\nu, \sigma) \quad (58)$$

up to relative errors of order d_E^{-2} . Replacing each k -fold folded gate $V_i^{(k)}$ in Eq. (41) by its Haar average, the averaged k -OTOC becomes a sum over contributions labeled by $2t$ -tuples of permutations $\circ = \sigma_1, \nu_1, \dots, \sigma_t, \nu_t = \square \in S_k$ of the form

$$\sum_{\circ} \prod_{i=1}^{t-1} W(\nu_i, \sigma_{i+1}; d = d_E d_C) \text{diagram of the chain of gates and permutations}. \quad (59)$$

Collecting powers of d_E , each $2t$ -tuple of permutations contributes a term with a prefactor scaling as $d_E^{\Delta \ell(\sigma_1, \dots, \nu_t)}$, where

$$\Delta \ell(\sigma_1, \dots, \nu_t) = -2k(t-1) - 1 + \sum_{i=1}^t |\sigma_i^{-1} \nu_i| + \sum_{i=1}^{t-1} |\nu_i^{-1} \sigma_{i+1}|. \quad (60)$$

The first sum originates from the overlaps ${}_E(\sigma_i | \nu_i)_E$ in Eq. (59), and the second sum stems from the scaling of the Weingarten functions (58).

Equation (60) can be given a geometric interpretation. The Cayley distance between two permutations σ and ν , i.e., the number of transpositions it takes to convert one

permutation into another, is given by $\ell(\sigma, \nu) = k - |\sigma^{-1}\nu|$. We can hence identify $\ell(\sigma_1, \dots, \nu_t) = \ell(\circ = \sigma_1, \nu_1) + \ell(\nu_1, \sigma_2) + \dots + \ell(\sigma_t, \nu_t = \square)$ as the Cayley length of the path which starts at $\circ$, ends at $\square$, and visits all intermediate permutations σ_i and ν_i in the prescribed order. Crucially, an alternative definition of noncrossing permutations is that they lie on a geodesic from the identity to the cyclic permutation [101]: A permutation σ is noncrossing if and only if it satisfies

$$\ell(\circ, \sigma) + \ell(\sigma, \square) = \ell(\circ, \square) = k - 1, \quad (61)$$

which is the minimal value the lhs can take. This is a statement of the more general triangle inequality for (noncrossing) permutations,

$$\ell(\sigma, \rho) + \ell(\rho, \nu) \leq \ell(\sigma, \nu), \quad (62)$$

with equality if and only if ρ lies on a geodesic from σ to ν , i.e., $\sigma \subseteq \rho \subseteq \nu$. Combining these, we find that

$$\Delta\ell(\sigma_1, \dots, \nu_t) = \ell(\circ, \square) - \ell(\sigma_1, \dots, \nu_t) \leq 0, \quad (63)$$

with equality if and only if all permutations are noncrossing and form a geodesic from the identity permutation $\circ$ to the cyclic permutation $\square$, i.e., a multichain [101]. In the limit $d_E \rightarrow \infty$, only these terms are nonvanishing and we recover Eq. (50).

In the thermodynamic limit, the sum over $(k!)^{2t}$ permutations arising within the Weingarten calculus hence simplifies to a sum over approximately t^{k-1} [102] multichains on the noncrossing lattice. Not only does this observation vastly reduce the complexity of the problem, it also gives a first hint toward the free-probability structure underlying the k -OTOC dynamics. This connection is strengthened by the observation that the Weingarten functions, when restricted to noncrossing permutations, directly return the Möbius functions appearing in the definition of free cumulants.

While we focused on the averaged k -OTOC, the concentration properties of the Haar measure imply that typical individual realizations of the circuit are also described by Eq. (50) (see Appendix B for details). Remarkably, Eq. (50) remains valid when allowing for additional structure in the environment. Replacing the Haar-random unitaries by a random brickwork circuit with large local dimension leaves the k -OTOCs unchanged, hinting toward a more general applicability of the presented structure (see Appendix E).

The remainder of this work analyzes Eq. (50) and its implications. In Sec. III, we establish the late-time behavior of the k -OTOCs on the level of individual multichains, identifying the slowest-decaying contributions and their combinatorial interpretation. In this way, we present a physical picture of the decay channels of the k -OTOC and the effect of increasing k . In Sec. IV, we show how the full summation can be recast in an influence-matrix formalism, resulting in a Markovian picture for the

k -OTOC. We explicitly recover the structure of free probability: The eigenmodes of the Markovian transfer matrix can be identified with (mixed) free cumulants, returning the dynamical behavior postulated by full ETH.

Before doing so, it is instructive to compare the structure of Eq. (50) with the results of Ref. [31] on deep thermalization in this model, which helps to make clear the differences between deep and full thermalization. In the context of deep thermalization, projective measurements are performed on the environment E , where the resulting projected ensemble of states on A is expected to form a maximally Haar-random ensemble at late times [30]. In order to form a k -design, the first k moments of this projected ensemble need to agree with the moments of the Haar ensemble expressed in terms of permutations of the form (42). Within deep thermalization, there is no preferred “ordering” between the k replicas, such that the relevant symmetry group is S_k . Within full thermalization, however, the different replicas have a preferred ordering and are only invariant under shifts: The choice of k -OTOC as observable introduces boundary conditions corresponding to the identity and the cyclic permutation. The relevant permutations are restricted to be noncrossing $\text{NC}(k)$, since noncrossing permutations can be connected to the identity and cyclic permutation through a minimal number of transpositions, minimizing the (Cayley) distance to these boundaries [i.e., satisfying a geodesic condition, also made apparent in the noncrossing partition lattice (24)]. The noncrossing partitions are additionally required to change between different time steps, starting at the identity and ending at the cyclic permutation, thereby giving rise to the summation over multichains in Eq. (50) and a nontrivial interaction between the different replicas. In contrast, in deep thermalization, the permutations do not change in time, and for a fixed permutation, the replicas do not interact.

We also note that a similar expression involving paths on the noncrossing partition lattice appeared in Ref. [75] for k -OTOCs in dual-unitary circuits, showing how these OTOCs spread with a maximum velocity, using a similar approach as in Ref. [85] for 2-OTOCs. Eigenmodes of a light-cone transfer matrix were identified and labeled by paths on the noncrossing partition lattice, but the complicated overlaps between these eigenmodes required taking the limit of an infinite-dimensional Hilbert space for each site and no connection with full ETH was made.

III. FREE INDEPENDENCE AT LATE TIMES

We now consider the dynamical behavior of the averaged k -OTOC at late times, which will also help to make clear the role of the different terms in Eq. (50). In particular, we show that all k -OTOCs between traceless observables exponentially decay to zero at late times with the same rate if $k \geq 2$. We derive these results from the combinatorial structure of multichains in the noncrossing lattice and their

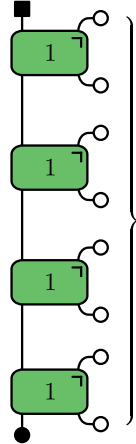
interplay with the quantum channels $\mathcal{M}_{\nu\sigma}$ and their spectra. We moreover characterize the steady state of the k -OTOC dynamics for arbitrary observables. In the free-probability language, the decay of the k -OTOCs gives rise to free independence between time-evolved and static observables. Correspondingly, the steady state is described by the correlations between freely independent observables.

A. Example: $k = 1$

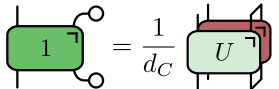
We first consider the simplest example $k = 1$ for which the k -OTOC simplifies to a standard (time-ordered) correlation function,

$$C_{ab}^{(k=1)}(t) = \lim_{d_E \rightarrow \infty} \mathbb{E} \frac{1}{D} \text{Tr}[A(t)B]. \quad (64)$$

In this limit, the underlying combinatorial structure simplifies considerably as there is only a single (non-crossing) partition $\circ = \square = (1)$ and a single corresponding quantum channel $\mathcal{M} = \mathcal{M}_{\circ\circ}$. There is hence only a single multichain, and all terms from the Weingarten function yield a factor of $W_{\circ\circ} = 1$. Consequently, Eq. (50) reduces to

$$C_{ab}^{(k=1)}(t) = \frac{1}{d_A} \text{Tr}[b \mathcal{M}^t(a)] = \frac{1}{d_A} \quad (65)$$


All dynamics follow from the action of $\mathcal{M}$ as defined by

$$\mathcal{M} = \text{[diagram of box 1]} = \frac{1}{d_C} \text{[diagram of box U]}. \quad (66)$$


Written out explicitly, $\mathcal{M}$ acts as

$$\mathcal{M}(a) = \frac{1}{d_C} \text{Tr}_C[U(a \otimes \mathbb{1}_C)U^\dagger] \quad (67)$$

corresponding to a unital quantum channel, i.e., a trace-preserving completely positive map which preserves the identity [103]. The long-time behavior of the correlation function depends on the spectral properties of this quantum channel. From unitality, we have that $\mathcal{M}(\mathbb{1}_A) = \mathbb{1}_A$, leading to a guaranteed eigenoperator with eigenvalue 1. All other eigenvalues are bounded in magnitude by 1.

If the identity is the only eigenoperator with eigenvalue with modulus 1, at late times we can replace $\mathcal{M}$ with a projector on this operator, leading to thermalization to the infinite-temperature state $\mathbb{1}_A$. If the operators are traceless, this projector vanishes since $\text{Tr}(\mathbb{1}_A a) = \text{Tr}(a) = 0$. If the operators are not traceless, the steady state follows from this projector as

$$\lim_{t \rightarrow \infty} C_{ab}^{(k=1)}(t) = \frac{1}{d_A^2} \text{Tr}(a) \text{Tr}(b) = \varphi(a) \varphi(b), \quad (68)$$

with $\varphi(\bullet) = \text{Tr}(\bullet)/d_A$ the normalized trace.

The dynamics of eigenoperators of $\mathcal{M}$ is particularly straightforward. For a right eigenoperator a_λ with eigenvalue λ , i.e., $\mathcal{M}(a_\lambda) = \lambda a_\lambda$, we can choose $a = a_\lambda$ and $b = b_\lambda$ the corresponding left eigenoperator to obtain the correlation function as

$$C_{ab}^{(k=1)}(t) = \frac{1}{d_A} \text{Tr}[b_\lambda \mathcal{M}^t(a_\lambda)] = \frac{1}{d_A} \lambda^t \text{Tr}[b_\lambda a_\lambda] = \lambda^t, \quad (69)$$

where we have taken this eigenoperator to be Hilbert-Schmidt normalized as $\text{Tr}(b_\lambda a_\lambda)/d_A = 1$. If $|\lambda| < 1$, this autocorrelation function decays to the zero ergodic value. General traceless local observables a and b can be expanded in eigenoperators of $\mathcal{M}$ that are necessarily orthogonal to $\mathbb{1}_A$, such that the largest (in modulus) of these eigenvalues dominates the late-time behavior. In the following, we will always denote this eigenvalue by λ . Alternatively, the above can be formulated in terms of the operator norm of $\mathcal{M}$ restricted to the space orthogonal to $\mathbb{1}_A$, i.e., the space of traceless operators. For normal $\mathcal{M}$ and real $\lambda \geq 0$ both coincide and give rise to an upper bound for traceless a and b :

$$|C_{ab}^{(k=1)}(t)| \leq \lambda^t \|a\|_2 \|b\|_2, \quad (70)$$

with $\|\cdot\|_2$ the Hilbert-Schmidt (Frobenius) norm of the local observables a and b . For non-normal $\mathcal{M}$, the eigenvalue λ needs to be replaced by an appropriate norm. For simplicity, we will formulate all of the following discussion in terms of the subleading eigenvalue, which we also assume to be unique, real, and nonnegative. We note, however, that our results hold in full generality upon replacing λ by the appropriate norm.

The dynamics is ergodic and mixing if λ is strictly smaller than 1. While for concreteness we have taken all gates U to be identical, this conclusion directly extends to the case where U is time dependent, as long as a nonvanishing fraction of the gates are sufficiently ergodic. A similar quantum channel construction appears in the calculation of light-cone correlation functions in dual-unitary circuits, for which the inside of the causal light cone acts as a random environment, and the ergodicity of dual-unitary circuits can be classified in terms of the eigenspectrum of a similar quantum channel [104]. The same classification applies in this model and we repeat it here for convenience.

- (i) Noninteracting: All d_A^2 eigenvalues equal 1. All dynamical correlations remain constant.
- (ii) Nonergodic: More than one but less than d_A^2 eigenvalues are equal to 1; some dynamical correlations decay to a nonzero constant.
- (iii) Ergodic and nonmixing: All nontrivial eigenvalues are different from 1, but there exists at least one nontrivial eigenvalue with unit magnitude. All correlations oscillate around a time-averaged value corresponding to the ergodic value.
- (iv) Ergodic and mixing: All nontrivial eigenvalues lie within the unit disk, and all dynamical correlations decay (exponentially) to their ergodic value.
- (v) Maximally ergodic: All $d_A^2 - 1$ nontrivial eigenvalues equal zero. Dynamical correlations decay instantaneously to the ergodic value.

As an aside, we note that it is possible to obtain a condition on the operator entanglement of the unitary U that guarantees mixing and, when maximized, maximally ergodic dynamics. Choosing $d_A = d_C = d$ for convenience, the operator Schmidt decomposition of U is given by

$$U = \sum_{j=1}^{d^2} \sqrt{\gamma_j} X_j \otimes Y_j, \quad (71)$$

where X_j and Y_j are single-site operators acting on A and C , respectively, orthonormalized such that $\text{Tr}(X_j^\dagger X_k) = \text{Tr}(Y_j^\dagger Y_k) = \delta_{jk}$. The linear entropy of the gate U , or operator entropy in short, is defined as [105–108]

$$E(U) = 1 - \frac{1}{d^4} \sum_{j=1}^{d^2} \gamma_j^2 = 1 - \frac{1}{d^2} \left[\text{diagram of a box labeled 2 with four external legs} \right], \quad (72)$$

where the folded gate acts on two replicas, such that $\circ = (1)(2)$ and $\square = (12)$.

This operator entropy is bounded as $0 \leq E(U) \leq 1 - 1/d^2$. As shown in Appendix C, in general and in Ref. [108] for the case of equal Hilbert-space dimensions, there exists a threshold value $E_* \equiv 1 - 2/d^2$ such that the channel is guaranteed to be mixing if $E(U) > E_*$. Intuitively, if the operator is sufficiently entangled, at each time step quantum information is guaranteed to “leak” into the environment resulting in correlations that decay to the ergodic value. The proof is based on a similar proof from Ref. [106] bounding the ergodicity of dual-unitary gates in terms of their entangling power.

Furthermore, for the maximal value $E(U) = 1 - 1/d^2$ the corresponding quantum channel is guaranteed to be maximally ergodic. In this limit, U is dual unitary, which enforces all information to leak into the environment after a single time step, consistent with the use of dual-unitary circuits as minimal models of maximal quantum chaos (for a recent review, see Ref. [96]). Dual-unitary gates are

unitary along the horizontal (spatial) direction [104,109], which can be graphically represented as

$$\left[\text{diagram of a box labeled 1 with four external legs} \right] \propto \left[\text{diagram of two vertical lines with circles at the ends} \right]. \quad (73)$$

This condition directly implies that $\mathcal{M}$ is a projector on the identity. Interestingly, the above bound implies that the ergodic behavior of dual-unitary gates is stable for gates “close to” dual unitarity, where the proximity to dual unitarity is quantified by the operator entanglement. The operator entanglement previously appeared as a measure for proximity to dual unitarity in the study of 2-OTOCs [110], similar to the appearance of the closely related entanglement velocity in the study of entanglement dynamics [111]. As discussed in the remainder of this paper, ergodicity on the level of correlation functions guarantees freeness at late times, such that ergodicity on the level of k -OTOCs is structurally stable near dual unitarity. The structural stability of dual unitarity is more generally an open question [112,113].

In the remainder of this paper, we focus exclusively on the ergodic and mixing case corresponding to a subleading eigenvalue $\lambda < 1$.

B. Example: $k = 2$

We now consider $k = 2$, returning the usual OTOC

$$C_{ab}^{(k=2)}(t) = \lim_{d_E \rightarrow \infty} \mathbb{E} \frac{1}{D} \text{Tr}[A(t)BA(t)B], \quad (74)$$

and show how the underlying combinatorial structure first emerges. In this case, there are two partitions, both non-crossing, $\circ = (1)(2)$ and $\square = (12)$, and multichains are characterized by domain walls between $\circ$ and $\square$. The resulting 2-OTOC follows as a sum over such domain-wall configurations

$$C_{ab}^{(k=2)}(t) = \frac{d_C}{d_A} \sum_{s=1}^t \left[\text{diagram of a chain of 2-gates with } t-s \text{ domain walls} \right] - \frac{1}{d_A} \sum_{s=1}^{t-1} \left[\text{diagram of a chain of 2-gates with } s \text{ domain walls} \right]. \quad (75)$$

The first sum runs over domain walls at half-integer time steps characterized by $\sigma_i = \circ$ and $\nu_i = \square$, whereas the second sum runs over domain walls at integer time steps corresponding to $\nu_i = \circ$ and $\sigma_{i+1} = \square$. The latter are accompanied by a nontrivial factor of $W_{\circ\square} = -1/d_C$. Such domain-wall configurations are ubiquitous in the study of 2-OTOCs in quantum-circuit models, where for random quantum circuits the dynamics of the 2-OTOC was originally understood through an effective model describing the dynamics of such domain walls as a biased random walk [82,114,115].

Similar to what was discussed in the previous section, the 2-OTOC decays to zero in the long-time limit provided the quantum channel $\mathcal{M}$ is ergodic and the operators are traceless. This is again best seen when taking a and b be a right and left eigenoperator of the single particle channel $\mathcal{M}$ with eigenvalue λ , respectively. It follows from the factorization of $\mathcal{M}_{\circ\circ}$ and $\mathcal{M}_{\square\square}$ that

$$\mathcal{M}_{\circ\circ}|a_{\circ}) = \lambda^2|a_{\circ}), \quad (b_{\square}|\mathcal{M}_{\square\square} = \lambda^2(b_{\square}|. \quad (76)$$

These identities can be used to “propagate” the boundaries through the sequences of transfer matrices to return

$$C_{ab}^{(k=2)}(t) = \lambda^{2t-2} t \frac{d_C(b_{\square}|\mathcal{M}_{\square\square}|a_{\circ})}{d_A} - \lambda^{2t}(t-1) \frac{(b_{\square}|a_{\circ})}{d_A}, \quad (77)$$

indicating exponential decay of OTOCs.

The decay of the OTOC as $C_{AB}^{(k=2)}(t) \propto t\lambda^{2t}$ can be argued to hold for generic traceless operators. Graphically, the factorization of $\mathcal{M}_{\circ\circ}|a_{\circ})$ can be made apparent by noting that both $\mathcal{M}_{\circ\circ}$ and $|a_{\circ})$ factorize in the same way,

$$\mathcal{M}_{\circ\circ}|a_{\circ}) = \begin{array}{c} \text{diagram of } \mathcal{M}_{\circ\circ} \text{ and } |a_{\circ}) \end{array} = \frac{1}{d_C^2} \begin{array}{c} \text{diagram of } \mathcal{M}_{\circ\circ} \text{ and } |a_{\circ}) \end{array}, \quad (78)$$

where the first and second layers are decoupled from the third and fourth layers, returning two copies of $\mathcal{M}(a)$:

$$\begin{array}{c} \text{diagram of } \mathcal{M}_{\circ\circ} \text{ and } |a_{\circ}) \end{array} = \begin{array}{c} \text{diagram of } \mathcal{M}_{\circ\circ} \end{array} \otimes \begin{array}{c} \text{diagram of } |a_{\circ}) \end{array}. \quad (79)$$

A similar factorization holds for $(b_{\square}|$ and $\mathcal{M}_{\square\square}$,

$$(b_{\square}|\mathcal{M}_{\square\square} = \begin{array}{c} \text{diagram of } (b_{\square}|\mathcal{M}_{\square\square} \end{array} = \frac{1}{d_C^2} \begin{array}{c} \text{diagram of } (b_{\square}|\mathcal{M}_{\square\square} \end{array}, \quad (80)$$

where the first and fourth layers are now decoupled from the second and third layers, such that the right-hand side factorizes as

$$\begin{array}{c} \text{diagram of } (b_{\square}|\mathcal{M}_{\square\square} \end{array} = \begin{array}{c} \text{diagram of } (b_{\square}|\mathcal{M}_{\square\square} \end{array} \otimes \begin{array}{c} \text{diagram of } (b_{\square}|\mathcal{M}_{\square\square} \end{array}. \quad (81)$$

The latter factorization returns the former when shifting by “half a replica,” i.e., by regrouping the backward sheet of the original replica i with the forward sheet of replica $i+1$. This picture has a precise formulation on the level of the noncrossing partitions as it corresponds to mapping a noncrossing partition σ to its Kreweras complement $\sigma^* = \sigma^{-1}\square$; see Appendix A. For $k=2$, this yields $\circ^* = \square$ and $\square^* = \circ$.

This factorization fails when acting with $\mathcal{M}_{\square\circ}$. However, for a total number of t time steps there can be only a single domain wall and corresponding “off-diagonal” map $\mathcal{M}_{\square\circ}$, such that the remaining $t-1$ other maps are “diagonal” maps $\mathcal{M}_{\circ\circ}$ or $\mathcal{M}_{\square\square}$. Using the same arguments as in the previous section, we can bound

$$\|\mathcal{M}_{\circ\circ}^s|a_{\circ})\|_2 \leq \lambda^{2s}\|a\|_2^2, \quad (82)$$

$$\|(b_{\square}|\mathcal{M}_{\square\square}^{t-s}\|_2 \leq \lambda^{2(t-s)}\|b\|_2^2. \quad (83)$$

By combining these two bounds, it directly follows that every term in Eq. (75) decays as $\propto \lambda^{2t}$. Since the total number of terms scales as t , we expect that the 2-OTOC decays to the ergodic zero value $\propto t\lambda^{2t}$, consistent with Eq. (77).

