

Maximum Entropy Principle in Deep Thermalization and in Hilbert-Space Ergodicity

Daniel K. Mark¹, Federica Surace², Andreas Elben², Adam L. Shaw², Joonhee Choi³,
Gil Refael², Manuel Endres², and Soonwon Choi^{1,*}

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

²*California Institute of Technology, Pasadena, California 91125, USA*

³*Department of Electrical Engineering, Stanford University, Stanford, California, USA*



(Received 2 April 2024; revised 14 August 2024; accepted 28 August 2024; published 25 November 2024)

We report universal statistical properties displayed by ensembles of pure states that naturally emerge in quantum many-body systems. Specifically, two classes of state ensembles are considered: those formed by (i) the temporal trajectory of a quantum state under unitary evolution or (ii) the quantum states of small subsystems obtained by partial, local projective measurements performed on their complements. These cases, respectively, exemplify the phenomena of “Hilbert-space ergodicity” and “deep thermalization.” In both cases, the resultant ensembles are defined by a simple principle: The distributions of pure states have maximum entropy, subject to constraints such as energy conservation, and effective constraints imposed by thermalization. We present and numerically verify quantifiable signatures of this principle by deriving explicit formulas for all statistical moments of the ensembles, proving the necessary and sufficient conditions for such universality under widely accepted assumptions, and describing their measurable consequences in experiments. We further discuss information-theoretic implications of the universality: Our ensembles have maximal information content while being maximally difficult to interrogate, establishing that generic quantum state ensembles that occur in nature hide (scramble) information as strongly as possible. Our results generalize the notions of Hilbert-space ergodicity to time-independent Hamiltonian dynamics and deep thermalization from infinite to finite effective temperature. Our work presents new perspectives to characterize and understand universal behaviors of quantum dynamics using statistical and information-theoretic tools.

DOI: [10.1103/PhysRevX.14.041051](https://doi.org/10.1103/PhysRevX.14.041051)

Subject Areas: Quantum Information, Statistical Physics

I. INTRODUCTION

The second law of thermodynamics is a foundational concept in statistical physics with far-reaching implications across many areas of studies, from quantum gravity to information theory and computation [1–5]. At its core is the idea of a *maximum entropy principle* [6–10], that a macroscopic system in equilibrium is described by the state with maximum entropy, subject to physical constraints such as global energy and particle number conservation. It is remarkable that such a simple principle enables highly accurate predictions about a large system without knowledge of the vast majority of its degrees of freedom.

In quantum systems, the maximum entropy principle [11,12] is often justified by one of the following two

conditions relevant to many conventional experiments [Fig. 1(a)]. First, one has access to only local or spatially averaged observables [13–16]. Second, one can measure only observables averaged over long time intervals or at late, equilibrium times without fine-grained temporal resolution [15–17]. In both cases, measurement outcomes become largely agnostic to microscopic details of the initial state or system dynamics and are well described by thermal density matrices. It is this *inability* to obtain detailed, microscopic information from traditional experiments that renders quantum states to be in a probabilistic ensemble and ultimately empowers the maximum entropy principle to make accurate and practical predictions of accessible observables.

Modern quantum experiments challenge such traditional settings [Fig. 1(b)]. Quantum devices built using atomic or superconducting qubit platforms have realized and probed diverse phenomena ranging from nonequilibrium dynamics [18–22] to exotic phases of matter [23–27]. These experiments have the capability to fully interrogate the entire quantum system, enabling the measurement of arbitrary, nonlocal multiparticle correlation functions at the

*Contact author: soonwon@mit.edu

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

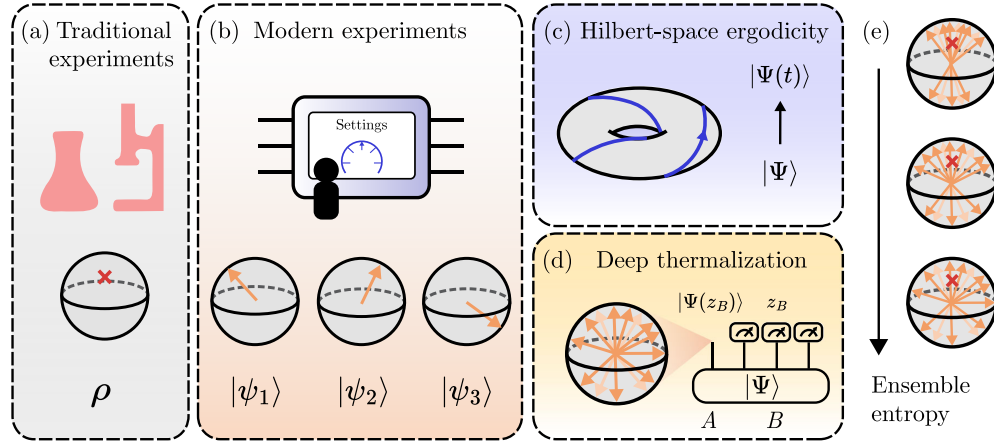


FIG. 1. Overview of this work. In this work, we study ensembles of states generated by natural dynamics. (a) Traditional experiments on quantum systems can measure only the average properties of such ensembles, described by a density matrix ρ (red cross in Bloch sphere). (b) In contrast, modern experimental capabilities are able to generate and investigate ensembles of labeled quantum states and study their higher statistical properties. We specifically study two ensembles of states in this work: (c) the *temporal ensemble* of states generated by taking an initial state $|\Psi_0\rangle$ and evolving it to different times under a (fixed) Hamiltonian H and (d) the *projected ensemble* of states $\{|\Psi(z_B)\rangle\}$ generated by projective measurement (performed on a large subsystem) of a single state $|\Psi\rangle$ resulting in many different outcomes z_B . These ensembles, respectively, exemplify the phenomena of *Hilbert-space ergodicity* and *deep thermalization*. (e) Our key finding is that these ensembles of states have *maximum entropy*, subject to constraints arising from dynamics. For example, quantum thermodynamics constrains projected ensembles to have a fixed first moment ρ . While many ensembles of states share the same first moment ρ (red crosses), the (unnormalized) projected ensemble is the one with the maximum ensemble entropy (bottom).

fundamental quantum projection limit. They are also able to deterministically evolve quantum many-body states with a high repetition rate, enabling the study of quantum dynamics with an unprecedented level of accuracy and details. Existing maximum entropy principles were not designed to capture such microscopic measurement outcomes, which are nonlocal and time resolved. Therefore, we set out to develop a new approach that may provide an effective statistical model for such data. Here, we report the discovery of a generalized maximum entropy principle for quantum many-body systems that is well aligned with modern capabilities of experiments.

We specifically study two settings, the *temporal* and *projected ensembles*. The former considers the trajectory of quantum states under time evolution, while the latter considers ensembles of states generated by partial projective measurements performed on the system [Figs. 1(c) and 1(d)]. In both settings, the average states, or first moments, of the ensembles are described by density matrices ρ which have been extensively studied in the literature [12–17]. Here, certain universal phenomena apply, such as the fact that the reduced density matrices of small subsystems are well described by thermal Gibbs states [14,15] or dephasing in the energy eigenbasis, which leads to the so-called diagonal ensemble as an effective description of equilibrium properties [12,13,17]. However, the temporal and projected ensembles of generic systems have not been studied beyond their first moments. Our results extend this universality to higher moments, finding that the first moment uniquely fixes the higher moments.

We find a new type of *maximum entropy principle* that describes ensembles of states in the above settings [Fig. 1(e)]. The entropy in question is an *ensemble entropy*, which quantifies how the states in an ensemble are spread out over the relevant Hilbert space. This is distinct from the conventional von Neumann entropy, which depends on only the average state (density matrix) of the ensemble. In broad strokes, our main finding is that, in many physical situations, the ensembles of pure states maximize the ensemble entropy up to physical constraints such as energy conservation or local thermalization, thus establishing a new maximum entropy principle. Special cases of our ensembles of states have been discussed in the literature under the names of “deep thermalization” [28–38] and “Hilbert-space ergodicity” [39,40], respectively, in particular, at infinite effective temperature. Here, we present a unified principle that holds under general conditions.

Ultimately, our work is a theory of fluctuations: the variations in observables over states in an ensemble, which go fundamentally beyond ensemble-averaged observables that are typically studied. In companion experimental work [41], we use insights gained from this work to predict and verify universal forms of fluctuations in a variety of quantities, interpolating from global to local, and in closed as well as open system dynamics.

Our findings have several implications. First, we confirm that Hamiltonian dynamics uniformly explores its available Hilbert space. Specifically, the temporal trace of a many-body quantum state satisfies the maximum entropy principle under the constraint of energy conservation.

For sufficiently long evolution, the trajectory of the state is constrained only by the population of the energy eigenstates: All other degrees of freedom are randomly distributed with maximum entropy. This improves our understanding of thermalization as a dynamical phenomenon [42,43] and its relationship to static notions of thermal equilibrium [44,45]. While thermalization has traditionally concerned the dynamics of local quantities—e.g., how such quantities reach their equilibrium values—here we study the dynamics of the *global* state, confirming and formalizing our intuition about ergodicity in Hilbert space [39,46]. We find necessary and sufficient conditions for Hilbert-space ergodicity and find that curiously, just as in classical dynamics, a system can be ergodic without being chaotic. In addition to exploring its unique signatures and properties in the many-body setting, Hilbert-space ergodicity serves as the core result enabling our other results.

Second, we generalize notions of deep thermalization [28–30] away from infinite temperature. Deep thermalization refers to the observation that, in generic many-body dynamics, not only does the density matrix of a local subsystem reach thermal equilibrium, so do the higher moments of the projected ensemble. Additionally, it has been pointed out that the higher moments may take a longer time than lower moments to converge, implying the possibility of sustained thermalization (in nonlocal quantities) even when local observables have converged [30,31,36]. The equilibrium values of such higher moments have considerable structure. At infinite effective temperature, the projected ensemble [28,29] has been found to approach the *Haar ensemble* [47,48], a paradigmatic uniform random ensemble of quantum states. We extend this phenomenology to finite temperature, identifying finite-temperature generalizations of the Haar ensemble with maximum entropy which describe our ensembles of states. Our result should be taken as complementary to but distinct from conventional theories of quantum thermalization such as the eigenstate thermalization hypothesis (ETH) [12,15]. It does not rely on the ETH but, when used in conjunction, implies that all higher moments of the projected ensemble reach thermal equilibrium with universal, Gibbs-like structure. For the first time, we rigorously derive expressions for the statistical properties of the projected ensembles generated by ergodic Hamiltonian dynamics, which had previously been only numerical observation [28,29]. Specifically, we describe the projected ensemble in terms of the generalized Scrooge ensemble. This ensemble reduces to the Scrooge ensemble in the simplest case that the measurement outcomes are uncorrelated with the energy. In the further limit of infinite temperature, we recover the Haar ensemble, proving deep thermalization in generic infinite-temperature Hamiltonian dynamics, which had remained an open conjecture. Our proofs rely on our results about the temporal ensemble, intricately linking the two ensembles.

Third, our approach predicts and, where relevant, confirms the existence of a universality in natural many-body states, revealed in their higher moments. Not only do mean quantities captured by density matrices ρ exhibit universal thermal behavior, so do higher-order moments such as the variance and skewness, as dictated by our maximum entropy principle. Higher-order quantities beyond local observables have received recent attention in nonequilibrium dynamics [49,50], and our results suggest the broader applicability of higher moments—the maximum entropy principle holds not only at the level of density matrices [51,52], but may also hold at the level of many-body wave functions [53–55]. This possibility has been recently investigated for energy eigenstates in Ref. [55]. Our results support such an approach based on the maximum entropy principle and suggest that they may be more widely applicable; here, we primarily apply them to time-evolved many-body states.

Finally, we draw further connections between thermalization and quantum information theory. Ensembles of quantum states are natural objects for quantum communication: Fundamental results such as the Holevo bound investigate the use of such ensembles to transmit information [56]. Our results indicate that ensembles of states that result from many-body dynamics are both maximally difficult to compress, distinguish, or use for information transmission, providing a new perspective on the nature of information scrambling and hiding in many-body dynamics [57–61].

The maximum entropy principle has repeatedly exceeded expectations in its ability to describe complex, interacting systems [6–10]. Most results have so far been restricted to classical or single-particle quantum physics. Our work represents a first step to adapting this principle to the modern setting of many-body physics. While we were able to firmly establish this principle in certain settings, we find that it may apply more generally in many-body physics. Indeed, see Ref. [62] for a recent application of this principle to many-body physics to full-counting statistics in systems with a $U(1)$ symmetry. These developments hint at broader applications of the maximum entropy principle in many-body physics.

II. SUMMARY OF RESULTS

In this work, we study ensembles of quantum states. These ensembles $\mathcal{E}$ are collections of pure quantum states $|\Psi_j\rangle$, labeled by an index j , with an associated probability distribution p_j , that is, $\mathcal{E} = \{p_j, |\Psi_j\rangle\}$. Quantum state ensembles can be discrete (i.e., the index j takes discrete values, such as a discrete set of measurement outcomes) or continuous (the index j takes continuous values, such as time). In situations where the label is not important, we simply write $\mathcal{E} = \{P(\Psi), |\Psi\rangle\}$.

Quantum state ensembles have traditionally been studied in quantum information for tasks such as quantum

communication [63]. However, until recently they have not been necessary in the context of many-body physics. Under traditional measurements, observations made on an ensemble of states depend on only the first moment $\rho_{\mathcal{E}}^{(1)} \equiv \int d\Psi P(\Psi) |\Psi\rangle\langle\Psi|$. This is because, without the ability to determine the index j , measurements of an observable O would yield its expected value, averaged over the ensemble, i.e., $\int d\Psi P(\Psi) \langle\Psi|O|\Psi\rangle = \text{tr}(O\rho_{\mathcal{E}}^{(1)})$. Therefore, in this setting, one can operate only at the level of mixed states $\rho_{\mathcal{E}}^{(1)}$ [12–17].

Modern experiments challenge this assumption and may involve full knowledge of the ensemble $\mathcal{E}$. In modern quantum devices, one can determine the index j in a variety of settings. For example, in the ensembles we study, j will be the evolution time, which can be precisely controlled, or j will be the outcome of a projective measurement which is recorded by the experimenter. When the index j can be determined, we can study a wider range of properties, such as the *fluctuations* of the expectation values $\langle\Psi_j|O|\Psi_j\rangle$. The simplest such quantity is the variance $\int d\Psi P(\Psi) \langle\Psi|O|\Psi\rangle^2 - [\int d\Psi P(\Psi) \langle\Psi|O|\Psi\rangle]^2$. This quantity depends on more than the first moment $\rho_{\mathcal{E}}^{(1)}$. Specifically, it depends on the second moment $\rho_{\mathcal{E}}^{(2)} \equiv \int d\Psi P(\Psi) |\Psi\rangle\langle\Psi|^{\otimes 2}$, since the first term is equal to $\int d\Psi P(\Psi) \langle\Psi|O|\Psi\rangle^2 = \text{tr}(\rho_{\mathcal{E}}^{(2)} O^{\otimes 2})$. In general, the statistical properties of quantum state ensembles can be systematically characterized through their moments

$$\rho_{\mathcal{E}}^{(k)} \equiv \int d\Psi P(\Psi) |\Psi\rangle\langle\Psi|^{\otimes k}, \quad (1)$$

for positive integer k [64].

Recent work has illustrated that rich physical phenomena are encoded in the higher moments of state ensembles. As a concrete but nonexhaustive example, higher moment quantities are necessary to detect the measurement-induced phase transition in monitored quantum circuits [65–68]. In this setting, an ensemble of pure states is generated by the different outcomes of midcircuit measurements. A phase transition is observed in quantities such as the bipartite entanglement entropy averaged over states in the ensemble. Their average k th Rényi entropies can be computed from the k th moment of the ensemble, through the quantity $\int d\Psi P(\Psi) \text{tr}_A[\text{tr}_B(|\Psi\rangle\langle\Psi|)^k]$, where A and B are a bipartition of the system.

Without reference to a specific observable, the fluctuations of quantum state ensembles can be studied by examining how the states are distributed in Hilbert space. We quantify this spread through an *ensemble entropy*, which is the Shannon entropy of the probability distribution $P(\Psi)$. For continuous ensembles, this requires the choice of a reference distribution, in our case, the Haar ensemble. Thus, for continuous quantum state ensembles, the

ensemble entropy is given as the (negative) Kullback-Leibler (KL) divergence between $P(\Psi)$ and the Haar ensemble

$$\begin{aligned} \text{Ent}(\mathcal{E}) &\equiv -\mathcal{D}(\mathcal{E}||\text{Haar}) \\ &\equiv - \int d\Psi P(\Psi) \log_2 \left(\frac{P(\Psi)}{P_{\text{Haar}}(\Psi)} \right). \end{aligned} \quad (2)$$

We find that the ensembles we study are statistically described by universal continuous [69] state ensembles with maximum ensemble entropy, up to constraints imposed by energy conservation or effective constraints imposed by thermalization. We note that the ensemble entropy differs from the traditionally considered von Neumann entropy $S(\rho_{\mathcal{E}}^{(1)}) = -\text{tr}[\rho_{\mathcal{E}}^{(1)} \log_2 \rho_{\mathcal{E}}^{(1)}]$ which depends on only the average state $\rho_{\mathcal{E}}^{(1)}$. Instead, $\text{Ent}(\mathcal{E})$ depends on the distribution of the individual states $P(\Psi)$, which is determined by all moments of $\mathcal{E}$, and has distinct information-theoretic properties (Fig. 2). We also remark that, while similar notions have been studied in the past, e.g., in Refs. [53,54,70–72], this work explicitly shows the link between this entropy measure and ensembles of states obtained from dynamics, the first to do so to our knowledge.

