Temporal Entanglement Transition in Chaotic Quantum Many-Body Dynamics

Ilya Vilkoviskiy,[1,2] Michael Sonner,[2,3,4] Qi Camm Huang,[5] Wen Wei Ho,[5,6] Alessio Lerose,[2,7,8] and Dmitry A. Abanin[1,9,10]

[1]*Department of Physics, Princeton University, Princeton, New Jersey 08544, USA*
[2]*Department of Theoretical Physics, University of Geneva, Quai Ernest-Ansermet 30, 1205 Geneva, Switzerland*
[3]*Max Planck Institute for the Physics of Complex Systems, 01187, Dresden, Germany*
[4]*Department of Physics, University of California, Berkeley, California 94720, USA*
[5]*Department of Physics, National University of Singapore, Singapore 117551*
[6]*Centre for Quantum Technologies, National University of Singapore, Singapore 117543*
[7]*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3PU, United Kingdom*
[8]*Institute for Theoretical Physics, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium*
[9]*École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland*
[10]*Google Research, Brandschenkestrasse 150, 8002 Zürich, Switzerland*

(Received 30 November 2025; revised 13 August 2026; accepted 5 October 2026; published 4 November 2026)

Temporal entanglement (TE) of an influence matrix (IM) has been proposed as a measure of the complexity of simulating dynamics of local observables in many-body systems. Foligno *et al.* Phys. Rev. X **13**, 041008 (2023) recently argued that the TE in chaotic 1D quantum circuits obeys linear (volume law) scaling with evolution time. To reconcile this apparent high complexity of IM with the rapid thermalization of local observables, here we study the relation between TE, non-Markovianity, and local temporal correlations for chaotic quantum baths. By exactly solving a random-unitary bath model and bounding distillable entanglement between future and past degrees of freedom, we identify a regime where TE is extensive and reflects genuine non-Markovianity. This memory, however, is entirely contained in highly complex temporal correlations, and its effect on few-point temporal correlators is negligible. An IM coarse-graining procedure, reducing the allowed frequency of measurements of the probe system, results in a transition from volume- to subvolume- (area or logarithmic) law TE scaling. We demonstrate the generality of this TE transition in 1D circuits by analyzing the kicked Ising model analytically at dual-unitary points, as well as numerically away from them. This finding indicates that dynamics of local observables are fully captured by a subvolume-law IM. We provide evidence that the compact IM MPS obtained via standard compression algorithms accurately describes local evolution.

DOI: 10.1103/s39j-czpf

I. INTRODUCTION

Quantum chaos and thermalization play a foundational role in many-body physics. Fueled by progress in quantum simulation and computing, which opened the door to their experimental studies, these phenomena continue to attract much interest and generate new lines of research. Theoretically, two classes of quantum circuits—random unitary [1] and dual-unitary (DU) [2–4] circuits—have been identified as analytical testbeds to study entanglement dynamics and operator growth under generic quantum dynamics, as they generally do not harbor any conservation laws [5–8]. Intuitively, under time evolution, a probe system interacting with such a chaotic system quickly becomes maximally entangled with it, and its reduced density matrix quickly evolves to a featureless mixed state (i.e., it thermalizes to an infinite-temperature state). This process is accompanied by a rapid delocalization of the information about the initial state of the probe system into the chaotic bath, which can no longer be retrieved by acting locally on the probe.

The dynamics of local observables in quantum many-body dynamics can be viewed through the lens of open quantum systems: one considers a small local subsystem and views it as a probe that interacts repeatedly with the rest of the system, viewed as a bath. At each step, the probe system, which we will assume to be a qudit, can be acted on by arbitrary physical operations, including preparation in a fresh state and measurement in any basis. This description gives rise to a multi-time quantum channel [9]

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

(or a many-body temporal *state*, via Choi duality [10]) which is closely related to the Feynman-Vernon influence functional [11] of a quantum bath. In the context of quantum information theory, this object is known as a process tensor [9], while in the context of many-body dynamics, it has been dubbed the influence matrix (IM) [12].

The entanglement properties of the IM, which we refer to as *temporal entanglement* (TE), provide insights into non-Markovianity [9,13–15] and into the complexity of simulating dynamics of local observables in many-body systems in different regimes [16–21], leading to new computational algorithms [22–25]. Scaling of TE with evolution time is of central importance: in particular, an area-law or logarithmic scaling with the evolution time T strongly suggests that the IM may be efficiently represented as a matrix-product state (MPS) [12,16,18]. Physically, compressing the IM of a large bath is to be expected if only a limited amount of information backflows from the bath to affect the subsequent dynamics of the probe system. In integrable, and especially in noninteracting many-body baths, information is carried away by ballistic quasiparticle excitations, leading to a favorable temporal entanglement scaling [17,24,26–29]. In this case efficient methods have been developed to compress the IM to MPS form, providing an accurate simulation of the dynamics of local observables [30–33], supported by rigorous justification [34,35].

In contrast, for chaotic quantum many-body systems, a precise connection between the entanglement properties of the IM and the possibility of efficiently approximating temporal correlators of local observables remains largely elusive. A recent study of TE in 1D chaotic quantum unitary circuits with brickwork architecture by Foligno *et al*. [36] established a volume-law scaling of TE with evolution time for dual-unitary circuits, in agreement with numerical results for random brickwork circuits, suggesting that for these chaotic quantum baths the memory required to describe the IM increases exponentially with evolution time. On the other hand, these systems undergo rapid thermalization, characterized by a finite relaxation timescale τ_{th} for local observables. This intuition agrees with the claim of Markovianization of the process tensor in chaotic systems [15,37]. Naively, one may thus expect that the object controlling the dynamics of local observables—the IM—should be well approximated using an MPS with a finite range of correlations, of the order of τ_{th} [12]. However, this expectation appears to be at odds with the conclusion of Ref. [36] that the IM has exponential complexity. The fact that generic 1D quantum circuits exhibit volume-law scaling of TE raises a question of the physical origin of this scaling: does it correspond to an operational protocol that probes the information recoverability between past and future degrees of freedom?

Even with the presence of such extensive temporal entanglement scaling, does it pose a fundamental obstruction to an efficient numerical simulation of the IM if we restrict attention to only practically relevant observables, such as few-point temporal correlations?

The goal of this paper is to investigate the connection between the volume-law TE of quantum chaotic systems' IMs, non-Markovianity, and temporal correlators of local observables. Focusing on 1D local quantum circuits, we argue that, while volume-law TE can reflect genuine non-Markovianity of dynamics, this high complexity is only manifest in temporal correlation functions involving a high frequency of measurements of the probe system. Instead, temporal correlation functions involving a sufficiently low (but finite) frequency of measurements of the probe system, including all the physically relevant few-point correlation functions, can be captured by a modified IM with subvolume-law scaling of TE. This IM is generated by a coarse-graining procedure, which effectively projects out high-frequency correlators while retaining an exact description of low-frequency correlators. In particular, we establish that there is a *finite* critical coarse-graining density $0 < n_{\mathrm{cg}}^{\star} < 1$ at which volume-law TE scaling of the IM sharply transitions to a subvolume-law TE scaling.

A. Main results

Before describing our main results, we recall some basic properties of IMs, with a more detailed discussion provided in Sec. II. Physically, an IM can be identified with a quantum state of T pairs of qudits, with local Hilbert-space dimension d: Starting with a product state of T maximally entangled qudit pairs (Bell pairs), we let one qudit of each pair sequentially interact with the bath (see Fig. 1). The final state of the $2T$ qudits is, up to a factor of d^{-T}, identical to the IM.

Knowledge of the IM allows one to describe an experiment where a probe system, prepared in some state, interacts with the bath and is measured after each interaction step, i.e., probe a *multi-time correlation function*. In particular, one can consider correlations between the first p qudit pairs ("past" subsystem, $\mathcal{P}$), and the remaining $f = T - p$ pairs (the "future" subsystem, $\mathcal{F}$); the size of correlations determine the degree of Markovianity of the bath, with small (large) correlations corresponding to the Markovian (non-Markovian) regimes. The operator-space entanglement entropy of the associated density matrix of the IM, called the *temporal entanglement* (TE), has been proposed to characterize the Markovian property of a bath.

However, in practice, one is often interested in a more restricted setup of *few-point* temporal correlation functions, with a given unitary evolution of the probe between interaction steps with the bath. Such correlators are obtained by projecting the IM onto a hyperplane determined by the local evolution of the probe qudit, as well as the chosen observables (see below for details). A central question addressed in this work is to elucidate how TE and Markovianity properties of chaotic quantum baths depend

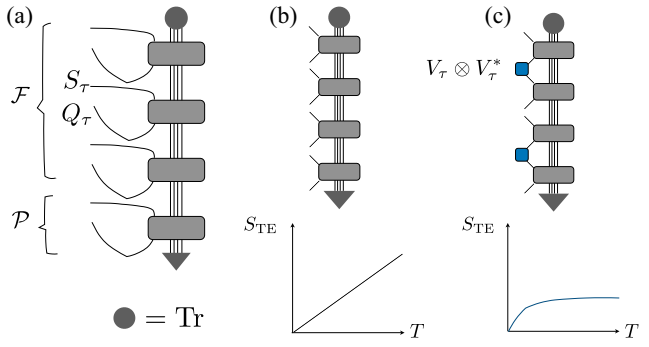


FIG. 1. Schematic of a general influence matrix (IM) and temporal entanglement (TE) transition. (a) IM state is obtained by preparing T qudit pairs (each comprising a "system" qudit S_τ and a "reference" qudit Q_τ) in maximally entangled Bell states and then sequentially bringing the qudits $S_1, S_2, \ldots, S_T$ to interact with a given bath. The IM state is the state of these $2T$ qudits at the end of the procedure (i.e., the state obtained by tracing over the bath degrees of freedom). (b), (c) IM before (b) and after (c) the coarse-graining operation, which amounts to inserting unitary evolution of the probe qudit at a fraction of the time steps. As we show, beyond a critical coarse-graining density, TE undergoes a transition from volume-law to subvolume-law (see main text for details).

on the allowed frequency of observations of the probe system.

Our analysis proceeds as follows. First, to develop intuition, we study a tractable toy model of thermalization where a probe system of Hilbert dimension d (taken to be 2 in most of the paper) is repeatedly coupled to a large structureless bath with Hilbert-space dimension $D_{\mathcal{B}}$, modeled via the repeated application of a random unitary acting on the $dD_{\mathcal{B}}$-dimensional space [see Fig. 1(a)]. We compute TE, which is a function of the ratio of evolution time T and bath size $b = \log_d D_{\mathcal{B}}$,

$$r = \frac{T}{b}, \tag{1}$$

and find two opposite regimes: In the limit $T \ll b$, or equivalently $r \ll 1$, the IM is close to a product state, implying that no experiment performed on the future system ($\tau > p$) can retrieve information about the past $\tau \leq p$. Thus, as previously found in Ref. [15], this limit is genuinely Markovian. Conversely, in the opposite limit $r \gg 1$, we instead find the TE saturates the upper boundset by the finite bath size, $S_{\mathrm{TE}} \sim 2b$ (i.e., TE is extensive in b or T), suggesting that this limit is non-Markovian. These two regimes are separated by a sharp transition at $r = 1/2$. A similar transition is found in a dynamical version of the toy model, where the number of bath degrees of freedom is taken to grow linearly in time as $b(\tau) = 2v_B\tau$, mimicking the locality constraints in 1D unitary circuits. Here, the TE is found to saturate for $v_B > 1$ and to grow extensively for $0 < v_B < 1$.

We then make a connection to the black-hole quantum information retrieval problem [38,39]. We apply quantum information techniques, such as decoupling theorems [40], to show that for $r \gg 1$, one can distill b Bell pairs between $\mathcal{P}$ and $\mathcal{F}$. Since the number of distilled Bell pairs puts a lower bound on the *distillable entanglement* [41,42] between the two parties, this shows the $r \gg 1$ limit is indeed associated with dynamics being genuinely non-Markovian: in particular, both TE and the number of distillable Bell pairs scale extensively here. At the same time, we argue that this non-Markovianity is encoded in highly nonlocal temporal correlations involving measurements at an extensive number of times. By contrast, few-point correlation functions decay rapidly to random fluctuations of order $1/D_{\mathcal{B}}$ and therefore do not exhibit any discernible signature of non-Markovianity.

Our analysis of the toy model suggests a simple picture explaining why volume-law scaling of TE is seen ubiquitously in more physical extended 1D brickwork circuits: due to locality of interactions, a chaotic bath prepared in an initial short-range-correlated state (e.g., fully mixed state or a product state), has an effective Hilbert-space dimension seen by a probe which grows in time as $D_B(t) \sim 2^{2v_B t}$, v_B being the butterfly velocity. A rough estimate of the numbers of temporal and bath degrees of freedom thus places the system in the non-Markovian regime, $r \approx \frac{1}{2v_B} > \frac{1}{2}$, where the inequality follows from the constraint $v_B < 1$ imposed by the assumption that the brickwork circuit is composed of two-local gates (as most past works assume [3,6,43,44]). In such a setting, we therefore expect TE to grow linearly with T.

An interesting possibility to curb such extensive growth of TE, further suggested by this analysis, is to consider the effect of a restricted setup by reducing the allowed frequency of operations that we can perform on the probe system. Mathematically, preventing access to the probe at a given time step is implemented by a local projection onto the unitary evolution of the probe; we refer to this projection as temporal *coarse-graining* of the IM state introduced earlier in Ref. [15]. Suppose, in particular, that the probe system's evolution at each step is set by the unitary operator V_τ, see Figs. 1(b), 1(c). Then we can impose this evolution on the probe system at a fraction $1 - n_{\mathrm{cg}}$ of time steps, effectively restricting the frequency of measurements we can access. This corresponds to projecting the IM on the remaining $n_{\mathrm{cg}}T$ probes, producing a new, coarse-grained IM state with a reduced temporal Hilbert-space dimension $D_T^{\mathrm{cg}} = d^{2n_{\mathrm{cg}}T}$. By construction, coarse-graining preserves the values of few-point temporal correlations exactly. The role of coarse-graining has been studied in Ref. [15] for the toy random unitary bath model considered in our Sec. III, and also in Ref. [37] where the

authors considered the process tensor of a Hamiltonian system, with probe interventions at discrete times with tunable separation δt playing the role of n_{cg}^{-1}.

The central result of the present paper is our claim that coarse-graining can in fact lead dramatically to a *TE transition* of the IM from volume-law to subvolume-law scaling at a finite coarse-graining density $n_{\mathrm{cg}} < n_{\mathrm{cg}}^{\star}$, where the threshold value $n_{\mathrm{cg}}^{\star}$ is system-specific, i.e., a *complexity transition* in the influence of the bath on the local system, see Figs. 1(b) and 1(c). Intuitively, coarse-graining effectively renormalizes $r \to r n_{\mathrm{cg}}$, thereby eliminating the volume-law contribution whenever $n_{\mathrm{cg}} < n_{\mathrm{cg}}^{\star} \simeq 1/(2r)$. We demonstrate this transition rigorously in the random-unitary toy model and argue that an analogous transition occurs in generic quantum circuits in 1D.

Namely, we investigate spatially extended quantum bath models in 1D with local interactions, exemplified by the Floquet kicked Ising model. Following Ref. [36], we first consider dual-unitary points of this model with nonsolvable initial conditions. We can analytically demonstrate that the TE transitions from a volume to area-law at an arbitrarily weak coarse-graining (i.e., $n_{\mathrm{cg}}^{\star} = 1$). We also show that, as expected, local observables $\langle O_\tau \rangle$ rapidly decay to zero, with irregular fluctuations decaying in magnitude of order $2^{-T/2}$. These fluctuations are very sensitive to the initial state and cannot be efficiently captured for arbitrarily long times by an MPS representation of the IM with a moderate bond dimension. However, owing to the rapid decay of these fluctuations, observables can be approximated to absolute accuracy ϵ using a truncated bond dimension that grows only polynomially in $1/\epsilon$.

Turning to more generic circuits, we numerically study the kicked Ising model away from the dual-unitary points and observe a TE transition upon coarse-graining. In contrast to random-unitary and dual-unitary circuits, below the transition the TE obeys a log-law rather than an area-law. This model also exhibits a much richer phenomenology: in particular, certain parameter regimes host prethermal, nearly conserved edge operators with parametrically slow temporal decay [45]. We show that the behavior of TE under coarse-graining is sensitive to these prethermal effects and differs qualitatively between the corresponding regimes.

Finally, the subvolume-law scaling of TE below the transition point suggests the existence of an effective low-bond-dimension IM that accurately reproduces few-point correlation functions. To test this expectation in a concrete setting, we study the convergence of these correlation functions with increasing bond dimension in the kicked Ising model and find that relatively modest bond dimensions are indeed sufficient to faithfully capture the thermalization dynamics.

All together, our work provides a unifying framework to understand and relate the concepts of temporal entanglement within the influence matrix, non-Markovianity of a bath, and temporal correlations. It also potentially provides a practical route to reducing the complexity needed to simulate a quantum chaotic bath, through the idea of coarse-graining.

The rest of the paper is organized as follows: in Sec. II, we review the properties of the IM. In Sec. III, we analytically study TE and distillable entanglement for the case of a structureless random unitary bath and discuss TE transition upon coarse-graining. Further, in Sec. IV, we consider TE transitions in dual-unitary circuits with nonsolvable initial conditions. After that, we turn to a more generic, non-dual-unitary kicked Ising model, discovering a phenomenology consistent with predictions from the structureless random unitary bath model. We argue that the IM can be significantly compressed into a compact MPS form, still capturing dynamics of local observables accurately. Finally, in Sec. VI we summarize the main results and outline directions for future research.

II. SETTING THE STAGE

A. IM formalism

We start by reminding the reader of the definition of the IM and reviewing its key properties [9,12] [see Fig. 1(a)]. An IM can be viewed as a multi-time generalization of a quantum channel: Given a many-body bath $\mathcal{B}$ with Hilbert-space dimension $D_{\mathcal{B}}$, at each time step $\tau = 1, \ldots, T$ a new probe qudit interacts with $\mathcal{B}$ via a unitary transformation U_τ acting on the composite $dD_{\mathcal{B}}$-dimensional space. This unitary contains both the bath's internal dynamics and qudit-bath interactions. At the end, bath degrees of freedom are traced out. The resulting object, which we denote by $\mathcal{I}$, is a quantum channel with T inputs and T outputs, causally related in a temporal sequence, see Fig. 2(a).

Explicitly, the IM elements are given by

$$\mathcal{I}_{q_1,s_1\cdots q_T,s_T}^{\bar{q}_1,\bar{s}_1\cdots\bar{q}_T,\bar{s}_T} = \mathrm{tr}_{\mathcal{B}}\Big((U_T)_{q_T}^{s_T}\cdots(U_1)_{q_1}^{s_1}\rho_{\mathcal{B}}^{0}(U_1^{\dagger})_{\bar{q}_1}^{\bar{s}_1}\cdots(U_T^{\dagger})_{\bar{q}_T}^{\bar{s}_T}\Big), \tag{2}$$

with indices q_τ, s_τ $[\bar{q}_\tau, \bar{s}_\tau]$ associated with the input and output state of the τth probe qudit on the forward [backward] branch of time evolution, respectively. In this equation, we have defined the conditioned operators $(U_\tau)_{q_\tau}^{s_\tau} = \langle s_\tau | U_\tau | q_\tau \rangle$, acting on $\mathcal{B}$ only.

As illustrated in Fig. 1(a), the multi-time quantum channel $\mathcal{I}$ can be mapped to a $2T$-qudit (mixed) quantum state via Choi's channel-state duality [10]. At each step τ, we prepare a maximally entangled (Bell pair) state of two qudits—a "reference" qudit Q_τ and "system" qudit S_τ, and we let the latter interact with the bath. This results in a state $\rho_{\mathcal{I}}$ of $2T$ qudits, which is dual to the IM channel: The density matrix associated with $\rho_{\mathcal{I}}$ has exactly the same elements as the IM $\mathcal{I}$ up to normalization,

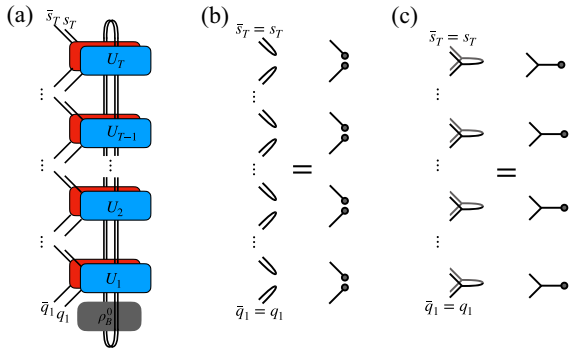


FIG. 2. (a) Tensor network representation of the IM: gray box at the bottom corresponds to an initial density matrix ρ_B^0, blue gates correspond to the forward action of unitary gates U_i, while red gates correspond to the corresponding backward action. This is identical to Fig. 1 except that there all forward and backward diagrams (gates, legs, etc.) are overlaid ("folded representation") while here all diagrams are explicitly shown ("unfolded representation"). In this paper, all diagrams drawn will predominantly be in the folded representation. (b), (c) Two illustrative perfectly Markovian IMs. (b) Perfect depolarizer solution in the unfolded (left) and folded (right) representation. (c) Perfect dephaser solution in unfolded (left) and folded (right) representation. Both IMs exhibit area-law scaling of their TE.