Compared to the $k=1$ case, the decay is now twice as fast due to the two copies of the channel $\mathcal{M}$ in $\mathcal{M}_{\circ\circ}$ and $\mathcal{M}_{\square\square}$ at each time step, respectively. Additionally, due to the increasing number of multichains, the exponential decay is accompanied by a linear in t prefactor. This linear corrections are a first hint toward an underlying Jordan structure, which we discuss in later sections.

If the observables are not traceless, the steady-state value can be calculated by repeatedly subtracting the appropriate traces, i.e., by computing the OTOC for the traceless observables $a_i - \varphi(a_i)$ and $b_i - \varphi(b_i)$, to return

$$\begin{aligned} \lim_{t \rightarrow \infty} C_{ab}^{(k=2)}(t) = & -\varphi(a_1)\varphi(a_2)\varphi(b_1)\varphi(b_2) \\ & + \varphi(b_1b_2)\varphi(a_1)\varphi(a_2) \\ & + \varphi(a_1a_2)\varphi(b_1)\varphi(b_2). \end{aligned} \quad (84)$$

We will return to this result in later sections.

C. General k

We now turn our attention to the decay of general k -OTOCs. For traceless observables, we will argue that they exhibit the same exponential decay $\propto \lambda^{2t}$ as in the $k = 2$ case, albeit with a modified polynomial prefactor due to the growing number of multichains. In Eq. (50), it is always possible to identify a subset of multichains whose contribution decays as $\propto \lambda^{2t}$, and these dominate the dynamics at late times.

For $k = 2$, there are only two permutations, i.e., the identity of the cyclic permutation guaranteeing a “matching” between the permutations in the linear maps $\mathcal{M}_{\nu\sigma}$ and one of the two boundaries, such that the decay directly follows from the product of two quantum channels $\mathcal{M}$ acting on the operators a (b) from the right (left). For $k > 2$, the summation can, however, involve permutations σ that are distinct from both the identity and cyclic permutation, and no such factorization can be performed. This lack of factorization results in a slower decay compared to the scaling $\propto \lambda^{kt}$ which would naively be expected.

For concreteness, we first illustrate this argument for $k = 3$ and $\sigma = (12)(3)$ and consider a diagonal path of the form $(b_{\square}|\mathcal{M}_{\sigma\sigma}^t|a_{\circ})$. While not technically part of the summation in Eq. (50) due to the different boundary conditions, this example helps to illustrate the general argument. Graphically,

$$(b_{\square}|\mathcal{M}_{\sigma\sigma}^t|a_{\circ}) \propto \quad (85)$$

where for concreteness we have fixed $t = 3$. Neither the action on a nor b factorizes fully, but we can identify a

sequence of the form $\mathcal{M}^t(a)$ in the third replica (starting from the bottom). We can isolate this term to factorize Eq. (85) as

$$= \quad \times \quad (86)$$

Here the multiplication should be interpreted as performing the appropriate contractions at the top. The absolute value of this expression can be bounded as the product of the operator norms of both terms. Since $\mathcal{M}^t(a)$ decays exponentially with t , the norm of the second term can again be bounded by $\lambda^t \|a\|_2$ (when reintroducing the appropriate prefactors of d_C).

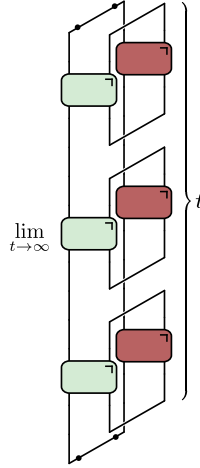
The first term on the right-hand side can now be bounded by identifying a sequence where the quantum channels act to the left on b (starting from the top), which can be factorized as

$$= \quad \times \quad (87)$$

Multiplication should here be interpreted as performing the appropriate contractions at the bottom. The norm of the second term on the right-hand side, i.e., the sequence of quantum channels acting on b from the left, can again be bounded as $\lambda^t \|b\|_2$. Combining this bound with the previous

bound, we recover an exponential decay as λ^{2t} from the two sequences of quantum channels.

No further simplifications can be made, since at no other point can we identify sequences of the quantum channel acting on a traceless operator. Indeed, the remaining term generally does not vanish at late times but rather decays to a constant. If a^2 and b^2 are not traceless, then contracting this term with the identity at the top and the bottom returns



$$\lim_{t \rightarrow \infty} \text{Tr}(a^2) \text{Tr}(b^2) = \mathcal{O}(1), \quad (88)$$

again when reintroducing the appropriate factors of d_C .

Taking these results together, we find that we can bound Eq. (85) as

$$|(b_{\square} | \mathcal{M}_{\sigma\sigma}^t | a_{\circ})| \leq \lambda^{2t} \|a\|_2^k \|b\|_2^k \quad (89)$$

for $\sigma = (12)(3)$.

This argument can now be extended to generic non-crossing partitions and the k -OTOC [Eq. (50)]. For sequences of diagonal quantum channels, we have the bound

$$\|(b_{\square} | \mathcal{M}_{\sigma\sigma}^t | a_{\circ})\|_2 \leq \lambda^{[n(\sigma) + n(\sigma^*)]t} \|a\|_2^k \|b\|_2^k, \quad (90)$$

where $n(\sigma)$ is the number of singletons (cycles of length one) in σ , and $n(\sigma^*)$ is the number of singletons in its Kreweras complement. The former number $n(\sigma)$ describes the number of quantum channels acting from the right (bottom) on a , i.e., the number of sequences $\mathcal{M}^t(a)$ that can be identified, whereas the latter number $n(\sigma^*)$ describes the number of quantum channels acting from the left (top) on b . For $\sigma = (12)(3)$, we have $n(\sigma) = n(\sigma^*) = 1$, returning the previous result for Eq. (85) establishing a decay $\propto \lambda^{2t}$. The diagonal channels corresponding to the identity and the cyclic permutation always decay as λ^{kt} , since $n(\circ) = n(\square^*) = k$ and $n(\square) = n(\circ^*) = 0$. The number of singletons in a permutation σ satisfies $n(\sigma) + n(\sigma^*) \geq 2$, where the inequality is satisfied for the permutations in the middle of the noncrossing partition lattice. For any $k > 2$, there are

permutations in which the combined number of singletons equals 2, independent of k , such that the slowest decay is given by $\propto \lambda^{2t}$. In Fig. 1(b), these correspond to the permutations in the center.

These arguments can be directly extended to bound the contribution $(b_{\square} | \mathcal{M}_{\Sigma} | a_{\circ})$ for a generic nondiagonal multichain $\Sigma = (\sigma_1 \subseteq \nu_1 \subseteq \dots \subseteq \nu_t)$ of length $2t$ and $\mathcal{M}_{\Sigma}$ the corresponding product of channels. As there are at most k distinct partitions along the multichain at most $k-1$ of the channels $\mathcal{M}_{\nu_i \sigma_i}$ are off diagonal, i.e., $\nu_i \neq \sigma_i$. For t sufficiently large, the majority of channels are hence diagonal. By definition of the order of the permutations, the above argument can be directly repeated, such that we can upper bound the contribution from the whole multichain

$$|(b_{\square} | \mathcal{M}_{\Sigma} | a_{\circ})| \leq \lambda^{m(t-k+1)} \|a\|_2^k \|b\|_2^k, \quad (91)$$

where $t - k + 1$ and $m = \min\{n(\sigma) + n^*(\sigma), \sigma \in \text{NC}(k)\}$ are the minimal number of diagonal channels along the chain and the minimal possible number of singletons within this chain, respectively. We illustrate the structure of singletons in Fig. 1 for two multichains and their Kreweras duals. The multichain depicted in Fig. 1(a) gives rise to a subleading contribution, as can be seen by following the singletons in their diagrammatic representation. The singleton (1) and (4) in $\circ$ are contained only in $\circ$ and, by the above arguments, each gives rise to a contribution $\propto \lambda$. The singleton (3) is contained in both $\circ$ and ν and hence contributes a factor $\propto \lambda^{s+1}$. In contrast, the singleton (2) is

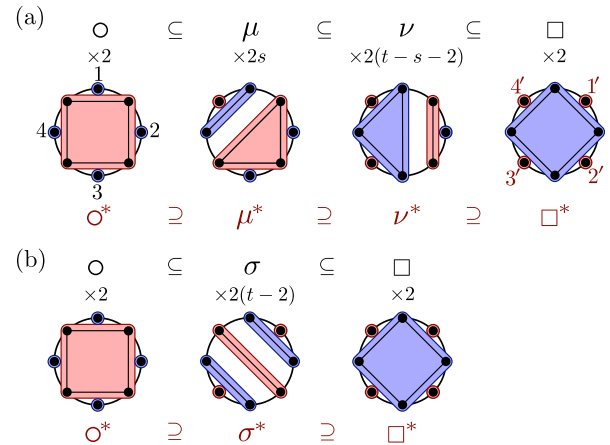


FIG. 1. Examples of multichains giving (a) subleading and (b) leading contributions to the decay of the averaged k -OTOC for $k = 4$. The multichains (top, blue) of length t are given by (a) $(\circ \subseteq \circ \subseteq \mu \subseteq \dots \subseteq \nu \subseteq \dots \subseteq \nu \subseteq \square \subseteq \square)$ and (b) $(\circ \subseteq \circ \subseteq \sigma \subseteq \dots \subseteq \sigma \subseteq \square \subseteq \square)$ ordered from left to right, and with the intermediate partitions occurring repeatedly as indicated in the figure. Their respective Kreweras duals (bottom, red) are ordered from right to left. For simplicity, both multichains are chosen such that the corresponding sequence of channels contains only diagonal channels $\mathcal{M}_{\rho\rho}$.

contained in $\circ$, μ , and ν and hence contributes a factor $\propto \lambda^{t-1}$. These are the contributions that can be estimated from the singletons in the original multichain and correspond to contracting single replica channels $\mathcal{M}$ in the networks above from bottom to top. The dual multichain gives rise to additional contributions scaling with powers of λ , which correspond to contracting the networks above from top to bottom. Following the singletons now from right to left, we find that the singletons (1') and (2') in $\square^*$ contribute a factor $\propto \lambda$, the singleton (3') contributes a factor $\propto \lambda^{t-s-1}$, and (4') contributes $\propto \lambda^{t-1}$. In total, the multichain gives rise to a contribution $\propto \lambda^{3t+2} \sim \lambda^{3t}$. The resulting exponent can also be directly obtained as half the total number of singletons along the chain. Repeating the same calculation for the multichain depicted in Fig. 1(b) instead gives a contribution $\propto \lambda^{2t+4} \propto \lambda^{2t}$, the slowest possible decay. Note that the occurrence of off-diagonal channels $\mathcal{M}_{\sigma\nu}$ with $\sigma \neq \nu$ or a fixed number of $\circ$ ($\square$) at the beginning (end) of the chain will change the exponent of λ only by a constant, while keeping the asymptotic decay with t unchanged. For this reason, there is not a single slowest-decaying multichain but a multitude of them, which all lead to the same asymptotic decay λ^{2t} and obey the bound (91).

Each multichain additionally comes with a prefactor $\prod_{i=1}^{t-1} W_{\nu_i \sigma_{i+1}}$ from the Weingarten functions (52). Again, at most $k-1$ terms in this product are different from unity, and those terms are bounded in modulus by the maximum value the Möbius function takes. This prefactor can hence be bounded by a constant that depends only on k and does not scale with t .

The above bound applies for every multichain, and we can get an estimate for the decay by combining the total number of multichains scaling as t^{k-1} , with the slowest decay scaling as λ^{2t} . Putting all of the above together, it follows that there exists a constant K_k such that

$$|C_{ab}^{(k)}(t)| \leq K_k t^{k-1} \lambda^{2t} \|a\|_2^k \|b\|_2^k. \quad (92)$$

At late times, all k -OTOCs hence decay at least exponentially up to polynomial corrections. At sufficiently late times, the above bound returns an exponential decay $\propto \lambda^{2t}$, and by considering eigenoperators a_λ of the channel $\mathcal{M}$ this asymptotic decay is saturated (at least for $d_A \geq 3$; see below for $d_A = 2$).

This scaling is compared with numerical simulations in Fig. 2 for a generic operator and for an eigenoperator of the quantum channel, showing excellent agreement. Both from the theoretical bound and the numerical data, it follows that the decay of the k -OTOC is the same for all $k \geq 2$, implying that higher-order OTOCs with $k > 2$ cannot be decomposed into two-point functions, highlighting the chaotic

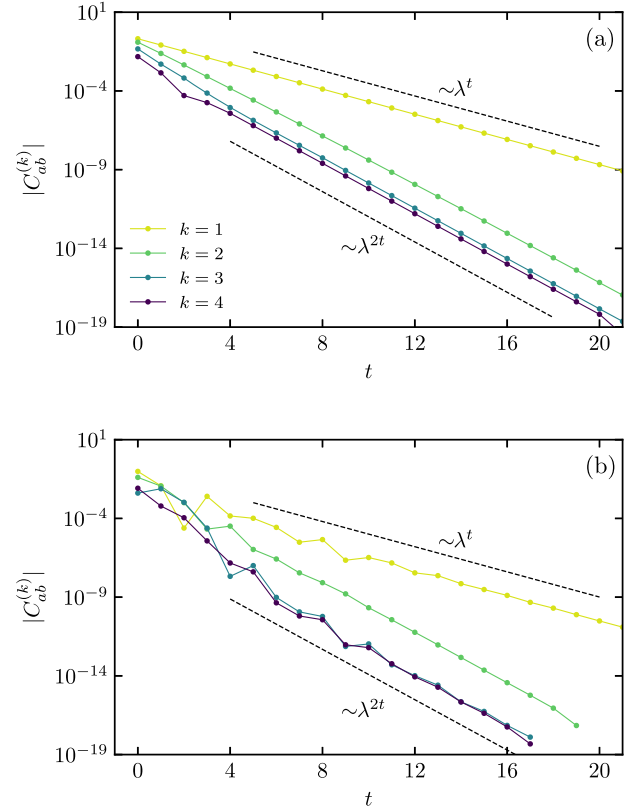


FIG. 2. k -OTOC for $d_A = d_C = 3$ for (a) left and right eigenoperators a_λ and b_λ and (b) generic traceless a and b for different k and a gate U with $\lambda = 0.398$. The asymptotic decay $\propto \lambda^{2t}$ ($k = 1$) and $\propto \lambda^{2t}$ ($k \geq 2$), respectively, are depicted as dashed lines.

nature of our model. This result can be seen as the dynamical manifestation of higher-order correlations in the full ETH ansatz.

This bound can be further sharpened when the subsystem A hosts a single qubit, i.e., $d_A = 2$, and all the local observables $a_1 = \dots = a_k = a$ and $b_1 = \dots = b_k = b$ are the same. In this case, being Hermitian and traceless directly implies $a^2 = b^2 \propto \mathbb{1}_A$. Consider again the example of Eq. (85): Fixing $a = a_\lambda$ and $b = b_\lambda$ to be decaying left and right eigenoperators of the quantum channel and propagating these through the multichain, the total expression is proportional to $\text{Tr}(b_\lambda a_\lambda b_\lambda) \propto \text{Tr}(a_\lambda) = 0$, using that the subleading eigenoperator a_λ is orthogonal to the leading one $\mathbb{1}_A$ and therefore traceless. This relation gives rise to additional cancellations and leads to an asymptotic decay of k -OTOC of the asymptotic form

$$C_{ab}^{(k)}(t) \propto \lambda^{kt}, \quad (93)$$

showing faster decay if $k \geq 3$. For a general sequence of diagonal quantum channels $\mathcal{M}_{\sigma\sigma}$, the decay is

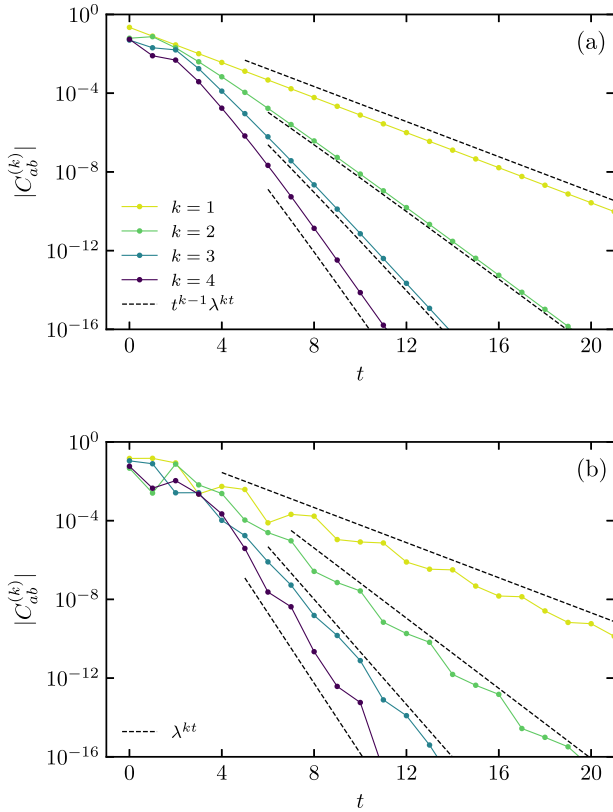


FIG. 3. k -OTOCs for $d_A = d_C = 2$ for (a) left and right eigenoperators a_λ and b_λ and (b) generic traceless a and b for different k and a gate U with $\lambda = 0.358$. The asymptotic decay (a) $\propto t^{k-1} \lambda^{kt}$ and (b) $\propto \lambda^{kt}$ is depicted as dashed lines.

determined by the unique eigenvector of the linear map $|(a_\lambda)_\sigma\rangle$ with eigenvalue λ^k , i.e., the permutation operator $|\sigma\rangle$ where every replica is dressed with an eigenoperator a_λ [as in Eqs. (46) and (47)]. This behavior is illustrated in Fig. 3.

D. Steady state

Having established that the k -OTOCs between traceless observables decays to zero at late times, we now turn our attention to the steady-state value of the k -OTOCs once the constraint on having zero trace is lifted. An example of which is given in Eq. (84) for $k = 2$, whereas we provide a systematic derivation and interpretation for arbitrary k in the following. More precisely, we characterize the limit $\lim_{t \rightarrow \infty} C_{ab}^{(k)}(t)$ in terms of generalized moments φ_{σ^*} and free cumulants κ_ν between the observables [cf. Eqs. (19) and (27)]. While this expression can be directly obtained from free probability or by repeatedly subtracting the nontraceless part of the operator, it is instructive to see how these expressions appear here.

To obtain the steady state, we observe that allowing for $\varphi(a_i) \neq 0 \neq \varphi(b_i)$ leads to nonvanishing overlaps with the leading eigenoperator $\mathbb{1}_A$ of the channel $\mathcal{M}$. The steady state is governed by the leading eigenoperators of the channels $\mathcal{M}_{\nu\sigma}$. By unitarity of the local gates U , it follows that the leading left and right eigenoperators are permutation states obeying

$$\mathcal{M}_{\nu\sigma}|\sigma\rangle = d_C^{-k+|\nu^{-1}\sigma|}|\sigma\rangle \quad (94)$$

and

$$(\nu|\mathcal{M}_{\nu\sigma} = d_C^{-k+|\nu^{-1}\sigma|}(\nu|. \quad (95)$$

For the diagonal channels $\nu = \sigma$, such that the leading left and right eigenoperators coincide and the leading eigenvalue is 1. Replacing the channels $\mathcal{M}_{\nu\sigma}$ along each multichain by the projector onto the leading eigenoperators, which after proper biorthonormalization reads

$$\mathcal{M}_{\nu\sigma} = d_C^{-k+|\nu^{-1}\sigma|} d_A^{-|\nu^{-1}\sigma|} |\sigma\rangle(\nu| + \dots, \quad (96)$$

returns the stationary state. Here the dots represent subleading terms (bi)orthogonal to the leading eigenoperators, which give rise to decaying contributions to the k -OTOC.

The steady state can be obtained by replacing the channels $\mathcal{M}_{\nu\sigma}$ by the above projection, simplifying the resulting overlaps between permutation states, and using the combinatorial properties of the Möbius function $\mu(\nu, \sigma)$. Inserting the definition of the Weingarten functions (52) and repeatedly using $|\sigma^{-1}\rho| + |\rho^{-1}\nu| = k + |\sigma^{-1}\nu|$ for $\sigma \subseteq \rho \subseteq \nu$, for each multichain $\Sigma = \circ \subseteq \nu_1 \subseteq \dots \subseteq \sigma_{t-1} \subseteq \square$ the contribution to Eq. (50) at late time reads

$$C_{ab,\Sigma}^{(k)}(t) = \left(\prod_{i=1}^{t-1} \mu(\sigma_i, \nu_i) \right) \varphi_{\nu_1}(a_1, \dots, a_k) \varphi_{\sigma_{t-1}^*}(b_1, \dots, b_k) \quad (97)$$

up to subleading terms which vanish as $t \rightarrow \infty$. Here we additionally used that the overlaps of the permutation states with the boundary conditions yield the moments, e.g., $d_A^{-|\nu|}(\nu|a_\circ) = \varphi_\nu(a_1, \dots, a_k)$. To perform the summation over all multichains, we first perform the sum over all multichains with fixed $\nu_1 = \nu$ and $\sigma_{t-1} = \sigma$, since

$$\sum_{\substack{\dots \subseteq \sigma_i \subseteq \nu_i \subseteq \dots \\ \nu_1 = \nu, \sigma_{t-1} = \sigma}} \prod_{i=1}^{t-1} \mu(\sigma_i, \nu_i) = \mu(\sigma, \nu), \quad (98)$$

where we repeatedly use the defining property of the Möbius function [Eq. (54)], here repeated for convenience

$$\sum_{\sigma \subseteq \rho \subseteq \nu} \mu(\nu, \rho) = \delta_{\nu, \sigma}. \quad (99)$$

The steady state for the k -OTOC follows as

$$\begin{aligned} \lim_{t \rightarrow \infty} C_{ab}^{(k)}(t) &= \sum_{\nu \subseteq \sigma} \mu(\nu, \sigma) \varphi_{\nu}(a_1, \dots, a_k) \varphi_{\sigma^*}(b_1, \dots, b_k) \\ &= \sum_{\sigma \in \text{NC}(k)} \varphi_{\sigma^*}(b_1, \dots, b_k) \kappa_{\sigma}(a_1, \dots, a_k), \end{aligned} \quad (100)$$

where we have used that the free cumulants are related to the moments through the Möbius function [see Eq. (26)].