In the following, we demonstrate this result by focusing on two specific settings—temporal and projected

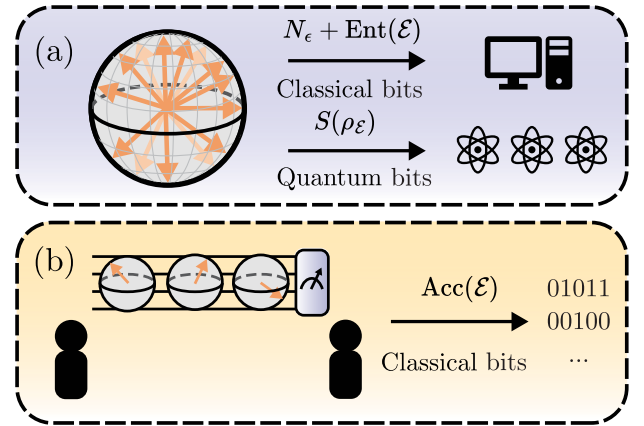


FIG. 2. Information-theoretic implications of our work. Our results indicate that two tasks are maximally difficult for our ensemble of states: (a) information compression of an ensemble of states $\mathcal{E}$. The ensemble entropy quantifies the task of classically representing M states from $\mathcal{E}$ up to precision ϵ (denoted here by a fine grid partitioning the Hilbert space into 2^{N_ϵ} elements). This task requires $M \cdot [N_\epsilon + \text{Ent}(\mathcal{E})]$ classical bits. This classical storage task differs from the analogous quantum task: Storing the same M states in quantum memory requires only $MS(\rho_{\mathcal{E}})$ qubits, where the $S(\rho_{\mathcal{E}})$ is the von Neumann entropy of the average state $\rho_{\mathcal{E}}$. (b) We also consider a task originally studied by Holevo [56]: sending classical information by sending states from $\mathcal{E}$ over a quantum channel. This channel capacity is given by the accessible information $\text{Acc}(\mathcal{E})$, which our ensembles of states uniquely minimize.

TABLE I. Summary of key results in this work. We describe the key ensembles considered in this work: the temporal ensemble and the projected ensembles generated by measurements in a basis $\{z_B\}$ under two different cases—when $\{z_B\}$ is *energy nonrevealing* and when it is general. Their statistical descriptions are probability distributions which have maximum entropy under appropriate constraints. We finally discuss their measurable signatures and information-theoretic properties.

Ensemble	Statistical description	Maximum entropy principle	Measurable signature	Information theoretic properties
Temporal ensemble	Random phase ensemble (Sec. IV)	Maximum entropy with fixed energy populations $\{ c_E ^2\}$	Porter Thomas (PT) dist. in outcome probabilities $p_z(t)$ (Theorem 3)	Constant mutual information [Eq. (43)] $I(Z; T) \approx (1 - \gamma)/\ln 2$
Projected ensemble: Energy nonrevealing basis	Scrooge ensemble (Sec. V)	Maximum entropy with fixed first moment ρ_A	PT dist. in joint probabilities $p(o_A, z_B)$ [Eq. (56)]	Minimal mutual information [Eq. (61)] $I(O_A; Z_B) \approx Q(\rho)$
General measurement basis	Generalized Scrooge ensemble (Sec. VI)	Collection of maximum entropy ensembles with fixed first moments $\rho_d(z_B)$	PT dist. in normalized probabilities $p_{o_A}(z_B)/p_{d,o_A}(z_B)$ [Eq. (72)]	Minimal interaction information [Eq. (76)] $I(O_A; Z_B; T) \approx \mathbb{E}_{z_B} Q[\bar{\rho}(z_B)]$

ensembles—and their relationship. Finally, we discuss implications of our results, summarized in Table I.

A. Temporal ensembles are random phase ensembles

The first ensemble of states we study is the *temporal ensemble*, consisting of quantum states generated by time evolution of an initial state $|\Psi_0\rangle$ under a Hamiltonian H [Fig. 3(a)], with probabilities being uniform on the time interval $t \in [-\tau/2, \tau/2]$:

$$\begin{aligned} \mathcal{E}_{\text{Temp}}(\tau) &\equiv \{(1/\tau, |\Psi(t)\rangle)\}_{t=-\tau/2}^{\tau/2} \\ &\equiv \{(1/\tau, \exp(-iHt)|\Psi_0\rangle)\}_{t=-\tau/2}^{\tau/2}. \end{aligned} \quad (3)$$

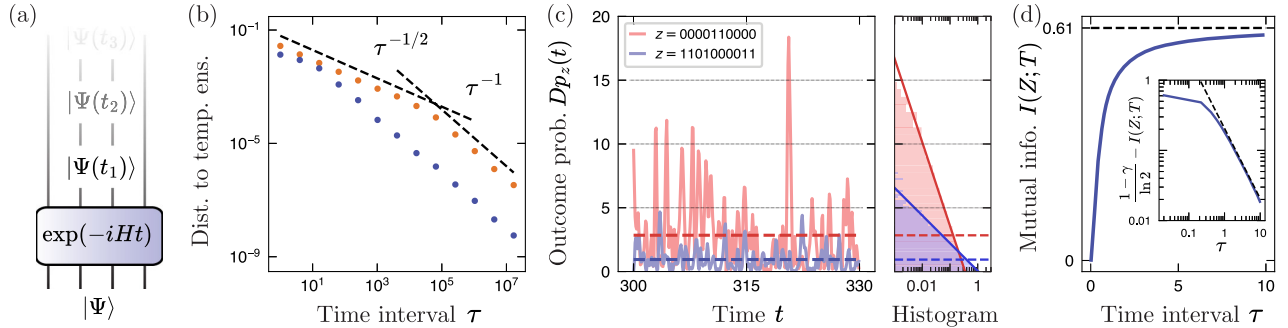


FIG. 3. Temporal ensembles from natural dynamics. (a) The temporal ensemble is the collection of states formed by time evolving a given state $|\Psi\rangle$ by a Hamiltonian H , for various times t . (b) Temporal ensembles formed over a finite interval of width τ approach the infinite-time ensemble as τ increases: The distance (here, the Frobenius norm) between their k th moment decreases as $1/\tau^{1/2}$, before decreasing as τ^{-1} . Data shown here for $k = 2$, for a state evolving under the mixed field Ising model (MFIM) in red, and a GUE-random Hamiltonian in blue. (c) Signatures of the temporal ensemble can be measured through the fluctuations of quantities such as the outcome probability $p(z, t) \equiv |\langle z|\Psi(t)\rangle|^2$ (normalized by a factor of D). Their histogram of values over time reveals a universal, Porter-Thomas (exponential) distribution. Histograms and time traces for two values of z shown in red and blue. (d) The mutual information $I(Z; T)$ as a function of the time interval τ of the ensemble. $I(Z; T)$ quickly saturates to the theoretical value of $(1 - \gamma)/\ln 2$, agreeing with the predicted $[1 - \gamma - \sqrt{\pi}/(2\sigma_H\tau)]/\ln 2$ [inset, dashed line, Eq. (44)]. In (c),(d), data are obtained from quench evolution in the MFIM, for the initial state in Eq. (37) with $\theta = 0.6$, and the units of time are determined by the $O(1)$ Hamiltonian parameters of Eq. (36).

time-dependent Hamiltonian dynamics cannot be formulated using conventional approaches based on the statistical properties of (quasi)energy eigenstates but instead can be characterized by the statistical properties of their trajectories. The central idea of Refs. [39,40] is to test if a time-evolved state uniformly visits all ambient points in the Hilbert space, a property dubbed (complete) Hilbert-space ergodicity (see also Ref. [76], which explores similar ideas under the name of “cyclic ergodicity” in quantum and classical systems).

Our work generalizes the notion of Hilbert-space ergodicity [39] to systems with energy conservation. The presence of energy conservation explicitly disallows ergodicity over all of Hilbert space. Nevertheless, we find that many-body dynamics under time-independent Hamiltonian evolution is still Hilbert-space ergodic, in the sense that the evolved states uniformly explore their available Hilbert space, which is constrained due to energy conservation. In order to understand the effects of energy conservation, consider the simple case of a single qubit initially polarized in the x direction, subject to a magnetic field along the z axis. Over time, the qubit periodically traces out an equatorial circle on the Bloch sphere instead of uniformly covering the sphere. In contrast, quantum dynamics without such conservation laws, specifically, a quasiperiodically driven system, can uniformly explore Hilbert space [39].

As in classical dynamics [77], we find that a quantum many-body system may be Hilbert-space ergodic without being quantum chaotic (i.e., being integrable). We show that Hilbert-space ergodicity holds as long as the system satisfies the *no-resonance conditions* (Definition 3), colloquially the absence of higher-order resonances in the spectrum (Definition 3). This condition is widely assumed to be true in chaotic systems [13,17,74,78], and we numerically observe that it is also true in certain integrable systems such as the XXZ model but not in others such as the transverse field Ising model.

To prove our claim of Hilbert-space ergodicity, we show that, in the $\tau \rightarrow \infty$ infinite interval limit, the temporal ensemble is equal to the *random phase ensemble* [46,79]. This is the ensemble of states with fixed amplitudes $|c_E| \equiv |\langle E | \Psi_0 \rangle|$ in some basis $\{|E\rangle\}$, here the energy eigenbasis, and with complex phases $\arg(c_E)$ independently random from the uniform distribution $\text{Unif}([0, 2\pi))$. In a time-evolved state, the amplitudes $|c_E|^2$ are fixed by the initial state and Hamiltonian, and the phases are the only degrees of freedom.

Theorem 1 (informal). The infinite-time temporal ensemble obtained by evolving an initial state is equal to a random phase ensemble if the Hamiltonian H satisfies all k th no-resonance conditions. Conversely, for almost every initial state, if the infinite-time temporal ensemble is equal to a random phase ensemble, the Hamiltonian H satisfies all k th no-resonance conditions.

The above theorem is somewhat of a folklore result [46,75,78], often stated without proof. Here, we provide a full statement and proof, for several reasons. First, Theorem 1 has a natural interpretation in terms of a maximum entropy principle. The random phase ensemble has maximum ensemble entropy [Eq. (2)] on this restricted Hilbert space, and, in Sec. IV B, we quantify the rate of convergence of the temporal ensemble to the random phase ensemble with increasing interval τ . Next, this leads to a novel type of universality in the many-body setting, which we outline below.

Unlike the above example of a two-level system, the exponentially larger Hilbert space in many-body systems leads to behavior much more akin to ergodicity in the full Hilbert space. Specifically, we show that the trajectory appears pseudorandom in the following sense: In the limit of large Hilbert-space dimension D , the temporal ensemble is *statistically equal* to the Haar ensemble, *deformed* as follows:

$$|\Psi(t)\rangle^{\text{stat}} \approx \sqrt{D\rho_d} |\phi\rangle, \quad |\phi\rangle \sim \text{Haar}, \quad (4)$$

where $\rho_d \equiv \sum_E |E\rangle \langle E | \Psi_0 \rangle \langle \Psi_0 | E \rangle \langle E|$ is the diagonal ensemble. By “statistically equal,” we mean that Eq. (4) is not true on the level of wave functions, but the moments $\rho^{(k)}$ of the ensembles are very close. To show this, we develop a simple analytic expression for the moments of the temporal ensemble that holds in the $D \rightarrow \infty$ limit (Sec. IV A). One consequence of this pseudorandomness is our finding that the measurement probabilities, upon appropriate rescaling, universally follow the Porter-Thomas distribution (below and Sec. IV C), a hallmark of random quantum states which had been observed in quantum dynamics without conservation laws [80]. We discuss information-theoretic aspects of this result and also prove an analog of the ergodic theorem, a result in the theory of dynamical systems [81], where we relate the statistics defined over time and configuration space (discussed below and in Sec. IV E). This ergodic theorem is essential for translating our results from the temporal ensemble to the projected ensemble, to which we now turn.

B. Projected ensembles are (generalized) Scrooge ensembles

Not only do temporal ensembles satisfy the maximum entropy principle, so do *projected ensembles*. Projected ensembles, introduced in a modern, many-body context by Refs. [28,29] (previously studied in Ref. [82] and under the name of “geometric quantum states” in Refs. [54,70–72]), are ensembles of states obtained from a single, larger many-body state $|\Psi\rangle$. To do so, we partition the system into two parts A and B , such that the Hilbert space factorizes (for simplicity of discussion) as $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. We then measure B in a basis $\{|z_B\rangle\}$. Each measurement outcome

$|z_B\rangle$ projects $|\Psi\rangle$ into a distinct pure state $|\Psi(z_B)\rangle \in \mathcal{H}_A$ [Fig. 1(d)], defined as

$$|\Psi(z_B)\rangle \equiv (\mathbb{I}_A \otimes \langle z_B|)|\Psi\rangle / \sqrt{p(z_B)}, \quad (5)$$

where the normalization constant is the probability of obtaining the outcome z_B , $p(z_B) \equiv \langle \Psi | (\mathbb{I}_A \otimes |z_B\rangle\langle z_B|) | \Psi \rangle$. This defines the projected ensemble:

$$\mathcal{E}_{\text{Proj}} = \{p(z_B), |\Psi(z_B)\rangle\}_{z_B}. \quad (6)$$

For concreteness, we primarily study global states $|\Psi\rangle = \exp(-iHt)|\Psi_0\rangle$ obtained from time evolution by a generic, ergodic many-body Hamiltonian H , which rapidly reach thermal equilibrium at a temperature set by the energy of the initial state [51,52]. We also study eigenstates of H that exhibit similar thermal properties.

In Refs. [28,29], it was found that when the state $|\Psi\rangle$ is obtained from time evolution under a chaotic many-body Hamiltonian at infinite effective temperature, the projected ensemble of small subsystems is very well approximated by the Haar ensemble. This phenomenon is known as *deep thermalization* [30]. This has since been rigorously shown in a variety of settings, including deep random circuits [29,36], dual-unitary models [30–32,83,84], free-fermion models [33], or Hamiltonian dynamics with certain assumptions [34]. Various aspects including the rate of deep thermalization [30,31,36,84], the effect of symmetries [35,37], and the presence of magic in the global state [38] have been investigated. The majority of what we know so far is restricted to the setting of infinite-temperature dynamics.

Understanding behavior at finite temperature is intricately linked to understanding the effects of energy conservation and requires a conceptual leap from the infinite-temperature case. Indeed, infinite-temperature dynamics can occur in the absence of any conservation law, in which there are models of dynamics such as random unitary circuits (RUCs) which are easily analyzed. Therefore, more is known of dynamics at infinite temperature than at finite temperature. As an example, the linear growth of entanglement entropy can be rigorously proven for models of dynamics at infinite temperature (for example, in RUCs [85]) but remains a matter of conjecture and numerical observation in the presence of energy conservation [86,87], save for simple models such as quadratic free-fermion Hamiltonians [88]. In this work, we set out to study the projected ensembles generated by finite-temperature dynamics.

At finite temperature, projected ensembles are not close to the Haar ensemble. To see this, note that the first moment of the projected ensemble $\rho_{\mathcal{E}}^{(1)} \equiv \sum_{z_B} p(z_B) |\Psi(z_B)\rangle\langle \Psi(z_B)| = \text{tr}_B[|\Psi\rangle\langle \Psi|]$ is precisely equal

to the reduced density matrix (RDM). In turn, only at infinite temperature is the RDM equal to the first moment of the Haar ensemble, the maximally mixed state $\mathbb{I}_A/D_A$. This follows from a standard result in quantum thermalization [74] that the reduced density matrix ρ_A is well described by a Gibbs state, i.e., $\rho_A \propto \exp(-\beta H_A)$, where H_A is the Hamiltonian truncated to the subsystem A and β^{-1} is an effective temperature. It has remained an open question as to how to describe these projected ensembles at finite temperature.

Our results at finite temperature are threefold. First, on the conceptual level, we identify an organizing principle for projected ensembles under general conditions: the maximum entropy principle. In our setting, this prescribes the Scrooge ensemble as the appropriate finite-temperature generalization of the Haar ensemble, but the principle may hold under more general conditions. Next, we elucidate new physical phenomena: We find a new, provable setting where ergodic quantum dynamics maximally hides information (elaborated on below), and we also describe the interplay between measurement basis and conserved quantities. This leads us to the generalized Scrooge ensemble as our most general description of projected ensembles. Finally, on the technical level, we develop the first proof of deep thermalization in time-independent Hamiltonian dynamics (at finite or infinite temperature), based on our results of the temporal ensemble.

For projected ensembles at finite temperature, the maximum entropy principle translates to our result: Under an “energy-nonrevealing” condition on the measurement basis $\{|z_B\rangle\}$ which we discuss below, the projected ensembles are described by the ensembles that have maximum entropy, over all ensembles with a given first moment. These ensembles, termed *Scrooge ensembles* [89–91], were introduced for their unique information-theoretic properties and have been studied in the context of projected ensembles in various limiting cases [29,31,90]. [92] Each density matrix ρ uniquely defines a Scrooge ensemble $\text{Scrooge}(\rho)$, and our result is that, at finite temperatures, the projected ensemble is the Scrooge ensemble corresponding to the appropriate thermal RDM.

Claim 1 (corollary of Theorem 4 below, informal). The projected ensemble is statistically described by the Scrooge ensemble, if the measurement basis $\{|z_B\rangle\}$ is energy nonrevealing.