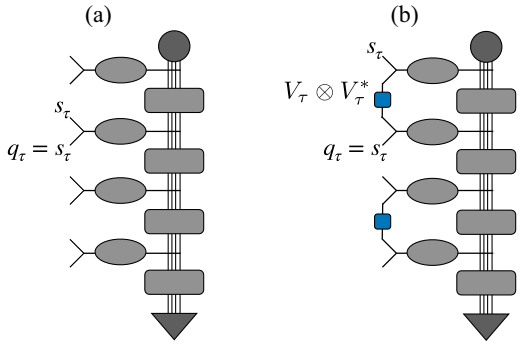


FIG. 3. (a) IM of the constrained type, arising from probe-bath interactions in product operator form $U = e^{-iH_{\text{probe}} \otimes H_{\text{bath}}}$ represented by oval gates. We use here the diagonal tensor notation as in Ref. [12]. (b) IM after coarse-graining procedure with $n_{\text{cg}} = 1/2$.

$$(\rho_{\mathcal{I}})^{\bar{q}_1,\bar{s}_1\cdots\bar{q}_T,\bar{s}_T}_{q_1,s_1\cdots q_T,s_T} = \frac{1}{d^T}\mathcal{I}^{\bar{q}_1,\bar{s}_1\cdots\bar{q}_T,\bar{s}_T}_{q_1,s_1\cdots q_T,s_T}. \tag{3}$$

When studying circuits in Secs. IV and V, we will instead consider a constrained type of IM arising from probe-bath interactions described by a (time-independent) product operator, i.e., $U = e^{-iH_{\text{probe}} \otimes H_{\text{bath}}}$. In this case, working in a basis where H_{probe} is diagonal, $H_{\text{probe}}|s\rangle = h_s|s\rangle$, the IM elements simplify as

$$\langle s_\tau|U|q_\tau\rangle = \delta_{q_\tau,s_\tau} e^{-ih_{s_\tau}H_{\text{bath}}} = \delta_{q_\tau,s_\tau} U_{s_\tau}, \tag{4}$$

where the conditioned bath operator U_s is here unitary, similarly to the original Feynman-Vernon influence functional [11]. Then, we can identify $q_\tau = s_\tau$ and $\bar{q}_\tau = \bar{s}_\tau$ in Eq. (2), which yields

$$\mathcal{I}^{\bar{s}_1\cdots\bar{s}_T}_{s_1\cdots s_T} = \text{tr}_{\mathcal{B}}\left(U_{s_T}\cdots U_{s_1}\rho_{\mathcal{B}}^0 U^\dagger_{\bar{s}_1}\cdots U^\dagger_{\bar{s}_T}\right). \tag{5}$$

In tensor notation, the general IM pictures in Figs. 1(b) and 1(c) simplify into the corresponding ones in Figs. 3(a) and 3(b). Such an IM naturally arises in applications to the kicked Ising model (see Ref. [12] for earlier work), and we therefore refer to it as an "Ising-type" IM.

B. Entanglement properties of the IM: Temporal entanglement

The IM provides a complete description of the influence of a many-body bath on a probe system. Intuitively, the IM encompasses the flow of information from earlier times (the past) to later times (the future) via the bath, see Fig. 1(a). A central question is to characterize this information flow and to find out whether the IM admits compression—in other words, whether the large unitary bath can be replaced by a smaller, potentially dissipative, effective bath, which still accurately approximates the probe system's dynamics. If the bath is Markovian (correlations between between past and future are small), then it is expected that compression can be performed well; otherwise if the bath is non-Markovian (correlations are large), then it is expected that such compression is not possible. Different measures of Markovianity based on the IM (or process tensor) have been introduced [13,14].

As mentioned before, temporal entanglement (TE) of the IM will be the key measure of complexity studied in this paper, following Refs. [12,16,19,24,26]. We view the IM (or its associated density matrix $\rho_{\mathcal{I}}$) as a *vector* $|\mathcal{I}\rangle$ in d^{4T}-dimensional space and study the entanglement properties of this vector, in particular its von Neumann entanglement entropy with a past-future bipartition $(\mathcal{P},\mathcal{F})$, Explicitly, this is given by

$$S_{\text{TE}} = -\text{Tr}_{\mathcal{P}}(\rho_{\mathcal{P}}\log_d\rho_{\mathcal{P}}), \qquad \rho_{\mathcal{P}} = \text{Tr}_{\mathcal{F}}\left(\frac{|\mathcal{I}\rangle\langle\mathcal{I}|}{\langle\mathcal{I}|\mathcal{I}\rangle}\right), \tag{6}$$

in contrast with other notions of IM entanglement that we investigate. In particular, the *scaling* of TE with T is an important aspect we are interested in (namely, how the amount of entanglement scales with the number of temporal degrees of freedom).

Indeed, a particularly simple illustrative example of an IM describing a perfectly Markovian (no correlations between past and future) bath is the *perfect depolarizer* [12],

$$(\mathcal{I}_{\text{depol}})^{\bar{q}_1,\bar{s}_1\cdots\bar{q}_T,\bar{s}_T}_{q_1,s_1\cdots q_T,s_T} = \frac{1}{d^T}\prod_{\tau=1}^{T}\delta_{q_\tau\bar{q}_\tau}\delta_{s_\tau\bar{s}_\tau}, \tag{7}$$

which can easily be seen to exhibit an *area-law* scaling of TE. Physically, this corresponds to a bath that transforms

any incoming impurity degree of freedom into a maximally mixed state, thereby completely erasing the memory of its prior history. Despite its simplicity, this state arises as an exact IM in a class of chaotic circuits known as dual-unitary circuits, for a family of solvable initial conditions [46]. A pictorial illustration of this state is provided in Fig. 2(b) both in unfolded (left) and folded (right) representation. The many-body state associated with this IM is the maximally mixed state $\rho_{\text{depol}} = \mathbb{1}_{d^{2T}}/d^{2T}$. A relative of the perfect depolarizer solution, specific to the Ising-type IM is the *perfect dephaser* solution:

$$\mathcal{I}_{T,\text{deph}} = \prod_{i=1}^{T} \delta_{s_i \bar{s}_i}, \tag{8}$$

illustrated in Fig. 2(c). Physically, this IM maps any incoming density matrix to its diagonal part, eliminating all off-diagonal coherences—hence the name "perfect dephaser." It arises, for example, as an exact IM of the kicked Ising model at a dual-unitary point and also exhibits area-law scaling of TE. Area-law scaling of TE was further found in some integrable systems [12,24,27–29].

In contrast, numerical simulations for generic chaotic non-integrable baths suggested a linearly increasing TE; this volume-law scaling was recently proven by Foligno *et al*. [36] for chaotic 1D dual-unitary circuits upon using non-solvable initial states, indicating a surprising fragility of the area-law depolarizer IM [Eq. (7)] solution of the perfectly Markovian quantum bath [47].

C. Temporal coarse-graining

The IM is a very general object, which contains significantly more information than the standard few-point temporal correlation functions, $\langle O_N(\tau_N)\cdots O_1(\tau_1)\rangle$, where O_j is an operator acting on the probe qudit and the number of probes is assumed few: $N = O(T^0)$.

Generally, such correlators can be computed by contracting the IM with quantum channels acting on the probe qudit only and inserting the operators $O_1, \ldots, O_N$ at the appropriate time steps. Assuming probe system's internal dynamics is given by a time-independent unitary $V_\tau = V$ [48], the temporal correlation function is explicitly given by

$$\begin{aligned}&\langle O_N(\tau_N)\cdots O_1(\tau_1)\rangle \\ &= \sum_{\{q_i,\bar{q}_i,s_i,\bar{s}_i\}} \mathcal{I}^{\bar{q}_1,\bar{s}_1\cdots\bar{q}_T,\bar{s}_T}_{q_1,s_1\cdots q_T,s_T} \prod_{k\neq\{\tau_i\}} V_{s_k,q_{k+1}} V^*_{\bar{s}_k,\bar{q}_{k+1}} \\ &\quad\times \prod_{k=1}^{N} [O_k(\tau_k)V]_{s_{\tau_k},q_{\tau_k+1}} V^*_{\bar{s}_{\tau_k},\bar{q}_{\tau_k+1}}.\end{aligned} \tag{9}$$

In this light, besides the task of globally approximating the IM channel $\mathcal{I}$ (or the dual state $\rho_{\mathcal{I}}$), another simpler task emerges: approximating the IM to the extent that a given relevant subset temporal correlation functions are accurately captured. On physical grounds, we are interested in describing processes where the frequency of "interrogations" of the probe is low; for instance, we could be interested in temporal correlations with a specified minimum temporal spacing between consecutive measurements of the probe. As we will show below, this task is significantly less demanding. To that end, we will consider a coarse-graining procedure for the IM, which amounts to inserting undisturbed unitary evolutions V_τ at a uniformly distributed fraction $(1 - n_{\text{cg}})$ of time steps τ. This way the number of IM degrees of freedom is reduced by a factor n_{cg}. (For a vectorized IM, coarse-graining is equivalent to projection onto a specific hyperplane set by the choice of V_τ.) This procedure is shown in Fig. 1(c) for a general IM and in Fig. 3(b) for the constrained-type IM of later relevance for the kicked Ising model. [In the latter case, for simplicity, we choose to restrict the operators V_τ to be diagonal, so as to preserve the constraint from Eq. (4).] By construction, a coarse-grained IM still allows to exactly compute low-frequency correlators. Crucially, we will show that below a critical coarse-graining threshold $n^\star_{\text{cg}}$ the IM complexity is greatly reduced, as TE collapses to a subvolume-law scaling. Such a coarse-grained IM can be effectively approximated by an MPS.

III. A TOY RANDOM UNITARY BATH MODEL

Although our ultimate goal is to understand spatially extended systems with a non-trivial locality structure, much of the intuition for their dynamics comes from the spreading and scrambling of initially local quantum information [5,6,43]. Once this information has spread over a sufficiently large region and become strongly scrambled, the detailed spatial structure within that region may become secondary to its collective role as an effective bath. Motivated by this picture, we begin by considering a toy model of a completely structureless random bath. This model isolates the dynamics of an already-scrambled region while abstracting away its detailed spatial organization, allowing us to study the relation between thermalization, temporal entanglement (TE), and coarse-graining in a particularly simple setting. It also provides a useful baseline for the local circuit models considered below, in which the effective bath emerges dynamically through information spreading.

We consider two variants of the random-bath model, which we call the "static" and "dynamical" models. In both variants, at each time step τ, a probe qudit interacts with a structureless bath of Hilbert-space dimension $D_{\mathcal{B}}(\tau) = d^{b(\tau)}$, where $b(\tau)$ denotes the effective number of bath qudits. The joint probe-bath evolution is generated by an independently drawn Haar-random unitary U_τ of dimension $dD_{\mathcal{B}}(\tau) \times dD_{\mathcal{B}}(\tau)$.

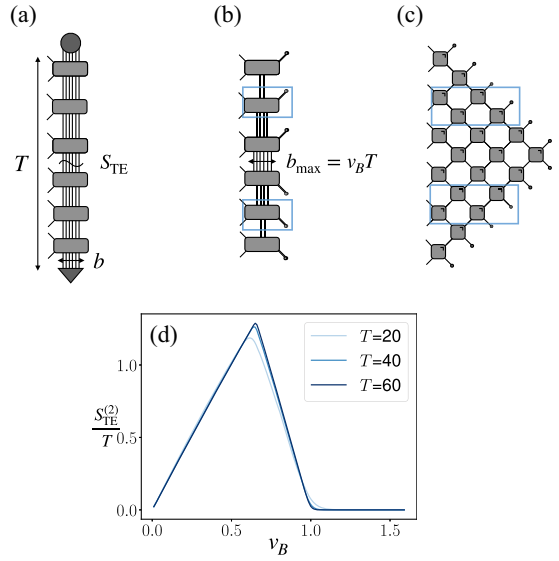


FIG. 4. Illustration of the IMs of the static (a) and dynamically growing (b) bath toy models, for $T = 6$ time steps and $b = 6$ bath qudits (a) or $b_{\max} = 6$ maximal number of bath qudits (b), with $v_B = 1$. For illustration, we also show the corresponding quantum circuit for an infinite-temperature initial state (c). The random-bath approximation consists of replacing the structured gate blocks, highlighted by blue boxes, with unstructured random unitary matrices. (d) Slope of TE scaling as a function of v_B for several values of T, showing the convergence to a limiting slope as $T \to \infty$ and a transition from volume-law to subvolume-law TE scaling at the critical value $v_B^\star = 1$.

In the static model, introduced in Ref. [15] and further studied in Ref. [49], the bath size is fixed throughout the evolution, $b(\tau) = b$. In the dynamical extension introduced here, we instead take $b(\tau) = 2v_B \min(\tau, T - \tau)$ so that the bath grows linearly until $\tau = T/2$ and then shrinks at the same rate. The parameter v_B plays the role of a butterfly velocity: it controls the rate at which new bath degrees of freedom become accessible, thereby mimicking the light-cone constraint on information propagation in a one-dimensional unitary circuit while leaving the bath itself structureless. More explicitly, for $\tau < T/2$, the bath is enlarged at each time step by adding $2v_B$ maximally mixed qudits. The probe and the enlarged bath are then acted upon by a Haar-random unitary U_τ of dimension $d^{2v_B\tau+1} \times d^{2v_B\tau+1}$. For $\tau > T/2$, the bath is correspondingly reduced by tracing out $2v_B$ qudits at each time step. The IMs for static and dynamical random-bath models, together with the local quantum-circuit models considered below, are illustrated in Figs. 4(a) and 4(b).

A. Overview of results

We give a high-level overview of the results contained within this section. First, in Sec. III B, we analyze TE [Eq. (6)] in these toy models using Weingarten calculus [50,51]. In the static version, its behavior is controlled by the ratio $r = T/b$ between the number of degrees of freedom of the probe system and of the bath, both assumed large, $T, b \gg 1$ [see Eq. (1)]. We show that TE transitions from an exponentially suppressed behavior for $r < 1/2$ to an extensive scaling (with T) for $r > 1/2$, reaching the maximum possible value $2b$ at $r = 1$. In the dynamical version, a similar behavior is found, with the role of r taken up by $1/(2v_B)$.

Further, we explain how the IM coarse-graining procedure alluded to in the introduction simply amounts to reducing the parameter $T \to n_{\mathrm{cg}} T$, and hence $r \to n_{\mathrm{cg}} r$ (static) or $v_B \to v_B / n_{\mathrm{cg}}$ (dynamical). This results in the claimed TE transition from volume-law to subvolume-law scaling.

In Sec. III C we investigate the relation between TE—which, by itself, is not a physical quantity—and an operationally well-defined notion of (non-)Markovianity of the probe system's dynamics, throughout the parameter space of this model. Focusing for definiteness on the static bath model, we first show that in the regime $r < 1/2$, where TE is exponentially suppressed, one can establish an upper bound on any possible connected correlation function between operators supported on $\mathcal{P}$ and $\mathcal{F}$, which ought to be smaller than $\epsilon = d^{-b(1/2-r)}$. This bound follows from the observation that the IM is ϵ-close in trace distance to the Markovian perfect depolarizer IM in Eq. (7). This fact was previously pointed out in Ref. [15], where the authors further bounded the probability of significant deviations from Markovianity.

To go beyond this Markovian regime, we study the relation between TE and non-Markovianity for $r > 1/2$. In particular, we provide intuition for the physical origin and meaning of the extensive scaling of TE in this by demonstrating the concomitant existence of an extensive amount of genuine quantum correlations between $\mathcal{P}$ and $\mathcal{F}$ in the form of shared Bell pairs, i.e., *distillable entanglement* [41,42].

More precisely, we show that $\mathcal{F}$ can retrieve at least $N_{\mathcal{PF}} = T \min\left(\frac{2r-1}{2r}, \frac{1}{r}\right)$ qudits of quantum information injected in the bath by $\mathcal{P}$. This result shows that the IM viewed as a many-body quantum state has extensive genuine quantum entanglement, and that this entanglement has a transparent physical interpretation as non-Markovian memory of the bath. For $r > 3/2$, in particular, the capacity of the bath to transmit quantum information from the past to the future is saturated, as $N_{\mathcal{PF}} = b$. Our argument may be viewed as a multi-time extension of the Hayden-Preskill protocol for the problem of information retrieval from evaporating black holes [39,52], as it reduces to it in the limit of a single time step.

B. Temporal entanglement

We start by analyzing the Rényi TE, defined as bipartite Rényi entanglement entropy of the IM vector,

$S_{\mathrm{TE}}^{(\alpha)} = \frac{1}{1-\alpha}\log_d \mathrm{Tr}\rho_{\mathcal{P}}^{\alpha}$, in the static bath model with constant b. We compute its behavior as a function of d and $D_{\mathcal{B}}$ using Weingarten calculus [50,51] for a pure initial state of the bath. The details of the calculation, along with the analysis of a fully mixed initial state of the bath, are provided in Appendix A 2. The result for $\alpha = 2$ and even bipartition $p = f = T/2$ is the following asymptotic TE formula:

$$S_{\mathrm{TE}}^{(2)} = -\log_d \frac{1 + D_{\mathcal{B}}^{4(r-1)} + \epsilon(D_{\mathcal{B}}, r)}{(1 + D_{\mathcal{B}}^{2r-1})^2} + o\left(\frac{1}{D_{\mathcal{B}}}\right). \quad (10)$$

Although the second Rényi entropy provides only a lower bound on the von Neumann TE [see Eq. (6)], it is sufficient to identify the regime in which TE grows. Conversely, in the regime of small TE, we derive an independent upper bound, thereby obtaining a clear picture of the overall behavior of TE.

The term $\epsilon(D_{\mathcal{B}}, r)$ in this equation is negligible for all r except in the regime $r \approx 1$, where it induces appreciable corrections of order $b^{-1} = (\log_d D_{\mathcal{B}})^{-1}$. This term can be computed exactly:

$$\epsilon(D_{\mathcal{B}}, r) = 2D_{\mathcal{B}}^{3(r-1)} + \left(\frac{3}{4} r \log_d D_{\mathcal{B}} + 2\right) D_{\mathcal{B}}^{2(r-1)} + 2D_{\mathcal{B}}^{(r-1)}. \quad (11)$$

Below we specialize to the case of qubits, $d = 2$. The exact behavior of $S_{\mathrm{TE}}^{(2)}$ as a function of r is illustrated in Fig. 5 for different values of $b = \log_2 D_{\mathcal{B}}$, illustrating the approach to the asymptotic limit for large b. Notably, this approach is relatively slow near $r \approx 1$, due to the previously mentioned $\mathcal{O}(1/b)$ corrections.

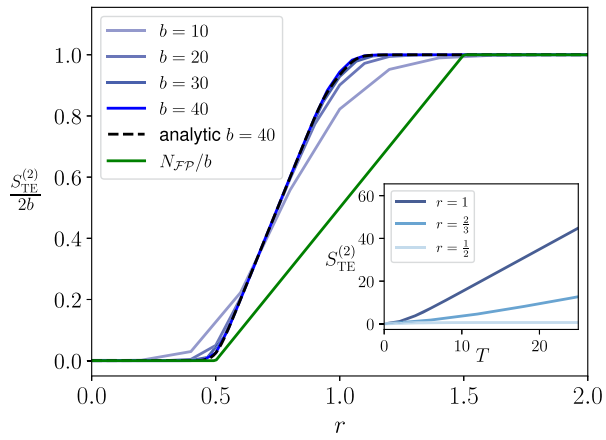


FIG. 5. Maximum temporal Rényi-2 TE for the structureless random unitary bath model, plotted as a function of r for different bath sizes $b = \log_2 \mathcal{D}_B$, fixed probe system dimension $d = 2$, and pure bath initial state. The dashed line illustrates the analytical prediction, see Eq. (10). The green curve gives a lower bound on distillable entanglement measured in units of b. *Inset:* Rényi-2 TE for $r = 1$ and coarse-graining parameters $rn_{\mathrm{cg}} = 1, 2/3, 1/2$. We observe a transition from volume-law to area-law scaling for $r^{\star} = 1/2$.

The above result indicates that there are three regimes of Rényi TE behavior, also visible in Fig. 5:

(i) For $r < 1/2$, TE is decaying to zero exponentially in b, $S_{\mathrm{TE}}^{(2)} \sim \frac{2d^{-b(1-2r)}}{\ln d}$;
(ii) For $1/2 < r \leq 1$, TE approaches the limiting behavior $S_{\mathrm{TE}}^{(2)} \sim 2(2r-1)b$ for large b;
(iii) For $r > 1$, TE saturates at the maximum possible value $2b$.