This expression describes the correlations between freely independent observables and corresponds to the correlations between time-evolved observables and static observables if we were to replace the evolution operator $\mathcal{U}(t)$ by a Haar-random unitary and subsequently taking the thermodynamic limit [71]. In other words, within the minimal model presented here, we recover the random-matrix result in the thermodynamic limit at late times.

E. Approach to equilibrium

Having established the decay of k -OTOCs between traceless observables as well as the emergence of the steady state for observables with nonzero trace, we combine both aspects in the following to obtain yet a different decaying behavior $\propto \lambda^t$ for arbitrary k . Provided the observables are no longer traceless, there will be nontrivial contributions from single sequences of quantum channels leading to a decay $\propto \lambda^t$ corresponding to a two-time correlation function being factored out. This decay is illustrated in Fig. 4 and will be discussed in detail in Sec. IV F. Interestingly, the presence of a single operator that is not traceless can lead to a two-step decay, a recent topic of interest in OTOC dynamics [116–119]. Choosing a single operator to be traceless, the steady-state value of the k -OTOC is still zero, but the decay to this steady state is modified from $\propto \lambda^{2t}$ to $\propto \lambda^t$. If this nontraceless part is sufficiently small, the term decaying $\propto \lambda^t$ is perturbative, leading to a two-stage relaxation of the OTOC, as illustrated in Fig. 4.

To conclude this section, we consider the maximally ergodic case when U is dual unitary [96]. In this case, the quantum channel (66) is a projector on the identity, and all

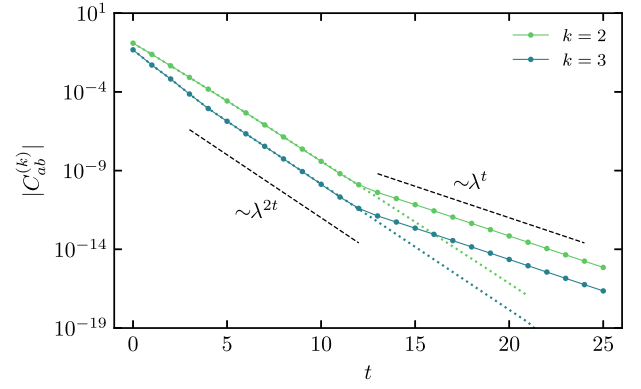


FIG. 4. k -OTOCs for $d_A = d_C = 3$ for $k = 2, 3$ and a gate U with $\lambda = 0.398$. The observables are left and right eigenoperators a_λ and b_λ , respectively, except for $a_0 = a_\lambda + \epsilon \mathbb{1}_A$ with $\epsilon = 10^{-4}$ ($k = 2$) and $\epsilon = 10^{-5}$ ($k = 3$). The decay at initial times $\propto \lambda^{2t}$ and at late times $\propto \lambda^t$ is represented by dashed lines, while the k -OTOCs for $\epsilon = 0$ are shown as dotted lines.

correlation functions for $k = 1$ reach their steady-state ergodic value after a single time step [see Eq. (73)]. For $k \geq 2$, the OTOC is nonvanishing after a single time step, since $(b_\square | \mathcal{M}_{\square \circ} | a_\circ) \neq 0$, but, in accordance with the previous discussion, for traceless observables the k -OTOC with $k \geq 2$ vanish identically after two time steps. Consider, e.g., again $k = 3$ and $\sigma = (12)(3)$, then

$$\mathcal{M}_{\sigma \circ | a_\circ} \propto \text{[Diagram 1]} = \text{[Diagram 2]} = 0, \quad (101)$$

where in the first equality we have used dual unitarity for the two gates in the third replica, corresponding to the singleton in σ , and in the second equality we have used the tracelessness. More generally, $\mathcal{M}_{\sigma \circ | a_\circ}$ vanishes whenever σ has a singleton, whereas $(b_\square | \mathcal{M}_{\square \nu^*}$ vanishes whenever ν^* has a singleton. Combined with the constraint that $\sigma \subseteq \nu$, which implies that every singleton in ν is a singleton in σ and that every singleton in σ^* is a singleton in ν^* , it immediately follows that either σ has a singleton or ν^* has a singleton, resulting in an identically vanishing k -OTOC after two time steps. For traceless observables and dual-unitary dynamics, it follows that

$$C_{ab}^{(k=1)}(t \geq 1) = 0, \quad C_{ab}^{(k=2)}(t \geq 2) = 0. \quad (102)$$

For operators that are not traceless, the nontraceless part can again be factored out, such that the steady-state value of Eq. (100) is reached after two discrete time steps. These results further cement dual-unitary circuits as maximally ergodic or chaotic models of many-body quantum dynamics.

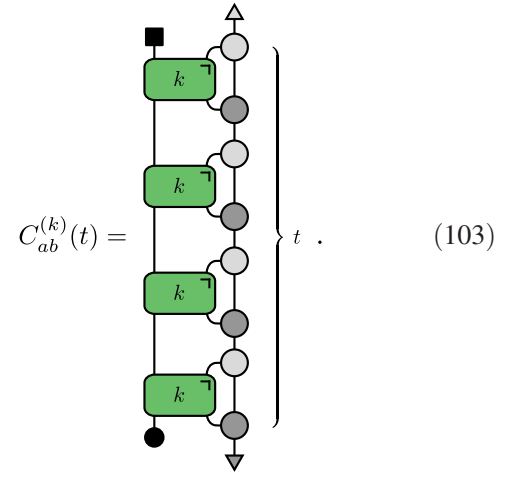
IV. MARKOVIAN INFLUENCE-MATRIX DYNAMICS

In this section, we show how the obtained expression for the k -OTOC [Eq. (50)] can be recast in an influence-matrix formalism, which captures the effect of the random-matrix environment on the local subsystem. This influence matrix corresponds to a matrix product state acting on the temporal Hilbert space of sites corresponds to the number of time steps. As opposed to influence matrices for correlation functions, where this influence matrix is unentangled, the minimal structure imposed by the random environment results in a low but nonvanishing (temporal) entanglement for k -OTOCs with $k \geq 2$. This entanglement can be bound by the necessary bond dimension, which for the k -OTOC corresponds to the number of noncrossing partitions of k elements, i.e., the Catalan numbers C_k . The auxiliary noncrossing partition degree of freedom effectively tracks the system's position in the noncrossing partition lattice, encoding how different replicas have interacted over time, i.e., which replicas have become dynamically correlated due to system-bath interactions.

In this way, the influence-matrix representation can be used to recast the k -OTOC dynamics as a Markovian process, provided we keep track of this additional noncrossing partition degree of freedom. This approach allows for numerical simplifications when expressing the summation in Eq. (50) and has the advantage that the leading eigenstates of this Markovian process can be constructed analytically. These eigenstates can be labeled by noncrossing partitions, which directly returns the late-time dynamics in terms of free cumulants indexed by noncrossing partitions. In this way, we recover the expansion of the k -OTOC predicted by full ETH, with the different free cumulants satisfying the expected factorization properties from free probability.

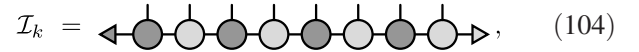
A. Influence-matrix representation

The summation over different paths in the expression for the k -OTOC [Eq. (50)] can be recast in a more compact form by treating the noncrossing partition degrees of freedom as states in an auxiliary Hilbert space. We now represent the k -OTOC of Eq. (50) as



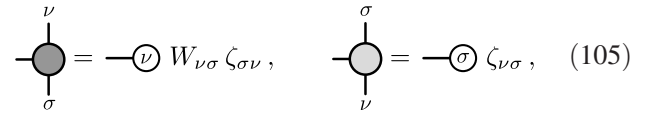
$$C_{ab}^{(k)}(t) = \text{[Diagram of matrix product state]} \quad (103)$$

The relevant influence matrix $\mathcal{I}_k$ corresponds to a matrix product state acting on the temporal Hilbert space of $2t$ sites as



$$\mathcal{I}_k = \text{[Diagram of matrix product state]} \quad (104)$$

here rotated by 90° for convenience. This matrix product state is defined in terms of tensors



$$\text{[Diagram of tensor]} = \text{[Diagram of tensor]} \quad (105)$$

where the states in this auxiliary space are labeled by noncrossing partitions such that the bond dimension equals C_k . The physical space is $\mathcal{H}_C^{\otimes 2k}$, i.e., the $(d_C)^{2k}$ -dimensional space of the replicas of C . These tensors enforce the noncrossing partitions to be nondecreasing in time by introducing the zeta function of the noncrossing lattice: $\zeta_{\sigma\nu} = 1$ if $\sigma \subseteq \nu$ and zero otherwise; see Appendix A. Moreover, the tensors introduce the Möbius functions $\mu(\nu, \sigma)$ encoded in the Weingarten functions $W_{\nu\sigma}$ at alternating time steps.

The boundary vectors are defined as



$$\text{[Diagram of vector]} = (d_A d_C)^{-1} \delta_{\sigma \square}, \quad \text{[Diagram of vector]} = d_C^{|\sigma|}. \quad (106)$$

These boundary vectors are chosen such that the partitions at the top and bottom correspond to the cyclic permutation and the identity permutation, respectively, and the prefactor from Eq. (50) has been distributed over these terms to simplify later expressions.

For the contraction at the bottom, we have that

$$\begin{array}{c} \nu \\ \uparrow \\ \text{---} \bigcirc \text{---} \\ \downarrow \end{array} = \text{---} \bigcirc d_C^k \delta_{\nu \circ}, \quad (107)$$

using the properties of the geodesics and the Möbius function,

$$\sum_{\sigma \subseteq \nu} \mu(\nu, \sigma) d_C^{-k+|\nu^{-1}\sigma|+|\sigma|} = \sum_{\sigma \subseteq \nu} \mu(\nu, \sigma) d_C^{|\nu|} = d_C^k \delta_{\nu \circ}. \quad (108)$$

For the contraction at the top, we directly find that

$$\begin{array}{c} \uparrow \\ \text{---} \bigcirc \text{---} \\ \nu \end{array} = \text{---} \square (d_A d_C)^{-1}. \quad (109)$$

B. Example: $k = 1$

For $k = 1$, there is only a single partition $\circ = (1)$, such that this matrix product state reduces to a product wave function on the temporal lattice. We can write

$$\mathcal{I}_{k=1} \propto \begin{array}{c} \downarrow \\ \bigcirc \end{array} \begin{array}{c} \downarrow \\ \bigcirc \end{array} \cdots \begin{array}{c} \downarrow \\ \bigcirc \end{array} \begin{array}{c} \downarrow \\ \bigcirc \end{array}, \quad (110)$$

where the number of sites corresponds to $2t$. Consequently, the diagrammatic representation in Eq. (104) reduces to Eq. (65). This influence matrix was previously obtained in Ref. [6], where it was termed the “perfect dephaser” limit. For such influence matrices, the environment cancels out all interference effects and acts as a perfectly Markovian bath. This perfect dephaser limit was analytically shown to be the influence matrix for dual-unitary dynamics, minimal models of ergodic many-body dynamics for which the correlation functions of local operators vanish identically after a single time step. In this limit, the temporal entanglement vanishes identically, and it was argued in Ref. [6] that temporal entanglement is expected to stay small even away from the perfect dephaser limit. This observation motivated the more general applicability of influence matrices away from this perfectly Markovian limit. For a detailed discussion of temporal entanglement in chaotic circuits, we refer the reader to Ref. [15].

C. Example: $k = 2$

For $k = 2$, there are two partitions, both noncrossing, $\circ = (1)(2)$ and $\square = (12)$. The influence matrix can be expanded in its components, where the only nontrivial term from the Weingarten functions is $W_{\circ\square} = -1/d_C$. The influence matrix describes a sum of domain-wall states, with weight depending on the parity of the position of the domain wall, i.e.,

$$\mathcal{I}_{k=2} \propto \sum_{s=1}^{2t-1} I_s \begin{array}{c} \downarrow \\ \bigcirc \end{array} \begin{array}{c} \downarrow \\ \bigcirc \end{array} \cdots \begin{array}{c} \downarrow \\ \bigcirc \end{array} \begin{array}{c} \downarrow \\ \square \end{array} \cdots \begin{array}{c} \downarrow \\ \square \end{array} \begin{array}{c} \downarrow \\ \square \end{array}, \quad (111)$$

where $I_s = 1$ if s is odd and $I_s = -1/d_C$ if s is even. Consequently, the tensor network representation of the influence matrix in Eq. (104) can be written as a sum over domain-wall configurations as was done in Eq. (75). In structured dual-unitary circuits, these influence matrices directly appear in the calculation of the OTOC [85,86,110,120]. An influence matrix of this form also appeared for the random phase model in Ref. [84] (with the identity and cyclic permutation corresponding to Gaussian and non-Gaussian diagrams in the language of Ref. [84]). A similar structure, away from solvable points, was observed in Ref. [121] and argued to underpin the generic dynamics of the 2-OTOC, again motivating the more general applicability of the developed framework.

D. Markovian dynamics

Markovian dynamics can be obtained by truncating the summation in Eq. (50) after a fixed number of time steps t and keeping track of the final noncrossing partition in the summations and the state returned by the action of the corresponding transfer matrix. This picture gives rise to a dynamical semigroup which generalizes the semigroup influence matrix developed for $k = 1$ [28]. Alternatively, this Markovian dynamics can be interpreted as truncating the matrix product state representation of Eq. (103) and performing a cut along the horizontal direction after a fixed number of time steps.

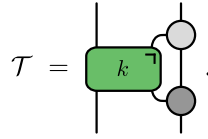
We define a state

$$|\varphi_t\rangle = \left\{ \begin{array}{c} \text{---} \square \text{---} \\ \text{---} \square \text{---} \\ \text{---} \square \text{---} \end{array} \right\}^t, \quad (112)$$

where the total number of folded gates equals t , which satisfies the dynamical rule

$$|\varphi_t\rangle = \mathcal{T}|\varphi_{t-1}\rangle, \quad (113)$$

with



$$\mathcal{T} = \text{Diagram} \quad (114)$$

Note that we use double brackets to emphasize that these states act on both the Hilbert space of k replicas and the auxiliary space. The transfer matrix $\mathcal{T}$ corresponds to the generator of the dynamical semigroup.

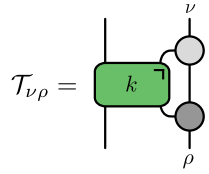
Fixing the index of the auxiliary space as, e.g., ν , we can write

$$|\varphi_t)\rangle = \sum_{\nu \in \text{NC}(k)} |\varphi_t)_{\nu}. \quad (115)$$

Written out explicitly, the dynamics of Eq. (113) for $|\varphi_t)_{\nu}$ is of the form

$$|\varphi_t)_{\nu} = \sum_{\rho \in \text{NC}(k)} \mathcal{T}_{\nu\rho} |\varphi_{t-1})_{\rho}, \quad (116)$$

where



$$\mathcal{T}_{\nu\rho} = \text{Diagram} = \begin{cases} \sum_{\rho \subseteq \sigma \subseteq \nu} \mathcal{M}_{\nu\sigma} W_{\sigma\rho} & \text{if } \rho \subseteq \nu, \\ 0 & \text{if } \rho \not\subseteq \nu. \end{cases} \quad (117)$$

The full k -OTOC can be written as

$$C_{ab}^{(k)}(t) = ((\psi_b | \mathcal{T}^t | \psi_a)), \quad (118)$$

with boundary vectors

$$((\psi_b | = (d_A d_C)^{-1} (b_{\square} |_{\square}, \quad (119)$$

$$|\psi_a)) = \sum_{\rho \in \text{NC}(k)} d_C^{|\rho|} |a_{\circ})_{\rho}. \quad (120)$$

For small values of $k = 1, 2$, these results again allow for a simple representation and connection with known results. For $k = 1$, we again have only a single noncrossing partition, we can identify $|\varphi_t)) = |\varphi_t)_{\circ} = |\varphi_t)$, and the transfer matrix $\mathcal{T}$ reduces to the quantum channel $\mathcal{M}$ acting on the vectorized operator. The dynamics of Eq. (113) returns the previously discussed quantum channel construction for correlation functions as

$$|\varphi_t) = \mathcal{M}_{\circ\circ} |\varphi_{t-1}) = \mathcal{M} |\varphi_{t-1}). \quad (121)$$

For $k = 2$, there are two noncrossing partitions, which are coupled through their dynamics as

$$|\varphi_t)_{\square} = \mathcal{M}_{\square\square} |\varphi_{t-1})_{\square} + \left(\mathcal{M}_{\square\circ} - \frac{\mathcal{M}_{\square\square}}{d_C} \right) |\varphi_{t-1})_{\circ}, \quad (122)$$

$$|\varphi_t)_{\circ} = \mathcal{M}_{\circ\circ} |\varphi_{t-1})_{\circ}. \quad (123)$$

It is a direct check that this expression is equivalent to Eq. (75).

This example illustrates a general property of the Markov process: The corresponding transfer matrix is naturally written in an upper-diagonal form, since $\mathcal{T}_{\nu\rho} = 0$ if $\rho \not\subseteq \nu$. The noncrossing partition degree of freedom can never decrease. In the same way that the eigenspectrum of an upper-diagonal matrix follows directly from its diagonal elements, the spectrum of the full matrix $\mathcal{T}$ follows from the spectrum of the diagonal elements. These diagonal elements here correspond to the diagonal transfer matrices $\mathcal{T}_{\nu\nu} = \mathcal{M}_{\nu\nu}$. Since these operators factorize in terms of the quantum channel $\mathcal{M}$ describing the correlation functions ($k = 1$), their spectrum is directly known, and the spectrum of the full Markovian dynamics directly follows as the union of these spectra. In this way, the timescales for the decay of the k -OTOC directly relate to the decay rate of the correlation functions. Note however that Jordan blocks can and generally will appear, as can be directly checked for $k = 2$, leading to the polynomial prefactors already observed in Eq. (75).

Before presenting a detailed analysis of the eigenspectrum of $\mathcal{T}$, we note that this Markovian dynamics can be written in an alternative way that makes the connection with free probability more explicit. Because of the alternation of transfer matrices and Möbius functions, the dynamics can also be written as a two-step process

$$|\kappa_t)_{\sigma} = \sum_{\nu \subseteq \sigma} W_{\sigma\nu} |\varphi_t)_{\nu}, \quad (124)$$

$$|\varphi_t)_{\nu} = \sum_{\sigma \subseteq \nu} \mathcal{M}_{\nu\sigma} |\kappa_{t-1})_{\sigma}, \quad (125)$$

with the initial condition corresponding to

$$|\kappa_{t=0})_{\sigma} = d_C^k |a_{\circ}). \quad (126)$$

The important observation is that Eqs. (124) and (125) strongly resemble the self-consistent relations between moments and free cumulants from Eqs. (27) and (26), up to unimportant factors of d_C . In the long-time limit where both states become time independent, this relation enforces that these states indeed satisfy the expected relations from free probability (see discussion below).

We note that the move from two sets of variables and an update rule that depends on a single time step to a single set of variables with a two-step update rule is reminiscent of the move from Hamiltonian to Lagrangian dynamics.

This analogy is strengthened by the observation that the Möbius function implements a discrete derivative on the noncrossing partition lattice, effectively returning the states $|\kappa_t)_\sigma$ as the conjugate variables of the states $|\varphi_t)_\nu$.

E. Steady state

We already argued that, for traceless operators a and b , the steady-state value of the k -OTOC vanishes for all k , indicating asymptotic freeness between $a(t)$ and b . The dynamics of Eq. (113) can additionally be used to make explicit the appearance of free probability and free cumulants under ergodic many-body dynamics.

For ergodic correlation functions, the quantum channel (66) has a unique leading eigenvalue 1 from unitality, with the corresponding eigenvector being the vectorized identity matrix. From the block structure of Eq. (117) and the surrounding discussion, it follows that the Markovian process for the k -OTOC has C_k leading eigenvalues 1, corresponding to the total number of diagonal blocks indexed by a noncrossing partition σ with corresponding operators $\mathcal{M}_{\sigma\sigma}$. The corresponding left and right eigenvectors can be directly obtained, which will in turn allow for obtaining the steady-state value of the k -OTOC for general operators through a projection on this eigenspace.

1. Eigenstates

The full derivation of the leading eigenstates is detailed in Appendix D, and here we cite only the final results. Each left and right eigenstate is labeled by a noncrossing partition, and we denote the right eigenstates as $|\varphi_\sigma)$ and the left eigenstates as $(\kappa_\nu|$. These eigenstates are given by

$$|\varphi_\sigma) = d_A^{|\sigma|-k} \sum_{\sigma \subseteq \rho} d_C^{|\rho|} |\sigma)_\rho, \quad (127)$$

$$(\kappa_\nu| = \sum_{\rho \subseteq \nu} (d_A d_C)^{-|\rho|} \mu(\nu, \rho) (\rho|_\rho. \quad (128)$$

Note that the subscript ρ denotes the auxiliary (bond dimension) degree of freedom, and $|\sigma)$ and $(\rho|$ present the vectorized permutation operators σ and ρ , respectively. The notation of $|\varphi_\sigma)$ and $(\kappa_\nu|$ is again motivated by the similarity of these definitions to the relation between moments and free cumulants, Eqs. (27) and (26), which will be made more explicit shortly. Overlaps between such states can be directly evaluated, where it is useful to observe that $|\varphi_\sigma)$ has only components ρ that lie “above” σ , whereas $(\kappa_\nu|$ has only components ρ that lie “below” ν . The overlap, hence, depends only on the noncrossing partitions ρ that lie on a path from σ to ν , i.e., $\sigma \subseteq \rho \subseteq \nu$, such that we can, e.g., evaluate

$$\begin{aligned} ((\kappa_\nu| \varphi_\sigma)) &= \sum_{\sigma \subseteq \rho \subseteq \nu} d_A^{|\sigma|-|\rho|-k} \mu(\nu, \rho) (\sigma|_\rho) \\ &= \sum_{\sigma \subseteq \rho \subseteq \nu} d_A^{|\sigma|+|\sigma^{-1}\rho|-|\rho|-k} \mu(\nu, \rho) \\ &= \sum_{\sigma \subseteq \rho \subseteq \nu} \mu(\nu, \rho) = \delta_{\nu\sigma}. \end{aligned} \quad (129)$$

In the second equality, we used that $(\sigma|_\rho) = d_A^{|\sigma^{-1}\rho|}$, in the third equality, we used that a triangle inequality (A7) is saturated for $\sigma \subset \rho \subset \nu$, and in the final equality, we used the defining property of the Möbius function (54).