The Scrooge ensemble may be essentially understood as a deformation of the Haar ensemble such that its first moment equals ρ [89] [Fig. 4(b)]. Specifically, we can form the Scrooge ensemble by first sampling Haar-random states $|\psi\rangle \sim \text{Haar}$ and then *deforming* them by a given density matrix ρ , i.e.,

$$|\Phi\rangle \sim \text{Scrooge}(\rho) \Leftrightarrow |\Phi\rangle \propto \sqrt{\rho}|\psi\rangle, \quad |\psi\rangle \sim \text{Haar}. \quad (7)$$

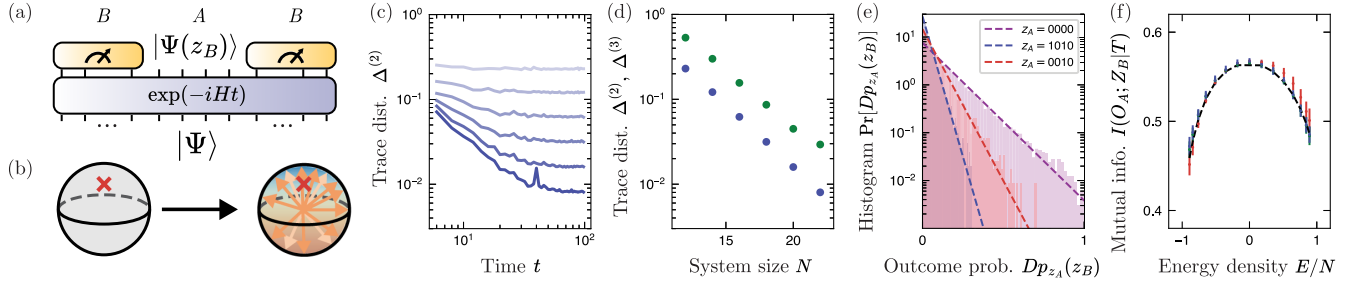


FIG. 4. Projected ensembles from natural dynamics. (a) The projected ensemble is formed by performing projective measurements on a large subsystem B of a many-body state, here obtained from time evolution of an initial state. (b) In this setting, the projected ensemble is described by the Scrooge ensemble, which is a unique probability distribution of states (color gradient) for each value of the first moment (red cross, in our setting equal to the reduced density matrix). (c) At constant energy density $E/N \approx 0.5$, the trace distance between the second moments of the projected and Scrooge ensembles decreases with quench evolution time before saturating to an equilibrium value which depends on system size. We time evolve a product state with the MFIM on a one-dimensional open chain of length N [Eq. (36)], from $N = 12$ (lighter) to $N = 22$ (darker) and fix the subsystem A to be four sites near the center of the chain. (d) The equilibrium trace distance decreases exponentially with system size for both second (blue) and third moments (green). (e) A signature of the Scrooge ensemble is the emergence of the Porter-Thomas distribution in measurement outcome distributions; here, we plot a histogram of the normalized outcome distributions $Dp_{z_A}(z_B)$, with fixed z_A and different values of z_B . Histograms for three different values of z_A (blue, red, and purple) follow PT distributions with different means $D\mathbb{E}_{z_B}[p_{z_A}(z_B)]$. (f) The mutual information $I(O_A; Z_B|T)$ (conditioned at fixed times) is an information-theoretic signature of the Scrooge ensemble: Its value is independent of the basis O_A and is equal to the subentropy $Q[\rho_A]$ [Eq. (61)] (black dashed line). Here, we plot $I(O_A; Z_B|T)$ for three different observables $O_A = \{X^{\otimes 4}, Y^{\otimes 4}, Z^{\otimes 4}\}$ (red, green, and blue, respectively), against the energy density E/N of the state, controlled by the polarization angle θ of the initial state [Eq. (37)].

The probability $p(\Phi)$ is further weighted by the norm $\langle \psi | \rho | \psi \rangle$, leading to the distribution of states

$$\mathcal{E}_{\text{Scrooge}}(\rho) = \left\{ \begin{array}{l} p(\Phi) = D\langle \psi | \rho | \psi \rangle, \\ |\Phi\rangle = \frac{\sqrt{\rho}|\psi\rangle}{\langle \psi | \rho | \psi \rangle^{1/2}} \end{array} \right\}_{|\psi\rangle \sim \text{Haar}}. \quad (8)$$

In Sec. VC, we give closed-form expressions for the resultant probability distribution function $p(\Phi)$ as well as its k th moments.

Scrooge ensembles do not always describe the projected ensemble. They do so only in the special case when the measurement basis z_B is uncorrelated with the energy, defined precisely in Sec. VID. Under these circumstances, we can treat the projected states $|\Psi(z_B)\rangle$ as independent, identical samples from a common distribution. However, this is not always true: When z_B is correlated with the subsystem energy, it is also correlated with the projected state $|\Psi(z_B)\rangle$, such that the energies $E_A \equiv \langle \Psi(z_B) | H_A | \Psi(z_B) \rangle$ and $E_B \equiv \langle z_B | H_B | z_B \rangle$ are anticorrelated. Under these general conditions, we introduce the *generalized Scrooge ensemble* to describe the projected ensemble (Sec. VI). Each projected state $|\Psi(z_B)\rangle$ should be treated as a sample from a *different* Scrooge ensemble for each z_B , i.e., where the first moment is not the RDM ρ_A but is a z_B -dependent mixed state $\bar{\rho}(z_B)$, which we may determine by the time average of the projected state, $\mathbb{E}_t[|\Psi(z_B, t)\rangle \langle \Psi(z_B, t)|]$. That is to say, instead of a collection of independent, identical samples from a fixed distribution, i.e., $|\Psi(z_B)\rangle \stackrel{\text{i.i.d.}}{\sim} \text{Scrooge}[\rho]$, the generalized

Scrooge ensemble is a statistical superposition of this collection of Scrooge ensembles, such that $|\Psi(z_B)\rangle \sim \text{Scrooge}[\bar{\rho}(z_B)]$. Therefore, the generalized Scrooge ensemble is more accurately a collection of ensembles, where the projected ensemble is comprised of exactly one sample from each ensemble, but we abuse notation and refer to it as an ensemble. This is our second key result.

Theorem 4 (informal). At sufficiently long times t , [93] the projected ensemble generated by a time-evolved state $|\Psi(t)\rangle$ is approximately equal to the generalized Scrooge ensemble, which depends on the measurement basis $\{z_B\}$, Hamiltonian H , and initial state $|\Psi_0\rangle$, if the Hamiltonian H satisfies all k th no-resonance conditions. The converse direction holds for almost every initial state $|\Psi_0\rangle$.

When the measurement basis $\{z_B\}$ is *uncorrelated* with the global energy, the generalized Scrooge ensemble reduces to the Scrooge ensemble (Claim 1). Specifically, this happens when the mixed state $\bar{\rho}(z_B)$ is independent of z_B (up to a rescaling), giving a concrete definition for “energy-nonrevealing” bases.

Despite the complexity of the generalized Scrooge ensemble, Theorem 4 remains a nontrivial statement with no free parameters. The reduced density matrices $\{\bar{\rho}(z_B)\}$ are determined by the initial state, Hamiltonian, and measurement outcomes $\{z_B\}$. Furthermore, Theorem 4 is testable: We can fix a typical point in time and compute the k th moment of the projected ensemble. While each individual state $|\Psi(z_B, t)\rangle$ is a single sample from a random ensemble and exhibits large fluctuations between points in time, in evaluating the moments of the projected ensemble,

we sum over a large number of such projected states. Fluctuations over the projected ensemble are averaged away, and the value of the low moments does not depend on the point in time. Theorem 4 states that this k th moment is equal to the weighted sum of the k th moments of the Scrooge ensembles $\text{Scrooge}[\bar{\rho}(z_B)]$, which we corroborate with numerical evidence.

We prove this result using our knowledge of the temporal ensemble. We show that the temporal trajectory of the projected state $|\Psi(z_B, t)\rangle$ is a Scrooge ensemble (here, because of the subsystem projection, there are no hard constraints on the amplitudes, as there were in the case of the global state). We then utilize our ergodic theorem (Sec. IV E) to translate our results for a distribution over time to a distribution over the outcomes z_B . Previously, deep thermalization had been proven only in certain special models of quantum dynamics [31,83] and was numerically observed to arise in projected ensembles generated by Hamiltonian time evolution. Our results enable us to definitely establish that deep thermalization occurs in generic Hamiltonian dynamics and also describe its properties beyond infinite temperature.

While Theorem 4 concerns time-evolved states $|\Psi(t)\rangle$, we numerically observe that the generalized Scrooge ensemble also describes the projected ensembles obtained from energy eigenstates, suggesting that our maximum entropy principle holds for a broader class of many-body states.

In projected ensembles, the maximum entropy principle sheds light on the nature of correlations between system and bath. Specifically, the projected ensemble reveals correlations between the subsystem A and its complement B , which we quantify in terms of the mutual information. We find that the *nontrivial* correlations between them are minimal and basis independent. Here, nontrivial correlations mean any amount of correlation (mutual information) that is beyond the thermal correlations that remain after infinite-time averaging. In other words, any meaningful correlations between A and B originate only from energy conservation and other symmetries if they exist. Below and in Secs. V H and V IC, we discuss consequences of these correlations for tasks such as communication and compression.

C. Implications and discussion

Below, we summarize several broader points that we found applicable to both the temporal and projected ensemble.

1. Ergodic theorem

Remarkably, the temporal and projected ensembles are closely related. We prove that a version of the ergodic theorem holds in many-body dynamics. This enables us to translate statements about distributions over time (temporal ensemble) into statements about distributions over

measurement outcomes, at fixed, typical points in time. In particular, we use this to prove Theorem 4, a statement about the projected ensembles at typical points in time. In practice, for quantities of interest, we observe the ergodic theorem to hold at short, physically relevant timescales.

2. Operational meaning of ensemble entropy

In this work, we use the ensemble entropy to define our maximum entropy principle. We also identify its operational interpretation: The ensemble entropy quantifies the information required to classically represent an ensemble of states. Our maximally entropic ensembles are, therefore, the most difficult to compress, formalizing our intuition that the states generated by natural, chaotic many-body dynamics lack overall structure that can be exploited for compression.

Specifically, the ensemble entropy quantifies the number of classical bits required to store an ensemble of states up to numerical precision ϵ [Fig. 2(a)]. This should be contrasted with the minimum number of *qubits* required to store the ensemble in quantum memory, which was established by Schumacher to depend only on the first moment $\rho_\mathcal{E}$ of the ensemble through the von Neumann entropy $S(\rho_\mathcal{E})$ [94].

3. Information-theoretic consequences

Our work introduces concrete, observable ways in which generic quantum dynamics “maximally hides” information. One question of study in quantum dynamics is to how information evolves in many-body quantum systems [95,96]. This may address fundamental questions such as the tensions between the apparent irreversibility of quantum dynamics and its unitarity and may also have bearing on applications including quantum error correction [97]. In many existing works, the notion of information hiding has been probed only indirectly through proxy quantities such as the entanglement entropy [95] and out-of-time-ordered correlators [96].

Here, we introduce the accessible information as a good measure to quantify information hiding in a specific, operational task and also prove that, in generic Hamiltonian dynamics, this information measure is minimized. Not only are our ensembles of states maximally difficult to classically store, they are also maximally difficult to distinguish by measurement, as quantified by their *accessible information* [63,89]. The accessible information quantifies the maximum rate of classical information that can be transmitted by sending states from the ensemble over a quantum channel (in terms of bits of information per channel use) [Fig. 2(b)]. It is upper bounded by the well-known Holevo bound [56] and lower bounded by Jozsa, Robb, and Wootters [89]. They found that the Scrooge ensemble achieves this lower bound, dubbed the *subentropy* $Q(\rho)$ [Eq. (61)]. They dubbed the ensemble the “Scrooge ensemble” because it is uniquely “stingy” in the amount of information revealed under measurement. Therefore, the

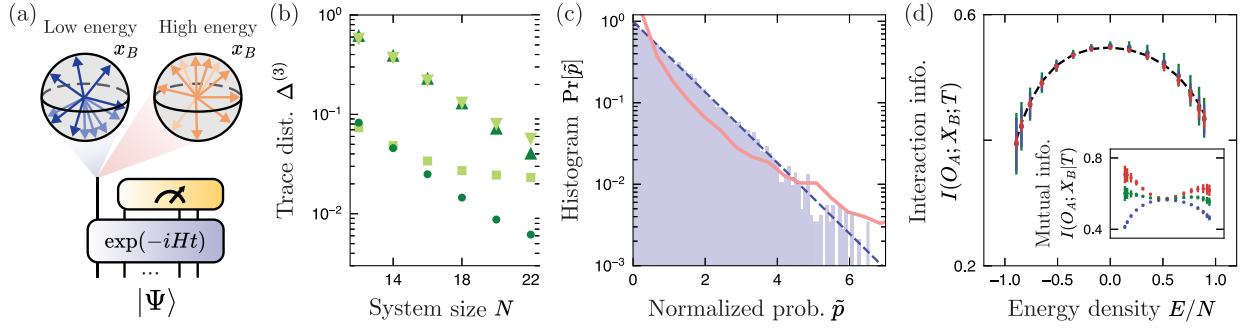


FIG. 5. Generalized Scrooge ensembles describe projected ensembles. (a) When the measurement basis on B (here denoted X_B) is correlated with the total energy of the system, the projected states $|\Psi(x_B)\rangle$ are correlated with the measurement outcome x_B , here illustrated by differences between outcomes with “low” (orange) and “high” (blue) energies. This leads to deviations from the Scrooge ensemble, and the projected ensemble is instead described by a generalized Scrooge ensemble (Claim 2). (b) We verify this conjecture numerically with the trace distance $\Delta^{(3)}$ between the third moments of the projected ensemble and either the Scrooge ensemble (squares and downward triangle) or the generalized Scrooge ensemble (circles and upward triangles). We consider projected ensembles of states time evolved (for sufficiently long times) under the MFIM [Eqs. (36) and (37)] and measured in the X_B basis. The difference between the two ensembles is more obvious for subsystems of two qubits (squares and circles) than for subsystems of four qubits (upward and downward triangles), due to the larger bath relative to the system size. While the distance to the Scrooge ensemble saturates to a nonzero value, the distance to the generalized Scrooge ensemble continues to decrease. (c) Consistent with the generalized Scrooge ensemble, the histogram (blue) of the rescaled probabilities $\tilde{p}(x_B) \equiv p_{x_A}(x_B)/\mathbb{E}_t[p_{x_A}(x_B)]$ follows a PT distribution, while the bare probabilities $Dp_{x_A}(x_B)$ (red line) do not, with $x_A = 0000$ plotted here as a representative. (d) As with the mutual information in Scrooge ensembles, in generalized Scrooge ensembles the interaction information $I(O_A; X_B; T)$ displays a universal behavior, independent of basis O_A , and agrees with the *weighted sum of subentropies* [Eq. (76)] (dashed line). Inset: In contrast, the mutual information $I(O_A; X_B|T)$ is no longer independent of the basis O_A , here illustrated in red, green, and blue, respectively, by $O_A = \{X^{\otimes 4}, Y^{\otimes 4}, Z^{\otimes 4}\}$, due to time-averaged (thermal) correlations between the measurement bases O_A and X_B .

projected ensemble not only has maximum ensemble entropy, it also minimizes the amount of information that can be extracted. These information-theoretic properties are directly measurable [Figs. 4(f) and 5(d)] but have not yet been observed in experiment.

In the case of temporal ensembles, the accessible information quantifies the amount of information that can be gained about the evolution time of the state. We find that it takes on a universal value $(1 - \gamma)/\ln 2 \approx 0.61$, where γ is the Euler-Mascheroni constant [Fig. 3(d)].

4. Measurable signature: Universal Porter-Thomas distribution

Our maximum entropy principle selects distributions of states that are “closest” to the Haar ensemble, satisfying certain constraints. The relationship between our ensembles of states to the Haar ensemble gives rise to a measurable signature of our maximum entropy principle: the *Porter-Thomas (PT) distribution* [98]. Originally identified in nuclear physics settings, the PT distribution has received renewed attention in many-body physics. This distribution represents the distribution of overlaps between Haar-random states $|\psi\rangle$ with a fixed (but arbitrary) state $|\Phi\rangle$, i.e., the probability $p(\Phi) \equiv |\langle\Phi|\psi\rangle|^2$ of measuring $|\Phi\rangle$ from a Haar-random state $|\psi\rangle$ [99]. Not only is the PT distribution a measurable signature of Haar-random states, we find that it is also present in many natural many-body

states. In companion work [41], we experimentally and theoretically investigate the emergence of the Porter-Thomas distribution in global quantities, its relationship to local quantities, and how the Porter-Thomas distribution is modified in the presence of noise, using our results to learn about the noise in an experiment.

D. Organization of paper

The rest of the paper is organized as follows. In Sec. III, we provide an overview of the central object of our work—ensembles of pure states and their information-theoretic and measurable properties. We then proceed to study our two ensembles of states: the temporal ensemble (Sec. IV) and the projected ensemble in a special and a general setting (Secs. V and VI, respectively). Finally, in Sec. VII, we discuss the PT distribution in the properties of many-body eigenstates.

III. ENSEMBLES OF PURE STATES

Ensembles of pure states generated by quantum many-body dynamics are the central objects of study in this work. In this section, we provide an overview of ensembles of pure states and their properties. After defining normalized and unnormalized ensembles of states, we introduce two quintessential ensembles: the Haar and Gaussian ensembles, which we use as the reference distributions for our ensemble entropy. We then discuss the quantities

we use to analyze our ensembles of states: the k th statistical moments, ensemble entropy, mutual information, and PT distribution.

Ensembles of pure states are collections of pure states in a Hilbert space of dimension D , weighted by a probability distribution.

Definition 1. An ensemble of pure states is a probability distribution of states $|\Psi_j\rangle$ weighted by probabilities p_j :

$$\mathcal{E} = \{p_j, |\Psi_j\rangle\}; \quad (9)$$

when the p_j 's are omitted, we take the states $|\Psi_j\rangle$ to be uniformly distributed over indices j [100].

Our distributions may be continuous or discrete. In the continuous case, the distribution is a measure over normalized wave functions, which we can specify in terms of its components $\Psi_z \equiv \langle z | \Psi \rangle$, subject to the constraint that $\sum_z |\Psi_z|^2 = 1$:

$$P(\Psi)d\Psi = \int P(\{\Psi_z\})\delta(\|\Psi\|^2 - 1)\prod_z d^2\Psi_z, \quad (10)$$

where $\prod_z d^2\Psi_z$ is the Euclidean measure on $\mathbb{C}^{\otimes D}$.