We note that extensivity of higher Rényi TEs implies extensivity of the regular von Neumann TE, due to monotonicity with the Rényi index α.

In the regime $r < 1/2$, the probe system's dynamics is nearly Markovian [15], and the IM is well approximated by a perfect depolarizer IM, such that the corresponding density matrix $\rho_{\mathcal{I}}$ is close to a maximally mixed state:

$$\rho_{\mathrm{depol}} = \frac{1}{d^{2T}} \mathbb{1}_{d^{2T}}. \quad (12)$$

In Appendix B [see Eq. (B13) and Ref. [15]], we estimate the trace distance from the maximally mixed state, obtaining the bound

$$\|\rho_{\mathcal{I}} - \rho_{\mathrm{depol}}\|_1 < d^{-b(1/2-r)}. \quad (13)$$

This indicates that all correlations of the probe system are suppressed in the limit $b \to \infty$. In contrast, for $r > 1/2$, where TE is proportional to T, the dynamics is non-Markovian, as discussed in the next Subsection.

Furthermore, the random unitary bath model allows to illustrate the effects of coarse-graining by elementary means, as previously noted in Ref. [15]. Since the product of Haar random unitaries is also a Haar-random unitary, coarse-graining the temporal axis (with an arbitrary V_τ) simply amounts to reducing the number of time steps $T \to n_{\mathrm{cg}} T$, leading to a rescaling of the parameter r:

$$r \to n_{\mathrm{cg}} r. \quad (14)$$

Thus, if the initial value $r = r_0$ satisfies the inequality $r_0 > 1/2$, such that the IM is initially an extensively entangled vector, coarse-graining beyond the critical value

$$n_{\mathrm{cg}}^{*} = (2r_0)^{-1} \quad (15)$$

brings the IM down to the regime $r < 1/2$, with exponentially small TE, exponentially small temporal correlators, and close to the Markovian perfect depolarizer "fixed-point" in trace distance. The bound (13) also guarantees that distilling even a single Bell pair between subsystems $\mathcal{P}$ and $\mathcal{F}$ is impossible.

In order to capture the local structure of the quantum circuits, we now turn to the dynamically growing random

unitary bath model. Figure 4(d) shows the dependence of the asymptotic slope of TE scaling on the parameter v_B. We observe a transition from volume-law to area-law TE scaling at the critical value $v_B^\star = 1$. Coarse-graining reduces the number of time steps, effectively rescaling $v_B \to v_B/n_{\mathrm{cg}}$ (see Appendix A 2 b), hence leading to a transition from volume-law to area-law TE at the critical value

$$n_{\mathrm{cg}}^\star = v_B. \tag{16}$$

This result in the dynamical version of the toy model leads to an intriguing proposition for the extended 1D quantum circuits it is supposed to mimic: that coarse-graining of their IMs (which reduces v_B too) may similarly lead to a transition in their temporal entanglement scaling. We shall explore this idea in Secs. IV and V.

C. Non-Markovianity and transmission of quantum information through the bath

Before exploring the effect of coarse-graining in extended 1D quantum circuits, we explore here within the toy model the relationship between (non-)Markovianity of the probe system's dynamics and its IM's TE—the latter quantity we remind the reader does not have a straightforward operational meaning. The aim of this subsection is to ground the expectation that TE indeed serves as a good proxy for the degree of correlations between the past and future. Again, we divide the temporal degrees of freedom into past $\mathcal{P}$ and future $\mathcal{F}$, as in Fig. 1(a), containing $2p$ and $2f$ qudits, respectively, such that

$$p + f = T. \tag{17}$$

Analogously to the previous subsection, we will assume that b and T, as well as p, f are large, and the ratio r is kept fixed.

Correlations between disjoint subsystems can be rigorously estimated using quantum information techniques, in particular decoupling theorems for noisy quantum channels [40]. Physically, the problem we consider is reminiscent of the problem of information retrieval from evaporating black holes [38,39], with the difference that here we are interested in information transfer from past to the future via the bath. At the technical level, we will look for an optimal temporal partition that maximizes distillable entanglement—that is, the number of shared Bell pairs—between $\mathcal{P}$ and $\mathcal{F}$. Below we provide an overview of the results along with intuitive explanations; details can be found in Appendix C.

As discussed in Sec. II A, it is convenient to divide the degrees of freedom in subsystem $\mathcal{P}$ into a single "reference" set $Q_{\mathcal{P}}$ of p qudits, and a single "system" set $S_{\mathcal{P}}$. The two sets are initially in a maximally entangled state. An analogous grouping into $Q_{\mathcal{F}}$ and $S_{\mathcal{F}}$ is defined for

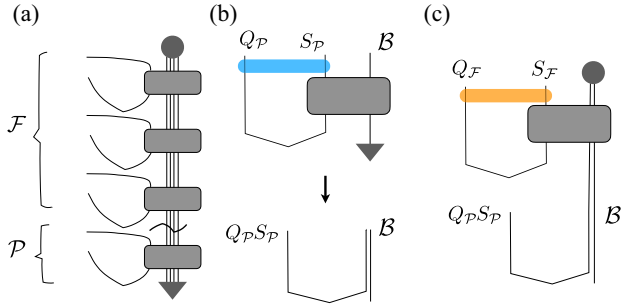


FIG. 6. Illustration of the estimate of distillable entanglement. (a) We divide the temporal system into past $\mathcal{P}$ and future $\mathcal{F}$ subsystems and think of evolution as consisting of two steps. (b) Top: in the first step, the bath gets entangled with the $\mathcal{P}$ qubits. Bottom: there exists a distillation protocol, acting only on the $\mathcal{P}$ degrees of freedom (blue box), which distills $N_{\mathcal{BP}}$ Bell pairs, provided the bath is sufficiently large, $b > 2p$. (c) In the second step, the Bell-pair partners in $\mathcal{B}$ get entangled with $\mathcal{F}$ qubits. As explained in the text, provided $\mathcal{F}$ is large enough, this allows to distill Bell pairs between $\mathcal{P}$ and $\mathcal{F}$ via a distillation protocol acting on $\mathcal{F}$ only (orange box).

subsystem $\mathcal{F}$, as illustrated in Fig. 6. Our goal is to estimate the number of Bell pairs $N_{\mathcal{PF}}$ shared between $\mathcal{P}$ and $\mathcal{F}$. In order to do so, we break down our problem into two steps: (i) First, $S_{\mathcal{P}}$ gets entangled with the bath $\mathcal{B}$ by a sequence of random unitaries, and (ii) $\mathcal{B}$ gets entangled with $S_{\mathcal{F}}$ by another sequence of random unitaries. Here we will focus on the case where the bath is initially in a pure state; the case of a mixed initial state is briefly considered in Appendix C.

Our strategy is to first estimate the number $N_{\mathcal{BP}}$ of shared Bell pairs between $\mathcal{B}$ and $\mathcal{P}$ and then estimate how much of that information can be retrieved by $\mathcal{F}$. The basic piece of intuition that guides us, adapted from Refs. [39,53], is that the large random unitaries acting on bipartite unentangled states yield random pure states whose bipartite entanglement is close to maximal, leaving the smaller subsystem in a maximally mixed state containing no information. Here, we adapt this intuition to the multi-time process that generates an IM.

In the first step, the bath gets entangled with $\mathcal{P}$, and the number of shared Bell pairs is determined by the smaller Hilbert-space dimension: $N_{\mathcal{BP}} = \min(b, 2p)$. In the second step, $\mathcal{F}$ is entangled with the $\mathcal{PB}$ system that can now be thought of as $N_{\mathcal{BP}}$ shared Bell pairs and the remaining $\max(b-2p, 0)$ bath qudits, which are in a pure state.

In the subsequent step, as in the black-hole information retrieval problem, the information about the state of $\mathcal{P}$ is contained mostly in $\mathcal{F}$, provided the size of $\mathcal{F}$ is large enough. More precisely, for $b > 2p$, information can be retrieved by $\mathcal{F}$ provided $2f > b + 2p$ (this condition simply states that the size of $\mathcal{F}$ exceeds the size of $\mathcal{B}$ and $\mathcal{P}$ combined); since $f = T - p$, this implies an inequality $4p < 2T - b$, which can only be satisfied for

$r = T/b > 1/2$. The optimal choice of p, yielding a maximum transfer of information from past to future, is given by $p_{\text{opt}} = \frac{2T-b}{4}$, and therefore we can transmit

$$N_{\mathcal{PF}} = 2p_{\text{opt}} = \frac{2T-b}{2} = \frac{(2r-1)b}{2} \tag{18}$$

qudits from the past to the future. This expression holds as long as $N_{\mathcal{PF}}$ does not exceed b: indeed, at most b Bell pairs can be transmitted via the bath. This is saturated at $T = 3b/2$, or, equivalently, $r = 3/2$.

To sum up, the number of Bell pairs between $\mathcal{P}$ and $\mathcal{F}$, that is, the distillable entanglement, is lower bounded by

$$N_{\mathcal{PF}} \gtrsim \begin{cases} 0, & \text{for } r < 1/2, \\ \frac{(2r-1)b}{2}, & \text{for } 1/2 < r < 3/2, \\ b, & \text{for } r > 3/2. \end{cases} \tag{19}$$

Comparing this result with the estimate of TE from Eq. (10) and Fig. 5, we conclude that, although the bound suggests a slope for the number of shared Bell pairs that is smaller by a factor of 2, both temporal and distillable entanglement begin to grow *concomittantly* at $r = 1/2$. This suggests that, although TE does not explicitly measure any physical quantity, it still provides a useful measure diagnosing both distillable entanglement and quantum non-Markovianity.

Now, the fact that past and future temporal subsystems share Bell pairs indicates the existence of strong quantum correlations. At the same time however, as pointed out in Ref. [15], such correlations are highly non-local (involving observables on many degrees of freedom in $\mathcal{P}$ and $\mathcal{F}$ at the same time) and detecting them would generally require measuring complex, multi-time operators. As a result, the impact of these correlations on *few-point* temporal correlators of local observables is expected to be negligible. This expectation is indeed consistent with the coarse-graining procedure described previously, which brings the IM close to the perfect depolarizer fixed-point, and suggests that coarse-graining is a useful tool to eliminate complicated (and hence physically irrelevant) correlators and hence effectively reduce the complexity of the IM, as we will explore next.

We close this subsection by outlining the technical steps leading to the result (19). The decoupling theorems in Ref. [40], colloquially, state that Bell pairs can be transmitted by a quantum channel, provided the state of the reference system (in the standard quantum information terminology) is decoupled from environment, where the quality of the decoupling is measured by the trace distance from a product state. We adapt those results to our setting as follows:

(i) The first step, as discussed above, is to show that $\mathcal{P}$ forms $\min(b, 2p)$ Bell pairs with the bath $\mathcal{B}$. This can be demonstrated by bounding the trace distance between the reduced density matrix of the smaller of subsystems $\mathcal{P}$, $\mathcal{B}$ and the maximally mixed state. The calculation proceeds by first computing the Haar-averaged Frobenius distance using Weingarten calculus and then applying the following inequality to relate it to the trace distance:

$$\|\rho\|_1 \leq \sqrt{\text{Dim}(\rho)}\|\rho\|_2, \tag{20}$$

where $\text{Dim}(\rho)$ is the dimension of the Hilbert space on which ρ acts.

(ii) When this distance is smaller than ϵ, Ulhman's theorem imples that the subsystems $\mathcal{P}$ and $\mathcal{B}$ are in a maximally entangled state with fidelity $F \geq 1 - \epsilon$.

(iii) Finally, to demonstrate that the information contained in the $\mathcal{BP}$ system is transmitted to system $\mathcal{F}$, we use the decoupling lemma from Ref. [40], which states that a sufficient condition for distilling Bell pairs between $\mathcal{P}, \mathcal{F}$ is provided by the decoupling of the degrees of freedom in $\mathcal{P}$ and the "outgoing" bath degrees of freedom following the interaction with $\mathcal{F}$. The decoupling is demonstrated by estimating the trace distance between the bipartite state on $\mathcal{BP}$ (obtained by tracing out $\mathcal{F}$) and the maximally mixed product state, with the same strategy as above.

For $r < 1/2$, step (i) above suffices to obtain the bound in Eq. (13). For $r > 1/2$, we show that at least $N_{\mathcal{PF}} - s$ Bell pairs are shared between $\mathcal{F}$ and $\mathcal{P}$ with fidelity $|1 - F| < d^{-s}$; thus, in the limit where both the bath $\mathcal{B}$ and the probe system $\mathcal{PF}$ are large, Eq. (19) holds asymptotically. When $r > \frac{3}{2}$, the maximum possible number of qudits, b, is shared between $\mathcal{F}$ and $\mathcal{P}$ through $\mathcal{B}$.

Summarizing this subsection, for a structureless random unitary bath model we showed that extensive TE does signal extensive distillable entanglement between the probe system's past and future, and hence genuine non-Markovian memory [54]. Yet, the coarse-graining argument indicates that arbitrary temporal correlation functions with up to $n^*_{\text{cg}}T$ operator insertions are extremely small, as indicated by the bounds in Eq. (13). Thus, the extensive IM entanglement of the random unitary bath in the regime $r > 1/2$ stems entirely from highly complex multi-time temporal correlations involving more than $n^*_{\text{cg}}T$ interrogations of the probe.

We emphasize that the fixed-point Markovian description applies—with exponentially small corrections—to a large family of temporal correlators that are still highly complex, associated with an extensive number $N = nT$ of operator insertions, with $0 \leq n < n^\star_{\text{cg}}$, all the way down to few-point correlators with $N = \mathcal{O}(1)$.

IV. TEMPORAL ENTANGLEMENT TRANSITION IN DUAL-UNITARY CIRCUITS

Equipped with the intuition of how coarse-graining can dramatically change the complexity structure of the IM in the random unitary bath models above, next we focus on coarse-graining's effect on the TE properties of IMs of 1D Floquet circuits. In this section, we first study chaotic dual-unitary circuits, whose special property of "dual-unitarity" allow for a number of analytical results [3,4,7,8,46,55] despite these systems being generically non-integrable and chaotic. Dual-unitarity can be colloquially described as the property that the "space-time rotated" version of the brickwork circuit (i.e., when viewed along the spatial direction, rather than the time direction) also describes unitary evolution [2–4]. In particular, the key results involving dual-unitary circuits relevant for us is that for a family of solvable initial states of the bath [46], their IMs can be shown to take the perfect depolarizer form of Eq. (12) (or perfect dephaser depending on the particular model considered), signifying that such a bath leads to perfectly Markovian dynamics [12]. Further, Ref. [36] analyzed TE for dual-unitary circuits with generic, non-solvable initial states and proved that von Neumann TE exhibits volume-law scaling with T, whereas all higher Rényi TE exhibit area-law scaling.

Here, we analytically demonstrate the occurrence of a TE transition upon coarse-graining dual-unitary baths. For definiteness, we will consider a specific circuit family, the kicked Ising model. The unitary dynamics U that defines the bath's influence matrix is given by the Floquet operator

$$U = \prod_{j\geq 0} P_{j+1/2} \prod_{j\geq 1} W_j \tag{21}$$

$$P_{j+1/2} = e^{iJ\sigma_j^z \sigma_{j+1}^z}, \qquad W_j = e^{ih\sigma_j^z} e^{ig\sigma_j^x}, \tag{22}$$

where $\sigma_j^{x,y,z}$ are Pauli matrices acting on site $j = 0, 1, 2, \ldots$ of a semi-infinite qubit chain. For $h = 0$, the kicked Ising model dynamics is mappable to a free Majorana chain, while for $h \neq 0$ it is expected to be chaotic [56]. The dual-unitary property is realized for $|J| = |g| = \pi/4$ or $3\pi/4$ and arbitrary h.

The probe system is chosen to be the boundary qubit at $j = 0$, while all other qubits at $j \geq 1$ constitute the bath. As discussed in Sec. II A, due to the product-operator structure of qubit-qubit interactions in the kicked Ising model, its IM takes a diagonal form with respect to input-output degrees of freedom [see Eq. (4)]. Defining the reduced Floquet operator

$$U_s = e^{iJs\tilde{\sigma}_1^z} \prod_{j\geq 1} P_{j+1/2} \prod_{j\geq 1} W_j, \tag{23}$$

conditioned on the state $s = \pm 1$ of the probe qubit at $j = 0$, the IM is defined by Eq. (5).The IM elements have a transparent physical interpretation as generalized Loschmidt echoes, given by two different time-dependent boundary magnetic fields in the forward-in-time and backward-in-time evolutions [12].

The initial states considered are simple product pure states

$$\rho_{\mathcal{B}}^0 = |\psi_0\rangle\langle\psi_0| = \otimes_{j=1}^{\infty} \rho_j^0. \tag{24}$$

For definiteness, we take all the bath qubits pointing along the direction $(\cos\phi, \sin\phi, 0)$, i.e.,

$$\rho_j^0 = \frac{1}{2}\begin{pmatrix} 1 & e^{i\phi} \\ e^{-i\phi} & 1 \end{pmatrix}. \tag{25}$$

At dual-unitary points of the kicked Ising model, this initial state is solvable for $\phi = 0, \pi$, whereupon its IM is provided by a perfect dephaser solution from Eq. (8). For parameters $\phi \neq 0, \pi$, the initial state is non-solvable; this will be the focus of our subsequent analysis.

A. IM structure and entanglement

In this section, we review the argument from Ref. [36] for the linear scaling of TE and adapt it to the case of kicked Ising model. The discussion is technical, but it is important to isolate the mechanism responsible for the growth of TE and provides the key ingredients needed to determine how temporal coarse-graining modifies the structure of the influence matrix. We present only the steps required for this purpose and refer the reader to Ref. [36] for the full derivation.

First, we decompose the IM in terms of IMs associated with shorter evolution time windows. To obtain this decomposition, we first note that the components of IM which are diagonal at the last time step simplify as

$$\delta_{s_T,\bar{s}_T} \mathcal{I}^{\bar{s}_1\cdots\bar{s}_{T-1},\bar{s}_T}_{s_1\cdots s_{T-1},s_T} = \delta_{s_T,\bar{s}_T} \mathcal{I}^{\bar{s}_1\cdots\bar{s}_{T-1}}_{s_1\cdots s_{T-1}} \tag{26}$$

(no summation implied). This property, which follows straightforwardly from Eq. (5), implies that the last time step's diagonal component of the IM $\mathcal{I}_T$ for evolution time T factorizes as tensor product of $\mathcal{I}_{T-1}$ and a single-qubit perfect dephaser. Using this observation and recursively decomposing the IM into its diagonal and off-diagonal components at the final time step, we obtain the following decomposition:

$$\mathcal{I}_T = \sum_{k=0}^{T} \tilde{\mathcal{I}}_{T-k} \otimes \mathcal{I}_{k,\mathrm{deph}}, \qquad \tilde{\mathcal{I}}_{T-k} = \pi_{T-k}\mathcal{I}_{T-k}. \tag{27}$$

In this equation we have defined the single-qubit projector π on the off-diagonal components $s \neq \bar{s}$, i.e.,

$$(\pi_\tau \mathcal{I}_T)^{\bar{s}_1\cdots\bar{s}_T}_{s_1\cdots s_T} = \delta_{s_\tau,-\bar{s}_\tau} (\mathcal{I}_T)^{\bar{s}_1\cdots\bar{s}_T}_{s_1\cdots s_T}. \tag{28}$$

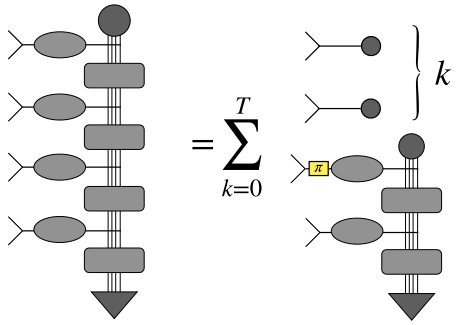


FIG. 7. Due to the recursive structure of the IM, as expressed by Eq. (26), an IM at any time T necessarily contains IMs corresponding to all shorter times. By recursively inserting the resolution of the identity $\mathbb{1}_{s,\bar{s}}^{s',\bar{s}'} = \delta_{s',\bar{s}'}\delta_{s,\bar{s}} + \pi_{s,\bar{s}}^{s',\bar{s}'}$, we obtain the decomposition shown here.

The kth term in the decomposition (27) is the tensor product of two factors: the perfect-dephaser IM, cf. Eq. (8), acting over the final k time steps, and the bath IM for the shorter evolution time $T-k$, projected at its final time step onto the off-diagonal subspace $s_{T-k} \neq \bar{s}_{T-k}$, as illustrated in Fig. 7.

We now analyze TE scaling of the IM viewed as a 2^{2T}-dimensional vector $|\mathcal{I}_T\rangle$. We define the TE and the "reduced density matrix" according to Eq. (6).