These states hence present a biorthogonal eigenbasis satisfying

$$((\kappa_\nu| \varphi_\sigma)) = \delta_{\nu\sigma}. \quad (130)$$

2. Steady-state value

For ergodic dynamics, these states exhaust the full eigenspace of nondecaying operators, and at late times we can write

$$\lim_{t \rightarrow \infty} T^t = \sum_{\sigma \in \text{NC}(k)} |\varphi_\sigma) ((\kappa_\sigma|. \quad (131)$$

In order to calculate the steady-state value of the k -OTOC (50), we note that the overlap of these eigenstates with the boundary states encoding the operators a and operators b [Eq. (119)] corresponds to

$$((\psi_b| \varphi_\sigma)) = \varphi_{\sigma^*}(b_1, \dots, b_k), \quad (132)$$

$$\begin{aligned} ((\kappa_\sigma| \psi_a)) &= \sum_{\nu \subseteq \sigma} \mu(\nu, \sigma) \varphi_\nu(a_1, \dots, a_k) \\ &= \kappa_\sigma(a_1, \dots, a_k). \end{aligned} \quad (133)$$

Here we have used that the moments of a follow from the overlaps with the permutation operators $\varphi_\nu(a) = d_A^{-|\nu|} (\nu| a_\circ)$, and that the Möbius functions express the free cumulants in terms of the moments.

Taking the above results together, it immediately follows that

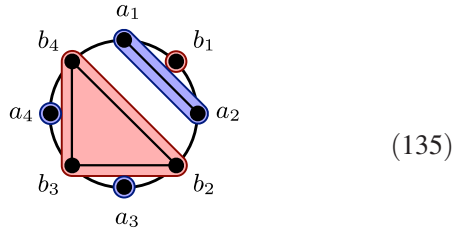
$$\begin{aligned} \lim_{t \rightarrow \infty} C_{ab}^{(k)}(t) &= \lim_{t \rightarrow \infty} ((\psi_b| T^t | \psi_a)) \\ &= \sum_{\sigma \in \text{NC}(k)} \varphi_{\sigma^*}(b_1, \dots, b_k) \kappa_\sigma(a_1, \dots, a_k), \end{aligned} \quad (134)$$

in accordance with Eq. (100).

3. Asymptotic freeness

This expression is generally expected when $a_i(t)$ and b_j are asymptotically free [71]. The different contributions

can be graphically illustrated, where, e.g., choosing $\sigma = (12)(3)(4)$ for $k = 4$ returns



$$= \varphi(b_1)\varphi(b_2b_3b_4)\kappa_2(a_1, a_2)\kappa_1(a_3)\kappa_1(a_4).$$

Here we illustrate σ in blue and σ^* in red, where both factorize in terms of free cumulants and moments, respectively. All such terms are illustrated in Fig. 5 for $k = 4$, which also makes apparent the Kreweras complement of each noncrossing partition. The interpretation of Eq. (134) as asymptotic freeness can be made explicit by noting that operators are free if their mixed free cumulants vanish and by expanding the k -OTOC in free cumulants. Assuming $a_i(t)$ and b_j to be free, we have, with a slight abuse of notation, that $\kappa_n(\dots, a_i(t), \dots, b_j, \dots) = 0$ whenever a free cumulant contains both an a and a b . We can use this property in the expansion of the k -OTOC to write

$$\begin{aligned} & \lim_{t \rightarrow \infty} \varphi(a_1(t)b_1a_2(t)b_2, \dots, a_k(t)b_k) \\ &= \sum_{\pi \in \text{NC}(2k)} \kappa_\pi(a_1(t), b_1, a_2(t), b_2, \dots, a_k(t), b_k) \\ &= \sum_{\sigma \in \text{NC}(k)} \sum_{\nu \subseteq \sigma^*} \kappa_\sigma(a_1, \dots, a_k) \kappa_\nu(b_1, \dots, b_k) \\ &= \sum_{\sigma \in \text{NC}(k)} \kappa_\sigma(a_1, \dots, a_k) \varphi_{\sigma^*}(b_1, \dots, b_k). \end{aligned} \quad (136)$$

In the second equality, we have used that all noncrossing partitions on $2k$ elements in which a and b are not

connected can be written as the form $\pi = \sigma \cup \nu$, with both σ and ν noncrossing partitions on k elements and where $\nu \subseteq \sigma^*$, factorized the free cumulant accordingly, and used $\kappa_\sigma(a_1(t), \dots, a_k(t)) = \kappa_\sigma(a_1, \dots, a_k)$ following from $\varphi(a(t)) = \varphi(a)$. In the final equality, we have used the expansion of the moments in terms of free cumulants. Returning to Eq. (134), we find that we have a decomposition of all nondecaying terms in terms of free cumulants, which satisfy the expected factorization properties from full ETH and free probability.

F. Approach to equilibrium

Full ETH and free probability more generally predict that all terms in the k -OTOC can be given an interpretation as free cumulants—not just the static terms. As discussed in the previous section, the leading eigenstates of $\mathcal{T}$ [Eq. (127)] are in one-to-one correspondence with the nonvanishing mixed cumulants determining the steady-state value of the OTOC. We here show that subleading eigenstates, which contribute terms that decay in time, can be given a similar interpretation, and the resulting terms in the dynamics correspond to free cumulants in which a and b are in the same block. We illustrate this correspondence by constructing the exact eigenstates of $\mathcal{T}$ corresponding to the leading nontrivial eigenvalue. As indicated in Sec. III E, these eigenstates with eigenvalue λ govern the approach toward equilibrium $\propto \lambda^t$. While for traceless observables these eigenstates will have vanishing overlaps with the boundary conditions, leading to a decay $\propto \lambda^{2t}$, these terms are nonvanishing for a nontraceless operator and simplify the following analysis. The construction of these subleading eigenstates and the correspondence with mixed cumulants can be directly extended to general eigenstates, albeit in a slightly more involved way.

1. Eigenstates

For an ergodic quantum channel $\mathcal{M}$ with leading nontrivial eigenvalue λ , such that the correlation functions

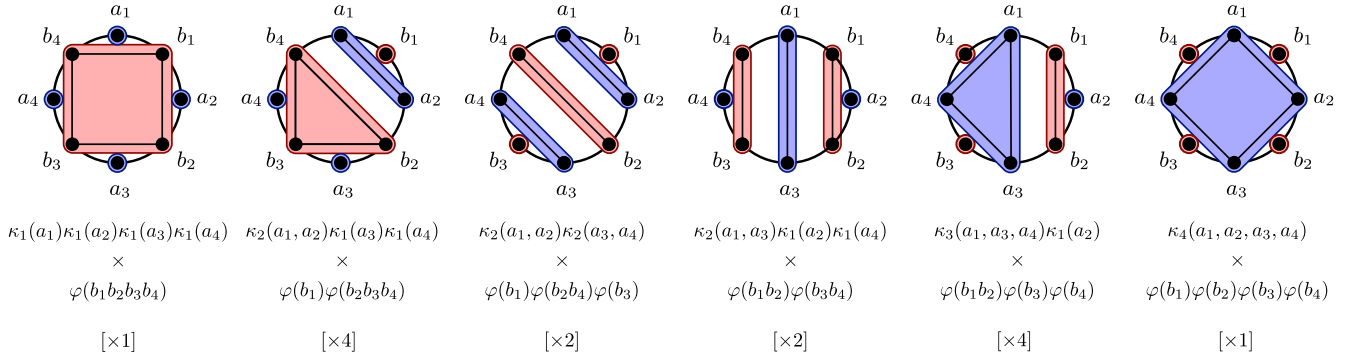


FIG. 5. Graphical illustration of the terms determining the steady-state value of the k -OTOC for $k = 4$. The shaded blue areas denote the free cumulants labeled by a noncrossing partition σ on the set $\{a_1, a_2, a_3, a_4\}$; the shaded red areas denote the moment labeled by its complement σ^* on the set $\{b_1, b_2, b_3, b_4\}$. The [$\times n$] indicates that there are n equivalent arrangements of this diagram. Based on a similar figure in Ref. [61].

decay as λ^l , the transfer matrix $\mathcal{T}$ has the same leading nontrivial eigenvalue λ . The corresponding right and left eigenstates a_λ and b_λ can be used to construct appropriately dressed left and right eigenstates of $\mathcal{T}$ with (not necessarily the subleading) eigenvalue λ . In the previously defined leading eigenstates, one of the contractions in the permutation operators is replaced by the right (left) eigenoperator a_λ (b_λ),

$$(|\varphi_\sigma^m\rangle) = a_{\lambda,m}|\varphi_\sigma\rangle, \quad ((\kappa_\nu^n| = ((\kappa_\nu|b_{\lambda,n}, \quad (137)$$

where $m, n = 1, \dots, k$ denote the contraction to be dressed, and $a_{\lambda,m}$ acts on all permutations in the expansions of these states as by dressing the m th contraction with this operator. These operators remain left and right eigenstates (see Appendix D) of $\mathcal{T}$ and present the complete set of $C_k \times k$ eigenoperators with eigenvalue λ . However, they are no longer biorthogonal.

Again fixing the right states as $|\varphi_\sigma^m\rangle$, states in which the same contraction is dressed again satisfy the expected orthogonality relations

$$((\kappa_\nu^n|\varphi_\sigma^n)) = \delta_{\nu\sigma}. \quad (138)$$

However, as detailed in Appendix D, for $m \neq n$ the state $((\kappa_\nu^n|$ additionally has a unit overlap with states $|\varphi_\sigma^m\rangle$, where (i) m and n are in the same cycle of ν , and (ii) all cycles of σ equal the cycles of ν except for the cycle that contains m and n , which in σ is split in two in such a way that the Kreweras complement of this cycle consists solely of singletons and a single transposition (m, n) . This σ is uniquely defined and will be denoted as $\tilde{\nu}(m, n)$. This permutation is constructed in such a way that m and n are in the same cycle in $\tilde{\nu}(m, n)^{-1}\nu$ and in different cycles in all $\tilde{\nu}(m, n)^{-1}\rho$ with $\tilde{\nu}(m, n) \subseteq \rho \subset \nu$. In all other cases, these states are orthogonal. Additionally, left and right eigenstates corresponding to different eigenvalues are automatically orthogonal.

2. Biorthogonalization

From the above relation, a biorthogonal basis can be defined iteratively. The free cumulant states in which n is a singleton within ν are already properly biorthogonalized with respect to the moment states, and we define these as

$$((\tilde{\kappa}_\nu^n| = ((\kappa_\nu^n| \quad \text{if } n \text{ is a singleton in } \nu \quad (139)$$

and

$$((\tilde{\kappa}_\nu^n| = ((\kappa_\nu^n| - \sum_{m \neq n} ((\tilde{\kappa}_{\tilde{\nu}(m,n)}^m| \quad \text{otherwise,} \quad (140)$$

where the summation $m \neq n$ runs over all elements that are in the same cycle of ν as n . This relation can be evaluated recursively: The states are defined when n is a singleton in

ν , and when n is not part of a singleton in ν , then n is necessarily part of a smaller cycle of $\tilde{\nu}(m, n)$. In this way, states where n is part of a transposition in ν can first be defined, after which states in which n is part of a cycle of length three can be defined, and so on.

These states satisfy the expected biorthogonality, where

$$((\tilde{\kappa}_\nu^n|\varphi_\sigma^m)) = \delta_{mn}\delta_{\nu\sigma}. \quad (141)$$

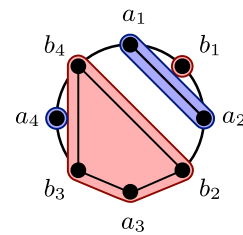
Abstractly, the above construction of eigenstates is the same as for the leading ones if we replace the noncrossing lattice $\text{NC}(k)$ by the partially ordered set $\text{NC}(k) \times \{1, 2, \dots, k\}$, where the second component keeps track of the dressing. The partial order is defined by declaring $(\mu, n) \subseteq (\nu, m)$ if $\mu \subseteq \nu$, and m and n are in the same cycle of $\mu^{-1}\nu$, and the recursive construction above corresponds to computing the Möbius function for this extended partially ordered set.

3. Projectors as mixed free cumulants

The corresponding projectors again have a direct interpretation as mixed free cumulants. Considering for concreteness $\sigma = (12)(3)(4)$ and fixing $n = 3$ such that the singleton construction applies, the relevant overlaps directly return

$$\begin{aligned} & ((\psi_b|\varphi_{(12)(3)(4)}^{n=3}))((\tilde{\kappa}_{(12)(3)(4)}^{n=3}|\psi_a)) \\ &= \varphi(b_1)\varphi(b_2a_3b_4b_1)\kappa_2(a_1, a_2)\kappa_1(a_4)\varphi(b_3a_3). \end{aligned} \quad (142)$$

The operator b_λ effectively “annihilates” the part of the operator a_n that has an overlap with a_λ from the free cumulants and inserts a_λ in between b_{n-1} and b_n in the moments. This result can be made explicit by choosing $a_3 = a_\lambda$, in which case the resulting contribution to the dynamics can be graphically represented as

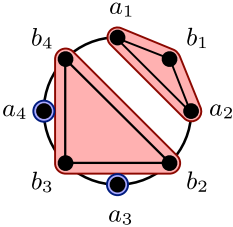


$$= \varphi(b_1)\varphi(b_2a_3b_3b_4)\kappa_2(a_1, a_2)\kappa_1(a_4).$$

This argument directly extends to eigenstates in which n is not part of a singleton. Consider again $\sigma = (12)(3)(4)$ but $n = 2$. The corresponding projector evaluates to

$$\begin{aligned} & ((\psi_b|\varphi_{(12)(3)(4)}^{n=2}))((\tilde{\kappa}_{(12)(3)(4)}^{n=2}|\psi_a)) \\ &= \varphi(a_\lambda b_1)\varphi(b_2b_3b_4)\kappa_1(a_3)\kappa_1(a_4) \\ & \quad \times [\varphi(a_1b_\lambda a_2) - \varphi(b_\lambda a_1)\varphi(a_2) - \varphi(a_1)\varphi(b_\lambda a_2)]. \end{aligned} \quad (144)$$

The operator b_λ now annihilates part of the combined operator $a_2 a_1$, i.e., the product of all operators in the same cycle, and reinserts these in the moments, merging this term with the cycle containing b_1 . Note that the additional terms in the last line, which arise from the biorthogonalization, here appear to avoid double-counting contributions from the singletons. If we choose $a_2 a_1 = a_\lambda$ and assume that a_1 and a_2 have a vanishing overlap with b_λ , this contribution can be graphically represented as



$$= \varphi(a_1 b_1 a_2) \varphi(b_2 b_3 b_4) \kappa_1(a_3) \kappa_1(a_4). \quad (145)$$

More generally, these eigenstates annihilate a full cycle of a 's and insert it in a cycle of b 's.

We find that the slowest-decaying modes again correspond to mixed cumulants, where the decaying mode of the quantum channels introduces free mixed cumulants that contain a single block in which both a 's and b 's appear. The contribution from these mixed cumulants decays exponentially in time as λ^t , as necessary for the emergence of free independence. These results are consistent with free cumulants κ_n in which a and b do not mix not decaying, and free cumulants κ_n in which a single cycle of a 's is merged with a noncrossing partition of b 's decaying $\propto \lambda^t$.

The construction of eigenstates directly extends to eigenvalues λ^m via dressing by multiple operators, where now however no two replicas within a single cycle can be dressed (see Appendix D). Because of this restriction, the eigenstates no longer form a complete basis, consistent with the appearance of Jordan blocks. The corresponding projectors again allow for an interpretation in terms of mixed free cumulants, where additional blocks mixing a and b appear.

V. CONCLUDING REMARKS

We have presented a full characterization of the dynamics of higher-order OTOCs in a minimal quantum-circuit model for local unitary dynamics. In doing so, we uncovered a dynamical picture in terms of free cumulants from free probability. Free probability was recently shown to underpin higher-order extensions of the ETH and, even though there is no meaningful notion of eigenstates in our discussed model, we reproduce its predicted decomposition of higher-order OTOCs in free cumulants and establish its predicted applicability of free probability beyond Hamiltonian or Floquet dynamics. In recent years,

quantum circuits have gained attention as minimal models for many-body quantum dynamics and were fundamental in uncovering the dynamics of scrambling and entanglement, but results on higher-order OTOCs are limited and no connection with full ETH was made until now. In this way, this work constitutes the first analytical result on full ETH in ergodic many-body dynamics, as well as recasting the eigenstate-based picture in a dynamical picture.

For the OTOC dynamics, all decay timescales were exactly characterized, and it was shown how ergodicity on the level of correlation functions here implies ergodicity on the level of the OTOCs, i.e., the emergence of freeness. We showed how in this model all k -OTOCs decay at the same rate $\propto \lambda^{2t}$, twice as fast as the correlation functions $\propto \lambda^t$ and significantly slower than the naive expectation $\propto \lambda^{kt}$. As a special limit, it was shown how dual-unitary gates return maximally ergodic dynamics, where both the correlation functions and k -OTOCs reach their ergodic steady-state value (almost) immediately. Furthermore, ergodicity was shown to be structurally stable away from this limit—more generally an open problem in many-body quantum chaos.

A fundamental result is the identification of a multi-replica influence matrix for higher-order OTOCs, capturing the effects of multiple copies of the ergodic environment. This influence matrix reproduces the perfect dephaser limit of a perfectly Markovian bath in the limit of time-ordered correlation functions, extending the perfect Markovianity to higher-order OTOCs. The general appearance of Markovian baths throughout statistical physics and many-body dynamics, combined with the combinatorial structure of noncrossing partitions encoded in this Markovian influence matrix, suggests the more general applicability of full ETH and free probability in many-body quantum dynamics. Eigenmodes of the Markovian process labeled by noncrossing partitions—the fundamental objects in both free probability and full ETH—could be identified with free cumulants. Through these eigenmodes we explicitly obtained steady-state values for the OTOCs consistent with time-evolved observables becoming free with respect to static observables, and we explicitly identified the decaying eigenmodes as mixed free cumulants.

Having established these results, we now discuss future directions and questions raised. The analytically established decay of the 2-OTOC at twice the rate of the correlation functions is consistent with previous numerical studies [122]. However, the identical decay rates for higher-order OTOCs differ from the hierarchy of timescales predicted using arguments from large deviation theory and numerically observed in a semiclassical system [59]. It remains an open question to identify the physical mechanism responsible for these differing predictions, whether this is the semiclassical limit, locality, the choice of observables, the non-Markovianity, or something else entirely. Identifying the origin of this difference is crucial in establishing the applicability of these distinct predictions.

Preliminary numerical explorations additionally indicate that the special role of qubit observables, with faster decay of higher-order OTOCs, can still be observed in more generic many-body quantum systems. The full extent to which this applies however remains an open question.

These results represent a starting point for general studies of quantum ergodicity, chaos, and operator scrambling, and can be readily probed in current gate-based quantum computing platforms. In all such cases, our results provide a first benchmark. A main step in this direction is the identification of structured Floquet models of many-body quantum dynamics which provably exhibit the same k -OTOCs dynamics. Such models allow for direct comparison of the eigenstate-based ETH prediction and the dynamical picture presented in this work, where higher-order correlations between matrix elements could be directly related to the corresponding higher-order OTOC dynamics. One such model was recently introduced in a follow-up work [123]. We also note that Ref. [75] previously established the role of the noncrossing partition lattice in structured dual-unitary circuit dynamics, but exact results required the limit of a diverging Hilbert-space dimension, limiting its applicability on the level of eigenstates. There has been significant interest in how random quantum circuits form designs [124–127], originating with Ref. [128], where it might similarly be possible to recover free-probability predictions for the k -OTOC using the formation of approximate k -designs [61]. Another main step is in establishing this structure in more generic systems, e.g., by perturbing away from solvable models, by extensive numerical investigations in paradigmatic models, or by experimental tests such as in Ref. [54]. A natural question here is the role of conservation laws, such as energy conservation. How does this affect the underlying structure and the emergence of freeness with respect to microcanonical or thermal states, does the approach to equilibrium change, and is there a hierarchy of relaxation times? In the presence of multiple conservation laws, it is *a priori* not clear if observables become free with respect to generalized Gibbs states or if freeness emerges only in individual charge sectors. Even more drastic changes to the free-probability description might be observed in integrable systems with an extensive number of conservation laws. In both cases, the lack of ergodicity implies that the dynamics cannot explore the full Hilbert space, limiting the random-matrix description. A natural first step here would be to extend the minimal model discussed in this work to include conservation laws by imposing a block structure on the random unitaries in the bath, following previous studies of the effect of conservation laws on quantum dynamics [129,130]. It is also possible to consider weaker forms of ergodicity breaking, which become apparent only for correlation functions of sufficiently high order. A simple example, readily applicable in our minimal setting, is by describing

the bath dynamics by an ensemble of unitaries which forms a k but not a $k + 1$ design, e.g., the Clifford group. In this case, we expect the dynamics and the stationary state of sufficiently high-order k -OTOCs to differ from our results. Given the many ways in which ergodicity can be broken, e.g., through many-body quantum scars or Hilbert-space fragmentation, we expect that different mechanisms could lead to ergodicity breaking that is apparent only at the level of higher-order OTOCs, where our results can again serve as a benchmark for ergodicity. Since the discussed model was originally introduced in the study of deep thermalization [31], another natural direction is to explore the interplay of deep and full thermalization, through, e.g., partial measurements or conditioned dynamics, for which there are no results yet on either the late-time dynamics or the corresponding timescales. As a final example, we note that, while the discussed model supports a notion of locality, it does so only in a minimal way. The spatiotemporal structure of higher-order OTOCs in general lattice systems remains an open question, and our results present a first step to a hydrodynamic description of free cumulants in Floquet and random quantum-circuit models (following, e.g., the approaches of Refs. [82–84,114,115]).