A. Unnormalized ensembles of states

We also find it useful to define distributions of *unnormalized* states. In this case, the distribution is simply a measure over D -dimensional complex vectors, defined in terms of its components $\tilde{\Psi}_z \equiv \langle z | \tilde{\Psi} \rangle$. For consistency, we use tildes throughout this work to denote unnormalized states $|\tilde{\Psi}\rangle$.

An unnormalized ensemble can be mapped to a normalized one by normalizing each state. This results in the probability distribution $P(\Psi)d\Psi = \int d\tilde{\Psi}P(\tilde{\Psi})\delta(\|\tilde{\Psi}\|/\|\tilde{\Psi}\| - |\Psi\rangle)$. For each state $|\Psi\rangle$, the normalized ensemble averages over the distribution of the norm $\|\tilde{\Psi}\|$ to give the measure $P(\Psi)$. Therefore, the unnormalized ensemble contains more information than the normalized one: specifically, information about fluctuations of the norm $\|\tilde{\Psi}\|$.

B. Gaussian and Haar-random ensembles of states

The ensemble entropy, which quantifies our maximum entropy principle, is defined with reference to the Haar (or Gaussian) ensemble. These ensembles are unique in being invariant under arbitrary unitary transformations and have been the subject of considerable study as paradigmatic distributions of random quantum states [48].

1. Random unnormalized states: The Gaussian ensemble

The Gaussian ensemble is a probability distribution over unnormalized states, where each component $\tilde{\Psi}_z \equiv \langle z | \tilde{\Psi} \rangle$ is an independent complex Gaussian variable:

$$P_{\text{Gauss}}(\tilde{\Psi})d\tilde{\Psi} = \prod_z \frac{D}{\pi} \exp[-D|\tilde{\Psi}_z|^2]d^2\tilde{\Psi}_z. \quad (11)$$

The normalization above is chosen so that the average norm is $\mathbb{E}[\langle \tilde{\Psi} | \tilde{\Psi} \rangle] = 1$.

2. Random normalized states: The Haar ensemble

The Haar ensemble is the unique distribution of normalized states that is invariant under any unitary transformation U . It is naturally related to the Gaussian ensemble: One can sample from the Haar ensemble by sampling states $|\Psi\rangle$ from the Gaussian ensemble and then normalizing them. This induces the measure

$$P_{\text{Haar}}(\Psi)d\Psi = \int P_{\text{Gauss}}(\tilde{\Psi})\delta(\|\Psi\| - \|\tilde{\Psi}\|/\|\tilde{\Psi}\|)d\tilde{\Psi}. \quad (12)$$

The Haar ensemble has been widely studied in quantum information science and utilized for a myriad of applications; see, e.g., Ref. [48]. In particular, the moments of the Haar ensemble play a key role, which we now discuss.

C. Statistical k th moments

An ensemble of states contains a large amount of information, beyond that of its density matrix. In this work, we systematically study and quantify ensembles of states by their higher-order moments.

Higher-order quantities have received less attention in many-body physics, because they have been inaccessible to traditional experiments, requiring a large number of controlled repetitions in order to access higher-order statistical properties. Modern quantum devices address these limitations with both high repetition rates and a high degree of controllability [28,101].

In this work, we find that not only do the average properties (first moments) thermally equilibrate, so too do the higher moments of our natural ensembles of states. We first define the k th moments.

The k th *moment* of an ensemble $\mathcal{E}$ is, for discrete ensembles,

$$\rho_{\mathcal{E}}^{(k)} \equiv \sum_j p_j |\Psi_j\rangle \langle \Psi_j|^{\otimes k}, \quad (13)$$

for continuous ensembles is

$$\rho_{\mathcal{E}}^{(k)} \equiv \int d\Psi P(\Psi) |\Psi\rangle \langle \Psi|^{\otimes k}, \quad (14)$$

and for unnormalized ensembles (denoting the moments with tildes) is

$$\tilde{\rho}_{\mathcal{E}}^{(k)} = \int d\tilde{\Psi} P(\tilde{\Psi}) |\tilde{\Psi}\rangle \langle \tilde{\Psi}|^{\otimes k}. \quad (15)$$

The k th moment of an ensemble can be used to describe the expectation value of any linear operator $O^{(k)}: \mathcal{H}^{\otimes k} \rightarrow \mathcal{H}^{\otimes k}$ which acts on k copies of the Hilbert

space. The ensemble average of $O^{(k)}$ can be obtained from the k th moment: $\mathbb{E}_{|\Psi\rangle \sim \mathcal{E}}[|\Psi\rangle^{\otimes k} O^{(k)} |\Psi\rangle^{\otimes k}] = \text{tr}[O^{(k)} \rho_{\mathcal{E}}^{(k)}]$. For example, $O^{(k)} = O^{\otimes k}$ describes the k th moment of the expectation values $\mathbb{E}_{\mathcal{E}}[|\Psi\rangle O |\Psi\rangle^k]$. Here, we use $\mathbb{E}_{\mathcal{E}}[f(\Psi)]$ to denote the ensemble average $\int d\Psi P(\Psi) f(\Psi)$.

The k th moment can be thought of as a generalization of the density matrix. Namely, the first moment $\rho^{(1)} = \sum_j p_j |\Psi_j\rangle \langle \Psi_j|$ is the conventional density matrix and describes the average state of the ensemble. All higher moments $\rho_{\mathcal{E}}^{(k)}$ are likewise positive semidefinite operators and, for ensembles of normalized states, have unit trace.

1. Moments of Haar ensemble

The higher moments of the Gaussian and Haar ensembles are particularly simple. The moments of the Haar ensemble have analytical expression [48]:

$$\rho_{\text{Haar}}^{(k)} \equiv \int_{\text{Haar}} d\Psi |\Psi\rangle \langle \Psi|^{\otimes k} = \frac{\sum_{\sigma \in S_k} \text{Perm}(\sigma)}{\prod_{i=0}^{k-1} (D+i)}, \quad (16)$$

with similar expression for the moments of the Gaussian ensemble. Here, $\text{Perm}(\sigma)$ is the operator that permutes k states by the permutation $\sigma \in S_k$, i.e., $\text{Perm}(\sigma)[|\Psi_1\rangle \otimes \cdots \otimes |\Psi_k\rangle] = |\Psi_{\sigma(1)}\rangle \otimes \cdots \otimes |\Psi_{\sigma(k)}\rangle$. This result follows from the Schur-Weyl duality [102]. Its simplicity enables the design of useful quantum algorithms such as classical shadow tomography [103,104] as well as to analytical solutions of models of quantum dynamics such as random quantum circuits [65]. We find that our ensembles of states have k th moments with similar structure as the Haar k th moments, enabling phenomena such as the emergence of the Porter-Thomas distribution.

2. Trace distance of higher moments

Higher moments also enable the systematic and quantitative comparison of two ensembles of states. Specifically, we compare their k th moments through their *trace distance*:

$$\Delta_{\text{Haar}}^{(k)} \equiv \frac{1}{2} \|\rho_{\mathcal{E}}^{(k)} - \rho_{\text{Haar}}^{(k)}\|_*, \quad (17)$$

where the trace (or nuclear) norm $\|\cdot\|_*$ of a matrix is the sum of the absolute values of its eigenvalues. In particular, the trace distance sets an upper bound on how well any k -copy observable $O^{(k)}$ can distinguish two ensembles of states [63].

References [28,29] found that, with increasing system size, projected ensembles of infinite effective temperature states—either obtained from quench dynamics of simple initial states or chosen from high-energy eigenstates—have closer k th moments to the Haar moment (in other words, projected ensembles form better approximate k -designs). Analogously, we find that the k th moments of the projected

and (generalized) Scrooge ensembles converge in trace distance with increasing system size [Figs. 4(b) and 5(b)].

D. Ensemble entropy

The central claim of our work is that natural many-body dynamics generates ensembles of states with maximum entropy. The entropy in question, the ensemble entropy, differs from the conventional von Neumann entropy $S(\rho) \equiv -\text{tr}[\rho \log(\rho)]$. This is not suitable as an entropy of an ensemble of states, since it depends on only the first moment ρ , and there are many ensembles of states which share the same average state ρ . Instead, our ensemble entropy is the Shannon entropy of the distribution of states over Hilbert space. For continuous distributions, this entropy must be defined relative to a reference distribution.

Definition 2. The *ensemble entropy* $\text{Ent}(\mathcal{E})$ is the negative of the KL divergence between $\mathcal{E}$ and the reference measure. For normalized ensembles, the reference distribution is the Haar measure [Eq. (2)]

$$\begin{aligned} \text{Ent}(\mathcal{E}) &\equiv -\mathcal{D}(\mathcal{E}||\text{Haar}) \\ &= - \int d\Psi P(\Psi) \log_2 \left(\frac{P(\Psi)}{P_{\text{Haar}}(\Psi)} \right), \end{aligned} \quad (18)$$

while, for unnormalized ensembles, the reference distribution is the Gaussian measure

$$\begin{aligned} \text{Ent}(\mathcal{E}) &\equiv -\mathcal{D}(\mathcal{E}||\text{Gauss}) \\ &= - \int d\tilde{\Psi} \tilde{P}(\tilde{\Psi}) \log_2 \left(\frac{P(\tilde{\Psi})}{P_{\text{Gauss}}(\tilde{\Psi})} \right). \end{aligned} \quad (19)$$

The ensemble entropy is nonpositive, achieving zero only when the two distributions are equal. See Refs. [53,54] for similar attempts to quantify the entropy of state ensembles.

1. Connection to von Neumann entropy

While the ensemble entropy may seem an unconventional notion of entropy, in fact it can be related to the von Neumann entropy of higher moments of the ensemble. For finite ensembles $\mathcal{E} = \{p_j, |\Psi_j\rangle\}$, the quantity

$$\lim_{k \rightarrow \infty} S(\rho_{\mathcal{E}}^{(k)}) \quad (20)$$

converges to the Shannon entropy $-\sum_j p_j \log_2 p_j$, which, in turn, converges to the ensemble entropy (plus a divergent constant), when the discrete ensemble is appropriately selected (see below). This follows from the fact that, with increasing number of copies k , two states $|\psi\rangle^{\otimes k}$ and $|\phi\rangle^{\otimes k}$ become increasingly distinguishable (orthogonal), since $|\langle \phi | \psi \rangle|^{2k}$ quickly vanishes when $|\langle \phi | \psi \rangle| \neq 1$. Therefore,

for any discrete ensemble $\{p_j, |\Psi_j\rangle\}$, the quantities p_j and $|\Psi_j\rangle^{\otimes k}$ converge to the eigenvalues and eigenvectors of $\rho^{(k)}$, respectively. A similar argument can be made for continuous distributions, and we obtain $\text{Ent}(\mathcal{E}) = \lim_{k \rightarrow \infty} S(\rho_{\mathcal{E}}^{(k)}) + C$, with a divergent constant C .

2. Ensemble entropy and information compression

The ensemble entropy has a simple operational meaning: It quantifies the minimum number of bits required to store the states in the ensemble in classical memory, up to a *fixed precision* ϵ . If storing a single such state requires N_ϵ classical bits, we find that storing a *large* number of states M drawn from the ensemble $\mathcal{E}$ requires $M[N_\epsilon + \text{Ent}(\mathcal{E})]$ classical bits [note that $\text{Ent}(\mathcal{E}) \leq 0$], as illustrated in Fig. 2(a). Our maximum entropy principle indicates that the required storage for many-body states generated by time evolution is maximal.

To see this, note that storing states up to accuracy ϵ amounts to partitioning the Hilbert space in a uniform ϵ net, i.e., into a discrete set of points which are a distance ϵ apart [3]. To store the states in the ensemble, each state in the continuous Hilbert space is assigned to one of these discrete points. Storing these states, thus, amounts to storing these labels—they may either be stored directly or compressed according to how often the labels appear. In the former case, each label requires N_ϵ bits, if there are 2^{N_ϵ} points in the net. In the latter, the optimal rate of compression is given by the relative entropy between the discrete probabilities on this ϵ net and the uniform distribution [3]. In the limit of small ϵ , the discrete and continuous distributions converge, and, therefore, the overall amount of compression achieved is equal to the ensemble entropy $\text{Ent}(\mathcal{E})$ per state, and the overall number of bits required to store M states is $M[N_\epsilon + \text{Ent}(\mathcal{E})]$.

Finally, we note that this task is different from the task of storing an ensemble of states in quantum memory, i.e., compressing the states into a smaller number of *qubits*. This can be done by Schumacher compression [94], with the compression rate given by the von Neumann entropy of the first moment $S(\rho_{\mathcal{E}})$. In contrast, the task of classical approximate storage depends on all moments of the ensemble.

E. Mutual information

In addition to the ensemble entropy, we study another information-theoretic property: the mutual information, which quantifies the information-theoretic correlations between two variables. We find that not only is the ensemble entropy maximized in our ensembles, the mutual information is also minimized. These two quantities are *a priori* unrelated and lead to distinct operational consequences.

The *mutual information* $I(X; Z)$ between random variables X and Z is defined as

$$\begin{aligned} I(X; Z) &= H(X) - H(X|Z) \\ &= -\sum_x p(x) \log_2[p(x)] \\ &\quad + \sum_z p(z) \sum_x p(x|z) \log_2[p(x|z)]. \end{aligned} \quad (21)$$

This can be interpreted as the information gain (reduction in entropy) of the variable X from measurements of the variable Z . In this work, we take the logarithm to base 2, which quantifies the mutual information in units of bits.

The mutual information is relevant in several information-theoretic tasks, including quantifying the maximum rate of information transmission through a noisy classical channel [89], where the variables X and Z are here the channel input and output, respectively. Below, we focus on the significance of mutual information to a quantum communication task.

Holevo studied the use of a noiseless quantum channel to transmit classical information [56]. Given an ensemble of (possibly mixed) states $\mathcal{E} = \{p_x, \rho_x\}$, Holevo considered a task where one party (Alice) transmits a classical message $\{x_1, x_2, \dots\}$ to another (Bob), by sending a sequence of states $\{\rho_{x_1}, \rho_{x_2}, \dots\}$. Bob decodes this message using measurements Z on the received states. When the quantum states $|\psi_{x_i}\rangle$ are not orthogonal, the symbols x_i cannot be unambiguously determined: The maximum classical transmission rate is, therefore, given by the *accessible information*

$$\text{Acc}(\mathcal{E}) \equiv \sup_Z I(X; Z), \quad (22)$$

the mutual information maximized over complete measurement bases Z .

Holevo provided an upper bound for the accessible information: the *Holevo- χ quantity* $\chi(\mathcal{E}) \equiv S(\rho_{\mathcal{E}}) - \sum_x p_x S(\rho_x)$ [56]. In an ensemble of pure states, the Holevo- χ quantity is equal to the von Neumann entropy $\chi(\mathcal{E}) = S(\rho_{\mathcal{E}})$. For a given first moment $\rho_{\mathcal{E}}$, the Holevo bound is saturated when the ensemble of states is formed using the eigenbasis of $\rho_{\mathcal{E}}$, i.e., $\mathcal{E} = \{p_j, |j\rangle\}$.

Jozsa, Robb, and Wootters [89] subsequently established a lower bound for the accessible information, dubbed the *subentropy* $Q(\rho_{\mathcal{E}})$. That is, for any ensemble of pure states $\mathcal{E}$ with a given first moment $\rho_{\mathcal{E}}$, the accessible information is bounded by

$$Q(\rho_{\mathcal{E}}) \leq \text{Acc}(\mathcal{E}) \leq S(\rho_{\mathcal{E}}). \quad (23)$$

Like the von Neumann entropy, the subentropy depends on only the first moment $\rho_{\mathcal{E}}$. For any ensemble $\mathcal{E}$, the subentropy is defined as the mutual information $I(X; Z)$ averaged over all complete projective measurement bases Z .

This is equivalent to averaging over unitaries U which rotate measurement bases and has expression [89,105]

$$Q(\rho_{\mathcal{E}}) \equiv -\sum_j \left(\prod_{i \neq j} \frac{\lambda_j}{\lambda_j - \lambda_i} \right) \log_2 \lambda_j, \quad (24)$$

where λ_j are the eigenvalues of $\rho_{\mathcal{E}}$. More recently, the subentropy has received attention in studies of the dynamics of information in many-body systems [61].

Jozsa, Robb, and Wootters established that the Scrooge ensemble saturates this lower bound (Sec. V [89]). Scrooge ensembles are, therefore, ensembles with minimum accessible information (or the most information “stingy” ensembles), leading to their name.

(Generalized) Scrooge ensembles describe projected ensembles; therefore, our projected ensembles have minimal accessible information. That is, in natural many-body systems, it is maximally hard to predict the measurement outcome of one subsystem from measurements on the other.

F. Porter-Thomas distribution

Finally, we turn to a measurable signature of our maximum entropy principle, the PT distribution. Originally studied in nuclear physics [98], the PT distribution has received renewed attention due to its presence in Haar-random many-body states, approximated by the output of deep random quantum circuits [99].

Concretely, a positive random variable X follows the Porter-Thomas distribution with mean μ [i.e., $X \sim \text{PT}(\mu)$] if its probability distribution function is

$$\Pr[X = x]dx = \exp(-x/\mu)dx/\mu. \quad (25)$$

In particular, the overlap of states $|\Psi\rangle$ drawn from the Haar-random distribution to any fixed state $|\Phi\rangle$ follows a PT distribution with $\mu = 1/D$ [80]:

$$\Pr_{|\Psi\rangle \sim \text{Haar}}[|\langle \Psi | \Phi \rangle|^2 = x]dx = \exp(-Dx)dDx. \quad (26)$$

Notably, this is a statement about the *probability distribution of probabilities* $|\langle \Psi | \Phi \rangle|^2$. While relatively unknown in physical settings, this has received attention in the statistical literature under several names, including the *fingerprint* and *histogram of the histogram* [106,107]. The fingerprint is sufficient to describe the properties of a distribution that do not depend on its labels, most notably its Shannon entropy.