First, using the orthogonality of the vectors $|\tilde{\mathcal{I}}_{T-k}\rangle$, we can write

$$\rho_{\mathcal{P}} = w\frac{|\mathcal{I}_p\rangle\langle\mathcal{I}_p|}{\langle\mathcal{I}_p|\mathcal{I}_p\rangle} + \sum_{k=p+1}^{T} w_k\tilde{\rho}_k, \tag{29}$$

where $\tilde{\rho}_k \propto \mathrm{Tr}_{\mathcal{F}}(|\tilde{\mathcal{I}}_k\rangle\langle\tilde{\mathcal{I}}_k|)$ are normalized density matrices with unit trace. The weights w_k and w are defined consistently with the normalization requirements of each term: Defining the squared norm of each vector

$$N_k^2 = \langle\tilde{\mathcal{I}}_k|\tilde{\mathcal{I}}_k\rangle\langle\mathcal{I}_{\mathrm{deph,T-k}}|\mathcal{I}_{\mathrm{deph,T-k}}\rangle = \langle\tilde{\mathcal{I}}_k|\tilde{\mathcal{I}}_k\rangle 2^{T-k}, \tag{30}$$

the weights can be expressed as

$$w_k = \frac{N_k^2}{\sum_{k=0}^{T} N_k^2}, \qquad w = \sum_{k=1}^{p} w_k. \tag{31}$$

Evaluating the weights in the decomposition (29) plays a central role in the computation of TE. It is natural to expect that the matrix elements of $\tilde{\mathcal{I}}_k$ are, effectively, exponentially small (in k) random numbers. Indeed, to gain an intuitive understanding of these scalings, we use the Ising-specific structure of the IM and interpret its matrix elements as generalized Loschmidt echoes following Ref. [12]. Namely, the matrix elements correspond to a forward-in-time evolution subject to the time-dependent boundary field $h_\tau = h + s_\tau J$, followed by a backward-in-time evolution subject to the field $\bar{h}_\tau = h + \bar{s}_\tau J$, with $J = \pi/4$ in the present case. This can be seen directly by specializing Eq. (5) to a pure initial state:

$$\mathcal{I}_{s_1\cdots s_T}^{\bar{s}_1\cdots\bar{s}_T} = \langle\psi|U_{\bar{s}_1}^\dagger\cdots U_{\bar{s}_T}^\dagger U_{s_T}\cdots U_{s_1}|\psi\rangle. \tag{32}$$

Combined with a rapidly entangling evolution in chaotic dual-unitary circuits, this suggests that the generalized Loschmidt echo is, in effect, an overlap of two random vectors $v_1 = U_{s_T}\cdots U_{s_1}|\psi\rangle$ and $v_2 = U_{\bar{s}_T}\cdots U_{\bar{s}_1}|\psi\rangle$ in the bath portion that is causally connected with the boundary. This effective bath has Hilbert-space dimension 2^k, and hence the typical overlap is expected to be of order $2^{-k/2}$ in magnitude

$$\|\mathcal{I}_{s_1\cdots s_{T-1},s_T}^{\bar{s}_1\cdots\bar{s}_{T-1},\bar{s}_T}\| \sim 2^{-k/2}. \tag{33}$$

This expectation is consistent with the conclusions of Ref. [36] for generic dual-unitary circuits, and we verified it numerically, cf. Fig. 10.

We are now in a position to estimate the norm of $\tilde{\mathcal{I}}_k$ as

$$\langle\tilde{\mathcal{I}}_k|\tilde{\mathcal{I}}_k\rangle \approx C2^k, \tag{34}$$

where C is a constant that depends on the initial state (in particular, for solvable initial states, see Ref. [46], corresponding to a perfect dephaser solution, we have $C = 0$). This yields

$$N_k^2 \approx C2^T \quad \text{for } k \neq T, \tag{35}$$

$$N_0^2 = 2^T, \tag{36}$$

where $N_0^2 = 2^T$ is the contribution of the perfect dephaser part in the IM decomposition. We thus see that all the $T+1$ orthogonal vectors appearing in the IM decomposition have comparable norms. The overlap with the non-entangled perfect dephaser term ($k = T$), in particular, is responsible for the area-law higher Rényi entropies [36].

As the last ingredient, we note that the von Neumann entanglement entropy of a statistical mixture of states may be upper and lower bounded as follows [36]:

$$S_{\mathrm{mix}} \leq S_{\mathrm{TE}}(\rho_{\mathcal{P}}) \leq S_{\mathrm{mix}} + S_{\mathrm{cl}}, \tag{37}$$

where we introduced the mixture of entanglement entropies S_{mix} and the classical mixing entropy S_{cl}, defined as

$$S_{\mathrm{mix}} = \sum_{k=p+1}^{T} w_k S_{\mathrm{TE}}(\tilde{\rho}_k), \tag{38}$$

$$S_{\mathrm{cl}} = -\sum_{k=p+1}^{T} w_k \log_2 w_k - w\log_2 w, \tag{39}$$

respectively. In Eq. (37), the left-hand side inequality is a direct consequence of the concavity of the entanglement, and the right-hand side one follows from the property $\chi = S_{\mathrm{TE}}(\rho_{\mathcal{P}}) - S_{\mathrm{mix}} \leq S_{\mathrm{cl}}$ of Holevo's information (see for example Chapter 10.6 of Ref. [57]).

As a crucial step, Ref. [36] showed that the second Rényi entropy $S^{(2)}(\tilde{\rho}_k)$ takes the maximal possible entropy value, which constrains the von Neumann entropy to take the same value:

$$S_{\mathrm{TE}}(\tilde{\rho}_k) = S^{(2)}(\tilde{\rho}_k) = \min\left(\frac{k-p}{2}, p\right), \tag{40}$$

where p denotes the number of past time steps before the cut, as defined above Eq. (17). This leads to the extensive entropy scaling

$$S_{\mathrm{mix}} = \begin{cases} \frac{(T-p)(T-p+1)}{4(TC+1)} & p > \frac{1}{3}T, \\ \frac{p(2T-4p+1)}{2(TC+1)} & p < \frac{1}{3}T. \end{cases} \tag{41}$$

Finally, since the classical entropy is seen to scale subextensively (logarithmically) for long evolution times,

$$S_{\mathrm{cl}} \approx \frac{pC+1}{TC+1}\log_2\frac{TC+1}{pc+1} + \frac{f}{1+TC}\log_2(TC+1), \tag{42}$$

then TE asymptotically equals the result (41), i.e., $S_{\mathrm{TE}} \sim S_{\mathrm{mix}}$ for large T.

In Fig. 8 we report a numerically exact computation of the von Neumann TE as a function of p and T, obtained by constructing a truncation-free MPS representation of the IM using the light-cone growth algorithm (LCGA). This algorithm iteratively constructs the IM adding one time step per iteration. At each evolution time, the corresponding IM includes only the portion of the bath contained within the probe's past causal light cone. Thus, although the effective bath is finite at any fixed time, its size grows with the maximal evolution time, allowing the dynamics to be simulated directly in the thermodynamic limit. We refer the reader to Refs. [23,24], where this algorithm was first introduced. We observe volume-law scaling with T, as well as an asymmetric profile of TE as a function of $0 \leq p \leq T$ at fixed evolution time T, reported in the inset. Both properties are in agreement with the analytical estimates above.

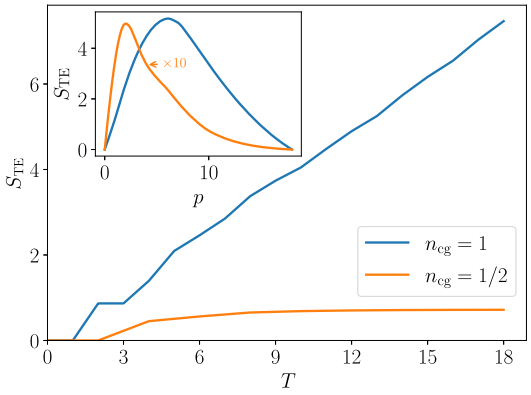


FIG. 8. Von Neumann TE before and after coarse-graining with density $n_{\mathrm{cg}} = 1/2$, for the dual-unitary kicked Ising model bath with $h = 0.5$ and initial pure state described in Eqs. (24) and (25) with tilt angle $\phi = 0.7$.

B. Coarse-graining and TE transition

We now turn to our central question: the effect of temporal coarse-graining on the IM. For simplicity, here we choose the simplest coarse-graining unitary $V_\tau = \mathbb{1}$; we expect, however, that our conclusions do not depend on the particular choice of V_τ. Under coarse-graining, the decomposition in Eq. (29) becomes

$$\rho^{\mathrm{cg}}_{\mathcal{P}} = w^{\mathrm{cg}} \frac{|\mathcal{I}_{p,cg}\rangle\langle\mathcal{I}_{p,\mathrm{cg}}|}{\langle\mathcal{I}_{p,cg}|\mathcal{I}_{p,cg}\rangle} + \sum_{k=n_{\mathrm{cg}}(p+1)}^{n_{\mathrm{cg}}T} \omega^{\mathrm{cg}}_{k/n_{\mathrm{cg}}} \tilde{\rho}_{k/n_{\mathrm{cg}}}, \tag{43}$$

where we assumed that any non-integer index is rounded to its closer integer neighbor; in the long-time limit, this is expected to introduce a small relative error suppressed as $1/T$. Crucially, different terms in the above equation "flow" differently under coarse-graining. As we discuss below, this implies that under an arbitrarily mild coarse-graining, TE collapses to an area-law scaling. This observation agrees with the heuristic guess in Eq. (16), which implies $n^{\star}_{\mathrm{cg}} = 1$ for systems with maximal butterfly velocity $v_B = 1$.

Let us first estimate the norms of the coarse-grained IM components. Our previous estimate of the magnitude of each matrix element of the IM still applies—in fact, matrix elements of the coarse-grained IM are a subset of matrix elements of the original IM. However, the total number of matrix elements has decreased from 2^{2T} to $2^{2n_{\mathrm{cg}}T}$. Thus, for the norms, we obtain [cf. Eq. (34) and also a diagrammatic comparison from Fig. 9]:

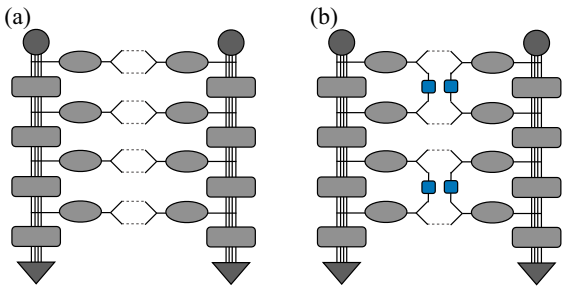


FIG. 9. Comparison between the (a) norm of IM from Eq. (34) and the (b) norm of coarse-grained IM from Eq. (45). The insertion of each vertical contraction (coarse-graining step) reduces the norm by the factor of 2^{-2}.

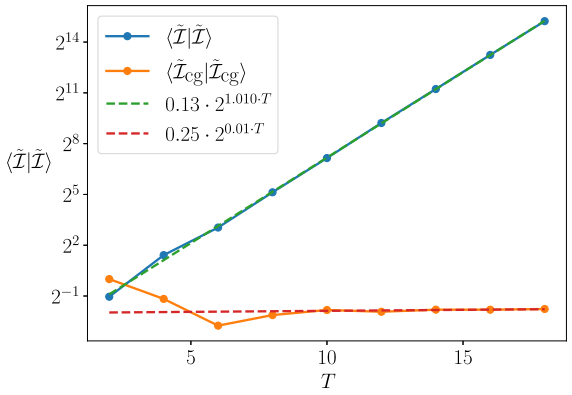


FIG. 10. Numerical calculation of the norm of the IM at dual-unitary kicked Ising model with $h = 0.5$ and initial tilt angle $\phi = 0.7$, before and after coarse-graining with density $n_{\rm cg} = 1/2$. The constants $C \simeq 0.13$ and $C_{\rm cg} = 0.25$ are not universal and depend both on ϕ and h.

$$\langle \mathcal{I}^{\rm cg}_{T-k,\rm deph}|\mathcal{I}^{\rm cg}_{T-k,\rm deph}\rangle = 2^{n_{\rm cg}(T-k)}, \tag{44}$$

$$\langle \tilde{\mathcal{I}}^{cg}_{k}|\tilde{\mathcal{I}}^{cg}_{k}\rangle \approx C_{\rm cg}2^{(2n_{\rm cg}-1)k}. \tag{45}$$

While Eq. (44) is exact, Eq. (45) is based on the heuristic Loschmidt-echo estimation of matrix elements from Eq. (33). To make sure that the norms $\langle\tilde{\mathcal{I}}^{cg}_{k}|\tilde{\mathcal{I}}^{cg}_{k}\rangle$ are not affected by possible neglected correlations between IM elements, we additionally verified Eq. (45) numerically, as reported in Fig. 10. The coarse-grained IM decomposition weights thus become

$$w^{\rm cg}_{k} = \tilde{C}_{T,cg}2^{-(1-n_{\rm cg})k}, \tag{46}$$

$$w^{\rm cg} = 1 - \sum_{k=n_{\rm cg}(p+1)}^{n_{\rm cg}T} w^{\rm cg}_{k/n_{\rm cg}}, \tag{47}$$

$$\tilde{C}_{T,cg} = \left(C_{cg}\frac{1-2^{-(1-n_{\rm cg})T}}{2^{\frac{1-n_{\rm cg}}{n_{\rm cg}}}-1}+1\right)^{-1}. \tag{48}$$

Crucially, from Eqs. (44) and (45), we see that the norm of the perfect dephaser components in the IM vector decomposition are suppressed exponentially *slower* than the orthogonal "random" components. Thus, coarse-graining remodulates the decomposition weights in favor of the perfect dephaser components. Such a remodulation is accompanied by a decrease in TE.

In order to prove the saturation of TE at a finite coarse-graining density, it is sufficient to use the right-hand side inequality from Eq. (38). We will separately bound the mixture of entanglement entropies and the classical entropy of the mixture. The former is given by

$$S_{\rm mix} = \sum_{k=n_{\rm cg}(p+1)}^{n_{\rm cg}T} w^{\rm cg}_{k/n_{\rm cg}} S_{\rm TE}(\tilde{\rho}^{\rm cg}_{k/n_{\rm cg}}). \tag{49}$$

Due to the exponential suppression of the weights in Eq. (46), it is sufficient to bound each term by the maximal possible entanglement entropy, that is, the number of qubits $2n_{\rm cg}p$,

$$S_{\rm TE}(\tilde{\rho}^{\rm cg}_{k}) < 2n_{\rm cg}p. \tag{50}$$

The sum in Eq. (49) then can be bounded as

$$S_{\rm mix} < p2^{-p(1-n_{\rm cg})}C_{\rm mix}, \tag{51}$$

where $C_{\rm mix} = \frac{2\log(2)n_{\rm cg}}{1-2^{-(1-n_{\rm cg})/n_{\rm cg}}}\tilde{C}_{T,\rm cg}$. Thus, the quantity $S_{\rm mix}$ is finite for arbitrarily large p and T, as soon as $n_{\rm cg} < 1$. In addition, the classical entropy from Eq. (37) is given by the sum

$$S^{\rm cg}_{\rm cl} = -\sum_{k=n_{\rm cg}(p+1)}^{n_{\rm cg}T} w^{\rm cg}_{k/n_{\rm cg}}\log_2 w^{\rm cg}_{k/n_{\rm cg}} - w^{\rm cg}\log w^{\rm cg}, \tag{52}$$

which is also convergent for large T, again thanks to the exponential decay of the coefficients $w^{\rm cg}_{k}$.

Putting everything together, we have shown that $S_{\rm TE}(\rho^{\rm cg}_{\mathcal{P}})$ stays finite for arbitrarily large T as soon as $n_{\rm cg} < 1$. This corresponds to the claimed transition to area-law scaling in the presence of an arbitrarily mild coarse-graining. This TE transition is illustrated in Fig. 8 for $n_{\rm cg} = 1/2$; we further verified this occurrence for other choices of $n_{\rm cg}$ (not shown).

Let us now discuss the consequences of our result for approximating the IM. Since the kicked Ising model IM is in diagonal form [see Eq. (4)] and we assumed a diagonal unitary evolution $V_\tau = \mathbb{1}$ for the probe spin, non-trivial correlation functions must involve off-diagonal observables such as $O = \sigma^x$ or σ^y. Two-point autocorrelation functions, for example, can be written as

$$\langle O(t_2)O(t_1)\rangle = \frac{1}{2}\delta_{t_1,t_2}\,\mathrm{Tr}(O^2) + R(t_1,t_2), \tag{53}$$

where the first term is determined by the perfect depolarizer component of the IM, whereas the remainder, $R(t_1,t_2)$, is a strongly fluctuating function arising from the highly entangled component $\tilde{\mathcal{I}}_{t_2}$.

Noting that two-point functions are obtained by coarse-graining all sites except the two insertions, and using Eq. (45) with $n_{\rm cg} = 0$, we estimate the magnitude of this term as $|R(t_1,t_2)| \lesssim 2^{-t_2/2}$ for $t_2 > t_1$, uniformly in t_1. Now, let us turn to the MPS approximation that allows us to compute the correlation function up to a fixed error ϵ. We may truncate $R(t_1,t_2)$ by setting it to zero for

$t_2 > 2\log_2(1/\epsilon) = t_2^{\max}$, and the corresponding IM can then be naturally represented as an MPS with bond dimension $2^{t_2^{\max}}$, yielding a polynomial relation between the absolute error and the required MPS bond dimension. Owing to the high entanglement of $\tilde{\mathcal{I}}_{t_2}$ [cf. Eq. (40)], we do not expect further compression to be compatible with the allowed error.

V. GENERIC FLOQUET CIRCUITS

In this section, we study TE in a more generic class of non-integrable Floquet circuits: the kicked Ising model away from dual-unitary points, which is considered a paradigmatic model of quantum chaos [16,56]. It is known, however, to provide an example of non-trivial thermalization dynamics of local observables. Specifically, its dynamics in an open chain (or in a semi-infinite chain, as considered here) exhibits four distinct regimes, which stem from the existence of prethermal 0- and π-edge modes [45,58,59]. Such modes are approximate integrals of motion localized at the boundary and give rise to parametrically slower, compared to a random or dual-unitary circuit, relaxation dynamics of a probe system coupled at the edge.

Below we study TE in three regimes: (I) a "plain vanilla" regime without edge modes, (II) a regime with an edge π-mode, and (III) a regime with an edge 0-mode. [We do not consider the regime where 0- and π-modes coexist, since it shows a similar phenomenology to (II).] We compute TE by numerically constructing the IM via the LCGA. Consistent with previous work and our analysis above, we generally find volume-law scaling of TE in all three regimes. As expected, upon coarse-graining, TE gets significantly reduced. For all parameter values we consider, TE transitions to a subvolume-law scaling below a finite critical coarse-graining threshold, consistent with the scenario outlined above.

The coarse-graining transition further suggests that few-point correlation functions should be captured by a compact MPS approximation to the influence matrix, consistent with the subvolume growth of TE. Motivated by this observation, we study the dynamics of local observables and quantify the errors introduced by replacing the full influence matrix—whose exact bond dimension grows exponentially with the evolution time—by a lower-bond-dimension approximation obtained through standard singular-value truncation.