While the model studied in this work is particularly simple, we expect the presented influence matrix to be relevant for more structured dynamics. The influence-matrix approach has proven successful for studying correlation functions, with the perfect dephaser limit as a natural Markovian limit, and our results provide a starting point for extending such methods to higher-order OTOCs. A first extension would be to go beyond the purely Markovian limit discussed here. In the minimal model, the assumption of an infinite bath precludes any information backflow and resulting non-Markovianity. Finite-size corrections can originate from subleading orders in the Weingarten calculus. Keeping the Hilbert-space dimension of the bath large but finite, such corrections on the thermodynamic limit can be systematically reintroduced and allow for a direct study of the corresponding non-Markovian effects. More generally, inspired by the successful implementation of the influence-matrix approach for studying time-ordered correlation functions [5,6], the multireplica influence matrix for k -OTOCs introduced here could provide the starting point for developing and benchmarking numerical schemes beyond exact solvability and perturbative approaches. As one example, effective semigroup influence matrices were recently shown to be able to accurately describe the dynamics of quantum impurity models for long times [28], and our work opens up this approach as a classical algorithm for the study of (higher-order) OTOC dynamics.

The minimal model can also be considered as the circuit equivalent of Lindblad dynamics, where a structured subsystem is coupled to a bath which rapidly reaches its equilibrium (random-matrix) state. It would be interesting to obtain a continuum-time limit of the presented dynamics

and study the interplay with external noise and dissipation, necessarily present in any experimental realization. Such a continuum limit would lead to an effective open-system dynamics for the k -OTOC that is distinct from simply replacing the unitary evolution operator for each observable by the corresponding Liouvillian, preserving the symmetries and structure from free probability. Correlations in the noise should be directly observable on the level of free cumulants and asymptotic freeness. This approach directly relates to the study of Ruelle-Pollicott resonances, a recent topic of active interest in quantum many-body dynamics [122,131–134]. In this approach, the decay of correlations in closed systems is related to the eigenspectrum of the evolution operator of a corresponding open system, similar to replacing the unitary evolution by a (contracting) quantum channel whose eigenspectrum can be studied. Ruelle-Pollicott resonances have recently been studied on the level of OTOCs [122,132], and this approach presents a natural way of replacing the unitary OTOC dynamics with an effective open-system description which preserves the expected structure from free probability.

Another direction of future work is the extension of these results to multipoint correlation functions $\langle A(t_1)A(t_2), \dots, A(t_k) \rangle$. The general applicability of full ETH [57,60] here again suggests the existence of a similar dynamical picture.

To conclude, we believe that these results pave the way for a systematic study of higher-order scrambling in quantum systems: on the fundamental level by clarifying the emergence of free cumulants and free probability, and on the practical level through the extension of the influence-matrix approach to OTOCs and their higher-order generalizations.

ACKNOWLEDGMENTS

We thank Gabriel O. Alves, John Chalker, Silvia Pappalardi, and Xhek Turkeshi for useful discussions. F.F. acknowledges support from the European Union's Horizon Europe program under the Marie Skłodowska Curie Action GETQuantum (Grant No. 101146632). P.W.C. acknowledges support from the Max Planck Society.

DATA AVAILABILITY

The numerical code for computing k -OTOCs using the Markovian process on the noncrossing lattice as well as the data and scripts for generating Figs. 2–4 are provided in Ref. [135].

APPENDIX A: A PRIMER ON FREE PROBABILITY

In this appendix we briefly review the relevant aspects of free probability. We refer the reader to Refs. [70,71,88] for more detailed introductions.

1. Noncommutative probability spaces

The main idea of free probability is to replace the commutative algebra formed by random variables on a classical probability space by elements of a possibly noncommutative unital $*$ -algebra $\mathcal{A}$. A prime example is the algebra of bounded linear operators on some Hilbert space, i.e., in the finite-dimensional case, the algebra of $D \times D$ complex matrices. The notion of computing expectation values is replaced by a unital and positive $*$ -linear functional φ on $\mathcal{A}$ obeying (i) $\varphi(1) = 1$, (ii) $\varphi(a^\dagger a) \geq 0$, and (iii) $\varphi(a^\dagger) = \varphi(a)^*$. Such a functional is called a state on $\mathcal{A}$. The canonical example for the above matrix algebra is $\varphi(\bullet) = \text{Tr}(\bullet)/D$, which further exhibits the property of being tracial, i.e., $\varphi(ab) = \varphi(ba)$. The algebra $\mathcal{A}$ equipped with the state φ is called a noncommutative probability space.

2. Freeness

The notion of free independence extends the notion of independence to noncommuting variables. Two subalgebras A and B of $\mathcal{A}$ are said to be freely independent if, for all $a_i \in A$ and $b_j \in B$,

$$\varphi(a_1 b_1 a_2 b_2 \dots a_k b_k) = 0, \quad (\text{A1})$$

provided $\varphi(a_i) = \varphi(b_j) = 0, \forall i, j$. Informally, two operators being freely independent means that their eigenbases appear maximally random with respect to each other. From this definition, it follows, e.g., that for nontraceless variables

$$\varphi(ab) = \varphi(a)\varphi(b), \quad (\text{A2})$$

as also expected for classically independent variables. The noncommutativity however prevents a similar factorization of $\varphi(abab) \neq \varphi(a^2)\varphi(b^2)$, as would be the case for commuting independent variables, but rather leads to

$$\varphi(abab) = \varphi(a^2)\varphi(b)^2 + \varphi(a)^2\varphi(b^2) - \varphi(a)^2\varphi(b)^2. \quad (\text{A3})$$

3. Noncrossing partitions

In order to describe the combinatorics of free probability, it is useful to introduce noncrossing partitions [136]. The noncrossing partitions of k elements are a subset of the partitions of k elements. Writing a partition as $\sigma = \{V_1, V_2, \dots, V_{|\sigma|}\}$, a partition is said to be noncrossing if, when the elements p_i, q_i and p_j, q_j belong to distinct blocks V_i and V_j , there are no “crossings” in which $p_i < p_j < q_i < q_j$. These noncrossing partitions are often represented graphically by arranging the elements on the circle and connecting elements that are in the same block. Focusing on $k = 4$ for concreteness, the partition (12)(34)

(read as $\{\{1,2\},\{3,4\}\}$) is noncrossing and $(13)(24)$ is crossing, as made explicit in their graphic representation:

$$(12)(34) = \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ 4 \quad 2 \\ \diagdown \quad \diagup \\ 3 \end{array}, \quad (13)(24) = \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ 4 \quad 2 \\ \diagdown \quad \diagup \\ 3 \end{array}. \quad (\text{A4})$$

A special role is played by the partition with k blocks, such that all blocks are singletons, and the partition with a single block

$$(1)(2)(3)(4) = \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ 4 \quad 2 \\ \diagdown \quad \diagup \\ 3 \end{array}, \quad (1234) = \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ 4 \quad 2 \\ \diagdown \quad \diagup \\ 3 \end{array}. \quad (\text{A5})$$

These correspond to the identity $\circ = (1)(2)\dots(k)$ and the cyclic shift permutation $\square = (12\dots k)$, respectively.

4. Permutations and partial order

We can identify partitions with permutations by interpreting the blocks of a partition σ , with the elements of the blocks arranged in ascending order, as the cycles of a permutation $\sigma \in S_k$ of k elements. We usually do not distinguish between both interpretations.

To characterize the relevant permutations, we consider the rank or Cayley weight of a permutation $\sigma \in S_k$ defined as $C(\sigma) = k - |\sigma|$, where $|\sigma|$ is the number of cycles in the cycle decomposition of σ . The rank corresponds to the smallest number of transpositions it takes to realize the permutation σ . The identity $\circ$ minimizes the rank as $C(\circ) = 0$, whereas the cyclic permutation $\square$ maximizes it as $C(\square) = k - 1$. The Cayley weight gives rise to a notion of distance in the symmetric group S_k or, more precisely, in its Cayley graph. The vertices of this graph are the permutation $\sigma \in S_k$, and σ to ν are connected by an edge if $\sigma^{-1}\nu$ is a transposition. The length of the shortest paths, i.e., the geodesics, connecting σ and $\nu \in S_k$ is determined by the Cayley weight of $\sigma^{-1}\nu$ as

$$\ell(\sigma, \nu) = C(\sigma^{-1}\nu) = k - |\sigma^{-1}\nu| \quad (\text{A6})$$

and provides a metric on S_k . Accordingly, this metric fulfills the triangle inequality

$$\ell(\sigma, \rho) + \ell(\rho, \nu) \leq \ell(\sigma, \nu), \quad (\text{A7})$$

with equality if and only if ρ lies on a geodesic from σ to ν . In terms of the number of cycles, the triangle inequality reads

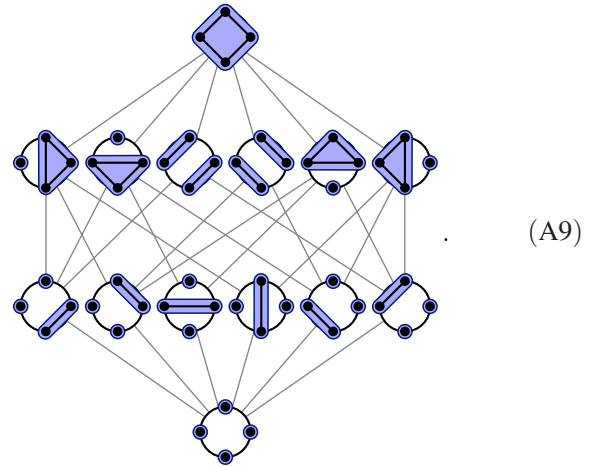
$$|\sigma^{-1}\rho| + |\rho^{-1}\nu| \leq k + |\sigma^{-1}\nu|. \quad (\text{A8})$$

Of particular interest are the geodesics from the identity $\circ$ to the cyclic shift $\square$, which are of length $\ell(\circ, \square) = C(\square) = k - 1$, and the permutations σ lying on such geodesics. The set of such permutations is denoted as the interval $[\circ, \square]$ and comes equipped with a partial order denoted by $\subseteq$, where $\nu \subseteq \sigma$ if and only if ν and σ lie on a common geodesics from $\circ$ to $\square$ and $\ell(\circ, \nu) = C(\nu) \leq \ell(\circ, \sigma) = C(\sigma)$. The resulting partially ordered set $[\circ, \square]$ is isomorphic to the noncrossing lattice [101].

5. The noncrossing partition lattice

This isomorphism is, as described above, simply given by interpreting the cycles of a permutation $\sigma \in [\circ, \square]$ as the blocks of a partition σ , which then turns out to be noncrossing. The noncrossing partitions inherit a partial ordering, where a noncrossing partition $\sigma \subseteq \nu$ if all blocks of σ are contained in the blocks of ν . This order is the dual of the usual refinement order of partitions.

The resulting ordering between noncrossing partitions can be represented in the noncrossing partition lattice, here illustrated for $k = 4$:

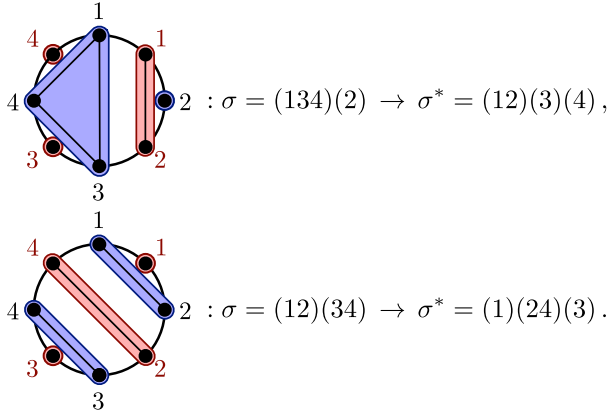


(A9)

Noncrossing partitions σ and ν are connected by a line if $\sigma \subseteq \nu$, and moving up a row in the lattice corresponds to increasing the rank of the noncrossing partitions by one.

6. Kreweras complement

The noncrossing partition lattice exhibits a symmetry that can be made explicit by defining the Kreweras complement [136], which inverts the noncrossing partition lattice [101]. The Kreweras complement σ^* of a noncrossing partition σ is defined as $\sigma^{-1}\square$ and can be graphically represented as the largest noncrossing partition on the “dual” of the elements on the circle, e.g.,



These satisfy $C(\sigma) + C(\sigma^*) = k - 1$ or, equivalently, $|\sigma| + |\sigma^*| = k + 1$.

7. Free cumulants

Mixed free cumulants are defined recursively through the combinatorics of noncrossing partitions as

$$\varphi_k(a_1, a_2, \dots, a_k) = \sum_{\sigma \in \text{NC}(k)} \kappa_\sigma(a_1, a_2, \dots, a_k) \quad (\text{A10})$$

with $\kappa_1(a) = \varphi(a)$ and $\varphi_k(a_1, a_2, \dots, a_k) = \varphi(a_1 a_2 \dots a_k)$. The sum is taken over all noncrossing partitions $\text{NC}(k)$ of the set $\{1, 2, \dots, k\}$, and κ_σ factorizes according to the blocks of $\sigma = \{V_1, \dots, V_{|\sigma|}\}$ as

$$\kappa_\sigma(a_1, a_2, \dots, a_k) = \prod_{V \in \sigma} \kappa_V(a_1, a_2, \dots, a_k), \quad (\text{A11})$$

where each block $V = \{V(1), \dots, V(s)\}$ of size $s = |V|$ defines a mixed free cumulant

$$\kappa_V(a_1, a_2, \dots, a_k) = \kappa_s(a_{V(1)}, \dots, a_{V(s)}). \quad (\text{A12})$$

Mixed free cumulants vanish whenever two or more variables are freely independent, which can also be taken as the definition of free independence. Returning to the example of $a \in A$ and $b \in B$ freely independent, we have that

$$\kappa_s(\dots, a, \dots, b, \dots) = 0, \quad \forall s. \quad (\text{A13})$$

8. Möbius inversion

Inverting the definition of the mixed free cumulants (A10) returns

$$\kappa_\sigma(a_1, a_2, \dots, a_n) = \sum_{\nu \subseteq \sigma} \varphi_\nu(a_1, \dots, a_n) \mu(\nu, \sigma), \quad (\text{A14})$$

where $\varphi_\nu(a_1, \dots, a_n)$ is defined analogously to Eq. (A11). The function $\mu(\nu, \sigma)$ acts as the Möbius function on the

noncrossing partition lattice and has an explicit expression as

$$\mu(\nu, \sigma) = \prod_{V \in \nu^{-1}\sigma} (-1)^{|V|-1} C_{|V|-1}. \quad (\text{A15})$$

For a fixed σ and ν , the Möbius function satisfies

$$\sum_{\sigma \subseteq \rho \subseteq \nu} \mu(\nu, \rho) = \delta_{\nu\sigma}, \quad (\text{A16})$$

which is, in fact, its defining property. More precisely, the above property can be read as the Möbius function being the inverse of the zeta function of the noncrossing lattice defined by

$$\zeta_{\nu\sigma} = \begin{cases} 1 & \text{if } \nu \subseteq \sigma, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A17})$$

APPENDIX B: WEINGARTEN CALCULUS AND HAAR AVERAGING

In this appendix, we provide the details for the derivation of the k -OTOC in the thermodynamic limit leading to Eq. (50). For convenience of the reader and for keeping this appendix self-contained, we repeat some of the formulas and arguments from Sec. II. We start by briefly reviewing Haar averages for the unitary group and Weingarten calculus, which allows for performing the average over the independent unitaries V_t entering Eq. (36) and ultimately determines the averaged k -OTOC for finite system sizes. We obtain the leading behavior in d_E , identifying the terms which survive in the thermodynamic limit $d_E \rightarrow \infty$. We then study fluctuations around this limiting averaged k -OTOC for single realizations and finite sizes and establish concentration bounds. These bounds imply that single realizations are close to the averaged result with a large probability and give rise to a notion of typicality, resulting in the stronger notion of almost sure convergence in the thermodynamic limit.

1. Haar averages and Weingarten calculus

We start by reviewing the Weingarten calculus [98,99] as a tool to compute the averages over the independent Haar-random unitaries V_t . From the independence of the V_t , the average factorizes and it suffices to compute the average $\mathbb{E}[V_t^{(k)}]$ of the k -folded gate in the replica picture. This average does not depend on the time step t , and we drop the index in the following. Moreover, the average is the vectorization of the k -fold twirl channel of the unitary group, i.e., the orthonormal projection onto the space which is invariant under permutations of the k replicas. In the vectorized (replica) picture, this space is spanned by all the permutation states $|\sigma\rangle_{CE}$ for $\sigma \in S_k$ [Eq. (42)]. These states are linearly independent and form a basis for the

From a statistical mechanics perspective, the terms $d_E^{\Delta\ell(\sigma_1, \dots, \nu_t)}$ resemble the Boltzmann weights of a one-dimensional model for permutation degrees of freedom subject to fixed boundary conditions $\sigma_1 = \circ$ and $\nu_t = \square$. The model's Hamiltonian corresponds to $-\Delta\ell$ and is bounded from below by 0, with the inverse temperature given by $-\ln d_E$. In this model, there is an associated energy cost $k - |\sigma_i^{-1}\nu_j| \geq 0$ if subsequent permutations are not identical, which is necessarily the case for some intermediate permutations due to the different top and bottom boundary conditions. In the limit $d_E \rightarrow \infty$, only the ground-state configurations, i.e., those with $\Delta\ell = 0$, contribute to the k -OTOC, as all other configurations will be suppressed in powers of d_E .

To identify these configurations, we note that the exponent $\Delta\ell$ has a geometric interpretation in the Cayley graph of the symmetric group S_k as detailed in Appendix A. It can be seen as a difference in length $\Delta\ell = \ell_0 - \ell_1$ between the length of a geodesic from the identity $\circ$ to the cyclic shift $\square$ given by $\ell_0 = \ell(\circ, \square) = k - 1$ and the length ℓ_1 of the shortest path that connects $\circ$ with $\square$ while visiting all permutations $\nu_1, \sigma_1, \dots, \sigma_{t-1}, \nu_t$ in the prescribed order. The length of such a path is given by $\ell_1 = k(2t - 1) - \sum_i |\nu_i^{-1}\sigma_i| - \sum_i |\sigma_i^{-1}\nu_{i+1}|$. In Fig. 6, we show a subgraph of the Cayley graph for $k = 4$ and indicate examples for such paths of various lengths. At time t , each path contains $2t$ permutations but might stay at the same permutation for some intermediate steps. This effect does not change the length of the path, and we show only the different permutations along the path. For instance, the solid blue path has intermediate permutations $\circ, (12)(3)(4), (12)(34),$ and $\square$. It corresponds to obtaining $\square$ from $\circ$ by repeated multiplication from the left with the transpositions indicated by the edges along the path yielding $\square = (24)(34)(12)\circ = (24)(34)(12)$.

This geometric observation allows for identifying the leading contribution to the averaged k -OTOC as follows: Since $\Delta\ell$ is nonpositive, it can be at most 0, which is satisfied when the permutations $\nu_1, \sigma_1, \dots, \sigma_{t-1}, \nu_t$ all lie on the same geodesic from $\circ$ to $\square$ with a nondecreasing distance to $\circ$. Note, however, that subsequent permutations along the path are allowed to be equal. The blue solid path depicted in Fig. 6 is an example for such a geodesic of length $k - 1 = 3$. This constraint forces all the permutations to correspond to the noncrossing partitions of Appendix A by identifying cycles of the permutations as blocks of a partition as is the case for the shown blue solid path. With this correspondence, the function $\mu(\nu, \sigma)$ from Eq. (B5) can be identified as the Möbius function of the noncrossing lattice. In contrast, the other two paths shown in Fig. 6 leave the noncrossing lattice by containing either a crossing permutation $(13)(24)$ (red dashed path) or permutations with cycles failing to have elements in ascending order such as (132) or (1342) (green dotted path). The intermediate permutations not corresponding to

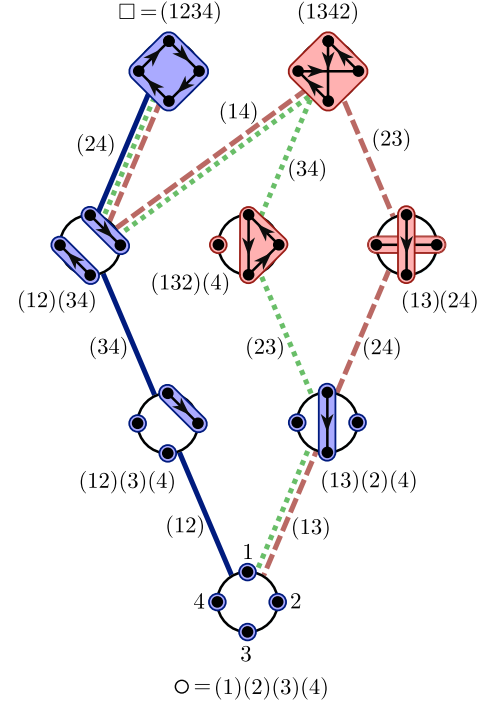


FIG. 6. Subgraph of the Cayley graph of the symmetric group for $k = 4$. We visualize the cycles of permutations by blocks and indicate the order within these cycles by arrows. For example, the cycle (132) is drawn as a block containing 1, 2, and 3 with arrows $1 \rightarrow 3 \rightarrow 2 \rightarrow 1$ indicating their order within the cycle. Permutations outlined in blue are associated with noncrossing partitions, while the others are not. We depict three different possible paths connecting $\circ$ to $\square$ distinguished by line style and color, neglecting repetitions at intermediate steps. The edges along the paths are labeled by the respective transpositions. For instance, the dashed green path corresponds to constructing $\square$ from $\circ$ via the factorization in transpositions given by $\square = (24)(14)(34)(23)(13)$.

noncrossing partitions can moreover not be connected to $\square$ along a path shorter than the ones shown.