The probability of probabilities can be studied in modern quantum experiments. For example, the PT distribution expected from Haar-random states has been predicted and subsequently experimentally verified in deep random quantum circuits [80,99] and many-body states at infinite temperature [29]. Even though the ensembles we study in

this work are not the Haar ensemble, we find that the maximum entropy principle leads to the same PT distribution, when observables are appropriately normalized. In companion work, we experimentally and theoretically study this object in detail [41].

IV. TEMPORAL ENSEMBLES

The most natural setting of quantum dynamics is time evolution under a time-independent Hamiltonian. It is an important endeavor to understand the phenomena that can arise in this setting. Despite the fact that it is simple to state, relatively few tools are available to analyze the dynamics under an interacting many-body Hamiltonian, and one must often turn to more analytically tractable models of quantum dynamics such as random quantum circuits [65].

In this section, we show universal statistical properties of the temporal ensemble—the time trace of a state evolving under Hamiltonian dynamics. Intuitively, Hamiltonian dynamics conserves energy, and its trajectory in Hilbert space cannot be unconstrained. In fact, the Schrödinger equation conserves the population $|c_E|^2$ of each energy eigenstate $|E\rangle$, where $|c_E|^2 \equiv |\langle \Psi_0 | E \rangle|^2$ for an initial state $|\Psi_0\rangle$. This imposes a total of D (the Hilbert-space dimension) constraints, and the only remaining degrees of freedom are the complex phases $\arg[\langle E | \Psi(t) \rangle]$. One might conjecture that these phases are uniformly and independently random, satisfying the maximum entropy principle.

Indeed, this expectation is correct: The temporal ensemble is (statistically) described by the random phase ensemble, the ensemble of states with fixed magnitudes and independent, uniformly random phases in the energy eigenbasis. This has been a folklore result in several works—see, e.g., Refs. [46,75,78]—however, it has to our knowledge not been presented with the same level of rigor as in this work. Hilbert-space ergodicity is a central concept in this work. While easy to state and extremely intuitive, this enables other results in this work, including our derivation of the statistical description of projected ensembles. Furthermore, we discover new signatures and properties of Hilbert-space ergodicity unique to the many-body setting where D is large. Below, we provide a careful description of Hilbert-space ergodicity, state our result with a necessary and sufficient condition—the k th no-resonance condition—and discuss its corollaries.

The first moments of the temporal ensemble are heavily used in the quantum thermalization literature. Seminal works in quantum chaos and thermalization introduced the *diagonal ensemble* $\rho_d \equiv \sum_E |c_E|^2 |E\rangle\langle E|$ to characterize observables at thermal equilibrium [51,52]. However, this captures only the average values of observables and does not address higher moments such as their variance over time. Previous work either bounds [13,17] or predicts the dependence of fluctuations with the total Hilbert-space dimension [74,108,109], but here we provide approximate formulas for all moments. We apply our result on the

temporal ensemble to a novel class of observables: global projective measurements, discovering universal behavior in its higher moments and information-theoretic properties.

We first formalize the above intuitive expectation and present the sketch of our proof. We define the *temporal ensemble* as the set of $|\Psi(t)\rangle$ explored by a *fixed* initial state $|\Psi_0\rangle$ time evolving under a Hamiltonian H over a time interval $t \in [-\tau/2, \tau/2]$:

$$\mathcal{E}_{\text{Temp}}(\tau) = \{\exp(-iHt)|\Psi_0\rangle\}_{t=-\tau/2}^{\tau/2}. \quad (27)$$

When $\tau \rightarrow \infty$, we equate this to the *random phase ensemble*, the distribution of states with fixed magnitudes in a given basis $|\langle j|\Psi\rangle| = |c_j| \equiv |\langle j|\Psi_0\rangle|$, with i.i.d. random phases $\phi_j \sim \text{Unif}([0, 2\pi))$:

$$\mathcal{E}_{\text{Rand Phase}} = \left\{ \sum_{j=1}^D |c_j| e^{i\phi_j} |j\rangle \middle| \phi_j \sim \text{Unif}([0, 2\pi)) \right\}.$$

This ensemble was studied in Ref. [46], which characterizes properties such as the entanglement of its typical states. Reference [79] develops a graphical calculus to perform the k -copy averages we discuss in this work, although we do not need its full generality here (see also Refs. [110,111]).

Our result, which equates the temporal ensemble (in the limit $\tau \rightarrow \infty$) with the random phase ensemble, relies on the following assumption.

Definition 3 (kth no-resonance condition). A Hamiltonian H satisfies the k th no-resonance condition [17,78,112,113] if for any two sets of k eigenvalues $\{E_{\alpha_j}\}_{j=1}^k$ and $\{E_{\beta_j}\}_{j=1}^k$ the equation

$$E_{\alpha_1} + E_{\alpha_2} + \dots + E_{\alpha_k} = E_{\beta_1} + E_{\beta_2} + \dots + E_{\beta_k} \quad (28)$$

is true if and only if the sets of indices $(\alpha_1, \alpha_2, \dots, \alpha_k)$ and $(\beta_1, \beta_2, \dots, \beta_k)$ are equal up to reordering.

In other words, the two sets of indices must be related by some permutation $\sigma \in S_k$, $\beta_j = \alpha_{\sigma(j)}$. It is easy to see that the k th no-resonance condition implies the k' th no-resonance conditions, for $k' < k$. The no-resonance conditions are considered generically satisfied and are used in works such as Refs. [13,17,74,78]: It is believed that if the condition does not hold in a given ergodic system, any small perturbation will generically break any k th resonances [112]. We note recent work [114] that studies near violations of the k th no-resonance condition, concluding that they are generically small and bounding their effects on certain quantities.

The no-resonance condition for $k = 1$ simply states that the energy eigenspectrum has no degeneracies. However, this may not be satisfied in generic Hamiltonians. The presence of non-Abelian symmetries ensures that there are degenerate multiplets of states. Strictly speaking, this violates all no-resonance conditions. However, for the

purposes of the temporal ensemble, we may disregard these degeneracies: The initial state $|\Psi_0\rangle$ projects the degenerate eigenspace onto a single eigenstate. Our conclusions below hold as long as the spectrum of H , with degeneracies removed, satisfies the k th no-resonance conditions [110]. We call this the *no-resonance condition modulo degeneracies*.

With the above definitions, we are ready to state our result equating the temporal and random phase ensembles.

Theorem 1. Given an initial state $|\Psi_0\rangle$ and a Hamiltonian H , the infinite-time temporal ensemble is equal to the random phase ensemble:

$$\lim_{\tau \rightarrow \infty} \mathcal{E}_{\text{Temp}}(\tau) = \mathcal{E}_{\text{Rand Phase}}, \quad (29)$$

with fixed energy populations $|c_E|^2 = |\langle \Psi_0 | E \rangle|^2$, if H satisfies all k th no-resonance conditions modulo degeneracies. The converse holds as long as $|\Psi_0\rangle$ has nonzero population on the eigenspace of every nondegenerate eigenvalue E , i.e., for almost every $|\Psi_0\rangle$ [115].

Proof sketch. Here, we sketch the idea of the proof, deferring its details to Appendix A. The general idea is to evaluate the k th moment $\rho_{\text{temp}}^{(k)}$ of the temporal ensemble and show that it matches the moments of the random phase ensemble. The evaluation of $\rho_{\text{temp}}^{(k)}$ is possible for an infinite interval $\tau = \infty$ based on the no-resonance condition Eq. (28), and the corresponding moments of the random phase ensemble can also be explicitly evaluated. We note that having matching moments does not always imply that two ensembles are equal to one another. Under certain conditions such as the (complex) Carleman condition [116], this equivalence can be made. The evaluated moments satisfy the complex Carleman condition, completing our proof. The converse direction can be easily shown by noting that if the k th no-resonance condition is not satisfied, the moment $\rho_{\text{temp}}^{(k)}$ will have additional terms that are not present in the random phase ensemble (as long as all coefficients c_E are nonzero), and, hence, the temporal and random phase ensembles will differ. ■

Ensemble entropy.—The ensemble entropy is always divergent in any distribution of states with fixed magnitudes, since the constrained space is a subdimensional manifold of the full Hilbert space. This can be intuitively understood from the following property of the KL divergence $\mathcal{D}(\mathcal{E}||\text{Haar})$. Assume we are given samples from $\mathcal{E}$ and promised that the underlying ensemble is either $\mathcal{E}$ or the Haar ensemble. The KL divergence quantifies how many samples one needs to correctly conclude that the ensemble is $\mathcal{E}$ and not the Haar ensemble [3]. In our case, we can rule out the Haar ensemble even with a *single* wave function $|\Psi(t)\rangle$ from the temporal ensemble. This is because the state $|\Psi(t)\rangle$ has the fixed amplitudes $|\langle E|\Psi(t)\rangle| = |c_E|$, which occurs with probability zero in the Haar ensemble. Therefore, given complete knowledge of the state $|\Psi(t)\rangle$,

the constraints $\{|c_E|\}$, and the promise that the underlying ensemble is either $\mathcal{E}$ or the Haar ensemble, this discrimination task can be done infinitely quickly, and, hence, the KL divergence is infinitely large. While artificial in practice, this operational meaning for the KL divergence accounts for its divergence in the temporal ensemble.

However, the maximum entropy principle still holds in a certain sense: We can separate the entropy into contributions from the phase and magnitude degrees of freedom; the contribution from the phase is maximized by independent, uniformly distributed phases (Appendix C). We find a constant, nondivergent contribution

$$\begin{aligned} \text{Ent}(\mathcal{E}_{\text{Temp}}) &\equiv -\mathcal{D}(\mathcal{E}_{\text{Temp}} \parallel \text{Haar}) \\ &= \sum_E \log(D|c_E|^2) + \lim_{\delta \rightarrow 0} D(\log(\delta/2) - 1). \end{aligned} \quad (30)$$

We interpret this expression as proportional to the volume $\prod_E |c_E|$ of the constrained space. Remarkably, this expression reappears in the ensemble entropy of the projected ensemble (Sec. V G).

A. Asymptotic product form

In this work, we are interested in the moments of the random phase ensemble. Here, we provide a simplified, approximate form of the k th moments, which enables their further analysis.

Theorem 2 (asymptotic product form). In the limit of small purity $\text{tr}(\rho_d^2)$, the random phase ensemble has the *product form*:

$$\rho_{\text{Temp}}^{(k)} = \rho_d^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma) + O[\text{tr}(\rho_d^2)]. \quad (31)$$

Here, $\rho_d \equiv \sum_E |c_E|^2 |E\rangle\langle E|$ is the *diagonal ensemble*, a common construction in the thermalization literature [51,52], and its purity $\text{tr}(\rho_d^2)$ is typically exponentially small in total system size.

Proof sketch. We provide the full proof of Theorem 2 in Appendix D and state only the key idea, which is the following: The k th moment of the temporal ensemble can be written as a sum over all lists of labels of energy eigenstates $\{\alpha_j\}$ and over all permutations of $\{\alpha_j\}$ (Theorem 1). Crucially, $\{\alpha_j\}$ is allowed to have repeated elements. Only if no elements are repeated will the number of unique permutations of $\{\alpha_j\}$ be $k!$. Using the same counting also for lists with repeated elements, the expression for the moments considerably simplifies:

$$\rho_{\text{Temp}}^{(k)} \approx \sum_{\sigma \in S_k} \sum_{\{\alpha_j\}}^k \bigotimes_{j=1}^k c_{\alpha_j} c_{\sigma(\alpha_j)}^* |E_{\alpha_j}\rangle \langle E_{\sigma(\alpha_j)}| \quad (32)$$

$$= \rho_d^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma). \quad (33)$$

Equation (32) is not exact, because it overcounts permutations when $\{\alpha_j\}$ contains repeated elements: There are additional correction terms to compensate for this overcounting. Careful consideration gives that the trace norm of all correction terms can be bounded by $O[\text{tr}(\rho_d^2)]$, with a k -dependent prefactor (Appendix D). ■

Theorem 2 explicitly demonstrates how the higher-order moments of the temporal ensemble depend on the first moment ρ_d . Specifically, the moments of the temporal ensemble have a structural resemblance to the moments of the Haar ensemble [Eq. (16)], up to a correction which is typically exponentially small, formalizing our intuition of maximally entropic ensembles as “distorted” Haar ensembles. This asymptotic behavior will also be common to projected ensembles.

B. Finite-time temporal ensembles

Theorem 1 studies the properties of temporal ensembles over infinite-time intervals. Here, we investigate how temporal ensembles over finite-time intervals converge to their infinite-interval values. We analytically study its convergence rate in random-matrix models and present numerical evidence that the above convergence rate also holds for realistic Hamiltonians. Specifically, we study the convergence of the finite-time k th moment $\rho_{\text{Temp}}^{(k)}(\tau)$ to its infinite-time limit $\rho_{\text{Temp}}^{(k)} \equiv \rho_{\text{Temp}}^{(k)}(\tau = \infty)$. This has expression:

$$\begin{aligned} \rho_{\text{Temp}}^{(k)}(\tau) &\equiv \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} dt |\Psi(t)\rangle \langle \Psi(t)|^{\otimes k} \\ &= \sum_{\{\alpha_j\}, \{\beta_j\}} \text{sinc}\left(\sum_j (E_{\alpha_j} - E_{\beta_j}) \frac{\tau}{2}\right) \left[\bigotimes_j c_{\alpha_j} c_{\beta_j}^* |E_{\alpha_j}\rangle \langle E_{\beta_j}| \right]. \end{aligned} \quad (34)$$

We consider the convergence in terms of both the trace norm and Frobenius norm, with the latter being analytically easier to study. For the first moment, i.e., $k = 1$, we numerically observe that the distance in Frobenius and trace norms both decay as τ^{-1} (see Appendix E). This holds for a generic chaotic model, the mixed field Ising model (the main model we study in this work, Sec. IV D), as well as for Hamiltonians sampled at random from the Gaussian unitary ensemble (GUE). We attribute this asymptotic $1/\tau$ scaling to the $1/x$ asymptotic behavior of the $\text{sinc}(x)$ function in Eq. (34). In Appendix F, we support these numerical findings by analytical calculations in random matrix theory, showing that the average of the squared Frobenius norm decays as τ^{-2} . This suggests that the average Frobenius norm decays as τ^{-1} , consistent with our numerical observations.

Interestingly, for higher-order moments $k = 2, 3$, we observe different behavior. For the mixed field Ising model

as well as individual samples from the GUE, the Frobenius distance decays at intermediate times as $\tau^{-1/2}$ before a crossover time, after which it decays as τ^{-1} [Fig. 3(b)]. This crossover happens in times which are exponentially long in system size, which we attribute to the presence of a minimum distance of gaps $E_i - E_j + E_k - E_l$ (for $k = 2$) for each individual Hamiltonian. While such minimum distance is exponentially small in system size, it has a finite, nonzero value $\Delta_{\min}$ for finite-sized systems. Only at times much longer than the inverse of this minimum gap does the asymptotic $1/(\Delta_{\min}\tau)$ behavior become clear. Numerically, we find that this crossover time increases significantly from $k = 2$ to $k = 3$ (Fig. 8 and Appendix E). This is consistent with our above argument, since the higher-order gaps will be exponentially smaller, leading to a longer crossover time.

Notably, we find that for the first moment, $k = 1$, the GUE ensemble mean closely resembles the behavior of individual instances and decays as τ^{-1} . In contrast, for higher moments, $k = 2$, the behavior of individual samples and the sample mean differs. With an increasing number of samples, the crossover is less pronounced. We relate these findings to statistical properties of the eigenvalues of GUE Hamiltonians. It is well known that the eigenvalues exhibit level repulsion; i.e., the probability to find two distinct eigenvalues close to each other $E_i - E_j \approx 0$ is vanishing. In contrast, as suggested by Ref. [114] and the analytical and numerical results presented in Appendix E, we find that the gaps $E_i - E_j + E_k - E_l$ of eigenvalues exhibit only weak repulsion, leading to the crossover timescale growing with the size of the sample as well as the order of k . Indeed, since the minimum gap size differs between instances of GUE Hamiltonians, we observe in Appendixes E and F that the crossover time also varies strongly between samples. Indeed, when averaging over Hamiltonians in the GUE to obtain the *root-mean-squared* (rms) Frobenius norm, the crossover is not present, and we find an approximate $1/\tau^{-1/2}$ decay (instead of $1/\tau$) of the rms Frobenius norm. If, instead, we take the median Frobenius norm, more indicative of typical behavior, a crossover to a $1/\tau$ late-time scaling apparently remains, illustrating further subtleties to this phenomenon.

Finally, it is important to note that the above relaxation times and crossover behavior concern highly atypical quantities—they describe the slowest k -copy features to relax to their equilibrium value and may not be observable in practice. Typical k -copy quantities may relax in considerably shorter, physically relevant timescales. For example, in the following section, we discuss the Porter-Thomas distribution as a signature of Hilbert-space ergodicity. In practice, these signatures are apparent at short times that grow weakly with system size [117], a crucial property that enables applications such as benchmarking in experimentally accessible times [28,101,110].

C. Porter-Thomas distribution as measurable signature of temporal ensembles

An immediate consequence of Theorem 2 is the emergence of the Porter-Thomas distribution (Sec. III F) in the probabilities of measuring a given outcome z , as a function of time. This result was originally presented in Ref. [110]. For clarity, we denote this distribution as $p_z(t) \equiv p(z, t) \equiv \langle z | \Psi(t) \rangle \langle \Psi(t) | z \rangle$, [118] which can be experimentally estimated by repeated measurements of the state $|\Psi(t)\rangle$.