A. TE across dynamical regimes

We consider a fully mixed initial state of the kicked Ising model bath. In the integrable limit $h = 0$ the kicked Ising model can be mapped onto a kicked Kitaev (free Majorana) chain [60], whose edge modes are exact integrals of motion. The phase diagram of the non-interacting kicked Ising model [17,61] is reported in Fig. 11. On breaking integrability, $h \neq 0$, the edge modes acquire a finite lifetime, which varies parametrically from infinity on approaching the top and bottom sides of the phase diagram to zero at the central point $g = J = \pi/4$ (dual-unitary point), qualifying the edge modes as prethermal [45]. Notably, for sizable fields h, the parameter region where prethermal edge modes are long-lived may differ somewhat compared to the non-interacting phase boundaries. We distinguish three distinct regimes, which are deformations of the regimes (I), (II), and (III) from Fig. 11, respectively. We choose the following representative parameter values in the three regimes:

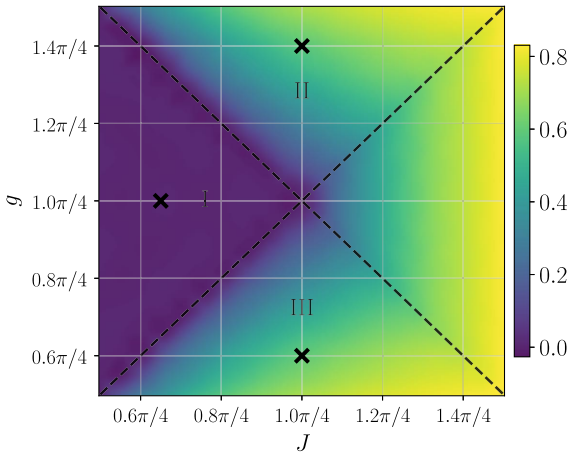


FIG. 11. Phase diagram of the integrable kicked Ising model ($h = 0$). Brighter color indicates a higher saturated value of the autocorrelator $\lim_{T\to\infty} |\langle \sigma_0^z(T)\sigma_0^z(0)\rangle|$, reflecting a more localized edge mode. The three crosses correspond to representative values of J and g chosen for numerical simulations.

$$\text{(I)}\ (g = \pi/4, J = 0.65\pi/4, h = 0.5), \qquad v_B = 0.61 \pm 0.05, \tag{54}$$

$$\text{(II)}\ (g = 1.4\pi/4, J = \pi/4, h = 0.5), \qquad v_B = 0.62 \pm 0.01, \tag{55}$$

$$\text{(III)}\ (g = 0.6\pi/4, J = \pi/4, h = 0.5), \qquad v_B = 0.62 \pm 0.02. \tag{56}$$

These three choices of J and g are marked by crosses in Fig. 11. For each of them, we numerically construct the IM via the LCGA up to times where TE is converged using a maximum bond dimension $\chi = 1024$. The three points have been chosen to have roughly equal values of the butterfly velocity v_B, which we computed independently by fitting the spatiotemporal spreading of the out-of-time-ordered correlator [6,62,63]

$$\mathcal{C}_{AB}(r,t) = \frac{1}{2}\,\mathrm{tr}([A_0(t), B_r(0)]^\dagger [A_0(t), B_r(0)]) \tag{57}$$

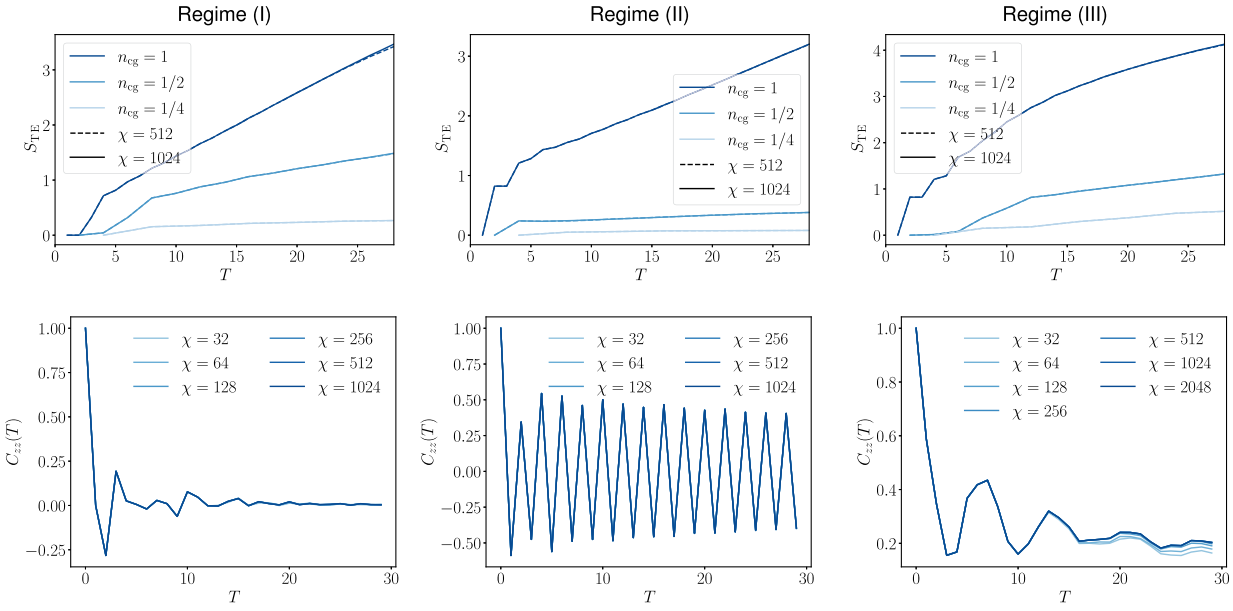


FIG. 12. Top panels: temporal entanglement entropy, computed using the LCGA, in different dynamical regimes of the nonintegrable KIM: (I) no edge modes, (II) edge π-mode, and (III) edge 0-mode. Bottom panels: The local temporal correlator $C_{zz}(T)$ in the three regimes shows qualitatively different behaviors. While it quickly decays to zero in the plain vanilla thermalization regime (I), the decay is slow in the other two regimes due to the overlap of the σ^z operator with the edge mode operator. In the main text we explain the observed quantitatively different behavior of coarse-grained TE in (II) and (III) and compare the convergence of observables with increasing χ.

averaged over Pauli operators $A, B \in \{\sigma^x, \sigma^y, \sigma^z\}$ using exact diagonalization for a chain of length $L = 12$ as explained in Appendix D.

The behavior of TE and the effect of coarse-graining is reported in the top panels of Fig. 12, and it follows the intuition developed above. In regime (I) observables decay to zero over a short timescale, similarly to random and dual-unitary circuits. Here, however, coarse-graining with density $n_{\mathrm{cg}} = 1/2$ reduces TE to a less steep but still volume-law scaling. A further round of coarse-graining, $n_{\mathrm{cg}} = 1/4$, finally yields a TE scaling consistent with a subvolume-law. This indicates that $1/4 < n^{\star}_{\mathrm{cg}} < 1/2$ for this parameter choice. The smaller threshold for the TE transition is expected given the reduced butterfly velocity $v_B \approx 0.61 \pm 0.05$ of this model compared to the dual-unitary value $v_B = 1$. The relation between the critical coarse-graining and v_B was examined in the toy model of a growing random unitary bath, which yielded Eq. (16); see also Appendix A 2 b. In a more structured bath model, such as a local unitary circuit, the simple dependence between v_B and the critical coarse-graining in Eq. (16) may give room to a more complicated relation; nevertheless, the basic picture of a competition between bath and temporal degrees of freedom remains valid and predicts a transition at a different $n^{\star}_{\mathrm{cg}}$ *at least* as small as v_B, in agreement with our numerical results. As discussed in Appendix E, tuning the model parameters closer to a dual-unitary point, $J = 0.8\pi/4$ with g and h fixed, where the butterfly velocity $v_B = 0.95 \pm 0.02$ is larger, we observe a clear collapse of TE to subvolume-law scaling after the first round of coarse-graining, which implies $1/2 < n^{\star}_{\mathrm{cg}} < 1$ for this parameter choice.

A more direct way to establish the existence of a transition is to vary n_{cg} continuously while keeping the model parameters g, h, and J fixed. For each value of n_{cg}, one can extract the linear coefficient v by fitting the TE to

$$S_{\mathrm{TE}} = vT + a \log T + b. \tag{58}$$

This ansatz is natural from the membrane perspective [64]: the linear term is associated with the membrane tension, while fluctuations of the membrane generate a subleading logarithmic correction. Its application to TE is, however, more subtle. As shown in Ref. [36] volume-law scaling can be established analytically at the dual-unitary point, whereas away from this point the membrane interpretation is less direct. Nevertheless, our numerical results indicate that the temporal entanglement is well described by Eq. (58).

Within this fitting procedure, the critical coarse-graining density $n^{\star}_{\mathrm{cg}}$ can then be identified as the value beyond which the fitted slope v vanishes. In practice, our ability to vary n_{cg} continously at a fixed evolution time T is limited due to the constraint that $n_{\mathrm{cg}}T$ is an integer number of time steps. It is more reliable to fix $n_{\mathrm{cg}} = 1/2$ and study v as a function of J. This allows us to determine the critical coupling $J^{\star}_{1/2}$ at which a coarse-graining density of $n_{\mathrm{cg}} = 1/2$ becomes sufficient to drive the IM from volume-law

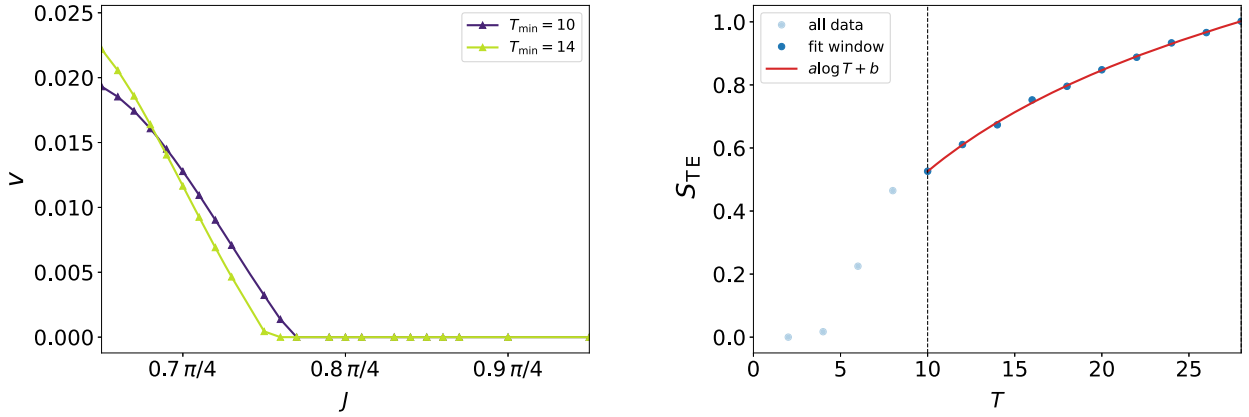


FIG. 13. (a) Linear fit coefficient as a function of J abruptly vanishes at $J^{\star}_{1/2} \simeq 0.77\frac{\pi}{4}$. The slope extraction was made for T in range $[T_{\min} = \{10, 14\}, T_{\max} = 28]$, (b) log fit for a critical $J^{\star}_{1/2}$ with $a \simeq 0.37, b \simeq -0.43$.

scaling into a subvolume-law, logarithmic regime. The corresponding results are shown on Fig. 13 and show a clear transition at $J^{\star}_{1/2} \simeq 0.77, \frac{\pi}{4}$.

Regimes (II) and (III) exhibit qualitatively similar phenomenology, but with striking quantitative differences, as shown in Fig. 12. In particular, in regime (II), even coarse-graining at $n_{cg} = 1/2$ is consistent with a subvolume-law scaling of the TE, whereas regime (III) exhibits a pronounced resilience of large TE. As we briefly discuss below, these differences can be qualitatively attributed to prethermal phenomena.

Next, we use the IM to compute the dynamics of a probe spin coupled to a half-chain kicked Ising bath. The dynamics is fully specified only once the boundary evolution of the probe spin is fixed. We choose the probe-spin kick to match the bulk kicks, setting $g_0 = g$ so that $V_\tau = e^{ig_0\sigma_x}$. This choice differs from the trivial boundary evolution, $V_\tau = \mathbb{1}$, used in defining the coarse-graining procedure. Although the same boundary protocol could in principle be used in both settings, we keep them distinct in order to probe more generic boundary dynamics. Nevertheless, as shown below, the resulting correlation functions exhibit behavior consistent with the temporal-entanglement evolution obtained under coarse-graining. Namely, the autocorrelator

$$C_{zz}(T) = \langle \sigma^z_0(T)\sigma^z_0(0)\rangle \tag{59}$$

behaves differently in the three regimes, see the bottom panels of Fig. 12. We note that in regime (I), $C_{zz}(T)$ exhibits irregular fluctuations around zero with exponentially decaying magnitude for $T > 0$, similar to the case of random and dual-unitary circuits. In contrast, $C_{zz}(T)$ is long-lived in regimes (II), (III), as can be expected from the sensitivity of the edge spin operator σ^z_0 to the prethermal edge mode [45]; we note that such slow decay for the edge π-mode was observed experimentally on a superconducting quantum processor [59]. The C_{xx} and C_{yy} correlators behaves similarly; we report the numerical results for C_{xx} in Appendix F.

The similarities in edge dynamics in regimes (II) and (III) (up to the presence or absence of the staggered sign-changing behavior) are not representative, however, of the markedly different bulk dynamics in the two regimes. This contrast, discussed below, leads us to a natural conjecture on the origin of the different behavior of TE under coarse-graining in the two regimes despite the similar v_B.

In regime (III) the Ising coupling J is larger than the transverse field g. Together with a symmetry-breaking longitudinal field this gives rise to domain-wall confinement [65–67]—a phenomenon that has been studied in detail in the Hamiltonian version of the model, but has a very similar phenomenology in the kicked Ising model [68]. In contrast with a symmetry-preserving integrability-breaking interaction, which conserves a prethermal edge mode but otherwise coexists with fast bulk dynamics, a symmetry-breaking confining field gives rise to slow dynamics uniformly across the system—a phenomenon that can be linked to an emergent strong Hilbert-space fragmentation in the limit of small transverse field [67,69]. It was shown that, throughout this regime, dynamics from certain highly excited states gives rise to long-lived oscillations of local observables, suppressed entanglement growth, and transient many-body localization of domain walls [67]. Recalling the discussion in Sec. III B, we note that, for the toy model, the critical coarse-graining density is proportional to the butterfly velocity. In a fragmented or quasi-fragmented Hilbert space, however, the dynamics may involve a hierarchy of approximately isolated subspaces, each characterized by its own effective growth rate, $D_k \sim d^{2v_k t}$. Rather than producing a single sharply defined coarse-graining threshold, this hierarchy may lead to a

broadened or multi-stage crossover, as different subspaces become progressively disentangled when n_{cg} becomes comparable to their respective effective rates v_k. Slowly thermalizing subspaces with small v_k, such as those associated with dynamics confined to a fixed domain-wall sector, may therefore keep being entangled at comparatively low coarse-graining densities.

This results in a longer thermalization scale, in particular linear fit with logarithmic correction from Eq. (58) suggests the domination of logarithmic contribution, signaling that the real thermodynamic behavior of TE might be observable at a much longer times only. This explains why even $n_{\mathrm{cg}} = 1/4$ does not eliminate the growth of TE, and why the convergence of observables with bond-dimension is slower in this regime. One intriguing aspect of our study is the unexpected robustness of prethermal physics in the kicked Ising model, even in a regime of strong integrability breaking. This observation motivates the investigation of less constrained models, such as spin-1 Ising models or brickwork quantum circuits, which we hope to pursue in future work.

In contrast, in regime (II) the kick can be viewed as a perturbation of the π-kick (realized for $g = \pi/2$), which perfectly flips all spins. Studies of prethermal discrete-time crystals in the kicked Ising model [68,70–73] have clarified that symmetry-breaking interactions such as the longitudinal field get "echoed out" at all orders in perturbation theory, effectively restoring the symmetry. Therefore, domain-wall confinement is fully suppressed dynamically, while integrability breaking persists in the form of effective symmetry-preserving interactions, giving rise to a long-lived prethermal edge mode coexisting with fast bulk dynamics. This argument explains why the resilience of high TE in regime (III) does not occur in regime (II), while the prethermal edge mode continues to affect edge temporal correlations over long timescales.

B. Local observables and IM compression

Finally, we turn to the important question regarding the possibility of using a compressed IM to approximate few-point local temporal correlators for general choices of probe-system dynamics V_τ. For concreteness, we again choose probe-spin kicks of equal strength to those in the bath, i.e., $g_0 = g$, and we will focus on the practical issue of determining the IM bond dimension necessary to approximate local autocorrelators of the kicked Ising model up to time T with either a given absolute or relative error. We quantify the latter via

$$\delta_\chi(T) = \min\left(\left|\frac{C^{\chi}_{zz}(T) - C^{\chi_{\max}}_{zz}(T)}{C^{\chi_{\max}}_{zz}(T)}\right|, 1\right), \tag{60}$$

which is truncated to a maximum value 1.

We first observe that the IM of the kicked Ising model bath with fully mixed initial state can be represented exactly by an MPS with bond dimension

$$\chi_{\mathrm{exact}} = 2^{T/2}. \tag{61}$$

We further expect that for $v_B < 1$, the exponential growth rate decreases accordingly, $\chi_{\mathrm{exact}} \approx 2^{v_B T/2}$.

Our analysis so far indicates that the volume-law TE scaling reflects highly complex temporal correlations corresponding to a high rate of interrogations of the probe system, and it has very little bearing on few-point temporal correlations. Coarse-graining provides a way to project out the IM components responsible for the complex correlations, giving rise to a new, subvolume-law IM. It is natural to expect that the "nonlocal" IM components are associated with the small Schmidt values of the IM vector. Then, we expect that truncation to the largest χ Schmidt values within LCGA performs an effective coarse-graining procedure, yielding a compressed IM $\mathcal{I}_\chi$ that retains accuracy on few-point temporal correlators while satisfying a subvolume-law scaling.

In Fig. 14 we report empirical tests of this expectation. In regime (I) local observables decay in around one Floquet period, subsequently displaying irregular fluctuations around zero, with exponentially decaying magnitude. Thus, if we set an absolute error threshold ϵ, we only need to approximate dynamics up to time $t(\epsilon) \propto |\log \epsilon|$. This corresponds to a polynomial scaling of the required bond dimension χ with ϵ. In general, the irregular fluctuations around zero of observables arise from complex operator growth process [74] and would require an exact description of the IM, so we do not expect further compression to be possible. Accordingly, we expect the *relative* error to rise rapidly to a value of order 1 as χ is reduced below χ_{exact}. We observed this behavior most clearly in the plain vanilla regime close to the dual-unitary point (see Appendix E). Slightly farther away from the dual-unitary point, for parameter as in Eq. (54), the absolute errors remain extremely small, and relative errors rise rapidly with T at fixed χ, as illustrated in Fig. 14.

In contrast, in regimes (II) and (III), an approximate localized integral of motion exists, leading to a slow decay of the autocorrelator. In this case, we expect that the slow regular decay $C_{zz}(T) \sim c\, 2^{-\gamma T}$ will be accompanied by irregular and rapidly decaying fluctuations similar to the regime (I); these fluctuations stem from the highly entangled components of the IMand would only lead to a small relative error for local correlators. We indeed found that in both regimes the IM's bond dimension can be significantly reduced while keeping high accuracy on the slow exponential decay. The relative and absolute errors for the autocorrelator computed for different values of χ are reported in Fig. 14. We conclude that, in line with our expectations, for chaotic baths naïve Schmidt value truncation projects out the IM components responsible for complex nonlocal temporal correlations and volume-law

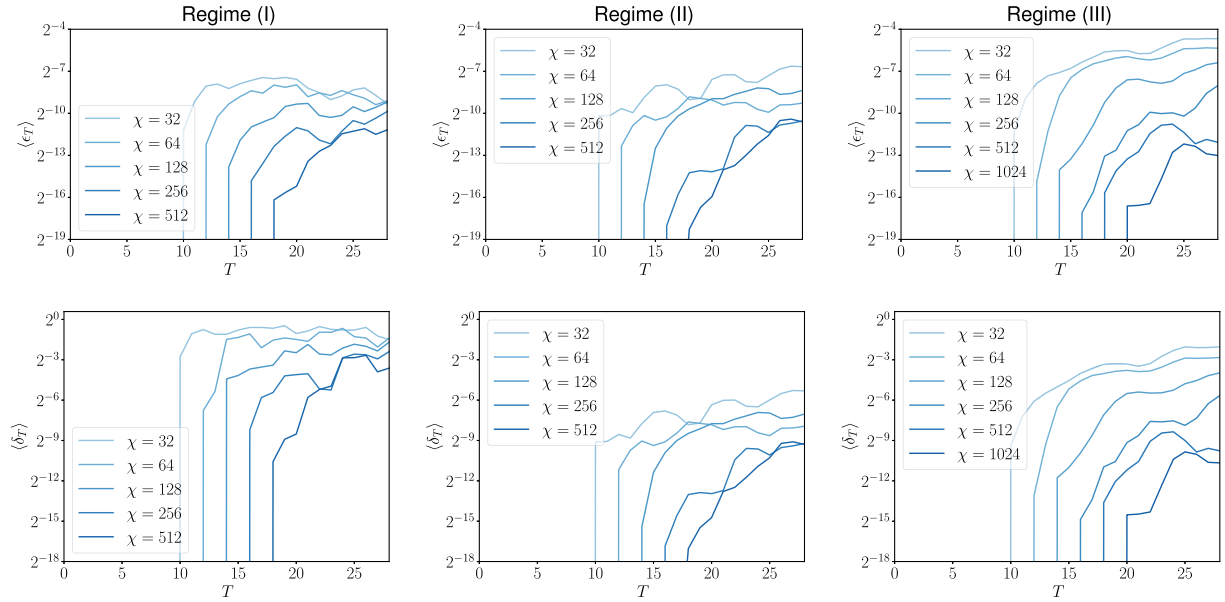


FIG. 14. Comparison of absolute (*top panels*) and relative (*bottom panels*) errors for different bond dimensions, with $\chi_{\max} = 1024$ in three regimes (I), (II), and (III) from left to right. In both figures, to make the graphs smoother and easier to interpret, we compute the moving average of the instantaneous error with the next two values.

TE scaling while keeping high accuracy on the edge (probe system) dynamics. For example, in regime (II), we find a bond dimension of just $\chi = 256$ is sufficient to capture dynamics of local observables with accuracy $\delta_T < 2^{-10}$ up to times $T \approx 30$.