To summarize, using the above description in terms of noncrossing partitions, the leading contributions to the k -OTOC are given in terms of nondecreasing paths on the noncrossing lattice starting at the smallest partition $\circ$ (the identity permutation) and ending at the largest partition $\square$ (the cyclic shift). Such a path is called a multichain and explicitly is given by the nondecreasing sequence

$$\Sigma = (\sigma_1 \subseteq \nu_1 \subseteq \sigma_2 \subseteq \nu_2 \dots \subseteq \sigma_t \subseteq \nu_t) \quad (\text{B9})$$

subject to the boundary conditions $\sigma_1 = \circ$ and $\nu_t = \square$. All contributions to the k -OTOC not coming from such multichains are subleading and are at least suppressed as d_E^{-2} . This suppression follows from the difference of the lengths of two paths on the Cayley graph with the same starting and end point being even. Combining these arguments, only the

multichains contribute in the thermodynamic limit $d_E \rightarrow \infty$ and we obtain

$$\mathbb{E}[c_{ab}^{(k)}(t; d_E)] = C_{ab}^{(k)}(t) + \mathcal{O}(d_E^{-2}), \quad (\text{B10})$$

where, as stated in the main text [Eq. (50)], and repeated here for convenience

$$C_{ab}^{(k)}(t) = d_A^{-1} d_C^{k-1} \sum_{\dots \subseteq \sigma_i \subseteq \nu_i \subseteq \sigma_{i+1} \subseteq \dots} \left(\prod_{i=1}^{t-1} W_{\nu_i \sigma_{i+1}} \right) \times (b_{\square} | \mathcal{M}_{\nu_1 \sigma_1}, \dots, \mathcal{M}_{\nu_{2\sigma_2}} \mathcal{M}_{\nu_1 \sigma_1} | a_{\circ}) \quad (\text{B11})$$

is the averaged k -OTOC in the thermodynamic limit. Using the abbreviations

$$\mathcal{M}_{\Sigma} = \mathcal{M}_{\nu_1 \sigma_1}, \dots, \mathcal{M}_{\nu_{2\sigma_2}} \mathcal{M}_{\nu_1 \sigma_1} \quad (\text{B12})$$

and

$$W_{\Sigma} = \prod_{i=1}^{t-1} W_{\nu_i \sigma_{i+1}}, \quad (\text{B13})$$

the above equation can be written in more compact form as

$$C_{ab}^{(k)}(t) = d_A^{-1} d_C^{k-1} \sum_{\Sigma} W_{\Sigma} (b_{\square} | \mathcal{M}_{\Sigma} | a_{\circ}). \quad (\text{B14})$$

Restricting the first and the last noncrossing partition to be $\circ$ and $\square$, respectively, results in a total number $\propto t^{k-1}$ multichains that need to be considered.

3. Concentration bounds and typicality

In the following, we discuss fluctuations of the finite-size k -OTOC $c_{ab}^{(k)}(t, d_E)$ around its average and around $C_{ab}^{(k)}(t)$. We show that fluctuations are strongly suppressed in d_E , and single realizations are close to the averaged result $C_{ab}^{(k)}(t)$ with high probability by deriving concentration bounds for the distribution of the k -OTOC. These bounds turn out to be strong enough to guarantee the convergence of the finite-size k -OTOCs to $C_{ab}^{(k)}(t)$ given by Eq. (50) not only in the averaged case, i.e., in distribution, but also for typical realizations, i.e., in the almost sure sense.

To establish these results, we start by considering the second moment of the finite-size k -OTOC given by

$$\mathbb{E}[c_{ab}^{(k)}(t; d_E)^2] = \frac{1}{D} \mathbb{E} \left(\text{Tr}[A_1(t) B_1 \dots A_k(t) B_k]^2 \right). \quad (\text{B15})$$

This average can again be computed by means of the general expression Eq. (B1) after replacing the number of replicas k by $2k$ and adjusting the boundary conditions accordingly. More precisely, the initial state $|a_{\circ})$ remains the same up to interpreting $\circ$ as the identity permutation on

$2k$ replicas and $a_i = a_{i+k}$, whereas the final state becomes $(b_{\square \otimes \square} | \otimes_{CE} (\square |^{\otimes 2}$ with $\square$ still being interpreted as the cyclic shift on k elements and $b_i = b_{i+k}$. The final state can be thought of as being associated with the permutation $\square \otimes \square = (1, 2, \dots, k)(k+1, k+2, \dots, 2k) \in S_{2k}$ of $2k$ elements. We then proceed as in the computation of the average. Replacing $V_t^{(2k)}$ by $\mathbb{E}[V^{(2k)}]$ we again obtain a sum over all $2t$ -tuples of permutations (on $2k$ replicas) with the first entry of the tuple being $\circ$, but the last entry now being $\square \otimes \square$. These tuples have a similar geometric interpretation as paths on the Cayley graph of the symmetric group connecting $\circ$ with $\square \otimes \square$ and visiting all intermediate permutations in the prescribed order. However, there are now two types of paths. The first type consists of paths which contain no edge in the graph corresponding to transpositions (i, j) with $i \leq k$ and $j > k$ or vice versa. These paths can be thought of as “products” of paths of permutations on the first k replicas with paths of permutations on the last k replicas. As these paths of the first type do not couple the first and the last k replicas, summing over all such paths contributes to the second moment with $C_{ab}^{(k)}(d; d_E)^2$, i.e., the average squared. Moreover, they contain the “products” of the geodesics which determine the thermodynamic limit. The second type of path from $\circ$ to $\square \otimes \square$ consists of paths which contain an edge corresponding to a transposition (i, j) of the above form. Multiplication by such a transposition creates a cycle with elements from both the first and the last k replicas for one of the intermediate permutations along the path. The final permutation $\square \otimes \square$, however, does not contain such a cycle, and hence there needs to be at least another edge along the path corresponding to the transposition (i', j') of the above form to split up the problematic cycle. Consequently, each such path is at least two steps longer than the geodesics of the product form, and hence the contribution from summing all paths of the second type gives rise to a $\mathcal{O}(d_E^{-2})$ contribution to the second moment only.

We obtain for the variance

$$\mathbb{E}[c_{ab}^{(k)}(t; d_E)^2] - \mathbb{E}[c_{ab}^{(k)}(t; d_E)]^2 = \mathcal{O}(d_E^{-2}), \quad (\text{B16})$$

and a similar estimate for the thermodynamic limit follows by using Eq. (B10) as

$$\mathbb{E}[c_{ab}^{(k)}(t; d_E)^2] - C_{ab}^{(k)}(t)^2 = \mathcal{O}(d_E^{-2}). \quad (\text{B17})$$

This estimate of fluctuations around the average allows for deriving concentration bounds. Explicitly, Eq. (B17) implies that there are constants $K_{t,k}$ depending on t and k , but independent from d_E , such that the right-hand side is upper bounded by $K_{t,k} d_E^{-2}$. By fixing arbitrary t_0 and taking the maximum of $K_{t,k}$ we get a uniform bound for all $t \leq t_0$

of the form $K_{t_0,k}/d_E^{-2}$ for an appropriate constant $K_{t_0,k}$. Using Markov's inequality, we obtain for every $\epsilon > 0$ and for all $t \leq t_0$,

$$\mathbb{P}\left(\left|c_{ab}^{(k)}(t, d_E) - C_{ab}^{(k)}(t)\right| > \epsilon\right) \leq \frac{K_{t_0,k}}{\epsilon^2 d_E^2}, \quad (\text{B18})$$

with the probability measure being the product Haar measure on t_0 independent copies of the unitary group. For instance, by choosing $\epsilon = d_E^{-1/2}$, the above bound implies that the probability to find deviations of the finite-size k -OTOC from the thermodynamic result for a single realization, i.e., $|c_{ab}^{(k)}(t, d_E) - C_{ab}^{(k)}(t)|$, larger than $d_E^{-1/2}$ is on the order $\mathcal{O}(d_E^{-1})$ and hence vanishes in the thermodynamic limit. In other words, typical realizations will be close to $C_{ab}^{(k)}(t)$ with high probability for all times $t \leq t_0$ for arbitrary chosen t_0 and d_E large enough.

The above concentration bounds can also be used to promote the convergence of the averaged k -OTOC in the thermodynamic limit to almost sure convergence. To this end, for fixed time $t \leq t_0$ we consider the sequence of k -OTOCs $C_{ab}^{(k)}(t; d_E)$ indexed by the dimension d_E of the environment as random variables on the probability space given as the infinite product of t_0 copies of the unitary group of dimension d_E equipped with the corresponding product of Haar measures. These random variables are independent for different d_E , and the probabilities in Eq. (B18) are summable in d_E . By the second Borell-Cantelli lemma for almost all d_E , one has $|c_{ab}^{(k)}(t, d_E) - C_{ab}^{(k)}(t)| \leq \epsilon$, and hence $c_{ab}^{(k)}(t, d_E)$ converges almost surely to $C_{ab}^{(k)}(t)$ as $d_E \rightarrow \infty$ [71].

Intuitively, the above concentration bounds and the stronger form of convergence can be interpreted as the dynamics of k -OTOCs for typical realizations of the minimal circuit model closely following $C_{ab}^{(k)}(t)$ as given in Eq. (50) even without averaging.

APPENDIX C: PROPERTIES OF THE QUANTUM CHANNEL

In this appendix, we briefly review some properties of the single-particle quantum channel $\mathcal{M}$ and derive a bound on the operator entanglement that guarantees mixing dynamics. The action of $\mathcal{M}$ is defined as

$$\mathcal{M}(a) = \frac{1}{d_C} \text{Tr}_C[U(a \otimes \mathbb{1}_C)U^\dagger], \quad (\text{C1})$$

where a is an operator acting on A , and U is a unitary matrix acting on A and C . This expression can be graphically represented in the folded representation as

$$\mathcal{M} = \text{[Diagram: A green box labeled '1' with two input wires on the left and two output wires on the right, representing the identity operator on a single replica.]}, \quad (\text{C2})$$

with the folded operator acting on a single replica and hence $\circ = (1)$.

From unitarity, it follows that $\mathcal{M}$ is trace preserving, i.e., $\text{Tr}[\mathcal{M}(a)] = \text{Tr}(a)$. The complete positivity follows from interpreting Eq. (C1) as a Stinespring dilation, such that $\mathcal{M}$ realizes a unital quantum channel. This operator is necessarily contracting, such that all eigenvalues lie either on or inside the unit circle. By taking the complex conjugate of Eq. (C1), it follows that if $\mathcal{M}(a_\lambda) = \lambda a_\lambda$ then $\mathcal{M}(a_\lambda^\dagger) = \lambda^* a_\lambda^\dagger$. The eigenvalues of $\mathcal{M}$ are either real, with Hermitian eigenoperators (up to an overall phase), or appear in complex conjugate pairs, with eigenoperators related through Hermitian conjugation.

From unitality, the quantum channel is guaranteed to have a trivial leading eigenvalue 1 with the identity matrix as eigenoperator $\mathcal{M}(\mathbb{1}_A) = \mathbb{1}_A$. As mentioned in the main text, it is possible to obtain a bound on the largest nontrivial eigenvalue λ from the operator entanglement. The proof of this bound is analogous to the proof from Ref. [106], which presented a bound on the entangling power for dual-unitary gates that guarantees the corresponding quantum channels for light-cone dynamics to be mixing (following Ref. [104]). A similar proof appeared in Ref. [108] for unitary gates with equal Hilbert-space dimensions, i.e., $d_A = d_C$.

The operator Schmidt decomposition of U is given by

$$U = \sum_{j=1}^{d^2} \sqrt{\gamma_j} X_j \otimes Y_j, \quad (\text{C3})$$

with $d = \min(d_A, d_C)$. Here, X_j and Y_j are single-site operators acting on A and C , respectively, orthonormalized such that $\text{Tr}(X_j^\dagger X_k) = \text{Tr}(Y_j^\dagger Y_k) = \delta_{jk}$. The linear entropy of the gate U , or operator entropy in short, is defined as [105–108]

$$E(U) = 1 - \frac{1}{(d_A d_C)^2} \sum_{j=1}^{d^2} \gamma_j^2 = 1 - \frac{1}{d_A^2} \text{[Diagram: A green box labeled '2' with two input wires on the left and two output wires on the right, representing the folded gate acting on two replicas.]}, \quad (\text{C4})$$

here expressed in terms of the folded gate acting on two replicas such that $\circ = (1)(2)$ and $\square = (12)$. The operator entropy is bounded as $0 \leq E(U) \leq 1 - 1/d^2$.

In order to relate the operator entropy to the eigenvalues of $\mathcal{M}$, we first define $\tilde{\mathcal{M}} = \mathcal{M} - |\circ\rangle\langle\circ|$, where $|\circ\rangle$ is the normalized vectorization of the identity, i.e., $(i, j|\circ) = \delta_{i,j}/\sqrt{d_A}$. This operator satisfies $\tilde{\mathcal{M}}|\circ\rangle = 0$ and otherwise

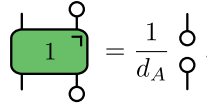
has the same nontrivial eigenvalues and eigenoperators as $\mathcal{M}$. We can rewrite Eq. (C4) as

$$E(U) = 1 - \frac{1}{d_A^2} \|\mathcal{M}\|_2^2 = 1 - \frac{1}{d_A^2} \left(1 + \|\tilde{\mathcal{M}}\|_2^2\right), \quad (\text{C5})$$

where in the first expression we have introduced the Frobenius norm of $\mathcal{M}$, which is in turn related to the Frobenius norm of $\tilde{\mathcal{M}}$. The latter norm bounds the largest nontrivial eigenvalue λ , such that we find

$$|\lambda| \leq \|\tilde{\mathcal{M}}\|_2 = \sqrt{d_A^2 [1 - E(U)] - 1}. \quad (\text{C6})$$

This bound becomes nontrivial when $E(U) > E^* = 1 - 2/d_A^2$, fixing $|\lambda| < 1$ and hence indicating ergodic and mixing dynamics. Note that this bound can be satisfied only when $d_A \leq d_C$, since otherwise E^* exceeds the maximally allowed value $1 - 1/d_C^2$. For $d_A = d_C$, a maximal operator entanglement implies that U is dual unitary [105]. Dual unitarity [104,109] then directly fixes



$$= \frac{1}{d_A} \text{ (diagram of two vertical lines connected by a horizontal line, with a small circle on each of the four wires)}, \quad (\text{C7})$$

in which case, $\mathcal{M} = |\circ\rangle\langle\circ|$ is a projector and $|\lambda| = 0$.

APPENDIX D: EIGENSTATE CONSTRUCTION

In this appendix, we construct the eigenstates of the Markovian matrix $\mathcal{T}$ [Eq. (116)]. We first derive the leading left and right eigenstates before performing a biorthogonalization that can be used to construct the projector on this leading eigenspace. We then discuss the construction of subleading eigenstates by dressing the leading eigenstates and show how they can be biorthogonalized.

1. Leading eigenstates

a. Left eigenstates

From the triangular structure of $\mathcal{T}$, it follows that the left eigenstates can be directly obtained from the left eigenstates of the diagonal elements $\mathcal{M}_{\rho\rho}$. These operators satisfy $(\rho|\mathcal{M}_{\rho\rho} = (\rho|$ with $(\rho|$ the permutation states defined in Eq. (42). We can use these states to construct left eigenstates of $\mathcal{T}$ denoted as $((L_\rho|$, which we index by a noncrossing partition ρ and write with a double bracket to highlight that these live in the joint Hilbert space of replicas and the auxiliary space of noncrossing partitions. These eigenstates have a single nonzero component in the auxiliary space and are given by

$$((L_\rho| = (\rho|_\rho. \quad (\text{D1})$$

Note that here ρ both labels the permutation state $(\rho|$ and indicates the auxiliary space. From the triangular structure, we need to check only the diagonal component, resulting in

$$(\rho|\mathcal{T}_{\rho\rho} = (\rho|\mathcal{M}_{\rho\rho} = (\rho|. \quad (\text{D2})$$

b. Right eigenstates

The right eigenstates can also be constructed, although they have a more involved structure. We can again use that the diagonal elements satisfy $\mathcal{M}_{\nu\nu}|\nu\rangle = |\nu\rangle$. Moreover, it can be directly checked that more generally

$$\mathcal{M}_{\nu\sigma}|\sigma\rangle = d_C^{-k+|\nu^{-1}\sigma|}|\sigma\rangle. \quad (\text{D3})$$

This identity can be used to construct right eigenstates $|R_\sigma\rangle\rangle$ of $\mathcal{T}$, which we again index by a noncrossing partition σ and write with a double bracket as

$$|R_\sigma\rangle\rangle = \sum_{\sigma \subseteq \rho} d_C^{|\rho|}|\sigma\rangle. \quad (\text{D4})$$

Note that on the right-hand side, $|\sigma\rangle$ indicates the permutation state, whereas the subscript ρ is the index in the auxiliary space. The summation runs over all noncrossing partitions ρ that lie above σ in the noncrossing partition lattice.

It is a direct check that the states (D4) are right eigenstates of $\mathcal{T}$. We can calculate the component ν of $\mathcal{T}|R_\sigma\rangle\rangle$ in the auxiliary space, where we separate the cases $\sigma \not\subseteq \nu$ and $\sigma \subseteq \nu$. For the former case, it follows that

$$\sum_{\sigma \subseteq \rho} \mathcal{T}_{\nu\rho} d_C^{|\rho|}|\sigma\rangle = 0 \quad \text{if } \sigma \not\subseteq \nu, \quad (\text{D5})$$

since the summation runs over all components ρ of $|R_\sigma\rangle\rangle$, which are only nonzero if $\sigma \subseteq \rho$, but $\mathcal{T}_{\nu\rho}$ is only nonzero if $\rho \subseteq \nu$, and fixing both to be nonzero implies $\sigma \subseteq \rho \subseteq \nu$ and hence $\sigma \subseteq \nu$. For the latter case, we have

$$\sum_{\sigma \subseteq \rho} \mathcal{T}_{\nu\rho} d_C^{|\rho|}|\sigma\rangle = \sum_{\sigma \subseteq \rho \subseteq \pi \subseteq \nu} \mu(\rho, \pi) d_C^{-k+|\rho^{-1}\pi|+|\rho|} \mathcal{M}_{\nu\pi}|\sigma\rangle. \quad (\text{D6})$$

Here we can first use that for a geodesic $\rho \subseteq \pi$ the triangle inequality becomes $|\pi^{-1}\rho| + |\rho^{-1}\pi| = k + |\pi^{-1}\pi| = k + |\pi|$, to write

$$\begin{aligned} \sum_{\sigma \subseteq \rho \subseteq \pi \subseteq \nu} \mu(\rho, \pi) d_C^{|\pi|} \mathcal{M}_{\nu\pi}|\sigma\rangle &= \sum_{\sigma \subseteq \pi \subseteq \nu} \delta_{\sigma, \pi} d_C^{|\pi|} \mathcal{M}_{\nu\pi}|\sigma\rangle \\ &= d_C^{|\sigma|} \mathcal{M}_{\nu\sigma}|\sigma\rangle = d_C^{-k+|\sigma|+|\sigma^{-1}\nu|}|\sigma\rangle \\ &= d_C^{|\nu|}|\sigma\rangle, \end{aligned} \quad (\text{D7})$$

where in the first equality, we used that Möbius functions satisfy $\sum_{\sigma \subseteq \rho \subseteq \pi} \mu(\rho, \pi) = \delta_{\sigma, \pi}$ for fixed σ and π and a summation over ρ ; see Eq. (A16). In the second to last equality, we applied Eq. (D3), and in the final equality, we again used that the triangle equality is saturated. Taking these results together, we recover that the states from Eq. (D4) are right eigenstates of the Markovian matrix T with eigenvalue 1.

c. Biorthogonalization

The previous constructions return C_k left and right eigenstates with eigenvalue 1, exhausting the leading eigenspace in the case of ergodic dynamics. However, these eigenstates are not biorthogonal. This biorthogonalization can be performed using the properties of the Möbius functions. For reasons that are discussed in the main text, we write the resulting left eigenstates as $((\kappa_\sigma|$ and the right eigenstates as $|\varphi_\sigma\rangle$,

$$|\varphi_\sigma\rangle = d_A^{|\sigma|-k} |R_\sigma\rangle = d_A^{|\sigma|-k} \sum_{\sigma \subseteq \rho} d_C^{|\rho|} |\sigma\rangle_\rho, \quad (\text{D8})$$

$$\begin{aligned} ((\kappa_\nu| &= \sum_{\rho \subseteq \nu} (d_A d_C)^{-|\rho|} \mu(\nu, \rho) ((L_\rho| \\ &= \sum_{\rho \subseteq \nu} (d_A d_C)^{-|\rho|} \mu(\nu, \rho) (\rho|_\rho. \end{aligned} \quad (\text{D9})$$

Using similar arguments as in our construction of the right eigenstates, it can be directly checked that these states satisfy

$$((\kappa_\nu| \varphi_\sigma)) = \delta_{\nu\sigma}. \quad (\text{D10})$$

To show this biorthogonality, we first note that

$$((L_\rho| R_\sigma)) = \begin{cases} d_A^{|\sigma|-1} d_C^{|\rho|} & \text{if } \sigma \subseteq \rho, \\ 0 & \text{if } \sigma \not\subseteq \rho. \end{cases} \quad (\text{D11})$$

We directly observe that $((\kappa_\nu| \varphi_\sigma)) = 0$ if $\sigma \not\subseteq \nu$. If $\sigma \subseteq \nu$, the Möbius functions enforce orthogonality since

$$\begin{aligned} ((\kappa_\nu| \varphi_\sigma)) &= d_A^{|\sigma|-k} \sum_{\sigma \subseteq \rho \subseteq \nu} (d_A d_C)^{-|\rho|} \mu(\nu, \rho) ((L_\rho| R_\sigma)) \\ &= \sum_{\sigma \subseteq \rho \subseteq \nu} d_A^{|\sigma|+|\sigma|-1-|\rho|-k} \mu(\nu, \rho) \\ &= \sum_{\sigma \subseteq \rho \subseteq \nu} \mu(\rho, \nu) = \delta_{\sigma\nu} \end{aligned} \quad (\text{D12})$$

using the same arguments as in Eq. (D7).

d. Overlaps

We now consider the overlap of these eigenstates with the boundary states $|\psi_a\rangle$ and $((\psi_b|$ [Eq. (119)], whose

definition we repeat here for convenience,

$$((\psi_b| = (d_A d_C)^{-1} (b_\square|_\square, \quad (\text{D13})$$

$$|\psi_a\rangle = \sum_{\rho \in \text{NC}(k)} d_C^{|\rho|} |a_\circ\rangle_\rho. \quad (\text{D14})$$

The relevant overlap of the right eigenstates can be directly calculated as

$$((\psi_b| \varphi_\sigma)) = d_A^{|\sigma|-k-1} (b_\square|_\square) = \varphi_{\sigma^*}(b_1, \dots, b_k), \quad (\text{D15})$$

where we used that $(b_\square|_\sigma) = d_A^{-|\sigma^*|} \varphi_{\sigma^*}(b_1, \dots, b_k)$ and $|\sigma^*| = k + 1 - |\sigma|$. For the left eigenstates, we similarly find

$$((\kappa_\sigma| \varphi_a)) = \sum_{\nu \subseteq \sigma} \mu(\sigma, \nu) d_A^{-|\nu|} (\nu| a_\circ) \quad (\text{D16})$$

$$= \sum_{\nu \subseteq \sigma} \mu(\sigma, \nu) \varphi_\nu(a_1, \dots, a_k) \quad (\text{D17})$$

$$= \kappa_\sigma(a_1, \dots, a_k), \quad (\text{D18})$$

now using that $\varphi_\nu(a_1, \dots, a_k) = d_A^{-|\nu|} (\nu| a_\circ)$.