Unlike local observables, whose expectation values quickly equilibrate in time— $\langle O(t) \rangle$ has exponentially small relative fluctuations about its equilibrium value $\mathbb{E}_t[\langle O(t) \rangle]$ [17]—global observables such as the projector $|z\rangle\langle z|$ never equilibrate. Theorem 3 indicates that they have $O(1)$ relative fluctuations which follow a universal, exponential distribution. We utilized this result in Ref. [110] to propose a protocol for many-body benchmarking in analog quantum simulators. Here, we use the PT distribution as a signature of our maximum entropy principle; it is a common signature among the temporal ensemble, projected ensemble, and beyond.

Theorem 3. Given an initial state $|\Psi_0\rangle$, a Hamiltonian H which satisfies the k th no-resonance condition, and a fixed state $|z\rangle$, the outcome distribution $p_z(t)$ follows an approximate Porter-Thomas (or exponential) distribution over time, with mean $p_{\text{avg}}(z) \equiv \mathbb{E}_t[p_z(t)]$. Specifically, the k th moments satisfy

$$\mathbb{E}_t[p_z(t)^k] = p_{\text{avg}}(z)^k [k! + O(D_\beta^{-1})], \quad (35)$$

where $D_\beta^{-1} \equiv \sum_{z,E} |\langle z | E \rangle|^4 |c_E|^4 / p_{\text{avg}}(z)$ is the inverse *effective dimension*, which captures the effective size of the Hilbert space explored during quench evolution that is accessible by measurements in $\{|z\rangle\}$. Our notation β associates D_β with the temperature of the initial state, which to a large extent controls the size of the Hilbert space explored.

Proof. Its k th moment can be easily obtained from the k th moment of the temporal ensemble, which gives the desired $\mathbb{E}_t[p_z(t)^k] \equiv \langle z |^{\otimes k} \rho_{\text{Temp}}^{(k)} | z \rangle^{\otimes k} \approx k! p_{\text{avg}}(z)^k$. This relation follows from Theorem 2. We conclude that $p_z(t)$ follows a Porter-Thomas distribution, [119] with a mean value equal to $p_{\text{avg}}(z) \equiv \mathbb{E}_t[p_z(t)] = \langle z | \rho_d | z \rangle$. This is a simple illustration of how the first moment $p_{\text{avg}}(z)$ uniquely determines all moments of the distribution $p_z(t)$. This relation is only approximate because of the correction terms in Theorem 2. We are able to (loosely) bound the effects of these correction terms on the k th moments of $p(z, t)$ in terms of the parameter D_β^{-1} . This is not a straightforward application of the bound provided in Theorem 2, which is too weak for our purposes. We refer

the reader to Ref. [110] for a detailed proof or to Appendix I for a very similar proof. ■

As a practical matter, verifying the Porter-Thomas distribution requires an exponential amount of information in the form of $p_{\text{avg}}(z)$. These can be obtained directly from the eigenstate overlaps $|c_E|^2$, available for systems which can be exactly diagonalized, but may also be approximated using statistical assumptions like Eq. (84). In related work [101], $p_{\text{avg}}(z)$ was obtained by classically simulating time evolution of the state and taking the time average $p(z, t)$ over a time interval. The length of this time interval required to obtain a good estimate of $p_{\text{avg}}(z)$ can be short. More precisely, as we show in Sec. IV F 1 and Appendix G, the autocorrelation time, which describes the timescale over which $p(z, t)$ fluctuates, decreases with system size N as $1/\sqrt{N}$, so if one wants to estimate $p(z, t)$ to precision ϵ , the size of the required time interval *decreases* with system size, of the order of $O[1/(\epsilon^2\sqrt{N})]$. Note, however, that, even though the length of the interval decreases with system size, the center of the interval does not: We have assumed that $p(z, t)$ exhibits Porter-Thomas-like fluctuations about $p_{\text{avg}}(z)$ —the timescale after which this is true is roughly the anticoncentration time, which, using results from RUCs [117], we expect to grow weakly with system size, as $O[\log(N)]$. The effects of energy conservation (such as diffusive modes which slowly relax) may change this scaling to polynomial scaling with small degree in N , although this has not been thoroughly investigated. Therefore, in order to estimate $p_{\text{avg}}(z)$ by time averaging requires classical simulation up to times which grow weakly with N . Finally, we may be able to estimate $p_{\text{avg}}(z)$ without such time averaging. $p_{\text{avg}}(z)$ is typically smoothly varying over quantities such as the energy expectation value $\langle z|H|z\rangle$, and we expect that it can be approximated with low computational cost.

D. Numerical model

We verify our theory predictions via explicit exact numerical simulations of interacting quantum many-body systems. As an illustration, we work with a paradigmatic model of many-body quantum chaos, the one-dimensional mixed field Ising model (MFIM) with open boundary conditions:

$$H_{\text{MFIM}} = \sum_{j=1}^N h_x X_j + h_y Y_j + \sum_{j=1}^{N-1} J X_j X_{j+1}, \quad (36)$$

where X_j and Y_j are Pauli matrices on site j . We use parameter values established in Ref. [120]: $(h_x, h_y, J) = (0.890, 0.9045, 1)$. Unless otherwise stated, we use this model in this and future sections for our numerical simulations.

In both the projected and temporal ensemble, we study the time evolution of the initial state

$$|\Psi_0\rangle = [\exp(i\theta X/2)|0\rangle]^{\otimes N}, \quad (37)$$

where the parameter θ is used to tune the state from infinite temperature ($\theta = 0$) to positive or negative temperature. Unless otherwise stated, we present data for $\theta = 0.6$, which has energy density $E/N \approx 0.51$, far from the center of the spectrum.

Here, we illustrate the content of Theorem 3 in Fig. 3(c), where we plot the probabilities $p_z(t)$ for two choices of z , which exhibit large fluctuations over time, about different average values $p_{\text{avg}}(z)$. The histograms of $\{p_z(t)\}$ show good agreement with the PT distributions $p_z(t) \sim \text{PT}[p_{\text{avg}}(z)]$ even over a relatively short time interval, confirming our analysis.

E. Ergodicity

Theorem 3 is a statement about the distribution of $p(z, t)$ over time [denoted $p_z(t)$ to make explicit that the variable z is held constant]. We can make an analogous statement about a distribution over z , holding t constant. At a fixed, late time t , the quantities $\tilde{p}_t(z) \equiv p(z, t)/p_{\text{avg}}(z)$ are approximately distributed according to the PT distribution with mean 1:

$$\tilde{p}_t(z) \sim \text{PT}(1). \quad (38)$$

This was originally derived in Ref. [110] by showing that the weighted k th moment $m^{(k)}(t) \equiv \sum_z p_{\text{avg}}(z) \tilde{p}_t(z)^k$ for a typical, late time t satisfies

$$m^{(k)}(t) = k! + O(D_\beta^{-1/2}). \quad (39)$$

We interpret this equivalence of distributions over time and over outcomes z as a kind of “ergodicity,” in the sense of the *ergodic theorem* in the study of classical dynamical systems [81], which states that, in an ergodic system, the average over time is equal to the “spatial” average over configurations. Such an equivalence is essential to our characterization of projected ensembles in Sec. VI.

F. Mutual information: Building a (bad) “clock”

The temporal ensemble has unique information-theoretic properties. In particular, we find near-universal behavior in the *mutual information* between the random variables Z and T which, respectively, represent the measurement outcomes z and evolution times t .

To illustrate the operational meaning of the mutual information, let us consider the following task. We are given several copies of the time-evolved state $|\Psi(t)\rangle$. We have knowledge of the initial state of the system as well as its Hamiltonian but not the duration of the time

evolution t , which is assumed to be uniformly distributed on some large interval $t \in [t_0, t_0 + \tau]$. Given multiple single-copy measurements performed in an optimal basis, how much information about t can we learn?

Before performing detailed analysis, we can make a prediction based on our maximum entropy principle: Since a temporal ensemble is maximally entropic, we expect that the amount of information must be small no matter which basis the measurements are performed. Indeed, our results imply that the (near) optimality is achieved by any generic measurement basis including the conventional configuration basis $|z\rangle\langle z|$. Furthermore, the information gain per measurement takes the universal value $(1 - \gamma)/\ln 2 \approx 0.6099$, independent of the Hamiltonian, initial state, measurement basis, and system size.

In order to formally define the mutual information $I(Z; T)$, we must make the following definitions. We let time be a uniformly distributed variable over a interval of (long but arbitrary) length τ . We then define a joint probability distribution over measurement outcomes z and times t , by $P(z, t) = p_z(t)/\tau$, which is normalized: $\int_{-\tau/2}^{\tau/2} dt \sum_z P(z, t) = 1$. We also need the marginal distributions $P(t) \equiv \sum_z P(z, t) = 1/\tau$ and $P(z) \equiv \int dt P(z, t) = p_{\text{avg}}(z)$ and the conditional distribution $P(t|z) = P(z, t)/P(z) = \tilde{p}_z(t)/\tau$.

Then, the mutual information can be computed as

$$I(Z; T) = H(T) - H(T|Z) \quad (40)$$

$$= - \int dt P(t) \log_2 [P(t)] + \sum_z P(z) \int_{-\tau/2}^{\tau/2} dt P(t|z) \log_2 P(t|z) \quad (41)$$

$$\approx \sum_z p_{\text{avg}}(z) \int_0^\infty d\tilde{p} \text{Pr}(\tilde{p}) \tilde{p} \log_2 \tilde{p} + O(D_\beta^{-1}) \quad (42)$$

$$\approx (1 - \gamma)/\ln 2, \quad (43)$$

where in Eq. (42) we use Theorem 3 to replace the integral over time with an integral over $\tilde{p}$, which follows a PT distribution, i.e., with probability density $\text{Pr}(\tilde{p}) d\tilde{p} = \exp(-\tilde{p}) d\tilde{p}$.

It is remarkable that this universal value does not depend on system size and, in particular, does not vanish with any parameter. This implies that, despite the fact that ergodic dynamics hides information about time t as strongly as possible, a finite amount (0.6099 bits) of information still cannot be concealed, in the large τ limit. In other words, our result implies that a certain level of temporal fluctuations in observables is inevitable in the ergodic dynamics of generic pure states, putting the lower bound on the ability of unitary dynamics to hide temporal information. We also note that estimating the evolution duration t is equivalent to

determining an overall strength scaling factor of a many-body Hamiltonian by evolving to a fixed time t , and, hence, this “clock” may equivalently be regarded as a sensor for the overall Hamiltonian strength.

1. Finite-time mutual information

We note that the mutual information discussed above should be distinguished from more conventional metrics for clocks such as sensitivity. The sensitivity at very small intervals Δt (i.e., in the limit of many measurements) depends instead on the energy uncertainty $\sigma_H \equiv (\langle \psi_0 | H^2 | \psi_0 \rangle - \langle \psi_0 | H | \psi_0 \rangle^2)^{1/2}$, which is not universal. The mutual information instead describes how quickly a large interval τ can be refined by measurements.

Our universal result above is valid in the limit of large time intervals τ , over which $p(z, t)$ follows the approximate PT distribution. At shorter intervals (but still at sufficiently late times t), the mutual information $I(Z; T)$ takes on a smaller value, due to $p(z, t)$ having a nonzero correlation time. This is illustrated in Fig. 3(d). In fact, $I(Z; T)$ approaches its late-time value as a power law:

$$I(Z; T) \approx \left(1 - \gamma - \frac{\sqrt{\pi}}{2\sigma_H\tau}\right) / \ln 2, \quad (44)$$

where $\sigma_H = (\langle \Psi_0 | H^2 | \Psi_0 \rangle - \langle \Psi_0 | H | \Psi_0 \rangle^2)^{1/2}$ is the uncertainty in energy of the initial state. This is illustrated in the inset in Fig. 3(d).

We present the full derivation in Appendix G. In brief, we obtain this result by estimating the autocorrelation time of $p(z, t)$, using approximations of the many-body spectrum. We find that the correlation time is inversely proportional to the energy uncertainty of the state σ_H . If we interrogate $p(z, t)$ when t is an integer multiple of the correlation time, we get an effective number $\sigma_H\tau$ of points in times for which $p(z, t)$ is distinct. This gives a mutual information that approaches the value $1 - \gamma$ at the rate $(\sigma_H\tau)^{-1}$. The coefficient $\sqrt{\pi}/2$ follows from a detailed calculation. This result complements Sec. IV B, providing a convergence timescale of finite-time temporal ensembles from the perspective of the mutual information. We leave to future work the analysis of the ultimate sensitivity at very small intervals $\tau \ll [\sqrt{\pi}/2\sigma_H(1 - \gamma)]$, at which point Eq. (44) breaks down.

Our results in this section describe universal, statistical features of the trajectory of a many-body state under chaotic many-body dynamics. They confirm an intuitive, folklore heuristic that chaotic many-body dynamics resembles random unitary dynamics. Here, we find signatures of Haar-random states such as the PT distribution in the time-evolved many-body states. We attribute this quasirandomness to our maximum entropy principle. This translates to our finding that these temporal trajectories can be used for certain protocols such as benchmarking

and sensing. Furthermore, their performance in these protocols is universal and independent of details such as the Hamiltonian, initial state, and measurement basis.

V. PROJECTED ENSEMBLES

Next, we turn to the projected ensemble, which is the second setting of our maximum entropy principle. When a many-body state is bipartitioned into a small system A and its complement B , the large subsystem B acts like a “thermal bath” to the smaller subsystem A . This observation is instrumental to the emergence of thermal equilibration in a closed quantum system. Nevertheless, the extent to which subsystem B truly acts as a bath is not fully characterized, owing to the complexity of the overall quantum many-body state. Projected ensembles offer a lens into the nature of correlations between subsystem and bath. Without the information of the bath, the subsystem is described by the reduced density matrix $\rho_A = \text{tr}_B(|\Psi\rangle\langle\Psi|)$. In order to go further, it is necessary to obtain information about the bath. In projected ensembles, this is done by projecting the bath onto definite states $|z_B\rangle$, typically by performing measurements in a product basis. This, in turn, projects $|\Psi\rangle$ into a pure state $|\Psi(z_B)\rangle \in \mathcal{H}_A$. Taken over all measurement outcomes, a single wave function $|\Psi\rangle$ generates a large ensemble of states $\{|\Psi(z_B)\rangle\}$. By studying the statistical properties of this ensemble of states, as well as any correlations between $|\Psi(z_B)\rangle$ and z_B , the projected ensemble construction provides a new way to study system-bath correlations.

Our maximum entropy principle states the following: There are correlations between $|\Psi(z_B)\rangle$ and z_B which are “thermal” in the sense that they arise from (and *only* from) energy conservation. Furthermore, there are random fluctuations over these thermal correlations. In Sec. VI, we find that these fluctuations are universal and follow a maximum entropy principle.

In this section, we study a special case. In certain measurement bases, the correlations between $|\Psi(z_B)\rangle$ and z_B vanish. Nevertheless, effects of energy conservation remain: A mixture of all projected states equals ρ_A , which is, up to subleading corrections, a thermal Gibbs state with temperature set by the bath [74]. Therefore, while the projected states $|\Psi(z_B)\rangle$ are effectively identical independent random samples from a distribution of states, the underlying distribution is not Haar random but has first moment ρ_A . In this section, we find that this distribution has maximum entropy under this first moment constraint, a distribution that has been previously studied under the name “Scrooge ensemble” [82,89,90].

Explicitly, the projected ensemble is defined as the set of states obtained by projectively measuring a pure state $|\Psi\rangle$ on a large subsystem B [Fig. 4(a)]. Specifically, the subsystem B is measured, giving a random outcome z_B , which occurs with probability $p(z_B)$:

$$p(z_B) \equiv \langle\Psi|(\mathbb{I}_A \otimes |z_B\rangle\langle z_B|)|\Psi\rangle. \quad (45)$$

This measurement outcome projects the initial state $|\Psi\rangle$ onto a state $|\Psi(z_B)\rangle$ in the Hilbert space of the remaining subsystem, denoted A :

$$|\Psi(z_B)\rangle \equiv (\mathbb{I}_A \otimes \langle z_B|)|\Psi\rangle / \sqrt{p(z_B)} \in \mathcal{H}_A. \quad (46)$$

The projected ensemble is the set of states

$$\mathcal{E}_{\text{Proj}} = \{p(z_B), |\Psi(z_B)\rangle\}. \quad (47)$$

While this object was first introduced in different contexts, e.g., in Refs. [70,90], the projected ensemble was later studied by some of the authors in Refs. [28,29] as a means of quantifying the notion of randomness in a *single* many-body state (as opposed to ensembles of them), by studying its projected ensemble.

A. Results at infinite temperature: Approximate k -designs

In Ref. [29], it was found that a wide class of many-body states induce projected ensembles that are statistically close to the Haar ensemble. This closeness was measured by the trace distance $\Delta_{\text{Haar}}^{(k)}$ between the k th moments of the projected ensemble and the Haar ensemble.

Specifically, it was proven that the projected ensembles are approximate k -designs when the global states are typical states drawn from the Haar ensemble or from a state k' -design (with $k' > k$). This was also numerically demonstrated in natural many-body states, obtained from either (a) time evolution of an initial state or (b) the eigenstates of a chaotic many-body Hamiltonian, close to infinite temperature, defined as states $|\Psi_0\rangle$ that satisfy $\langle\Psi_0|H|\Psi_0\rangle = \text{tr}(H)/D$. Subsequent work has rigorously established the emergence of the projected k -design in specific settings, such as in dual-unitary models [30–32,83,84], in free-fermion models [33], or under assumptions of the reduced density matrix [34]. Further work has studied the effects of symmetry [35,37], the effect of quantum magic in the global state [38] on the projected ensemble, and deep thermalization time-scales [30,31,36].

In the above studies, it was important that the initial state $|\Psi_0\rangle$ be at infinite temperature. This is because the first moment of the projected ensemble is the reduced density matrix $\rho_A \equiv \text{tr}_B[|\Psi\rangle\langle\Psi|]$. For thermal states, ρ_A is close to a Gibbs state [74]; only at infinite temperature is this equal to the first moment of the Haar ensemble: the maximally mixed state $\mathbb{I}/D_A$. Therefore, away from infinite temperature, the projected ensembles of such states cannot equal the Haar ensemble. It has remained an open question as to which ensemble describes the finite-temperature projected ensemble.