VI. DISCUSSION

In summary, we have studied TE properties of thermalizing many-body baths. We provided an intuitive explanation of the previously observed volume-law scaling of TE in 1D chaotic brickwork circuits [36] and its relation to non-Markovianity and local temporal correlation functions. We found that the volume-law scaling does reflect genuine non-Markovianity of dynamics, and, by employing quantum information techniques, we established a link with distillable entanglement between the past and the future. Yet, the volume-law TE manifests itself only in highly complex temporal correlation functions. Such correlations are effectively projected out through a coarse-graining procedure, leading to a sharp collapse of TE to a subvolume-law scaling, which we dubbed TE transition, at a *finite* critical density of coarse-graining. We have studied this transition both analytically and numerically in a toy random unitary bath model, as well as in 1D dual-unitary and generic Floquet circuits.

Furthermore, we considered a Floquet model that displays nontrivial thermalization dynamics, leading to a parametrically slow decay of local autocorrelators of a probe spin at the edge. In this model, we found that TE undergoes a similar transition from volume- to subvolume-law under coarse-graining. Turning to local observables, we demonstrated that the slow thermalization dynamics could be captured accurately with a compressed representation of the IM via a low-bond dimension MPS. This is consistent with the intuition that, when the system is quickly thermalizing apart from quasi-conserved edge operators, truncating the highly non-local temporal correlations responsible for volume-law TE does not significantly affect the accuracy of the description of few-point correlators. An unexpected high accuracy of the method of transverse contractions ("folding algorithm") for long-time dynamics of local observables with moderate bond dimensions—far beyond the times allowed by convergence of TE—was reported for certain instances of Hamiltonian (continuous-time) dynamics in the first explorations [16,22,26] and more recently for discrete-time Floquet dynamics [24]. We hypothesize that the leading Schmidt vectors carry information on the sub-volume-law coarse-grained IM, whereas the small Schmidt values are related to highly complex multi-time correlators, providing the underlying mechanism behind the empirically observed accuracy.

Our results suggest that, although the exact IM is highly entangled, for the goal of computing N-point temporal correlators, there exists a greatly compressed MPS representation of IM that captures dynamics of slowly decaying operators. In the future, it will be interesting to explore different strategies for constructing such effective MPS representations, either via a generalization of the coarse-graining procedure or via introducing weak dissipation, similar to the dissipation-assisted operator evolution (DAOE) approach [75].

Finally, we note that coarse-graining effectively increases the range of circuit gates by a factor of $1/n_{\mathrm{cg}}$. A direct prediction of our analysis is that IMs of k-local circuits with $k > 2$ obey subvolume-law scaling, provided the elementary gates are sufficiently entangling. We can apply the same argument to higher-dimensional baths: The bath size $\sim 2^{T^d}$ greatly exceeds the probe's accessible temporal Hilbert space $\sim 2^T$ whenever the dimensionality of the bath is $d > 1$. Thus, in chaotic higher-dimensional systems, we predict subvolume-law TE scaling. This suggests that their IM can be approximated by an MPS with a moderate bond dimension. Developing approximate algorithms for constructing MPS representation of higher-dimensional interacting quantum baths is an exciting direction for future research. A recently introduced influence-functional belief-propagation method for 2D dynamics provides a first promising step in that direction [76].

ACKNOWLEDGMENTS

We thank Matan Lotem, Tianci Zhou, Jacopo De Nardis, Bruno Bertini, and Sarang Gopalakrishnan for insightful discussions. This work was supported by a Brown Investigator Award (I. V. and D. A. A.) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 864597) (I. V., A. L., D. A. A.). M. S. is supported by the Swiss National Science Foundation (SNSF) through a Postdoc.Mobility fellowship (Grant No. P500PT_225372). W. W. H. is supported by the National Research Foundation (NRF), Singapore through the NRF Fellowship NRF-NRFF15-2023-0008 and through the National Quantum Office, hosted in A*STAR, under its Centre for Quantum Technologies Funding Initiative (S24Q2d0009). A. L. was supported by a Leverhulme-Peierls fellowship at the University of Oxford and by the Flemish Research Foundation (FWO) through an Odysseus Grant No. (project G0ASY25N). This research was supported in part by Grant No. NSF PHY-2309135 to the Kavli Institute for Theoretical Physics (KITP).

DATA AVAILABILITY

The data used in this manuscript are available at Ref. [77]. Additional data generation code is available on request.

APPENDIX A: CALCULATION OF TEMPORAL RÉNYI ENTROPY

In this appendix, we provide details on the calculation of the temporal Rényi entropy for the structureless random unitary bath, leading to the result in Eq. (10).

1. Definition of temporal entanglement

We start by revising the definition of the IM provided by Eq. (2). In order to define TE, we vectorize the IM and treat it as a pure state. Formally this can be done by introducing a vector $|\mathcal{I}\rangle$:

$$\langle s_T, \bar{s}_T \cdots s_1, \bar{s}_1 | \mathcal{I} \rangle = \mathcal{I}_{s_1 \cdots s_T}^{\bar{s}_1 \cdots \bar{s}_T}. \tag{A1}$$

This vector is not normalized: rather, its squared norm N^2 equals

$$N^2 = \langle \mathcal{I} | \mathcal{I} \rangle. \tag{A2}$$

Let us also introduce the corresponding unnormalized density matrix $\tilde{\rho} = |\mathcal{I}\rangle\langle\mathcal{I}|$, which satisfies

$$\mathrm{tr}\tilde{\rho} = N^2, \tag{A3}$$

and we denote its partial trace as

$$\tilde{\rho}_{\mathcal{F}} = \mathrm{tr}_{\mathcal{P}}(|\mathcal{I}\rangle\langle\mathcal{I}|). \tag{A4}$$

With these definitions, temporal Rényi entropy takes the standard expression

$$S^{(\alpha)} = \frac{1}{1-\alpha} \log_d \frac{\mathrm{tr}\tilde{\rho}_{\mathcal{F}}^{\alpha}}{N^{2\alpha}}. \tag{A5}$$

For brevity, we will henceforth drop d in the base of logarithms.

2. Weingarten calculation of Rényi TE for the structureless random unitary bath

a. Static bath. In this section we concentrate on the computation of the temporal Rényi entropy for the structureless random unitary bath, in the limit of large bath $D_{\mathcal{B}} \gg 1$. In principle, we are interested in Rényi entropy averaged over the Haar-random unitaries U_t:

$$\mathbb{E}[S^{(\alpha)}] = \frac{1}{1-\alpha} \mathbb{E}\left[\log \frac{\mathrm{tr}\tilde{\rho}_{\mathcal{F}}^{\alpha}}{N^{2\alpha}}\right]. \tag{A6}$$

Instead, we calculate a related but easier quantity, i.e., the annealed entropy:

$$S_{\mathrm{ann}}^{(\alpha)} = \frac{1}{1-\alpha} \log \mathbb{E}\left[\frac{\mathrm{tr}\tilde{\rho}_{\mathcal{F}}^{\alpha}}{N^{2\alpha}}\right]. \tag{A7}$$

The latter provides a good approximation to the former if fluctuations are small, as we discuss below. The calculation of averages can be done with the help of Weingarten calculus [51], with the central formula

$$\langle (\mathcal{U} \otimes \mathcal{U}^{\dagger})^{\otimes k} \rangle = \sum_{\tau,\sigma} \mathrm{Wg}_q(\tau^{-1}\sigma) |\tau\rangle_{\mathrm{t}} |\tau\rangle_{\mathcal{B}} \times {}_{\mathcal{B}}\langle\sigma|_{\mathrm{t}} \langle\sigma|, \tag{A8}$$

where the sum is taken over all possible permutations τ, σ of the set of k tensor copies. Here $q = d \times D_{\mathcal{B}}$ is the Hilbert-space dimension of the composite system of the probe and bath at a single time step, namely the dimensionality of $\mathcal{U}$; the states $|\tau\rangle_{\mathrm{t}}, |\tau\rangle_{\mathcal{B}}$ correspond to permutation of replica spaces in probe and bath spaces, they can be explicitly defined via matrix elements:

$$ {}_a\langle s_k, \bar{s}_k, \cdots, s_1, \bar{s}_1|\tau\rangle_a = \prod_{i=1}^{k} \delta_{s_{\tau(i)}, \bar{s}_i}, \tag{A9}$$

where $a = \mathrm{t}$ or $\mathcal{B}$; and s_i ($\bar{s}_i$) is the index of computational basis in the ith forward (backward) branch, respectively. These permutation states are neither orthogonal nor normalized: the scalar product of two such states is given by

$$ {}_a\langle\tau|\sigma\rangle_a = C_{\sigma,\tau} = q_a^{\#\mathrm{cycles}(\tau^{-1}\sigma)}, \tag{A10}$$

where $q_{\mathrm{t},\mathcal{B}} = d, D_{\mathcal{B}}$ is dimension of the probe system or bath.

Below, we will concentrate on the case $\alpha = 2$, for which the annealed entanglement is nothing but the logarithm of the average purity. Analysis for general α is similar. Moreover, one can estimate the fluctuations of N^2, concluding that they are suppressed as $\frac{1}{D_{\mathcal{B}}}$, allowing us to further approximate the entropy as

$$ S^{(2)}_{\mathrm{ann}} \simeq -\log \frac{\mathbb{E}[\mathrm{tr}\tilde{\rho}^2_{\mathcal{F}}]}{\mathbb{E}[N^4]}. \tag{A11}$$

We will establish this approximation later; for now, let us concentrate on numerator and denominator separately. Each of them involves four copies of the IM and can be calculated using Weingarten calculus. The two terms are illustrated diagrammatically in Fig. 15(a,b).

Based on these figures, we can compute the averaged quantities using the corresponding transfer matrices:

$$ \mathbb{E}[\mathrm{tr}(\tilde{\rho}^2_{\mathcal{F}})] = {}_{\mathcal{B}}\langle\varnothing|\mathbb{T}^f_{\pi_{\mathcal{F}}}\mathbb{T}^p_{\pi_{\mathcal{P}}}|\rho_0\rangle_{\mathcal{B}}, \tag{A12}$$

$$ \mathbb{E}[N^4] = {}_{\mathcal{B}}\langle\varnothing|\mathbb{T}^T_{\pi_{\mathcal{P}}}|\rho_0\rangle_{\mathcal{B}}, \tag{A13}$$

where we introduced two permutations $\pi_{\mathcal{P}} = (1\,2)(3\,4)$, $\pi_{\mathcal{F}} = (1\,4)(2\,3)$. The state ${}_{\mathcal{B}}\langle\varnothing|$ corresponds to the identity permutation, representing the final tracing over the bath degrees of freedom, while $|\rho_0\rangle_{\mathcal{B}}$ is the initial state of the bath. Two relevant choices are a pure state, $|\rho_0\rangle_{\mathcal{B}} = (|0\rangle_{\mathcal{B}} \otimes |0^*\rangle_{\mathcal{B}})^{\otimes 4}$—in this case ${}_{\mathcal{B}}\langle\sigma|\rho_0\rangle_{\mathcal{B}} = 1,\ \forall\ \sigma \in S_4$—or a maximally mixed state, $|\rho_0\rangle_{\mathcal{B}} = (\frac{1}{D_{\mathcal{B}}})^4|\varnothing\rangle_{\mathcal{B}}$,—in which case ${}_{\mathcal{B}}\langle\sigma|\rho_0\rangle_{\mathcal{B}} = \frac{1}{D_{\mathcal{B}}^{|\sigma|}}$. In the latter equation $|\sigma|$ denotes the minimal number of transpositions to obtain σ from the identity permutation, which can alternatively be written in terms of the number of cycles in σ as

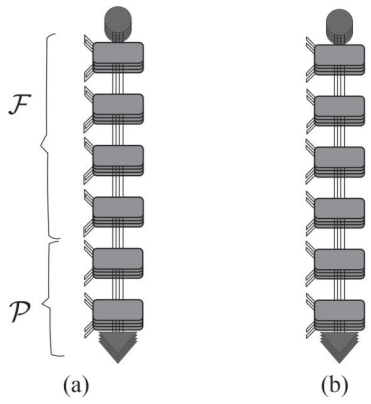


FIG. 15. (a) Pictorial definition of the numerator $\mathrm{tr}\tilde{\rho}^2_{\mathcal{F}}$. (b) Pictorial definition of the normalization term N^4.

$$ |\sigma| = 4 - \#\mathrm{cycles}(\sigma). \tag{A14}$$

For the long products Eqs. (A12) and (A13) under consideration, the transfer matrices are given explicitly as

$$ \mathbb{T}_{\mathrm{X}} = \sum_{\tau,\sigma} (\mathcal{T}_{\mathrm{X}})_{\tau,\sigma}|\tau\rangle_{\mathcal{B}\mathcal{B}}\langle\sigma|, \tag{A15}$$

where we defined component-wise

$$ (\mathcal{T}_{\mathrm{X}})_{\tau,\sigma} = \sum_{\tau,\sigma} \mathrm{Wg}_q(\tau^{-1}\sigma)_{\mathrm{t}}\langle\pi_{\mathrm{X}}|\tau\rangle_{\mathrm{tt}}\langle\sigma|\pi_{\mathrm{X}}\rangle_{\mathrm{t}}, \tag{A16}$$

and $\mathrm{X} \in \{\mathcal{P}, \mathcal{F}\}$ denotes the type of boundary condition imposed. Let us fix the basis of permutation states and switch to a matrix representation of transfer matrices provided by

$$ (C\mathcal{T}_{\mathrm{X}})_{\tau,\sigma} = \sum_{\lambda} {}_{\mathcal{B}}\langle\tau|\lambda\rangle_{\mathcal{B}} (\mathcal{T}_{\mathrm{X}})_{\lambda,\sigma}, \tag{A17}$$

which takes into account the inner products between permutation states. Both Eqs. (A12) and (A13) can be computed in terms of eigenvalues and eigenvectors of $(C\mathcal{T}_{\pi_{\mathcal{P}}})$ and $(C\mathcal{T}_{\pi_{\mathcal{F}}})$. Below, we examine these eigenvalues. In the limit of large $D_{\mathcal{B}}$, the leading order contribution is provided by diagonal terms of the Weingarten matrix $\mathrm{Wg}_q(\sigma^{-1}\sigma) = \mathrm{Wg}_q(\mathrm{id})$, for which the suppression $1/q^{4+|\lambda^{-1}\sigma|}$ is minimal. By again selecting the leading contribution in ${}_{\mathcal{B}}\langle\tau|\lambda\rangle_{\mathcal{B}}$, we obtain a condition $\sigma = \lambda = \tau$. Thus, the eigenvalues (to leading order) of the matrix $(C\mathcal{T}_{\mathrm{X}})$ are labeled by single permutations σ:

$$ E_\sigma = d^{4-2|\pi_{\mathrm{X}}\sigma^{-1}|} + o\left(\frac{1}{D_{\mathcal{B}}}\right). \tag{A18}$$

Examining the permutation group $S(4)$, we summarize the eigenvalues into a following table:

TABLE I. Eigenvalue spectrum and degeneracies.

Eigenvalue	Degeneracy
d^4	1
d^2	6
1	11
d^{-2}	6

It can be demonstrated that the leading eigenvalue d^4 remains unchanged to all orders in $D_{\mathcal{B}}$—in fact, this is a consequence of the unitarity condition of the Haar-random ensemble matrices U_τ. The remaining eigenvalues receive corrections.

According to Table I, it is sufficient to compute the leading order contributions to the terms proportional to d^{kT}, for $k = 1, \ldots, 4$. Omitting the details, we provide the results of an explicit expansion in $D_{\mathcal{B}}^{-1}$ using the Wolfram *Mathematica* package:

$$\mathbb{E}[\mathrm{tr}(\tilde{\rho}_{\mathcal{F}}^2)] = 1 + \frac{d^{4T}}{D_{\mathcal{B}}^4} + 2\frac{d^{4f+2p}}{D_{\mathcal{B}}^3} + \frac{d^{4f} + \frac{3}{4}d^{4p}f}{D_{\mathcal{B}}^2} + 2\frac{d^{2p} + d^{2f-2p}}{D_{\mathcal{B}}}, \quad \text{(A19)}$$

and

$$\mathbb{E}[\mathrm{tr}(N^4)] = \left(1 + \frac{d^{2T}}{D_{\mathcal{B}}}\right)^2. \quad \text{(A20)}$$

Armed with these results, let us go back to entanglement, given by Eq. (A11). In the regime $r < 1$ or $T < \log_d D_{\mathcal{B}}$, the numerator is dominated by 1, and the entanglement is symmetric with respect to f and p. In the regime $r > 1$, the leading term in the numerator is $\frac{d^{4T}}{D_{\mathcal{B}}^4}$ which is again symmetric. We conclude that the maximal entanglement is reached around $f = p$ in all regimes, with possible corrections at $r \simeq 1$. Thus, we have reproduced Eqs. (10) and(11) of the main text.

Additionally, by examining Eq. (A20) and the two-copy Weingarten calculation from Eq. (B11), we conclude the following equality:

$$\mathbb{E}[\mathrm{tr}(N^4)] = \mathbb{E}[\mathrm{tr}(N^2)]^2 + o(D_{\mathcal{B}}^{-1}). \quad \text{(A21)}$$

This suppression of fluctuations of $\mathrm{tr}(N^2)$ justifies the approximation in Eq. (A11).

b. Dynamically growing bath. In the case of a dynamically growing random unitary bath, we increase the bath dimension at every time step by feeding in the bath evolution a maximally mixed state of dimension d^{2v_B}. The modification of the transfer matrices from Eqs. (A15) and (A16) is given by the following formula:

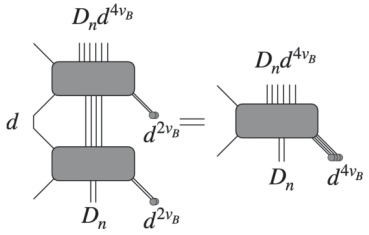


FIG. 16. Illustrating that the coarse-graining two consecutive gates is equivalent to a single gate with twice larger extra spin fed into the bath.

$$\mathbb{T}_{\mathrm{X}} = \sum_{\tau,\sigma} (\mathcal{T}_{\mathrm{X}})_{\tau,\sigma} \times |\tau\rangle_{\mathcal{B}\mathcal{B}}\langle\sigma|, \quad \text{(A22)}$$

with

$$(\mathcal{T}_{\mathrm{X}})_{\tau,\sigma} = d^{-4v_B} \sum_{\tau,\sigma} \mathrm{Wg}_q(\tau^{-1}\sigma)_{\mathrm{t}} \langle\pi_{\mathrm{X}}|\tau\rangle_{\mathrm{tt}} \langle\sigma|\pi_{\mathrm{X}}\rangle_{\mathrm{te}} \langle\sigma|\varnothing\rangle_{\mathrm{e}}, \quad \text{(A23)}$$

where we introduced an additional space e corresponding to the newly introduced spin to the bath at each time step. The transfer matrices now explicitly depend on n, and $D_n = d^{2v_B n}$. The average Rényi entropy can be computed along the same lines as before with the modified transfer matrices; the numerical results are presented in Fig. 4(c).

To estimate the effect of CG, we consider the following equation, which is most naturally represented diagrammatically, as shown in Fig. 16.

Mathematically, this equation can be formalized by introducing $\mathbb{T}(q, d_{\mathrm{e}})$, which acts on the temporal state as well as on the bath state:

$$\mathbb{T}(q, d_{\mathrm{e}}) = d^{-2v_B} \sum_{\sigma,\tau} \mathrm{Wg}_q(\tau^{-1}\sigma) |\tau\rangle_{\mathrm{t}} |\tau\rangle_{\mathcal{B}\,\mathcal{B}} \langle\sigma|_{\mathrm{t}} \langle\sigma|_{\mathrm{e}} \langle\sigma|\varnothing\rangle_{\mathrm{e}}. \quad \text{(A24)}$$

Here, d_{e} denotes the dimension of extra spin, and $q = dd_{\mathrm{e}}D_{\mathcal{B}}$ denotes the total dimension of the space. The pictorial equation from the figure reads as

$$\mathbb{T}(D_{n+1}d, d^{2v_B})\mathbb{T}(D_{n+2}d, d^{2v_B}) = \mathbb{T}(D_{n+2}d, d^{4v_B}), \quad \text{(A25)}$$

where we used $D_n = d^{2v_B n}$. This observation demonstrates that the coarse-graining reduces the number of time steps, effectively upscaling the $v_B \to \frac{v_B}{n_{\mathrm{cg}}}$.

APPENDIX B: ESTIMATION FOR THE NORMS OF IMS

In this section we provide details on the Weingarten calculations leading to the estimates for the density matrices $\rho_{\mathcal{I}}$ and $\rho_{\mathrm{in,out}}$, used in Eqs. (13) and in Appendix C, respectively. To obtain these estimates, we first use

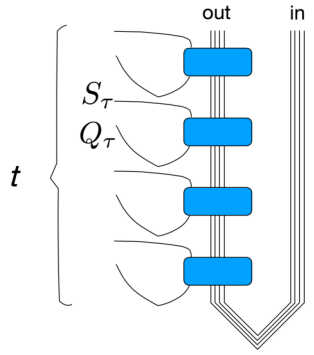


FIG. 17. Partition of the precursor state from Eq. (B1) into impurity degrees of freedom and bath degrees of freedom; the latter are also split into in and out degrees of freedom. The blue gates correspond to unitaries U_τ (forward branch only).