2. Subleading eigenstates

We now consider eigenstates with eigenvalue λ obtained by dressing these states with a right (left) eigenoperator $a = a_\lambda$ ($b = b_\lambda$) of the single-particle channel as

$$|\varphi_\sigma^m\rangle = a_{\lambda,m} |\varphi_\sigma\rangle, \quad ((\kappa_\nu^n| = ((\kappa_\nu| b_{\lambda,n}, \quad (\text{D19})$$

where $m, n = 1, \dots, k$ denote the contraction to be dressed, and a_λ (b_λ) acts on all permutations in the expansions of these states as by dressing the m th contraction with this operator; see Eq. (42) and the surrounding discussion. In this appendix, we will drop the subscript λ for ease of notation. For the left eigenstates, the proof that these are eigenstates is fully analogous to the proof for the leading eigenstates. For the right eigenstates, we need only the additional identity

$$\mathcal{M}_{\nu\sigma} a_m |\sigma\rangle = \lambda d_C^{-k+|\nu|-1} |\sigma\rangle a_m, \quad (\text{D20})$$

which can be directly verified.

a. Biorthogonalization

The overlaps can be calculated in a similar way as before, to return

$$((\kappa_\nu^n| \varphi_\sigma^m)) = \sum_{\sigma \subseteq \rho \subseteq \nu} d_A^{-|\sigma|-1} \mu(\nu, \rho) (\rho| b_n a_m |\sigma\rangle. \quad (\text{D21})$$

The contraction $(\rho|b_n a_m|\sigma)$ vanishes unless m and n are in the same cycle of $\sigma^{-1}\rho$, and hence of $\sigma^{-1}\nu$, since otherwise the contraction will contain factors $\text{Tr}(a)\text{Tr}(b) = 0$, whereas if m and n are in the same cycle, the contraction contains a factor $\text{Tr}(ab) = 1$.

Furthermore, if m and n are in the same cycle of $\sigma^{-1}\nu$, there exists a unique permutation $\tilde{\nu}(m, n)$ with $\sigma \subseteq \tilde{\nu}(m, n) \subseteq \nu$, for which there are no permutations below $\tilde{\nu}(m, n)$ in which m and n are in the same cycle of $\sigma^{-1}\tilde{\nu}(m, n)$. For all permutations $\tilde{\nu}(m, n) \subseteq \rho \subset \nu$, m and n are then necessarily in the same cycle. The overlap is fully determined by this permutation $\tilde{\nu}(m, n)$, since Eq. (D21) reduces to

$$((\kappa_\nu^n|\varphi_\sigma^m)) = \sum_{\tilde{\nu}(m, n) \subseteq \rho \subseteq \nu} \mu(\nu, \rho) = \delta_{\nu, \tilde{\nu}(m, n)}. \quad (\text{D22})$$

In other words, these states have unit overlap if and only if m and n are in the same cycle in $\sigma^{-1}\nu$ and in different cycles in all $\sigma^{-1}\rho$ with $\sigma \subseteq \rho \subset \nu$. In all other cases, these states are orthogonal.

For $m = n$, this overlap directly returns $\delta_{\nu\sigma}$. For $m \neq n$ but in the same cycle of ν , there is a single additional permutation for which this overlap is nonvanishing. This permutation is uniquely determined by ν and $m \neq n$, motivating our notation $\tilde{\nu}(m, n)$. This property can be proven by explicitly constructing a permutation that satisfies the above properties. For ease of notation, we drop the explicit dependence on m and n and write $\tilde{\nu}(m, n) = \tilde{\nu}$.

First, we note that m and n need to be part of the same cycle of ν , otherwise for all $\tilde{\nu} \subseteq \nu$, we have that m and n are part of different cycles in $\tilde{\nu}^{-1}\nu$. Second, we note that all cycles in $\tilde{\nu}$ that do not contain m or n should be identical to cycles in ν that also do not contain m or n . If this were not the case, we could merge two such cycles to create a new permutation ρ that satisfies $\tilde{\nu} \subset \rho \subset \nu$ and for which $\rho^{-1}\nu$ has m and n in the same cycle (since this cycle would not be impacted by merging the other cycles), violating our assumption there are no permutations ρ above $\tilde{\nu}$ for which m and n are in the same cycle of $\rho^{-1}\nu$. We hence find that $\tilde{\nu}$ is identical to ν , except that the single cycle of ν that contains m and n is split into two cycles, one of which contains m and the other of which contains n . We take this cycle to have ℓ elements and denote it as $\nu_c = (i_1 i_2 \dots i_m \dots i_n \dots i_\ell)$. This cycle can be identified with the cyclic permutation on ℓ elements. We hence want to find a way of splitting this permutation in two as $\tilde{\nu}_c = (\dots i_m \dots)(\dots i_n \dots)$, in such a way that i_m and i_n are in the same cycle in $\tilde{\nu}_c^{-1}\nu_c$ but in different cycles of $\rho_c^{-1}\nu_c$ for every $\tilde{\nu}_c \subset \rho_c \subseteq \nu_c$. We note that multiplication with the inverse cyclic permutation is equivalent to constructing the Kreweras complement, such that we hence require that i_m and i_n are in the same cycle of the Kreweras complement $\tilde{\nu}_c^*$ of $\tilde{\nu}_c$, but in different cycles for the Kreweras complement of ρ_c^* for $\tilde{\nu}_c \subset \rho_c$. The Kreweras

complement reverses the ordering, such that we need to have i_m and i_n in different cycles for every $\rho_c^* \subset \tilde{\nu}_c^*$. This demand uniquely fixes $\tilde{\nu}_c^*$ as the permutation in which i_m and i_n form a transposition (a cycle of length two), and all other indices form singletons (cycles of length one). This construction then uniquely determines $\tilde{\nu}$ given m , n , and ν .

We conclude this proof with some examples. For $\ell = 2$, we can write $\nu_c = (i_1 i_2)$ with, e.g., $i_m = i_1$ and $i_n = i_2$. We have that $\tilde{\nu}_c^* = (i_1 i_2)$, and the Kreweras complement of (1) (2) is (12), such that $\tilde{\nu}_c = (i_1)(i_2)$. For $\ell = 3$, we can have $\nu_c = (i_1 i_2 i_3)$, and fixing $i_m = i_1$ and $i_n = i_2$, we have that $\tilde{\nu}_c^* = (i_1 i_2)(i_3)$. Using that the Kreweras complement of (13)(2) is (12)(3), we have that $\tilde{\nu}_c = (i_1 i_3)(i_2)$. For $\ell = 4$ and $\nu_c = (i_1 i_2 i_3 i_4)$, we can have, e.g., $i_m = i_1$ and $i_n = i_2$, for which $\tilde{\nu}_c = (i_1 i_3 i_4)(i_2)$, or $i_m = i_1$ and $i_n = i_3$, for which $\tilde{\nu}_c = (i_1 i_4)(i_2 i_3)$.

b. Singletons

To illustrate the construction of biorthogonalized eigenstates, we first choose ν such that n is a cycle of length one (a singleton). In this case, the condition for the overlap to be nonvanishing simplifies. Since $\sigma \subseteq \rho$, every singleton in ρ needs to be a singleton in σ . As such, m and n can be only in the same cycle if $m = n$. This additionally implies that the lowest noncrossing partition in which m and n are in the same cycle is σ , i.e., $\tilde{\nu}(m, n) = \sigma$, such that the overlap reduces to

$$((\kappa_\nu^n|\varphi_\sigma^m)) = \delta_{mn}\delta_{\nu\sigma}. \quad (\text{D23})$$

We recover the expected biorthogonality. The free cumulant states in which n is a singleton within ν are already properly biorthogonalized with respect to the moment states.

c. Transpositions

We now consider ν such that n is part of a cycle of length two (a transposition). Denoting $\nu = (\dots)(n\tilde{n})(\dots)$ in which n forms a cycle of length two with $\tilde{n}$, the overlap (D22) is nonvanishing only if either $m = n$ or $m = \tilde{n}$. For $m = n$, we can directly use the previous result to find orthogonality, since $((\kappa_\nu^n|\varphi_\sigma^n)) = \delta_{\nu\sigma}$ since a and b are always guaranteed to be in the same cycle.

For $m = \tilde{n}$, we however have additionally nonvanishing overlaps. Since $\sigma \subseteq \nu$, there are now only two options: Either n and m form a transposition within $\sigma = (\dots)(nm)(\dots)$, or they are both singletons and $\sigma = (\dots)(n)(m)(\dots)$. The former option again leads to the expected orthogonality, since we can directly repeat the previous argument. In the latter case, the overlap is nonvanishing if and only if $\sigma = \tilde{\nu}(m, n)$, which here corresponds to ν with the transposition split into two singletons.

If n forms a transposition with m in ν , then the two nonvanishing overlaps are

$$((\kappa_\nu^n | \varphi_\nu^n)) = ((\kappa_\nu^n | \varphi_{\bar{\nu}(m,n)}^{\bar{n}})) = 1. \quad (\text{D24})$$

Here the recursive construction comes into play. Since the (left) cumulant states in which a singleton is dressed already satisfy the orthogonality, and the additional state with nonvanishing overlap is exactly a (right) state in which a singleton is dressed, we can define orthogonal left states for transpositions by simply subtracting the left singleton state from this transposition state. For partitions ν in which n and m form a transposition, biorthogonalized left eigenstates are given by

$$((\tilde{\kappa}_\nu^n | = ((\kappa_\nu^n | - ((\kappa_{\bar{\nu}(m,n)}^{\bar{n}} |. \quad (\text{D25})$$

These states satisfy

$$((\tilde{\kappa}_\nu^n | \varphi_\sigma^m)) = \delta_{mn} \delta_{\nu\sigma}. \quad (\text{D26})$$

d. General cycles

The previous construction can be directly extended to the case where n is part of a larger cycle of ν . Properly biorthogonalized right eigenstates can be defined as

$$((\tilde{\kappa}_\nu^n | = ((\kappa_\rho^n | - \sum_{m \neq n} ((\tilde{\kappa}_{\bar{\nu}(m,n)}^{\bar{n}} |, \quad (\text{D27})$$

where the summation runs over all indices m that are in the same cycle of ν as n . The corresponding permutation $\bar{\nu}(m, n)$ is defined above. Since n will always belong to a shorter cycle in this permutation, the construction can be done recursively, where if n is a singleton in ν we define $((\tilde{\kappa}_\nu^n | = ((\kappa_\nu^n |$. These states now define an orthonormal basis

$$((\tilde{\kappa}_\nu^n | \varphi_\sigma^m)) = \delta_{mn} \delta_{\nu\sigma}. \quad (\text{D28})$$

To conclude, we note that the above construction directly extends to states in which multiple contractions are dressed with (possibly different) eigenoperators of $\mathcal{M}$. The only restriction is that no two contractions within the same cycle can be dressed, since otherwise the equivalent of Eq. (D20) no longer holds. This constraint prevents construction of the full set of eigenstates that would be allowed from the eigenstates of the channels $\mathcal{M}_{\nu\sigma}$ and ultimately gives rise to the observed Jordan structure.

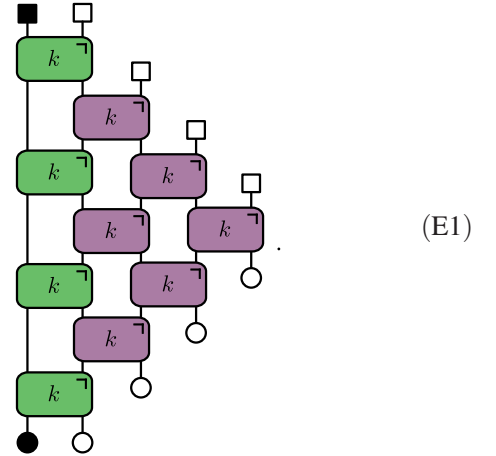
APPENDIX E: ADDITIONAL SPATIAL STRUCTURE

In this appendix, we briefly discuss a possible extension of the minimal quantum circuit that includes additional spatial structure beyond the simple three-site setup. While

such generalizations seem natural, we argue that these are already captured by the minimal circuit, indicating the wider applicability of the presented results.

We first note that, since we did not make any assumptions on the Hilbert space for A , e.g., we might take A as a finite one-dimensional lattice and U as a local quantum circuit which couples only the boundary of the lattice A to the bottleneck C . While this choice would influence the short-time dynamics, the late-time behavior does not depend on the specific structure of U but only on the spectral properties of the associated channel $\mathcal{M}$. As for generic choices of local gates that comprise the quantum circuit U , the resulting circuit will still be ergodic and mixing ($\lambda < 1$), all presented results remain valid but might apply at slightly later times. Intuitively, this shift in time can be understood from additional constraints on the dynamics from causality and might be interpreted as the time it takes for local observables in A to spread into the effective bath. Nevertheless, the minimal, three-site circuit accurately captured the dynamics at late times irrespective of the details within A and the concrete structure of U .

Another natural extension of the minimal setup is to allow for additional spatial structure in the effective bath while still preserving solvability of the model. This can be achieved by replacing the unitaries V_i by a random quantum circuit with a brickwork geometry. A natural choice consistent with causality is by choosing the environment dynamics as



We again consider Haar-random local gates, allowing for an exact solution in the limit of the local Hilbert-space dimensions going to infinity. Intuitively, the more involved environment should not change the dynamics as, in the minimal setting, a single Haar-random unitary V_t already perfectly scrambles any information leaking into the effective bath as $d_E \rightarrow \infty$.

This intuition can be made precise. To this end, consider E as a one-dimensional lattice of length L with all local Hilbert spaces of the same dimension d_E and replace V_t by a random two-layer brickwork circuit $V_t = V_t^{(e)} V_t^{(o)}$ with

$$V_t^{(e)} = W_t^{(C,1)} \otimes W_t^{(2,3)} \otimes \dots \otimes W_t^{(L-1,L)}, \quad (\text{E2})$$

$$V_t^{(o)} = \mathbb{1}_C \otimes W_t^{(1,2)} \otimes \dots \otimes W_t^{(L-2,L-1)} \otimes \mathbb{1}_L, \quad (\text{E3})$$

where we have chosen L to be odd and $\mathbb{1}_L$ denotes the identity on site L . This dynamics gives rise to a causal light cone for the OTOC, as in Eq. (E1). The $W_t^{(C,1)}$ are $d_C d_E$ -dimensional Haar-random unitaries acting on the bottleneck C and the first site of the environment, whereas the $W_t^{(x,x+1)}$ are d_E^2 -dimensional Haar-random unitaries acting on the neighboring sites x and $x+1$. All random unitaries are assumed to be independent.

In this setting, the ensemble-averaged k -OTOCs in the limit $d_E \rightarrow \infty$ can be computed by the same methods as in the minimal setup, and we report only the result here. In particular, the k -OTOC can again be written as a sum over L -tuples of multichains on the noncrossing lattice

$$C_{ab}^{(k)}(t) = d_A^{-1} d_C^{-1} \sum_{\Sigma_1, \dots, \Sigma_L} W_{\Sigma} \left(\prod_{x=2}^L \mu_{\Sigma_x} \right), (b_{\square} | \mathcal{M}_{\Sigma} | a_{\circ}) \quad (\text{E4})$$

similar to Eq. (50). Because of the two-layer brickwork layout and not using the causality indicated by Eq. (E1), the multichains Σ_x for $x \neq L$ now are of length $4t$ with the appropriate boundary conditions $\circ$ and $\square$. We denote them by $\Sigma_x = (\dots \subseteq \sigma_i^{(x-1,x)} \subseteq \nu_i^{(x-1,x)} \subseteq \sigma_i^{(x,x+1)} \subseteq \nu_i^{(x,x+1)} \subseteq \dots)$, keeping track from which averaged Haar-random gate $W_i^{(x,x+1)}$ the noncrossing partitions originate. This enforces constraints as due to the brickwork structure most partitions are part of multiple multichains. Because of the boundary conditions, the multichain $\Sigma_L = (\dots \subseteq \sigma_i^{(L-1,L)} \subseteq \nu_i^{(L-1,L)} \subseteq \sigma_{i+1}^{(L-1,L)} \dots)$ is only of length $2t$. The coupling to C is described by $\Sigma = (\dots \subseteq \sigma_i^{(C,1)} \subseteq \nu_i^{(C,1)} \subseteq \sigma_{i+1}^{(C,1)} \dots)$, which has no independent partitions as it is derived from Σ_1 and hence is not summed over. All other multichains enter exclusively via their weights related to the Möbius function

$$\mu_{\Sigma_x} = \prod_{i=1}^{t-1} \mu(\nu_i^{(x-1,x)}, \sigma_i^{(x-1,x)}). \quad (\text{E5})$$

We now utilize the defining properties of the Möbius function [Eq. (A16)] to simplify the above expression. To this end we fix $\Sigma_1, \dots, \Sigma_{L-1}$ and first perform the sum over Σ_L subject to the constraints imposed by Σ_{L-1} . That is, we sum only over the $\nu_i^{(L-1,L)} \equiv \nu$ and $\sigma_i^{(L-1,L)} \equiv \sigma$ subject to the constraints $\sigma_i^{(L-2,L)} \subseteq \nu_i^{(L-1,L)} \subseteq \sigma_i^{(L-1,L)} \subseteq \nu_{i+1}^{(L-2,L)}$. The corresponding sum yields

$$\sum_{\nu \subseteq \sigma} \mu(\nu, \sigma) = \sum_{\sigma} \delta_{\sigma_i^{(L-2,L)}, \sigma} = 1, \quad (\text{E6})$$

and, applying this result to all time steps, we find that performing the sum over Σ_L in Eq. (E4) simply returns a factor of 1. In other words, the k -OTOCs remain unchanged under replacing L by $L-1$ and, by iterating this argument, is hence completely independent of L . The only remaining sum runs over those noncrossing partitions not fixed by the above argument which are exactly those appearing in Σ . The k -OTOC for the extended effective bath comprising a random brickwork circuit is identical to the one obtained in the minimal setup.

-
- [1] M. Srednicki, *The approach to thermal equilibrium in quantized chaotic systems*, *J. Phys. A* **32**, 1163 (1999).
 - [2] J. M. Deutsch, *Eigenstate thermalization hypothesis*, *Rep. Prog. Phys.* **81**, 082001 (2018).
 - [3] 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).
 - [4] M. C. Bañuls, M. B. Hastings, F. Verstraete, and J. I. Cirac, *Matrix product states for dynamical simulation of infinite chains*, *Phys. Rev. Lett.* **102**, 240603 (2009).
 - [5] A. Lerose, M. Sonner, and D. A. Abanin, *Influence matrix approach to many-body Floquet dynamics*, *Phys. Rev. X* **11**, 021040 (2021).
 - [6] M. Sonner, A. Lerose, and D. A. Abanin, *Influence functional of many-body systems: Temporal entanglement and matrix-product state representation*, *Ann. Phys. (Amsterdam)* **435**, 168677 (2021).
 - [7] R. Feynman and F. Vernon, *The theory of a general quantum system interacting with a linear dissipative system*, *Ann. Phys. (N.Y.)* **24**, 118 (1963).
 - [8] A. Müller-Hermes, J. Ignacio Cirac, and M. C. Bañuls, *Tensor network techniques for the computation of dynamical observables in one-dimensional quantum spin systems*, *New J. Phys.* **14**, 075003 (2012).
 - [9] M. B. Hastings and R. Mahajan, *Connecting entanglement in time and space: Improving the folding algorithm*, *Phys. Rev. A* **91**, 032306 (2015).
 - [10] M. R. Jørgensen and F. A. Pollock, *Exploiting the causal tensor network structure of quantum processes to efficiently simulate non-Markovian path integrals*, *Phys. Rev. Lett.* **123**, 240602 (2019).
 - [11] M. Cygorek, M. Cosacchi, A. Vagov, V. M. Axt, B. W. Lovett, J. Keeling, and E. M. Gauger, *Simulation of open quantum systems by automated compression of arbitrary environments*, *Nat. Phys.* **18**, 662 (2022).
 - [12] M. Frías-Pérez and M. C. Bañuls, *Light cone tensor network and time evolution*, *Phys. Rev. B* **106**, 115117 (2022).
 - [13] G. Giudice, G. Giudici, M. Sonner, J. Thoenniss, A. Lerose, D. A. Abanin, and L. Piroli, *Temporal entanglement, quasiparticles, and the role of interactions*, *Phys. Rev. Lett.* **128**, 220401 (2022).
 - [14] J. Thoenniss, M. Sonner, A. Lerose, and D. A. Abanin, *Efficient method for quantum impurity problems out of equilibrium*, *Phys. Rev. B* **107**, L201115 (2023).