B. Projected ensembles at finite temperature: Scrooge ensemble

Our main result of this section is that, under appropriate conditions, the projected ensemble of a finite-temperature state is described by the Scrooge ensemble [29,31,82,89,90]. Introduced in Ref. [89], any density matrix ρ has a corresponding Scrooge ensemble, which we denote $\text{Scrooge}[\rho]$ [Fig. 4(b)]. This is most easily understood as a “ ρ distortion” of the Haar ensemble, defined by the probability distribution. The state $|\Psi\rangle$ has a higher probability when it has a larger overlap with the principal axes of ρ . We can sample from the Scrooge ensemble by “distorting” wave functions sampled from the Haar ensemble, as described in Sec. II. We can also describe the Scrooge ensemble in terms of its distribution function [90]:

$$\begin{aligned} P_{\text{Scr}}(\Psi)d\Psi &= \frac{D!}{2\pi^D \det \rho} \langle \Psi | \rho^{-1} | \Psi \rangle^{-D-1} P_{\text{Haar}}(\Psi)d\Psi \\ &= \frac{D!}{2\pi^D \prod_{m=1}^D \lambda_m} \left(\sum_{m=1}^D \frac{|\Psi_m|^2}{\lambda_m} \right)^{-D-1} P_{\text{Haar}}(\Psi)d\Psi, \end{aligned} \quad (48)$$

where $\{\lambda_m, |m\rangle\}$ is the eigensystem of ρ and $\Psi_m \equiv \langle m | \Psi \rangle$. In the context of the projected ensemble, the first moment ρ is the reduced density matrix ρ_A of $|\Psi\rangle$.

The relationship between the projected ensemble and Scrooge ensemble has been previously studied in certain limits. In the ideal limit in which the measurement basis is a *Haar-random basis* on B , Ref. [82] proves that the resulting projected ensemble is the Scrooge ensemble. Meanwhile, Ref. [31] studies the nonequilibrium dynamics of the projected ensemble in a certain model of random circuit dynamics. In this model, they find that, before the first moment ρ reaches its equilibrium value of the maximally mixed state $\mathbb{I}/D_A$, the higher moments of the projected ensemble can be related to the first moment ρ : They are precisely equal to the higher moments of the Scrooge ensemble $\text{Scrooge}[\rho]$, indicating that the Scrooge ensemble emerges in the transient dynamics of the particular model studied. Finally, Ref. [29] provides numerical evidence that the Scrooge ensemble is generated by energy eigenstates at finite temperature.

In this work, we focus on the projected ensembles generated by Hamiltonian dynamics at equilibrium, and we prove our results in Sec. VI. One novel effect we uncover is the effect of correlations between the measurement basis and the Hamiltonian generating the dynamics. While we provide a precise definition in Sec. VI, “correlations between measurement basis and the Hamiltonian” may be taken to be the distribution of expectation values $\langle z_B | H_B | z_B \rangle$, where H_B are the terms of the Hamiltonian H restricted onto the large subsystem B . We find that the projected states $|\Psi(z_B)\rangle$ are correlated with the energy $\langle z_B | H_B | z_B \rangle$. If there is a large spread of

values $\langle z_B | H_B | z_B \rangle$, the projected states are distinguishable. Conversely, if $\langle z_B | H_B | z_B \rangle$ is independent of z_B , the projected states are independent of their label z_B , and the projected ensemble can be regarded as a collection of identical samples from the same distribution. In brief, we claim the following.

Claim 1.—The projected ensemble is statistically described by the Scrooge ensemble, if the measurement basis $\{|z_B\rangle\}$ is energy nonrevealing.

When such correlations are not negligible, the projected ensemble is described by a *generalized Scrooge ensemble*, discussed in Sec. VI. There, we provide a proof of our general claim for the projected ensembles of time-evolved states, based on our results about the temporal ensemble in Sec. IV. These results are analogous to the effects of conserved quantities on projected ensembles, investigated in random circuits with symmetry at infinite temperature in Ref. [37]. In this work, the conserved quantity in question is energy, which shows two primary differences to the results in Ref. [37]: The correlations between energy and measurement basis can be highly nontrivial, and, even when these correlations are negligible, the ensemble is not the Haar ensemble.

C. Moments of the Scrooge ensemble

Our conventional understanding of many-body dynamics is that, on sufficiently small subsystems, a many-body state will rapidly reach thermal equilibrium; i.e., its reduced density matrix will be close to that of a Gibbs state at an appropriate temperature [74]. As with the temporal ensemble, our results imply that not only is the first moment of the projected ensemble (the reduced density matrix) a universal function of the temperature, so are all higher statistical moments.

We directly test Claim 1 by comparing the k th moments of the projected ensemble to those of the Scrooge ensemble. In order to do so, here we provide explicit expressions for the k th moment of the Scrooge ensemble, which have previously not been available in the literature.

Our expression for the k th moment is in terms of its matrix elements in (k copies of) the eigenbasis $\{\lambda_m, |m\rangle\}$ of ρ . We find that the nonzero elements of the Scrooge k th moment $\rho_{\text{Scr}}^{(k)}$ are those which permute the k copies. For a basis element $|\mathbf{m}\rangle \equiv |m_1, \dots, m_k\rangle$, $\langle \mathbf{m} | \rho_{\text{Scr}}^{(k)} | \mathbf{m}' \rangle$ is nonzero only if $|\mathbf{m}'\rangle$ is a permutation of $|\mathbf{m}\rangle$, i.e., $|\mathbf{m}'\rangle = |\sigma(\mathbf{m})\rangle$, $\sigma \in S_k$. Furthermore, the matrix elements depend only on $\mathbf{m}$ and are independent of the permutation σ :

$$\rho_{\text{Scr}}^{(k)} = \sum_{\mathbf{m}} \sum_{\mathbf{m}' = \sigma(\mathbf{m})} \rho_{\text{Scr}, \mathbf{m}}^{(k)} |\mathbf{m}\rangle \langle \mathbf{m}'|. \quad (49)$$

This structure is similar to the k th moments of the Haar ensemble [Eq. (16)]. However, unlike in the Haar

ensemble, the coefficients $\rho_{\text{Scr},m}^{(k)}$ are nontrivial and are given by

$$\rho_{\text{Scr},m}^{(k)} = \left(\prod_{i=1}^D \lambda_i^{-1} \right) \frac{\partial^k \Lambda_k}{\partial \mu_{m_1} \dots \partial \mu_{m_k}} \Big|_{\mu_i = \lambda_i^{-1}}, \quad (50)$$

where

$$\Lambda_k(\mu_1, \dots, \mu_D) \equiv \sum_{j=1}^D \frac{\mu_j^{k-2} \ln \mu_j}{\prod_{i \neq j} (\mu_j - \mu_i)}. \quad (51)$$

We derive this result in Appendix H. While this expression appears complicated, it reduces to a simple form in the limit that ρ is a sufficiently mixed state. We illustrate this for $k=2$. When $\sum_m \lambda_m^{2n} \ll 1$ for every $n > 1$, Eq. (50) simplifies to the product form (cf. Theorem 2)

$$\rho_{\text{Scr}}^{(2)} \approx \rho^{\otimes 2}(\mathbb{I} + \mathbb{S}), \quad (52)$$

where $\mathbb{I}$ and $\mathbb{S}$ are the identity and swap operators, respectively, permuting two copies of the Hilbert space $\mathcal{H}^{\otimes 2}$.

D. Numerical evidence: Trace distance between k th moments

We numerically support our claim by identifying the projected ensemble with the Scrooge ensemble. Detailed information about the model of our numerical simulations is provided in Sec. IV D. In Figs. 4(c) and 4(d), we compare the $k=2, 3$ moments of the Scrooge ensemble $\rho_{\text{Scr}}^{(k)}$ with those of the projected ensemble

$$\rho_{\text{Proj}}^{(k)} \equiv \sum_{z_B} p(z_B) |\Psi(z_B)\rangle \langle \Psi(z_B)|^{\otimes k}, \quad (53)$$

obtained from a time-evolved state $|\Psi(t)\rangle$ under the MFIM. As a function of quench evolution time t , the trace distance $\Delta^{(k)}$ decreases as a power law with t [29], before saturating at a value which depends on system size, illustrated for $k=2$ in Fig. 4(c). This saturation value decreases exponentially with increasing system size for both $k=2$ and $k=3$ [Fig. 4(d)].

E. Unnormalized ensembles of states

It is mathematically more convenient to treat the projected ensemble as an ensemble of the *unnormalized states* $|\tilde{\Psi}(z_B)\rangle \equiv (\mathbb{I}_A \otimes \langle z_B|) |\Psi\rangle = \sqrt{p(z_B)} |\Psi(z_B)\rangle$, such that the projection operation $|\Psi\rangle \langle \Psi| \rightarrow |\tilde{\Psi}(z_B)\rangle \langle \tilde{\Psi}(z_B)|$ acts linearly on the state $|\Psi\rangle \langle \Psi|$. The k th moment of the

ensemble of unnormalized states [Eq. (15), reproduced below] differs from that of the ensemble of normalized states [Eq. (14)]:

$$\begin{aligned} \tilde{\rho}^{(k)} &\equiv \frac{1}{D_B} \sum_{z_B} |\tilde{\Psi}(z_B, t)\rangle \langle \tilde{\Psi}(z_B, t)|^{\otimes k} \\ &\equiv \int d\tilde{\Psi} |\tilde{\Psi}\rangle \langle \tilde{\Psi}|^{\otimes k}. \end{aligned} \quad (54)$$

The primary reason for considering the unnormalized ensemble moments is that we will be able to derive analytical expressions for the corresponding moments of projected ensembles generated by time-evolved states.

In terms of unnormalized ensembles, our claim is that the projected ensemble is statistically described by the unnormalized-state analog of the Scrooge ensemble, which we refer to as the *unnormalized Scrooge ensemble* and denote $\tilde{\text{Scrooge}}[\rho]$. This is simply a distortion of the Gaussian ensemble of unnormalized states [Eq. (12) and Ref. [90]]. In the eigenbasis of ρ , the probability distribution function factorizes:

$$P_{\tilde{\text{Scrooge}}}(\tilde{\Psi}) d\tilde{\Psi} = \prod_{m=1}^D \frac{1}{\lambda_m \pi} \exp[-\lambda_m^{-1} |\tilde{\Psi}_m|^2] d^2 \tilde{\Psi}_m. \quad (55)$$

Note that we can map the unnormalized $\tilde{\text{Scrooge}}[\rho]$ into the normalized ensemble $\text{Scrooge}[\rho]$ by the procedure described in Sec. III.

Indeed, the k th moments of $\tilde{\text{Scrooge}}[\rho]$ are far simpler than those of $\text{Scrooge}[\rho]$. They are *exactly* of the product form:

$$\begin{aligned} \tilde{\rho}_{\text{Scr}}^{(k)} &= \int d\tilde{\Psi} \exp[-\lambda_j^{-1} |\tilde{\Psi}_j|^2] \sum_{\substack{m, m' \\ m' = \sigma(m)}}^k \prod_{i=1}^k |\tilde{\Psi}_{m_i}|^2 |m\rangle \langle m'| \\ &= \rho^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma). \end{aligned} \quad (56)$$

F. Porter-Thomas distribution in outcome probabilities

The simple k th moments of the unnormalized Scrooge ensembles lend themselves to a simple, measurable signature of our claim. We consider the joint probabilities $p_{o_A}(z_B) \equiv |\langle o_A | \tilde{\Psi}(z_B) \rangle|^2 = |\langle o_A, z_B | \Psi \rangle|^2$. Equation (56) predicts that if we fix o_A , the distribution of probabilities $p_{o_A}(z_B)$ over the bit strings z_B will be a PT distribution $p_{o_A}(z_B) \sim \text{PT}(\langle o_A | \rho_A | o_A \rangle)$. In Fig. 4(e), A is a subsystem of four qubits. We plot the histogram of $p_{o_A}(z_B)$ for three different outcomes $o_A = 0000, 1010$, or 0010 , all of which follow PT distributions.

In contrast to the joint probability $p_{o_A}(z_B)$, the distribution of the conditional probabilities $p(o_A | z_B) \equiv p(o_A, z_B) / p(z_B)$ is nonuniversal. While the moments of

the joint probability $p_{o_A}(z_B)$ correspond to the moments of the unnormalized Scrooge ensemble, the moments of the conditional probability $p(o_A|z_B) = |\langle o_A|\Psi(z_B)\rangle|^2$ are given by the moments of the normalized Scrooge ensemble [Eq. (50)]. The predicted distribution of $p(o_A|z_B)$ is nonuniversal, depends on the observable O_A , and, hence, is more challenging to verify, illustrating the conceptual utility of ensembles of unnormalized states.

G. Maximally entropic ensembles of unnormalized states

That the Scrooge ensemble describes the projected ensemble is no coincidence: It is known that the *unnormalized* Scrooge ensemble has maximum entropy, subject to the constraint that its first moment is ρ_A [90,121] (see Ref. [53] for a discussion of maximum entropy normalized ensembles). Showing this is a simple application of Lagrange multipliers, but we reproduce it here for completeness. We want to solve for the probability distribution P_{MaxEnt} which minimizes the relative entropy $D(\mathcal{E}||\text{Gauss})$, subject to the first moment constraint. This optimization may be independently solved in the eigenbasis of ρ_A ; i.e., P_{MaxEnt} factorizes in the eigenbasis. In each component, P_{MaxEnt} has exponential form: $P_{\text{MaxEnt}}(\tilde{\Psi}_j)/P_{\text{Gauss}}(\tilde{\Psi}_j) \propto \exp(-c_j|\tilde{\Psi}_j|^2)$. The unknown values c_j are solved by enforcing the first moment constraint. With $\rho_A = \sum_j \lambda_j |j\rangle\langle j|$, we have

$$\lambda_j = \int |\tilde{\Psi}_j|^2 P_{\text{MaxEnt}}(\tilde{\Psi}) d\tilde{\Psi} \quad (57)$$

$$= \frac{1}{Z} \left[\prod_{k \neq j} \frac{D}{D + c_k} \right] \frac{D}{(D + c_j)^2} = \frac{1}{D + c_j}, \quad (58)$$

where in the second line we have performed the relevant Gaussian integrals and we note that the normalization constant Z satisfies $Z = \prod_k D/(D + c_k)$. It follows that $c_j = \lambda_j^{-1} - D$, precisely equal to the unnormalized Scrooge ensemble in Eq. (55). This derivation is fairly general and may apply to multiple settings, e.g., recent work in the context of maximum entropy states under charge conservation [62].

The corresponding maximized value of the ensemble entropy is

$$\text{Ent}(\mathcal{E}) = -D(\mathcal{E}||\text{Gauss}) \quad (59)$$

$$\begin{aligned} &= \sum_j \int d\tilde{\Psi}_j P_{\text{Scr}}(\tilde{\Psi}_j) [(\lambda_j^{-1} - D)|\tilde{\Psi}_j|^2 + \log(D\lambda_j)] \\ &= \sum_j \log(D\lambda_j), \end{aligned} \quad (60)$$

identical to the (nondivergent part of the) ensemble entropy of the temporal ensemble [Eq. (30)], strongly suggesting a connection between the two ensembles.

H. Basis-independent mutual information

The Scrooge ensemble was originally studied for its information-theoretic properties: It is the ensemble of states with minimal *accessible information* (Sec. III E) [89]. Specifically, Jozsa, Robb, and Wootters [89] lower bounded the accessible information of any ensemble of states with first moment ρ by the *subentropy* $Q(\rho)$ [Eq. (61)] and found that the Scrooge ensemble achieves this lower bound. A nontrivial, testable consequence of their result is that the mutual information $I(O_A; Z_B|T)$ is equal to the subentropy, for any complete basis O_A :

$$I(O_A; Z_B|T) \approx Q(\rho_A). \quad (61)$$

Here, the *conditional* mutual information $I(O_A; Z_B|T)$ refers the mutual information between the variables O_A and Z_B of the probability distribution $p(o_A, z_B, t)$ at *fixed times* t , i.e., conditioning on the time T . As a reminder, we denote variables with uppercase letters and their values in lowercase. Although time does not play a role here (as long as the evolution time is sufficiently long), we primarily adopt this notation anticipating our result in Sec. VIC, where the dependence on time is important.

The operational meaning of the mutual information can be understood by considering the following scenario. Suppose Alice and Bob share an entangled quantum many-body state obtained from ergodic Hamiltonian dynamics of duration t . Bob performs a measurement on his (larger) portion of the state in the computational basis, resulting in a measurement outcome z_B . Alice performs the measurement of an arbitrary (potentially optimized) local observable leading to an outcome o_A . How much information do Alice and Bob share in this exercise? Or, equivalently, how much can Alice or Bob know about the measurement outcomes of the other from their own measurement outcomes? Our result implies that they know *as little as possible*, while some amount of information is inevitably shared due to energy conservation. Furthermore, this feature is universal, independent of the choice of Alice's observable as long as Bob's measurement basis realizes a Scrooge ensemble (we discuss the effects of measurement basis below).

We verify this property in Fig. 4(f). We choose three different measurement bases $O_A \in \{X^{\otimes 4}, Y^{\otimes 4}, Z^{\otimes 4}\}$, i.e., the X , Y , and Z bases on the subsystem A . We verify that the mutual information $I(O_A; Z_B|T)$ of a time-evolved state $|\Psi(t)\rangle$ is independent of O_A and agrees with the subentropy $Q(\rho_A)$ across a wide range of energy densities of the initial state $|\Psi_0\rangle$, i.e., over a wide range of temperatures of the reduced density matrix ρ_A .