Eq. (20) to bound the trace norm by the Frobenius (Hilbert-Schmidt) norm, paying a dimensional penalty for switching norms. The Frobenius norm then allows for a direct Weingarten calculation, which is related closely with the purities.

We start by fixing the notations. As discussed in Sec. III C, we group the $2T$ temporal degrees of freedom in a "reference" set Q and a "system" S each containing T qudits and initialize them in pairs of maximally entangled states $|\Phi^{\max}_{SQ}\rangle$. Analogously, we split the bath degrees of freedom into "in" and "out" parts, also preparing them in a maximally entangled state $|\Phi^{\max}_{\mathcal{B}_{\rm in}\mathcal{B}_{\rm out}}\rangle$. Then, we act on a system and bath out degrees of freedom by a sequence of unitary operators, preparing the precursor state that belongs to the tensor product of three states: in, t, out (see the illustration in Fig. 17):

$$|\Psi_{\rm in,t,out}\rangle = U_T \cdots U_1 |\Phi^{\max}_{\mathcal{B}_{\rm in}\mathcal{B}_{\rm out}}\rangle |\Phi^{\max}_{SQ}\rangle. \tag{B1}$$

Tracing out the bath degrees of freedom defines the density matrix $\rho_{\rm t}$, which is nothing but the IM in a density matrix form (i.e., its Choi dual $\rho_{\mathcal{I}}$). Alternatively, one can trace out the temporal degrees of freedom, defining the density matrix on bath "in" and "out" degrees of freedom $\rho_{\rm in,out}$. Below, we will provide explicit calculations of the purities of these corresponding density matrices.

1. Purity calculations

a. Structureless bath. Proceeding in analogy with Eqs. (A12) and (A13), we may define

$$\mathbb{E}[{\rm tr}\rho^2_{\mathcal{I}}] = d^{-2T}{}_{\mathcal{B}}\langle\varnothing|\mathbb{T}^T_{\rm S}|\rho_0\rangle_{\mathcal{B}}, \tag{B2}$$

$$\mathbb{E}[{\rm tr}(\rho^2_{\rm in,out})] = d^{-2T}{}_{\mathcal{B}}\langle\varnothing|\mathbb{T}^T_{\rm I}|\rho_0\rangle_{\mathcal{B}}, \tag{B3}$$

where the transfer matrices are defined in analogy with Eqs. (A15) and (A16), with a difference that in this section we consider only two replicas (not four). Hence, there are only two permutations: the identity permutation $\pi_{\rm I}$ and the transposition $\pi_{\rm S} = (1,2)$. We consider maximally mixed (MM) or pure initial states. The MM state is defined by $\rho_0 = \frac{1}{D^2_{\mathcal{B}}}|\varnothing\rangle$, and the pure initial state ρ_0 is characterized by a unit overlap with all permutation states: $\langle\sigma|\rho_0\rangle = 1$. The Weingarten matrix reads

$${\rm Wg}_q = \frac{1}{q^2-1}\begin{pmatrix} 1 & -q^{-1} \\ -q^{-1} & 1 \end{pmatrix}, \tag{B4}$$

and an explicit calculation provides the following expression for the transfer matrix components in the basis of permutation states:

$$\mathcal{T}_{\rm S} = \frac{d^2}{D^2_{\mathcal{B}}d^2-1}\begin{pmatrix} 1 & -D^{-1}_{\mathcal{B}} \\ -D^{-1}_{\mathcal{B}} & d^2 \end{pmatrix}. \tag{B5}$$

The purity from Eqs. (B11) and (B12) can be written as

$$\mathbb{E}[{\rm tr}\rho^2_{\mathcal{I}}] = d^{-2T}\langle\varnothing|(C\mathcal{T}_{\rm S})^T C|\sigma_{\rm in}\rangle, \tag{B6}$$

where we introduced the matrix of scalar products C:

$$C = \begin{pmatrix} D^2_{\mathcal{B}} & D_{\mathcal{B}} \\ D_{\mathcal{B}} & D^2_{\mathcal{B}} \end{pmatrix}. \tag{B7}$$

The matrix product $C\mathcal{T}_{\rm S}$ gives

$$C\mathcal{T}_{\rm S} = d^2\begin{pmatrix} \frac{D^2_{\mathcal{B}}-1}{D^2_{\mathcal{B}}d^2-1} & \frac{(d^2-1)D_{\mathcal{B}}}{D^2_{\mathcal{B}}d^2-1} \\ 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1-\frac{1-d^{-2}}{D^2_{\mathcal{B}}} & \frac{d^2-1}{D_{\mathcal{B}}} \\ 0 & d^2 \end{pmatrix}. \tag{B8}$$

The leading contribution to the Tth power is then given by

$$d^{-2T}(C\mathcal{T}_{\rm S})^T \sim \begin{pmatrix} d^{-2T} & \frac{1}{D_{\mathcal{B}}} \\ 0 & 1 \end{pmatrix}. \tag{B9}$$

Taking into account that $C|\sigma_{\rm in}\rangle$ equals to

$$\begin{cases} (1,1)^\top, & \text{for pure initial state} \\ (1, D^{-1}_{\mathcal{B}})^\top, & \text{for MM initial state,} \end{cases} \tag{B10}$$

one gets for the purity:

$$\mathbb{E}[{\rm tr}\rho^2_{\mathcal{I}}] = d^{-2T} + \frac{1}{D_{\mathcal{B}}}, \qquad \text{for pure initial state,} \tag{B11}$$

$$\mathbb{E}[{\rm tr}\rho^2_{\mathcal{I}}] = d^{-2T} + \frac{1}{D^2_{\mathcal{B}}}, \qquad \text{for MM initial state.} \tag{B12}$$

This immediately reproduces the following estimate for the Frobenius distances:

$$\mathbb{E}[\|\rho_{\mathcal{I}} - \rho_{\mathrm{PD}}\|_2] \leq \frac{1}{D_{\mathcal{B}}^{\xi}}, \tag{B13}$$

where $\xi = 1, 2$ for the case of pure and MM initial state, respectively. Combining this estimate with Eq. (20) we reproduce Eq. (13).

The calculation of the purity in the bath channel is analogous, with the difference that one has to consider $\mathcal{T}_{\mathrm{I}}$:

$$C\mathcal{T}_{\mathrm{I}} = d^2 \begin{pmatrix} 1 & 0 \\ \frac{(d^2-1)D_{\mathcal{B}}}{d^2 D_{\mathcal{B}}^2 - 1} & \frac{D_{\mathcal{B}}^2 - 1}{d^2 D_{\mathcal{B}}^2 - 1} \end{pmatrix} \sim d^2 \begin{pmatrix} 1 & 0 \\ \frac{d^2-1}{d^2 D_{\mathcal{B}}} & d^{-2} \end{pmatrix}. \tag{B14}$$

The purity of $\rho_{\mathrm{in,out}}$ from Eq. (C5) then equals

$$\begin{aligned} \mathbb{E}[\mathrm{tr}\rho_{\mathrm{in,out}}^2] &= d^{-2T} \begin{pmatrix} 0 & 1 \end{pmatrix} (C\mathcal{T}_{\mathrm{I}})^T \begin{pmatrix} D_{\mathcal{B}}^{-1} \\ 1 \end{pmatrix} \\ &= d^{-2T} + D_{\mathcal{B}}^{-2} + o\left(D_{\mathcal{B}}^{-2}\right). \end{aligned} \tag{B15}$$

APPENDIX C: INFORMATION RECOVERY PROTOCOL

In this appendix we describe the information transition protocol for the scenario described in Sec. III C. This protocol bounds the distilled entanglement and quantum non-Markovianity of structureless and Hayden-Preskill bath.

We perform the information recovery in two steps. At first step we encode the information from the state in the past $\mathcal{P}$ to an intermediate out$'$ state in the bath. At the second step we feed the intermediate out$'$ state back to the bath, denoting it as in$' =$ out$'$ and decode the information back to the future system registers $\mathcal{F}$. We call these two steps an *encoding* and *decoding*, respectively. In the next paragraphs, we first introduce the so-called *decoupling lemma*—our main tool to study the encoding/decoding properties, next we explain the decoding and encoding steps separately, and finally combine all together, providing the estimate in Eq. (19). In the end of this appendix, we discuss the concentration of the probabilistic measure, which allows us to estimate the probability of a large deviation of the *encoding/decoding* fidelity from the average value. We will observe that this deviation is suppressed in the scaling limit when both the bath and the observation time are large.

a. Decoupling lemma. The derivation in this appendix relies on an important lemma 1.1 from Ref. [40]. Assume that the initial bath density matrix has exactly the rank $D_{\mathrm{in}'}$. Consider the precursor tripartite pure state between the temporal space "t" that can be chosen as $\mathcal{P}$ or $\mathcal{F}$ and the bath in and out' spaces introduced in Eq. (B1).

The following lemma ensures the existence of a decoding map:

Lemma C1 (Decoupling lemma). Let $\rho_{\mathrm{in,out}} = \mathrm{tr}_t(|\Psi_{\mathrm{in},t,\mathrm{out}}\rangle\langle\Psi_{\mathrm{in},t,\mathrm{out}}|)$. Then there exists a decoding operator $\mathcal{D}^{t\to t'}$ acting on the temporal spin degrees of freedom, such that:

$$F\left(\Phi_{\mathrm{in},t'}^{\max}, \mathbb{1}_{\mathrm{in}} \otimes (\mathcal{D}^{t\to t'})\rho_{\mathrm{in},t}\right) \geq 1 - \epsilon, \tag{C1}$$

provided that:

$$\|\rho_{\mathrm{in,out}} - \pi_{\mathrm{in}} \otimes \pi_{\mathrm{out}}\|_1 < \epsilon, \tag{C2}$$

where $\pi_{\mathrm{in/out}}$ is the density matrix of a maximally mixed (MM) in/out state.

In the simplified Hayden-Preskill one step process, the encoding and decoding processes are symmetric, and the existence of both maps can be proven using the decoupling lemma. Indeed, specializing the purity calculation from Eq. (B12) to the case of $T = 1, d = D_{\mathrm{t}} = d^f$, we may estimate average trace distance from the left-hand side of Eq. (C2) bounding it by a Frobenius norm:

$$\begin{aligned} &\|\rho_{\mathrm{in,out}'} - \pi_{\mathrm{in}} \otimes \pi_{\mathrm{out}'}\|_1^2 \\ &\quad \leq D_{\mathcal{B}}^2 \|\rho_{\mathrm{in,out}'} - \pi_{\mathrm{in}} \otimes \pi_{\mathrm{out}'}\|_2^2 = \frac{D_{\mathcal{B}}^2}{D_t^2}, \end{aligned} \tag{C3}$$

$$\begin{aligned} &\|\rho_{\mathrm{in}',\mathrm{out}} - \pi_{\mathrm{in}'} \otimes \pi_{\mathrm{out}}\|_1^2 \\ &\quad \leq D_{\mathcal{B}}^2 \|\rho_{\mathrm{in}',\mathrm{out}} - \pi_{\mathrm{in}'} \otimes \pi_{\mathrm{out}}\|_2^2 = \frac{D_{\mathcal{B}}^2}{D_t^2}. \end{aligned} \tag{C4}$$

These two inequalities sets a bound on ϵ from Eq. (C2) in average and prove the existence of a perfect decoding of the initial state in the regime $D_{\mathcal{B}} \gg D_t$ for the majority of choices for the unitaries. More precisely, one can use the concentration of measure technique, to estimate the probability of a significant deviation of the fidelity from the average value. We will make this estimation at the end of this appendix [see Eq. (C20)], assuming for now that the average trace distance is a good approximation for the trace distance of any typical process.

b. Decoding map. The same lemma can be used to prove the existence of a decoding map for the random unitary bath. The direct Weingarten calculation from Eq. (B15) provides on average:

$$\begin{aligned} &\|\rho_{\mathrm{in}',\mathrm{out}} - \pi_{\mathrm{in}'} \otimes \pi_{\mathrm{out}}\|_1^2 \\ &\quad < D_{\mathcal{B}} D_{\mathrm{in}'} \left(\mathrm{tr}\rho_{\mathrm{in}',\mathrm{out}}^2 - \frac{1}{D_{\mathcal{B}} D_{\mathrm{in}'}}\right) = \frac{D_{\mathcal{B}} D_{\mathrm{in}'}}{d^{2f}}, \end{aligned} \tag{C5}$$

where we assumed that the internal in$'$ space might have smaller dimension than $D_{\mathcal{B}}$. The relevance of this assumption will become clear later when we consider the encoding algorithm.

This estimate guarantees that as long as $f > \frac{1}{2}\log_d(D_{\mathcal{B}} D_{\mathrm{in}'})$, it is possible to distill $\log_d(D_{\mathrm{in}'})$

entangled qudits between the bath initial state and the observer $\mathcal{F}$, with the fidelity $F > 1 - \frac{\sqrt{D_{\mathcal{B}} D_{\mathrm{in}'}}}{d^{2f}}$. Let us note that Hayden-Preskill setup is a particular example of a random unitary bath, with $f = 1$ and large dimension of spin $D_t = d^f$. Both setups lead to the same scaling of critical time beyond which the recovery becomes possible.

c. Encoding of an intermediate bath state. We first start by considering pure initial bath state. Let us first assume a very short time in the past $p \ll \frac{1}{2}b$. As we have a pure initial bath state, the whole evolution defines the pure two-partite precursor state $|\Psi_{t,\mathrm{out}'}\rangle$. We will prove now the possibility to distill $2p$ qudits between impurity and the out bath state. To do so, we use again the result from Eq. (13) to show that the impurity density matrix $\rho_{\mathcal{P}}$ is close to a maximally mixed state:

$$\|\rho_{\mathcal{P}} - \rho_{\mathrm{depol}}\|_1 < \frac{d^{2p}}{D_{\mathcal{B}}}. \tag{C6}$$

This trace distance gives the lower bound on the fidelity:

$$F(\rho_{\mathcal{P}}, \rho_{\mathrm{depol}}) > 1 - \|\rho_{\mathcal{P}} - \rho_{\mathrm{depol}}\|_1. \tag{C7}$$

According to Ulhmann's theorem, there exists a purification of a MM density matrix ρ_{PD} by a maximally entangled state $|\Phi^{\max}_{\mathcal{P},\mathrm{out}'}\rangle$ such that

$$|\langle \Psi_{\mathcal{P},\mathrm{out}'} | \Phi^{\max}_{\mathcal{P},\mathrm{out}'} \rangle|^2 > 1 - \frac{d^{2p}}{D_{\mathcal{B}}}. \tag{C8}$$

This proves the possibility to distill $2p$ qudits between past and out$'$ bath state. No encoding operator $\mathcal{D}_{\mathrm{enc}}$ is needed in this case. In case of $p \gg \frac{1}{2}b$, the estimate (C6) is not applicable. One however can simply reduce the number of legs in $\mathcal{P}$ by coarse-graining the $2f - b + 2s$ qudits. Which guarantees the distillation of $b - 2s$ Bell pairs with the fidelity $F = 1 - 2^{-2s}$.

The case of MM initial state provides a much weaker bound, stating that it is possible to distill b Bell pairs between the temporal spin and bath out state, as long as $p \gg b$ with the fidelity $F = 1 - \frac{D_{\mathcal{B}}^2}{d^{2p}}$ which is an implication of decoupling Lemma (C1).

d. Distillation of information between the future and the past. Summarizing the encoding and decoding steps, for the pure bath state, we may first decode $N_{\mathcal{PB}} = \min(2p, b)$ qudits between the past and the intermediate bath. Moreover, according to Eq. (C8), in the regime where $2p < b$ the out$'$ approximately belongs to a subspace of dimension d^{2p} making it possible to consider $D_{\mathrm{in}'} = d^{2p} < D_{\mathcal{B}}$ in Eq. (C5). Taking it into account, Eq. (C5) guarantees that having $f = \frac{1}{2}b + p$ time steps on $\mathcal{F}$ is enough to distill Bell pairs between $\mathcal{F}$ and in$'$. Thus, we provide a protocol of distilling $N_{\mathcal{PF}}$ Bell pairs with $N_{\mathcal{PF}} = T - \frac{1}{2}b$ time steps, reproducing Eq. (19).

e. Concentration of measure. In this paragraph, we estimate the probability of the trace distances to deviate from the average values estimated above. The main estimation tool is the Levy's lemma [78]:

Lemma C2 (Lévy concentration inequality). Let (X, dist, μ) be a metric probability space and let $f\colon X \to \mathbb{R}$ be L-Lipschitz, then for all $\varepsilon > 0$,

$$\mu(|f - \mathbb{E}_\mu f| \geq \varepsilon) \leq 2\, \alpha_X\left(\frac{\varepsilon}{L}\right).$$

Here $\alpha_X(\frac{\varepsilon}{L})$—is a concentration function defined below:

Definition C1 (Concentration function). Let (X, dist, μ) be a metric probability space. For $\varepsilon \geq 0$ and $A \subseteq X$, write

$$A_\varepsilon := \{x \in X : \mathrm{dist}(x, A) \leq \varepsilon\}.$$

The (Lévy) concentration function of X is

$$\alpha_X(\varepsilon) := \sup\{1 - \mu(A_\varepsilon) :\ \mu(A) \geq 1/2\}.$$

In our case, the manifold X is the T-fold Cartesian product of unitary groups of size $dD_{\mathcal{B}} \times dD_{\mathcal{B}}$, each equipped with the Haar probability measure, and endowed with the metric induced by the operator norm:

$$\mathrm{dist}(\boldsymbol{U}, \boldsymbol{V}) = \|\boldsymbol{U} - \boldsymbol{V}\|, \tag{C9}$$

where $\|A\|$ is the largest eigenvalue of $\sqrt{A^\dagger A}$. As proven in the Appendix of Ref. [15] the concentration function can be bounded as:

$$\alpha_X\left(\frac{\varepsilon}{L}\right) \leq e^{-\frac{\epsilon^2 (T+1) d D_{\mathcal{B}}}{4L^2}}. \tag{C10}$$

Therefore, we only need to estimate the Lipschitz constant L of the function

$$f(\boldsymbol{U}) = \|\rho_{\mathrm{in,out}} - \pi_{\mathrm{in}} \otimes \pi_{\mathrm{out}}\|_1. \tag{C11}$$

By definition, Lipschitz constant is defined through a bound connecting the difference between the function at different points and the distance between these points:

$$|f(\boldsymbol{U}) - f(\boldsymbol{V})| \leq L \|\boldsymbol{U} - \boldsymbol{V}\|. \tag{C12}$$

To estimate the lhs of this equation, let us recall that $\rho_{\mathrm{in,out}}$ is obtained by tracing out the temporal degrees of freedom; therefore, it may be represented as

$$\rho_{\mathrm{in,out}} = M_T \cdot \cdots \cdot M_1 \cdot |\Phi^{\max}_{\mathcal{B}_{\mathrm{in}},\mathcal{B}_{\mathrm{out}}}\rangle\langle \Phi^{\max}_{\mathcal{B}_{\mathrm{in}},\mathcal{B}_{\mathrm{out}}}|, \tag{C13}$$

where M_τ is a quantum channel defined from the unitary U_τ:

$$M_\tau \cdot \rho = \frac{1}{d}\sum_{i=1}^{d} \mathrm{Tr}_s U_\tau(|i\rangle_{ss}\langle i| \otimes \rho_B)U_\tau^\dagger. \tag{C14}$$

In this equation, the trace is taken over the temporal spin at timestep τ. To proceed with the bound, let us suppose that we have two sequences of quantum channels $\boldsymbol{M}$ and $\boldsymbol{W}$ corresponding to the unitaries $\boldsymbol{U}$, $\boldsymbol{V}$ respectively. Using the triangle inequality, we may estimate:

$$\begin{aligned}
&\|M_T \cdot \cdots \cdot M_1 - W_T \cdot \cdots \cdot W_1\|_1 \\
&\le \|(M_T - W_T)M_{T-1} \cdot \cdots \cdot M_1\|_1 \\
&\quad + \|W_T(M_{T-1} \cdot \cdots \cdot M_1 - M_{T-1} \cdot \cdots \cdot M_1)\|_1 \\
&\le \|M_T - W_T\|_1 + \|M_{T-1} \cdot \cdots \cdot M_1 - W_{T-1} \cdot \cdots \cdot W_1\|_1 \\
&\le \sum_{\tau=1}^{T} \|M_\tau - W_\tau\|_1.
\end{aligned} \tag{C15}$$

The distance between single M_τ and W_τ can be further bounded by a distance between corresponding unitaries. By definition from Eq. (C14), we have

$$M_\tau = \frac{1}{d}\sum_{i,j} U^i_{\tau,j} \cdot (U^j_{\tau,i})^\dagger, \tag{C16}$$

where $U^i_{\tau,j} \cdot (U^j_{\tau,i})^\dagger$ is a superoperator acting on the space of bath density matrices. Using the triangle inequality again we have:

$$\|M_\tau - W_\tau\|_1 \le \sum \|U^i_{\tau,j} \cdot (U^j_{\tau,i})^\dagger - V^i_{\tau,j} \cdot (V^j_{\tau,i})^\dagger\|_1, \tag{C17}$$

now, using the lemma 12 from Ref. [79] we have:

$$\begin{aligned}
&\|U^i_{\tau,j} \cdot (U^j_{\tau,i})^\dagger - V^i_{\tau,j} \cdot (V^j_{\tau,i})^\dagger\|_1 \le 2\|U^i_{\tau,j} - V^i_{\tau,j}\| \\
&\le 2\|U - V\|.
\end{aligned} \tag{C18}$$

This finally guarantees:

$$\|M_\tau - W_\tau\| \le 2d\|U_\tau - V_\tau\| \le 2d\|\boldsymbol{U} - \boldsymbol{V}\|. \tag{C19}$$

Putting everything together, we find $L \le 2Td$, and the probability to violate the average estimates from the first part of the appendix is suppressed as

$$P(|F - F_{\mathrm{av}}| > \varepsilon) < e^{-\frac{\varepsilon^2 D_B}{16}}. \tag{C20}$$

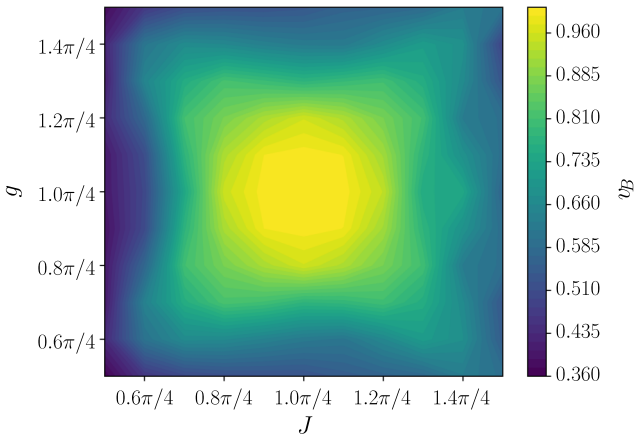


FIG. 18. Butterfly velocity v_B extracted from the out-of-time-order correlator front for system size $L = 10$.