- [15] A. Foligno, T. Zhou, and B. Bertini, *Temporal entanglement in chaotic quantum circuits*, *Phys. Rev. X* **13**, 041008 (2023).
- [16] J. Thoenniss, A. Lerose, and D. A. Abanin, *Nonequilibrium quantum impurity problems via matrix-product states in the temporal domain*, *Phys. Rev. B* **107**, 195101 (2023).
- [17] N. Ng, G. Park, A. J. Millis, G. K.-L. Chan, and D. R. Reichman, *Real-time evolution of Anderson impurity models via tensor network influence functionals*, *Phys. Rev. B* **107**, 125103 (2023).
- [18] R. Chen, X. Xu, and C. Guo, *Grassmann time-evolving matrix product operators for quantum impurity models*, *Phys. Rev. B* **109**, 045140 (2024).
- [19] M. Nayak, J. Thoenniss, M. Sonner, D. A. Abanin, and P. Werner, *Steady-state dynamical mean field theory based on influence functional matrix product states*, *Phys. Rev. B* **112**, 035103 (2025).
- [20] K. Klobas and B. Bertini, *Entanglement dynamics in Rule 54: Exact results and quasiparticle picture*, *SciPost Phys.* **11**, 107 (2021).
- [21] B. Bertini, K. Klobas, V. Alba, G. Lagnese, and P. Calabrese, *Growth of Rényi entropies in interacting integrable models and the breakdown of the quasiparticle picture*, *Phys. Rev. X* **12**, 031016 (2022).
- [22] B. Bertini, K. Klobas, and T.-C. Lu, *Entanglement negativity and mutual information after a quantum quench: Exact link from space-time duality*, *Phys. Rev. Lett.* **129**, 140503 (2022).
- [23] M. A. Rampp, S. A. Rather, and P. W. Claeys, *Entanglement membrane in exactly solvable lattice models*, *Phys. Rev. Res.* **6**, 033271 (2024).
- [24] B. Bertini, P. Calabrese, M. Collura, K. Klobas, and C. Rylands, *Nonequilibrium full counting statistics and symmetry-resolved entanglement from space-time duality*, *Phys. Rev. Lett.* **131**, 140401 (2023).
- [25] J. Yao and P. W. Claeys, *Temporal entanglement barriers in dual-unitary Clifford circuits with measurements*, *Phys. Rev. Res.* **6**, 043077 (2024).
- [26] B. Bertini, K. Klobas, M. Collura, P. Calabrese, and C. Rylands, *Dynamics of charge fluctuations from asymmetric initial states*, *Phys. Rev. B* **109**, 184312 (2024).
- [27] V. Link, H.-H. Tu, and W. T. Strunz, *Open quantum system dynamics from infinite tensor network contraction*, *Phys. Rev. Lett.* **132**, 200403 (2024).
- [28] M. Sonner, V. Link, and D. A. Abanin, *Semigroup influence matrices for nonequilibrium quantum impurity models*, *Phys. Rev. Lett.* **135**, 170402 (2025).
- [29] J. S. Cotler, D. K. Mark, H.-Y. Huang, F. Hernández, J. Choi, A. L. Shaw, M. Endres, and S. Choi, *Emergent quantum state designs from individual many-body wave functions*, *PRX Quantum* **4**, 010311 (2023).
- [30] J. Choi, A. L. Shaw, I. S. Madjarov, X. Xie, R. Finkelstein, J. P. Covey, J. S. Cotler, D. K. Mark, H.-Y. Huang, A. Kale, H. Pichler, F. G. S. L. Brandão, S. Choi, and M. Endres, *Preparing random states and benchmarking with many-body quantum chaos*, *Nature (London)* **613**, 468 (2023).
- [31] M. Ippoliti and W. W. Ho, *Solvable model of deep thermalization with distinct design times*, *Quantum* **6**, 886 (2022).
- [32] W. W. Ho and S. Choi, *Exact emergent quantum state designs from quantum chaotic dynamics*, *Phys. Rev. Lett.* **128**, 060601 (2022).
- [33] P. W. Claeys and A. Lamacraft, *Emergent quantum state designs and biunitarity in dual-unitary circuit dynamics*, *Quantum* **6**, 738 (2022).
- [34] H. Wilming and I. Roth, *High-temperature thermalization implies the emergence of quantum state designs*, *arXiv*:2202.01669.
- [35] M. Ippoliti and W. W. Ho, *Dynamical purification and the emergence of quantum state designs from the projected ensemble*, *PRX Quantum* **4**, 030322 (2023).
- [36] T. Bhore, J.-Y. Desaulles, and Z. Papić, *Deep thermalization in constrained quantum systems*, *Phys. Rev. B* **108**, 104317 (2023).
- [37] C. Liu, Q. C. Huang, and W. W. Ho, *Deep thermalization in Gaussian continuous-variable quantum systems*, *Phys. Rev. Lett.* **133**, 260401 (2024).
- [38] R.-A. Chang, H. Shrotriya, W. W. Ho, and M. Ippoliti, *Deep thermalization under charge-conserving quantum dynamics*, *PRX Quantum* **6**, 020343 (2025).
- [39] M. Lucas, L. Piroli, J. De Nardis, and A. De Luca, *Generalized deep thermalization for free fermions*, *Phys. Rev. A* **107**, 032215 (2023).
- [40] D. K. Mark, F. Surace, A. Elben, A. L. Shaw, J. Choi, G. Refael, M. Endres, and S. Choi, *Maximum entropy principle in deep thermalization and in Hilbert-space ergodicity*, *Phys. Rev. X* **14**, 041051 (2024).
- [41] S. Manna, S. Roy, and G. J. Sreejith, *Projected ensemble in a system with locally supported conserved charges*, *Phys. Rev. B* **111**, 144302 (2025).
- [42] A. Chan and A. De Luca, *Projected state ensemble of a generic model of many-body quantum chaos*, *J. Phys. A* **57**, 405001 (2024).
- [43] N. D. Varikuti and S. Bandyopadhyay, *Unraveling the emergence of quantum state designs in systems with symmetry*, *Quantum* **8**, 1456 (2024).
- [44] H. Shrotriya and W. W. Ho, *Nonlocality of deep thermalization*, *SciPost Phys.* **18**, 107 (2025).
- [45] W.-K. Mok, T. Haug, A. L. Shaw, M. Endres, and J. Preskill, *Optimal conversion from classical to quantum randomness via quantum chaos*, *Phys. Rev. Lett.* **134**, 180403 (2025).
- [46] X.-H. Yu, W. W. Ho, and P. Kos, *Mixed state deep thermalization*, *Phys. Rev. Lett.* **135**, 260402 (2025).
- [47] G. Lami, A. De Luca, X. Turkeshi, and J. De Nardis, *Quantum state design and emergent confinement mechanism in measured tensor network states*, *arXiv*:2504.16995.
- [48] J. Maldacena, S. H. Shenker, and D. Stanford, *A bound on chaos*, *J. High Energy Phys.* **08** (2016) 106.
- [49] B. Swingle, *Unscrambling the physics of out-of-time-order correlators*, *Nat. Phys.* **14**, 988 (2018).
- [50] N. Tsuji, T. Shitara, and M. Ueda, *Out-of-time-order fluctuation-dissipation theorem*, *Phys. Rev. E* **97**, 012101 (2018).
- [51] M. Fava, S. Biswas, S. Gopalakrishnan, R. Vasseur, and S. Parameswaran, *Hydrodynamic nonlinear response of interacting integrable systems*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106945118 (2021).

- [52] B. Doyon and J. Myers, *Fluctuations in ballistic transport from Euler hydrodynamics*, *Ann. Henri Poincaré* **21**, 255 (2020).
- [53] J. Myers, J. Bhaseen, R. J. Harris, and B. Doyon, *Transport fluctuations in integrable models out of equilibrium*, *SciPost Phys.* **8**, 007 (2020).
- [54] Google Quantum AI and Collaborators, *Observation of constructive interference at the edge of quantum ergodicity*, *Nature (London)* **646**, 825 (2025).
- [55] L. Foini and J. Kurchan, *Eigenstate thermalization hypothesis and out of time order correlators*, *Phys. Rev. E* **99**, 042139 (2019).
- [56] M. Brenes, S. Pappalardi, M. T. Mitchison, J. Goold, and A. Silva, *Out-of-time-order correlations and the fine structure of eigenstate thermalization*, *Phys. Rev. E* **104**, 034120 (2021).
- [57] S. Pappalardi, L. Foini, and J. Kurchan, *Eigenstate thermalization hypothesis and free probability*, *Phys. Rev. Lett.* **129**, 170603 (2022).
- [58] S. Jindal and P. Hosur, *Generalized free cumulants for quantum chaotic systems*, *J. High Energy Phys.* **09** (2024) 066.
- [59] E. Vallini and S. Pappalardi, *Long-time freeness in the kicked top*, *arXiv:2411.12050*.
- [60] S. Pappalardi, F. Fritzsche, and T. Prosen, *Full eigenstate thermalization via free cumulants in quantum lattice systems*, *Phys. Rev. Lett.* **134**, 140404 (2025).
- [61] M. Fava, J. Kurchan, and S. Pappalardi, *Designs via free probability*, *Phys. Rev. X* **15**, 011031 (2025).
- [62] G. O. Alves, F. Fritzsche, and P. W. Claeys, *Probes of full eigenstate thermalization in ergodicity-breaking quantum circuits*, *Quantum* **9**, 1949 (2025).
- [63] F. Fritzsche, T. Prosen, and S. Pappalardi, *Microcanonical free cumulants in lattice systems*, *Phys. Rev. B* **111**, 054303 (2025).
- [64] A. Chan, A. De Luca, and J. Chalker, *Eigenstate correlations, thermalization, and the butterfly effect*, *Phys. Rev. Lett.* **122**, 220601 (2019).
- [65] D. Hahn, D. J. Luitz, and J. Chalker, *Eigenstate correlations, the eigenstate thermalization hypothesis, and quantum information dynamics in chaotic many-body quantum systems*, *Phys. Rev. X* **14**, 031029 (2024).
- [66] D. L. Jafferis, D. K. Kolchmeyer, B. Mukhametzhanov, and J. Sonner, *Matrix models for eigenstate thermalization*, *Phys. Rev. X* **13**, 031033 (2023).
- [67] D. L. Jafferis, D. K. Kolchmeyer, B. Mukhametzhanov, and J. Sonner, *Jackiw-Teitelboim gravity with matter, generalized eigenstate thermalization hypothesis, and random matrices*, *Phys. Rev. D* **108**, 066015 (2023).
- [68] D. Voiculescu, *Free noncommutative random variables, random matrices and the Π_1 factors of free groups*, in *Quantum Probability and Related Topics* (World Scientific, Singapore, 1991), Vol. 6, pp. 473–487.
- [69] R. Speicher, *Free probability theory and random matrices*, in *Asymptotic Combinatorics with Applications to Mathematical Physics* (Springer, Berlin, Heidelberg, 2003), Vol. 1815, pp. 53–73.
- [70] J. A. Mingo and R. Speicher, *Free Probability and Random Matrices*, Fields Institute Monographs Vol. 35 (Springer, New York, NY, 2017).
- [71] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability* (Cambridge University Press, Cambridge, England, 2006), Vol. 13.
- [72] J. Novak and M. LaCroix, *Three lectures on free probability*, *arXiv:1205.2097*.
- [73] R. Speicher, *Free probability theory*, *Jahresber. Deutsch. Math.-Verein.* **119**, 3 (2017).
- [74] L. Hruza and D. Bernard, *Coherent fluctuations in noisy mesoscopic systems, the open quantum SSEP, and free probability*, *Phys. Rev. X* **13**, 011045 (2023).
- [75] H. J. Chen and J. Kudler-Flam, *Free independence and the noncrossing partition lattice in dual-unitary quantum circuits*, *Phys. Rev. B* **111**, 014311 (2025).
- [76] H. A. Camargo, Y. Fu, V. Jahnke, K. Pal, and K.-Y. Kim, *Quantum Signatures of Chaos from Free Probability*, *J. High Energy Phys.* **10** (2025) 138.
- [77] V. Jahnke, P. Nandy, K. Pal, H. A. Camargo, and K.-Y. Kim, *Free probability approach to spectral and operator statistics in Rosenzweig-Porter random matrix ensembles*, *J. High Energy Phys.* **12** (2025) 002.
- [78] D. A. Roberts and D. Stanford, *Diagnosing chaos using four-point functions in two-dimensional conformal field theory*, *Phys. Rev. Lett.* **115**, 131603 (2015).
- [79] D. A. Roberts, D. Stanford, and L. Susskind, *Localized shocks*, *J. High Energy Phys.* **03** (2015) 051.
- [80] R. N. Das, S. Dutta, and A. Maji, *Late time dynamics in susy saddle-dominated scrambling through higher-point otoc*, *Phys. Scr.* **99**, 085246 (2024).
- [81] J. Maldacena and D. Stanford, *Remarks on the Sachdev-ye-Kitaev model*, *Phys. Rev. D* **94**, 106002 (2016).
- [82] A. Nahum, S. Vijay, and J. Haah, *Operator spreading in random unitary circuits*, *Phys. Rev. X* **8**, 021014 (2018).
- [83] 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).
- [84] T. Yoshimura, S. J. Garratt, and J. T. Chalker, *Operator dynamics in Floquet many-body systems*, *Phys. Rev. B* **111**, 094316 (2025).
- [85] P. W. Claeys and A. Lamacraft, *Maximum velocity quantum circuits*, *Phys. Rev. Res.* **2**, 033032 (2020).
- [86] B. Bertini and L. Piroli, *Scrambling in random unitary circuits: Exact results*, *Phys. Rev. B* **102**, 064305 (2020).
- [87] D. A. Roberts and B. Yoshida, *Chaos and complexity by design*, *J. High Energy Phys.* **04** (2017) 121.
- [88] R. Speicher, *Lecture Notes on “Free Probability Theory”*, *arXiv:1908.08125*.
- [89] D. Voiculescu, *Limit laws for Random matrices and free products*, *Inventiones Mathematicae* **104**, 201 (1991).
- [90] B. Bertini, P. Kos, and T. Prosen, *Exact spectral form factor in a minimal model of many-body quantum chaos*, *Phys. Rev. Lett.* **121**, 264101 (2018).
- [91] B. Bertini, P. Kos, and T. Prosen, *Random matrix spectral form factor of dual-unitary quantum circuits*, *Commun. Math. Phys.* **387**, 597 (2021).
- [92] P. W. Claeys and G. De Tomasi, *Fock-space delocalization and the emergence of the Porter-Thomas distribution from dual-unitary dynamics*, *Phys. Rev. Lett.* **134**, 050405 (2025).

- [93] F. Fritzsche, M. F. I. Kieler, and A. Bäcker, *Eigenstate correlations in dual-unitary quantum circuits: Partial spectral form factor*, [Quantum](#) **9**, 1709 (2025).
- [94] E. Vallini, L. Foini, and S. Pappalardi, *Refinements of the eigenstate thermalization hypothesis under local rotational invariance via free probability*, [arXiv:2511.23217](#).
- [95] Note that the order of limits matters, since for finite systems there is no expected convergence in time due to revivals.
- [96] B. Bertini, P. W. Claeys, and T. Prosen, *Exactly solvable many-body dynamics from space-time duality*, [Rev. Mod. Phys.](#) **98**, 025001 (2026).
- [97] Because of unitarity of V_t or, equivalently, the strict causality in the brickwork circuit, the above k -OTOC does not depend on V_t .
- [98] B. Collins, *Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability*, [Int. Math. Res. Not.](#) **2003**, 953 (2003).
- [99] B. Collins and P. Śniady, *Integration with respect to the Haar measure on unitary, orthogonal and symplectic group*, [Commun. Math. Phys.](#) **264**, 773 (2006).
- [100] D. Weingarten, *Asymptotic behavior of group integrals in the limit of infinite rank*, [J. Math. Phys. \(N.Y.\)](#) **19**, 999 (1978).
- [101] P. Biane, *Some properties of crossings and partitions*, [Discrete Math.](#) **175**, 41 (1997).
- [102] R. Speicher, *Multiplicative functions on the lattice of non-crossing partitions and free convolution*, [Math. Ann.](#) **298**, 611 (1994).
- [103] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2010).
- [104] 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).
- [105] S. A. Rather, S. Aravinda, and A. Lakshminarayanan, *Creating ensembles of dual unitary and maximally entangling quantum evolutions*, [Phys. Rev. Lett.](#) **125**, 070501 (2020).
- [106] S. Aravinda, S. A. Rather, and A. Lakshminarayanan, *From dual-unitary to quantum Bernoulli circuits: Role of the entangling power in constructing a quantum ergodic hierarchy*, [Phys. Rev. Res.](#) **3**, 043034 (2021).
- [107] S. A. Rather, S. Aravinda, and A. Lakshminarayanan, *Construction and local equivalence of dual-unitary operators: From dynamical maps to quantum combinatorial designs*, [PRX Quantum](#) **3**, 040331 (2022).
- [108] S. Aravinda, S. Banerjee, and R. Modak, *Ergodic and mixing quantum channels: From two-qubit to many-body quantum systems*, [Phys. Rev. A](#) **110**, 042607 (2024).
- [109] S. Gopalakrishnan and A. Lamacraft, *Unitary circuits of finite depth and infinite width from quantum channels*, [Phys. Rev. B](#) **100**, 064309 (2019).
- [110] M. A. Rampp, R. Moessner, and P. W. Claeys, *From dual unitarity to generic quantum operator spreading*, [Phys. Rev. Lett.](#) **130**, 130402 (2023).
- [111] T. Zhou and A. W. Harrow, *Maximal entanglement velocity implies dual unitarity*, [Phys. Rev. B](#) **106**, L201104 (2022).
- [112] P. Kos, B. Bertini, and T. Prosen, *Correlations in perturbed dual-unitary circuits: Efficient path-integral formula*, [Phys. Rev. X](#) **11**, 011022 (2021).
- [113] J. Riddell, C. von Keyserlingk, T. Prosen, and B. Bertini, *Structural stability hypothesis of dual unitary quantum chaos*, [Phys. Rev. Res.](#) **6**, 033226 (2024).
- [114] 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).
- [115] M. P. Fisher, V. Khemani, A. Nahum, and S. Vijay, *Random quantum circuits*, [Annu. Rev. Condens. Matter Phys.](#) **14**, 335 (2023).
- [116] J. Bensa and M. Žnidarič, *Two-step phantom relaxation of out-of-time-ordered correlations in random circuits*, [Phys. Rev. Res.](#) **4**, 013228 (2022).
- [117] M. Žnidarič, *Two-step relaxation in local many-body Floquet systems*, [J. Phys. A](#) **56**, 434001 (2023).
- [118] C. Jonay and T. Zhou, *Physical theory of two-stage thermalization*, [Phys. Rev. B](#) **110**, L020306 (2024).
- [119] C. Jonay, C. Li, and T. Zhou, *Two-stage relaxation of operators through domain wall and magnon dynamics*, [Phys. Rev. B](#) **111**, 224304 (2025).
- [120] B. Bertini, P. Kos, and T. Prosen, *Operator entanglement in local quantum circuits I: Chaotic dual-unitary circuits*, [SciPost Phys.](#) **8**, 067 (2020).
- [121] K. Huang, X. Li, D. A. Huse, and A. Chan, *Out-of-time-order correlator, many-body quantum chaos, light-like generators, and singular values*, [arXiv:2308.16179](#).
- [122] J. Duarte, I. García-Mata, and D. A. Wisniacki, *Ruelle-Pollicott decay of out-of-time-order correlators in many-body systems*, [Phys. Rev. E](#) **113**, 024209 (2026).
- [123] F. Fritzsche, G. O. Alves, M. A. Rampp, and P. W. Claeys, *Free cumulants and full eigenstate thermalization from boundary scrambling*, [arXiv:2509.08060](#).
- [124] J. Haferkamp, *Random quantum circuits are approximate unitary t -designs in depth $O(nt^{5+o(1)})$* , [Quantum](#) **6**, 795 (2022).
- [125] T. Schuster, J. Haferkamp, and H.-Y. Huang, *Random unitaries in extremely low depth*, [Science](#) **389**, 92 (2025).
- [126] J. Riddell, K. Klobas, and B. Bertini, *Quantum state designs from minimally random quantum circuits*, [arXiv:2503.05698](#).
- [127] L. Grevink, J. Haferkamp, M. Heinrich, J. Helsen, M. Hinsche, T. Schuster, and Z. Zimborás, *Will it glue? On short-depth designs beyond the unitary group*, [arXiv:2506.23925](#).
- [128] A. W. Harrow and R. A. Low, *Random quantum circuits are approximate 2-designs*, [Commun. Math. Phys.](#) **291**, 257 (2009).
- [129] 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).
- [130] T. Rakovszky, F. Pollmann, and C. W. von Keyserlingk, *Diffusive hydrodynamics of out-of-time-ordered correlators with charge conservation*, [Phys. Rev. X](#) **8**, 031058 (2018).

- [131] T. Mori, *Liouvillian-gap analysis of open quantum many-body systems in the weak dissipation limit*, [*Phys. Rev. B* **109**, 064311 \(2024\)](#).
- [132] J. A. Jacoby, D. A. Huse, and S. Gopalakrishnan, *Spectral gaps of local quantum channels in the weak-dissipation limit*, [*Phys. Rev. B* **111**, 104303 \(2025\)](#).
- [133] C. Zhang, L. Nie, and C. von Keyserlingk, *Thermalization rates and quantum Ruelle-Pollicott resonances: Insights from operator hydrodynamics*, [arXiv:2409.17251](#).
- [134] U. Duh and M. Žnidarič, *Ruelle-Pollicott resonances of diffusive $U(1)$ -invariant qubit circuits*, [*SciPost Phys.* **20**, 061 \(2026\)](#).
- [135] F. Fritzsche and P. W. Claeys, *Data and numerical code for “free probability in a minimal quantum circuit model”*, [Data Set], Zenodo, [10.5281/zenodo.20325400](#) (2025)).
- [136] G. Kreweras, *Sur les partitions non croisées d’un cycle*, [*Discrete Math.* **1**, 333 \(1972\)](#).