APPENDIX D: BUTTERFLY VELOCITY

The front of the out-of-time-order correlator in Eq. (57) is defined operationally as the locus where the signal drops to a fraction p of its maximum. We estimate the butterfly velocity v_B as the speed of this front. For each parameter set, we compute v_B at thresholds p uniformly sampled in $p \in [0.3, 0.6]$; the reported value is the mean over p, and the error bar is the variance across these estimates. As expected, $v_B \to 1$ near the dual-unitary point, and it decreases as the g or J approach 0 or $\pi/2$. The resulting landscape is summarized in Fig. 18.

APPENDIX E: TE IN THE PLAIN VANILLA REGIME

In this Appendix, we report numerical results for an additional plain vanilla point, $g = \frac{\pi}{4}, j = 0.8\,\frac{\pi}{4}, h = 0.5$, for which the butterfly velocity is $v_B = 0.95 \pm 0.02$ as provided by Fig. 19. At such a high v_B, even a coarse-graining of $n = 1/2$ is sufficient to drive the IM into an subvolume-law regime. In contrast to regimes with nontrivial boundary modes, the coarse-grained IM is dominated by perfect dephaser components with suppressed correlations. This picture is consistent with a rapid decay to zero of $|C_{zz}|$ accompanied by large irregular relative fluctuations. The latter are not expected to be efficiently captured by MPS approximations of the IM.

APPENDIX F: Cxx CORRELATORS

In this appendix, we provide the numerical results for C_{xx} correlators in the three regimes (see Fig. 20. While different in magnitude, they demonstrate similar qualitative behavior.

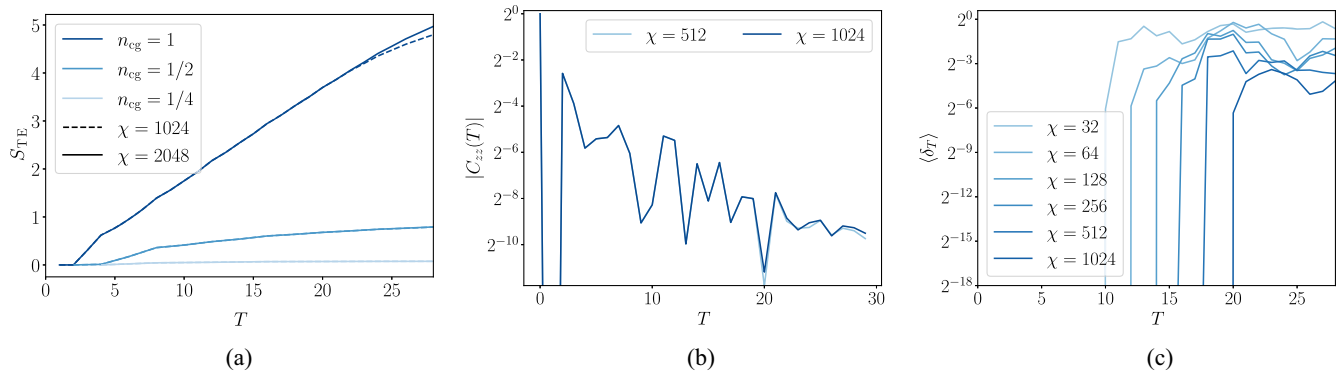


FIG. 19. Plain vanilla regime near the dual-unitary point, parameters $g = \frac{\pi}{4}$, $J = 0.8\,\frac{\pi}{4}$, $h = 0.5$. (a) (TE), (b) absolute correlator $|C_{zz}(T)|$, (c) moving average of relative error $\langle \delta_T \rangle$.

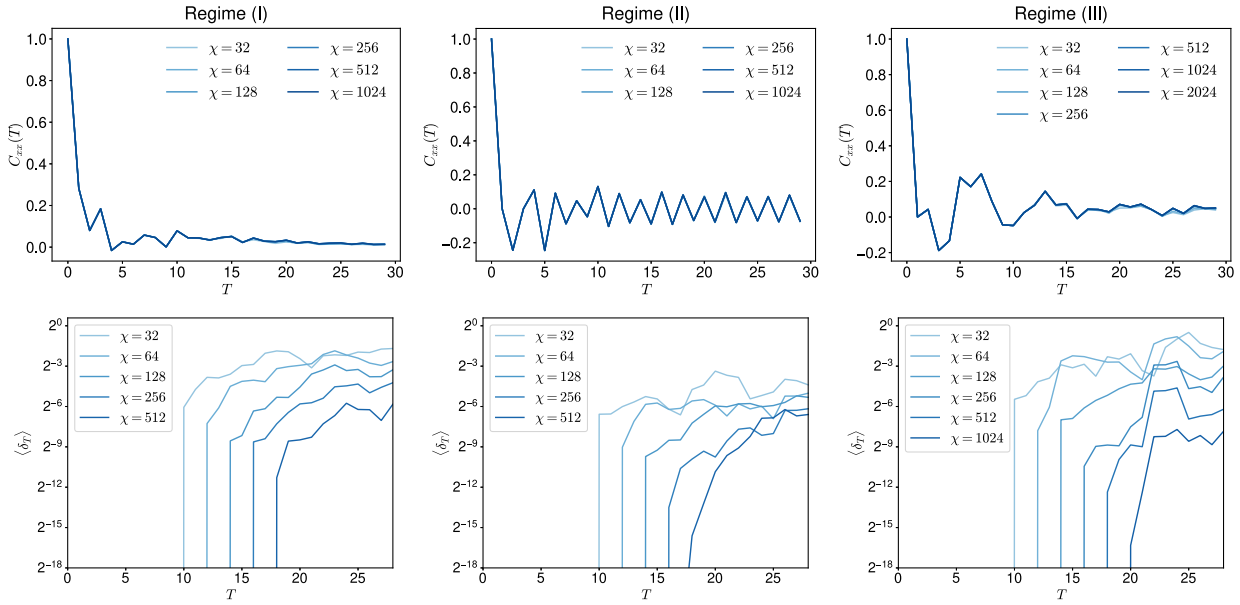


FIG. 20. Top panels: local temporal correlator $C_{xx}(T)$ in the three regimes shows qualitatively similar behavior to their C_{zz} counterparts. Bottom panels: relative error for the C_{xx} correlator and different bond dimensions.

[1] M. P. A. Fisher, V. Khemani, A. Nahum, and S. Vijay, *Random quantum circuits*, Annu. Rev. Condens. Matter. Phys. **14**, 335 (2023).

[2] M. Akila, D. Waltner, B. Gutkin, and T. Guhr, *Particle-time duality in the kicked Ising spin chain*, J. Phys. A **49**, 375101 (2016).

[3] B. Bertini, P. Kos, and T. c. v. Prosen, *Exact correlation functions for dual-unitary lattice models in* $1 + 1$ *dimensions*, Phys. Rev. Lett. **123**, 210601 (2019).

[4] B. Bertini, P. W. Claeys, and T. Prosen, *Exactly solvable quantum many-body dynamics from space-time duality*, Rev. Mod. Phys. **98**, 025001 (2026).

[5] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, *Quantum entanglement growth under random unitary dynamics*, Phys. Rev. X **7**, 031016 (2017).

[6] A. Nahum, S. Vijay, and J. Haah, *Operator spreading in random unitary circuits*, Phys. Rev. X **8**, 021014 (2018).

[7] B. Bertini, P. Kos, and T. c. v. Prosen, *Exact spectral form factor in a minimal model of many-body quantum chaos*, Phys. Rev. Lett. **121**, 264101 (2018).

[8] B. Bertini, P. Kos, and T. c. v. Prosen, *Entanglement spreading in a minimal model of maximal many-body quantum chaos*, Phys. Rev. X **9**, 021033 (2019).

[9] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, *Non-Markovian quantum processes: Complete framework and efficient characterization*, Phys. Rev. A **97**, 012127 (2018).

[10] M.-D. Choi, *Completely positive linear maps on complex matrices*, Linear Algebra Appl. **10**, 285 (1975).

[11] R. P. Feynman and F. L. Vernon, *The theory of a general quantum system interacting with a linear dissipative system*, Ann. Phys. (N.Y.) **281**, 547 (2000).

[12] A. Lerose, M. Sonner, and D. A. Abanin, *Influence matrix approach to many-body Floquet dynamics*, Phys. Rev. X **11**, 021040 (2021).

[13] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, *Operational Markov condition for quantum processes*, Phys. Rev. Lett. **120**, 040405 (2018).

[14] I. A. Luchnikov, S. V. Vintskevich, H. Ouerdane, and S. N. Filippov, *Simulation complexity of open quantum dynamics: Connection with tensor networks*, Phys. Rev. Lett. **122**, 160401 (2019).

[15] P. Figueroa-Romero, K. Modi, and F. A. Pollock, *Almost Markovian processes from closed dynamics*, Quantum **3**, 136 (2019).

[16] 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).

[17] A. Lerose, M. Sonner, and D. A. Abanin, *Scaling of temporal entanglement in proximity to integrability*, Phys. Rev. B **104**, 035137 (2021).

[18] M. Sonner, A. Lerose, and D. A. Abanin, *Influence functional of many-body systems: Temporal entanglement and matrix-product state representation*, Ann. Phys. (N.Y.) **435**, 168677 (2021).

[19] E. Ye and G. K.-L. Chan, *Constructing tensor network influence functionals for general quantum dynamics*, J. Chem. Phys. **155**, 044104 (2021).

[20] M. Sonner, A. Lerose, and D. A. Abanin, *Characterizing many-body localization via exact disorder-averaged quantum noise*, Phys. Rev. B **105**, L020203 (2022).

[21] S. Carignano, C. R. Marimón, and L. Tagliacozzo, *Temporal entropy and the complexity of computing the expectation value of local operators after a quench*, Phys. Rev. Res. **6**, 033021 (2024).

[22] M. B. Hastings and R. Mahajan, *Connecting entanglement in time and space: Improving the folding algorithm*, Phys. Rev. A **91**, 032306 (2015).

[23] M. Frías-Pérez and M. C. Bañuls, *Light cone tensor network and time evolution*, Phys. Rev. B **106**, 115117 (2022).

[24] A. Lerose, M. Sonner, and D. A. Abanin, *Overcoming the entanglement barrier in quantum many-body dynamics via space-time duality*, Phys. Rev. B **107**, L060305 (2023).

[25] S. Carignano, G. Lami, J. De Nardis, and L. Tagliacozzo, *Overcoming the entanglement barrier with sampled tensor networks*, arXiv:2505.09714.

[26] 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).

[27] K. Klobas, B. Bertini, and L. Piroli, *Exact thermalization dynamics in the "rule 54" quantum cellular automaton*, Phys. Rev. Lett. **126**, 160602 (2021).

[28] 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).

[29] 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).

[30] A. Strathearn, P. Kirton, D. Kilda, J. Keeling, and B. W. Lovett, *Efficient non-Markovian quantum dynamics using time-evolving matrix product operators*, Nat. Commun. **9**, 3322 (2018).

[31] 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).

[32] 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).

[33] M. Sonner, V. Link, and D. A. Abanin, *Semi-group influence matrices for non-equilibrium quantum impurity models*, arXiv:2502.00109.

[34] I. Vilkoviskiy and D. A. Abanin, *Bound on approximating non-Markovian dynamics by tensor networks in the time domain*, Phys. Rev. B **109**, 205126 (2024).

[35] J. Thoenniss, I. Vilkoviskiy, and D. A. Abanin, *Efficient pseudomode representation and complexity of quantum impurity models*, Phys. Rev. B **112**, 155114 (2025).

[36] A. Foligno, T. Zhou, and B. Bertini, *Temporal entanglement in chaotic quantum circuits*, Phys. Rev. X **13**, 041008 (2023).

[37] P. O'Donovan, N. Dowling, K. Modi, and M. T. Mitchison, *Diagnosing chaos with projected ensembles of process tensors*, PRX Quantum **7**, 020322 (2026).

[38] D. N. Page, *Information in black hole radiation*, Phys. Rev. Lett. **71**, 3743 (1993).

[39] P. Hayden and J. Preskill, *Black holes as mirrors: Quantum information in random subsystems*, J. High Energy Phys. 09, 120 (2007).

[40] P. Hayden, M. Horodecki, A. Winter, and J. Yard, *A decoupling approach to the quantum capacity*, Open Syst. Inf. Dyn. **15**, 7 (2008).

[41] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, *Purification of noisy entanglement and faithful teleportation via noisy channels*, Phys. Rev. Lett. **76**, 722 (1996).

[42] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, *Mixed-state entanglement and quantum error correction*, Phys. Rev. A **54**, 3824 (1996).

[43] 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).

[44] 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).

[45] D. V. Else, P. Fendley, J. Kemp, and C. Nayak, *Prethermal strong zero modes and topological qubits*, Phys. Rev. X **7**, 041062 (2017).

[46] L. Piroli, B. Bertini, J. I. Cirac, and T. c. v. Prosen, *Exact dynamics in dual-unitary quantum circuits*, Phys. Rev. B **101**, 094304 (2020).

[47] These authors also found that the IM vector for 1D chaotic baths has a high (i.e., only polynomially suppressed in T) overlap with a product state, resulting in area-law at the dual-unitary point or logarithmic scaling away from this point for all higher Rényi TEs.

[48] The extension to dissipative and time-dependent probe system's dynamics is straightforward.

[49] M. Ippoliti and W. W. Ho, *Solvable model of deep thermalization with distinct design times*, Quantum **6**, 886 (2022).

[50] P. W. Brouwer and C. W. J. Beenakker, *Diagrammatic method of integration over the unitary group, with applications to quantum transport in mesoscopic systems*, J. Math. Phys. **37**, 4904 (1996).

[51] B. Collins, S. Matsumoto, and J. Novak, *The Weingarten calculus*, Not. Am. Math. Soc. **69**, 1 (2022).

[52] B. Yoshida and A. Kitaev, *Efficient decoding for the Hayden-Preskill protocol*, arXiv:1710.03363.

[53] D. N. Page, *Average entropy of a subsystem*, Phys. Rev. Lett. **71**, 1291 (1993).

[54] Distillable entanglement provides a lower bound for the TE, but the opposite is not true in general. Nothing excludes the scenario of high TE and low distillable entanglement. For example this could happen in the system where correlations are almost classical but non-Markovian.

[55] P. W. Claeys and A. Lamacraft, *Ergodic and nonergodic dual-unitary quantum circuits with arbitrary local Hilbert space dimension*, Phys. Rev. Lett. **126**, 100603 (2021).

[56] H. Kim, T. N. Ikeda, and D. A. Huse, *Testing whether all eigenstates obey the eigenstate thermalization hypothesis*, Phys. Rev. E **90**, 052105 (2014).

[57] J. Preskill, *Quantum Shannon theory*, arXiv:1604.07450.

[58] D. J. Yates, F. H. L. Essler, and A. Mitra, *Almost strong $(0,\pi)$ edge modes in clean interacting one-dimensional Floquet systems*, Phys. Rev. B **99**, 205419 (2019).

[59] X. Mi *et al.*, *Noise-resilient edge modes on a chain of superconducting qubits*, Science **378**, 785 (2022).

[60] A. Y. Kitaev, *Unpaired Majorana fermions in quantum wires*, Phys. Usp. **44**, 131 (2001).

[61] M. Thakurathi, A. A. Patel, D. Sen, and A. Dutta, *Floquet generation of Majorana end modes and topological invariants*, Phys. Rev. B **88**, 155133 (2013).

[62] A. Bohrdt, C. B. Mendl, M. Endres, and M. Knap, *Scrambling and thermalization in a diffusive quantum many-body system*, New J. Phys. **19**, 063001 (2017).

[63] D. J. Luitz and Y. Bar Lev, *Information propagation in isolated quantum systems*, Phys. Rev. B **96**, 020406(R) (2017).

[64] T. Zhou and A. Nahum, *Entanglement membrane in chaotic many-body systems*, Phys. Rev. X **10**, 031066 (2020).

[65] M. Kormos, M. Collura, G. Takács, and P. Calabrese, *Real time confinement following a quantum quench to a non-integrable model*, Nat. Phys. **13**, 246 (2017).

[66] A. J. A. James, R. M. Konik, and N. J. Robinson, *Nonthermal states arising from confinement in one and two dimensions*, Phys. Rev. Lett. **122**, 130603 (2019).

[67] A. Lerose, F. M. Surace, P. P. Mazza, G. Perfetto, M. Collura, and A. Gambassi, *Quasilocalized dynamics from confinement of quantum excitations*, Phys. Rev. B **102**, 041118(R) (2020).

[68] M. Collura, A. De Luca, D. Rossini, and A. Lerose, *Discrete time-crystalline response stabilized by domain-wall confinement*, Phys. Rev. X **12**, 031037 (2022).

[69] Z.-C. Yang, F. Liu, A. V. Gorshkov, and T. Iadecola, *Hilbert-space fragmentation from strict confinement*, Phys. Rev. Lett. **124**, 207602 (2020).

[70] C. W. von Keyserlingk, V. Khemani, and S. L. Sondhi, *Absolute stability and spatiotemporal long-range order in floquet systems*, Phys. Rev. B **94**, 085112 (2016).

[71] D. V. Else, B. Bauer, and C. Nayak, *Prethermal phases of matter protected by time-translation symmetry*, Phys. Rev. X **7**, 011026 (2017).

[72] V. Khemani, R. Moessner, and S. Sondhi, *A brief history of time crystals*, arXiv:1910.10745.

[73] X. Mi, M. Ippoliti, C. Quintana, A. Greene, Z. Chen, J. Gross, F. Arute, K. Arya, J. Atalaya, R. Babbush *et al.*, *Time-crystalline eigenstate order on a quantum processor*, Nature (London) **601**, 531 (2022).

[74] T. Yoshimura, S. J. Garratt, and J. T. Chalker, *Operator dynamics in floquet many-body systems*, Phys. Rev. B **111**, 094316 (2025).

[75] T. Rakovszky, C. W. von Keyserlingk, and F. Pollmann, *Dissipation-assisted operator evolution method for capturing hydrodynamic transport*, Phys. Rev. B **105**, 075131 (2022).

[76] G. Park, J. Gray, and G. K.-L. Chan, *Simulating quantum dynamics in two-dimensional lattices with tensor network influence functional belief propagation*, Phys. Rev. B **112**, 174310 (2025).

[77] I. Vilkoviskiy, M. Sonner, Q. C. Huang, W. W. Ho, A. Lerose, and D. A. Abanin, *Plotting code and data for "temporal entanglement transition in chaotic quantum many-body dynamics"* (2026), https://zenodo.org/records/23201444.

[78] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs Vol. 89 (American Mathematical Society, Providence, RI, 2001).

[79] D. Aharonov, A. Kitaev, and N. Nisan, *Quantum circuits with mixed states*, in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98* (Association for Computing Machinery (ACM), New York, NY, USA, 1998), pp. 20–30, 10.1145/276698.276708.