Taken together, our results provide a surprising universality in the behavior of projected ensembles, even away from infinite temperature. When the basis Z_B is uncorrelated with the projected states $|\Psi(z_B)\rangle$, they are statistically described by independent samples from the Scrooge ensemble $|\Psi(z_B)\rangle \sim \text{Scrooge}[\rho_A]$. We discussed measurable signatures of the Scrooge ensemble: statistics of the outcomes probabilities $p_{o_A}(z_B)$ and universal behavior of the conditional mutual information $I(O_A; Z_B|T)$. However, we have not yet provided a derivation for the emergence of the Scrooge ensemble. Furthermore, our results in this section do not hold for generic measurement bases. In the following section, we resolve both issues by deriving a *generalized Scrooge ensemble* from the temporal ensemble.

VI. FROM TEMPORAL TO PROJECTED ENSEMBLES: THE EFFECT OF MEASUREMENT BASIS

As aforementioned, properties of the projected ensemble depend on the measurement basis on B : Our findings in Sec. V hold only for certain measurement bases. In this section, we seek to extend our description of projected ensembles to general bases. Instead of Scrooge ensembles, we find that projected ensembles are described by a generalized Scrooge ensemble, which we introduce, that accounts for correlations between z_B and the projected states $|\Psi(z_B)\rangle$.

For projected ensembles generated by from time-evolved states, we derive this result from properties of the temporal ensemble. Our finding is more general than time-evolved states: We numerically find that the projected ensembles arising from energy eigenstates are also well described by generalized Scrooge ensembles. We discuss measurable and information-theoretic signatures of the generalized Scrooge ensemble, which mirror that of the Scrooge ensemble. Finally, we identify conditions under which the generalized Scrooge ensemble reduces to the Scrooge ensemble. These results give a complete description of the projected ensembles generated by time-evolved states, proven for generic Hamiltonian dynamics, and broaden our understanding of deep thermalization.

As an illustrative example, we consider the MFIM. For convenience, we repeat its Hamiltonian [Eq. (36)] here:

$$H_{\text{MFIM}} = \sum_{j=1}^N h_x X_j + h_y Y_j + \sum_{j=1}^{N-1} J X_j X_{j+1}. \quad (62)$$

The projected ensemble is equal to the Scrooge ensemble when the subsystem B is measured in the Z_B basis. In other bases, e.g., the X_B basis, there are correlations between the measurement outcomes $|x_B\rangle$ and the projected states $|\Psi(x_B)\rangle$. In particular, their energies are anticorrelated: Measurement outcomes $|x_B\rangle$ with high energies

$\langle x_B | H_B | x_B \rangle$ are likely to result in projected states with lower energies $\langle \Psi(x_B) | H_A | \Psi(x_B) \rangle$ (and vice versa), where H_A and H_B are the terms of the Hamiltonian restricted to the subsystems A and B , respectively. Note that the energies we refer to here are local energies on subsystems A and B , and, therefore, such energetic correlations are not strict but exist in a statistical sense; see, e.g., Ref. [122]. Under these conditions, one must specify exponentially many probability distributions which take into account the correlations between x_B and $|\Psi(x_B)\rangle$ arising from energy or other conservation laws. This is in contrast to what we have discussed in Sec. V, where we could treat the projected states $|\Psi(x_B)\rangle$ with different labels (measurement outcomes) x_B as samples from the same probability distribution. To emphasize the effect of measurement bases, in this section we refer to such measurement outcomes and projected states as x_B and $|\Psi(x_B)\rangle$, respectively.

Special case: Magnetization-sectored bases.—To gain intuition about the effects of measurement basis, it is illustrative to consider a special case: one where the outcomes can be partitioned into sets (“sectors”), within which there are no further correlations between the outcomes and projected states. For example, the MFIM we consider has a Zeeman field in the Y axis. When this Zeeman field is sufficiently large, the sector M of bit strings $|y_B\rangle$ with the same magnetization have approximately the same energy on B but have different energies from bit strings from other sectors. One consequence of this is the following: Consider the “sector-resolved reduced density matrix,” defined as

$$\rho_A(M) = \frac{\sum_{y_B \in M} P(y_B) |\Psi(y_B, t)\rangle \langle \Psi(y_B, t)|}{\sum_{y_B \in M} P(y_B)}. \quad (63)$$

These mixed states $\rho_A(M)$ show systematic differences across magnetization sectors M . Accordingly, it is not appropriate to take the projected states from $|\Psi(y_B, t)\rangle$ as identical samples from the same ensemble of states. Rather, projected states corresponding to the same sector M are drawn from the same ensemble. As we show below, these are Scrooge ensembles, with different first moments:

$$\forall y_B \in M, \quad |\Psi(y_B)\rangle \sim \text{Scrooge}[\rho_A(M)]. \quad (64)$$

However, this sector picture is valid only in special cases. Our generalized Scrooge ensemble extends this observation to the general case.

A. Generalized Scrooge ensemble

Our key idea is to treat each projected state $|\Psi(x_B)\rangle$ as a random sample from a *different* Scrooge ensemble $\text{Scrooge}[\bar{\rho}(x_B)]$, where $\bar{\rho}(x_B)$ is a density matrix associated

to each outcome x_B . We define the density matrix $\bar{\rho}(x_B)$ be the time-averaged state

$$\begin{aligned}\bar{\rho}(x_B) &\equiv \mathbb{E}_t[|\tilde{\Psi}(x_B)\rangle\langle\tilde{\Psi}(x_B)|]/\mathbb{E}_t[|\tilde{\Psi}(x_B)\rangle\langle\tilde{\Psi}(x_B)|] \\ &= (\mathbb{I}_A \otimes \langle x_B|)\rho_d(\mathbb{I}_A \otimes |x_B\rangle)/p_d(x_B),\end{aligned}\quad (65)$$

$$\begin{aligned}p_d(x_B) &\equiv \mathbb{E}_t[\langle\tilde{\Psi}(x_B)|\tilde{\Psi}(x_B)\rangle] \\ &= \text{tr}[(\mathbb{I}_A \otimes \langle x_B|)\rho_d(\mathbb{I}_A \otimes |x_B\rangle)].\end{aligned}\quad (66)$$

$p_d(x_B)$ is the time-averaged probability of measuring the outcome x_B , and we express both $\bar{\rho}(x_B)$ and $p_d(x_B)$ in terms of the diagonal ensemble $\rho_d \equiv \mathbb{E}_t[|\Psi(t)\rangle\langle\Psi(t)|]$. Unlike the Scrooge ensemble, where the projected ensemble consists of i.i.d. samples from $\text{Scrooge}[\rho]$, here, each projected state is an independent sample from nonidentical distributions, $|\Psi(x_B)\rangle \sim \text{Scrooge}[\bar{\rho}(x_B)]$. We refer to this statistical superposition of probability distributions, each sampled from once, as the generalized Scrooge ensemble.

Claim 2.—The states $|\Psi(x_B)\rangle$ of the projected ensemble are sampled from the generalized Scrooge ensemble

$$|\Psi(x_B)\rangle \sim \text{Scrooge}[\bar{\rho}(x_B)]. \quad (67)$$

As stressed in Sec. II B, there are no free parameters in this statement: The initial state, Hamiltonian, and measurement outcomes $\{x_B\}$ fix the reduced density matrices $\bar{\rho}(x_B)$. Furthermore, Claim 2 can be tested through its low moments. Specifically, we fix a point in time and compute the k th moment of the projected ensemble. There are fluctuations arising from the fact that each Scrooge ensemble is sampled only once in the generalized Scrooge ensemble. However, these fluctuations average to an exponentially small value, since we are summing over a large number of projected states. Therefore, the value of the low moments is effectively independent of time—Claim 2 predicts that these moments are equal to the weighted k th moment of the Scrooge ensembles $\text{Scrooge}[\bar{\rho}(x_B)]$. We verify this numerically by comparing the k th moments of the normalized projected ensemble with the predicted k th moment

$$\rho_{\text{Gen Scr}}^{(k)} = \sum_{x_B} p_d(x_B) \rho_{\text{Scrooge}}^{(k)}[\bar{\rho}(x_B)]. \quad (68)$$

As illustrated in Fig. 5(b), the higher moments of the projected ensemble agree with $\rho_{\text{Gen Scr}}^{(k)}$ and differ from $\rho_{\text{Scrooge}}^{(k)}$ [Eq. (50)], with exponential convergence with increasing system size. Here, we calculate the average states $\bar{\rho}(x_B)$ by numerically time averaging $|\tilde{\Psi}(x_B)\rangle\langle\tilde{\Psi}(x_B)|$ over a long interval. Having numerically demonstrated our

claim, we proceed to derive it with results from the temporal ensemble.

1. Temporal ensemble of projected states

In order to derive results about projected ensembles, we first establish a result about the temporal ensemble of a projected state. Specifically, we consider the set of unnormalized states $|\tilde{\Psi}(x_B, t)\rangle$ for a fixed outcome x_B . We show that the time trajectory (temporal ensemble) is described by the unnormalized ensemble $\tilde{\text{Scrooge}}[\bar{\rho}(x_B)]$.

Lemma 1 (temporal k th moment of the projected state). Consider the temporal ensemble comprised of the (unnormalized) projected states $|\tilde{\Psi}(x_B, t)\rangle \equiv (\mathbb{I}_A \otimes \langle x_B|)|\Psi(t)\rangle$. The k th moment of this temporal ensemble is equal to the k th moment of the corresponding unnormalized Scrooge ensemble, up to small corrections:

$$\begin{aligned}\rho_{\text{Temp Proj}}^{(k)}(x_B) &\equiv \mathbb{E}_t[|\tilde{\Psi}(x_B, t)\rangle\langle\tilde{\Psi}(x_B, t)|^{\otimes k}] \\ &= p_d(x_B)^k \bar{\rho}_{\text{Scr}}^{(k)}[\bar{\rho}(x_B)] + O[\Delta_\beta(x_B) p_d(x_B)^k],\end{aligned}\quad (69)$$

where $\Delta_\beta(x_B) \equiv \text{tr}[(\mathbb{I}_A \otimes |x_B\rangle\langle x_B|)^{\otimes 2} \rho_d^{(2)}]/p_d(x_B)^2$ is a number, typically exponentially small in system size, characterizing the validity of our approximation and where $\rho_d^{(2)} \equiv |c_E|^4 |E, E\rangle\langle E, E|$.

Proof. The leading term follows from Theorem 2. However, the bound $O(\Delta_\beta(x_B))$ does not immediately follow from the bound given in Theorem 2, which is too weak and should be improved for this context.

We outline the proof here, with full details in Appendix I. We use an intermediate step in the proof of Theorem 2 in Appendix D. This step expresses the difference $\rho_{\text{Temp Proj}}^{(k)}(x_B) - p_d(x_B)^k \bar{\rho}_{\text{Scr}}^{(k)}[\bar{\rho}(x_B)]$ in terms of lists of energies $(E_1, \dots, E_k)$ —those with at least one repeated element. This allows us to replace two copies of the diagonal matrix $\rho_d^{\otimes 2} \equiv \sum_{E_1, E_2} |c_{E_1}|^2 |c_{E_2}|^2 |E_1, E_2\rangle\langle E_1, E_2|$ with the term $\rho_d^{(2)} \equiv \sum_E |c_E|^4 |E, E\rangle\langle E, E|$. The relative size of the correction term is then $\Delta_\beta(x_B)$, which is typically exponentially small in total system size (Appendix J). ■

2. Projected ensemble of time-evolved states

While Lemma 1 is interesting in its own right, we use our notion of ergodicity (Sec. IV E) to turn it into a statement about the projected ensemble at fixed (but generically late) points in time. The below theorem provides a rigorous basis for deep thermalization in Hamiltonian dynamics, improving upon previous numerical observations at infinite temperature [29].

Theorem 4. At typical (late) times t , the weighted k th moment $\bar{\rho}_{\text{Proj}}^{(k)}(t)$ of the unnormalized projected ensemble is

equal to the k th moment of the generalized Scrooge ensemble:

$$\tilde{\rho}_{\text{Proj}}^{(k)}(t) \equiv \sum_{x_B} \frac{|\tilde{\Psi}(x_B, t)\rangle \langle \tilde{\Psi}(x_B, t)|^{\otimes k}}{p_d(x_B)^{k-1}} \quad (70)$$

$$= \sum_{x_B} p_d(x_B) \tilde{\rho}_{\text{Scr}}^{(k)}[\tilde{\rho}(x_B)] + O(\Delta_\beta^{1/2}), \quad (71)$$

where $\Delta_\beta \equiv \sum_{x_B} p_d(x_B) \Delta_\beta(x_B)$, the weighted mean of $\Delta_\beta(x_B)$.

Proof. From Lemma 1, the *time-averaged* k th moment $\mathbb{E}_t[\tilde{\rho}_{\text{Proj}}^{(k)}(t)]$ is approximately equal to the weighted sum $\sum_{x_B} p_d(x_B) \tilde{\rho}_{\text{Scr}}^{(k)}[\tilde{\rho}(x_B)]$. We then show in Appendix J that the variance over time $\text{var}_t[\tilde{\rho}_{\text{Proj}}^{(k)}(t)]$ is exponentially small, and, therefore, the projected k th moment $\tilde{\rho}_{\text{Proj}}^{(k)}(t)$ is close to its temporal average. The temporal fluctuations are of size $\|\text{var}_t[\tilde{\rho}_{\text{Proj}}^{(k)}(t)]\| = O(\Delta_\beta)$. While the most general bound on Δ_β that we can prove is $\Delta_\beta \leq 1$, we argue in Appendix J that it is typically exponentially small. ■

Finally, we note that the k th moment described in Eq. (J1) is a slightly different moment than both the moments of the normalized and unnormalized projected ensembles. Each projected state is weighted by the factor $p_d(x_B)^{1-k}$ in order to have a quantity that can be bounded by a small quantity (Δ_β). The norm $\langle \tilde{\Psi}(x_B) | \tilde{\Psi}(x_B) \rangle$ fluctuates about the value $p_d(x_B)$. These fluctuations are not universal. [123] Accordingly, the average value of $\langle \tilde{\Psi}(x_B) | \tilde{\Psi}(x_B) \rangle^k / p_d(x_B)^k$ is equal to the k th moment of this distribution, which is larger than 1. With increasing size of the subsystem A , these values concentrate around 1, and, therefore, our weighted k th moment in Eq. (J1) converges to the normalized ensemble moment.

B. Porter-Thomas distribution in generalized Scrooge ensembles

In Secs. IV C and V F, we discussed the Porter-Thomas distribution as a measurable signature of the temporal and Scrooge ensembles, respectively. Perhaps unsurprisingly, the PT distribution is also present in the generalized Scrooge ensemble but only after suitable data processing. In Sec. V F, we established that, in the Scrooge ensemble, the joint probabilities $p_{o_A}(z_B)$ follow a PT distribution with mean $\mathbb{E}_{z_B}[p_{o_A}(z_B)] \equiv \langle o_A | \rho_A | o_A \rangle / D_B$. In the generalized Scrooge ensemble, since each projected state $|\tilde{\Psi}(x_B)\rangle$ is sampled from a different Scrooge ensemble, we must normalize the overlaps $p_{o_A}(x_B)$ by their respective time-averaged values $\mathbb{E}_t[p_{o_A}(x_B)] = \langle o_A | \tilde{\rho}(x_B) | o_A \rangle$ (see Sec. IV E). Having done so,

$$p(o_A, z_B) / \mathbb{E}_t[p(o_A, z_B)] \sim \text{PT}(1), \quad (72)$$

as illustrated in Fig. 5(c).

C. Interaction information

We next turn to the information-theoretic properties of the generalized Scrooge ensemble. Specifically, we consider the mutual information $I(O_A; X_B)$ between any basis $\{o_A\}$ on A and projective measurements $\{|x_B\rangle\}$. Unlike in the Scrooge ensemble, there are additional correlations between x_B and the projected states $|\Psi(x_B)\rangle$. With the generalized Scrooge ensemble, we find that the mutual information $I(O_A; X_B|T)$ [i.e., the mutual information of the distribution $p(o_A, z_B, t)$ at fixed, late times] can be written as follows:

$$I(O_A; X_B|T) \approx I(O_A; X_B) + \sum_{x_B} p_d(x_B) Q[\tilde{\rho}(x_B)], \quad (73)$$

where the first term $I(O_A; X_B)$ refers to correlations that exist even in the time-averaged distribution $\mathbb{E}_t[p(o_A, z_B, t)]$. These correlations are *nonuniversal* and are thermal in nature, arising from energy correlations between the bases $\{|o_A\rangle\}$ and $\{|x_B\rangle\}$. In contrast, the second term captures the correlations associated with fluctuations in time. These fluctuations in time are universal and are expressed as a weighted sum of subentropies $Q[\tilde{\rho}(x_B)]$. Equation (73) follows from the following:

$$\begin{aligned} I(O_A; X_B|T) - I(O_A; X_B) \\ = H(O_A|T) - H(O_A|X_B, T) - H(O_A) + H(O_A|X_B) \\ \approx H(O_A|X_B) - H(O_A|X_B, T) \end{aligned} \quad (74)$$

$$\approx \sum_{x_B} p_d(x_B) [H(O_A|X_B = x_B) - H(O_A|X_B = x_B, T)] \quad (75)$$

$$\approx \sum_{x_B} p_d(x_B) Q[\tilde{\rho}(x_B)]. \quad (76)$$

In Eq. (74), we use the fact that $H(O_A|T) \approx H(O_A)$, because the time-dependent reduced density matrix ρ_A is exponentially close to its time-averaged value $\text{tr}_B[\rho_d]$. We then approximate $H(O_A|X_B, T) = \sum_{x_B} p(x_B) H(O_A|X_B = x_B, T) \approx p_d(x_B) H(O_A|X_B = x_B, T)$. This is justified, because $p(x_B, t)$ fluctuates about $p_d(x_B)$. When summed over in $H(O_A|X_B, T)$, these fluctuations average away and we can replace $p(x_B)$ with $p_d(x_B)$. Equation (75) is a weighted sum of the mutual information for the conditional distributions $p(o_A, t|x_B)$. Theorem 4 allows us to equate this mutual information with the subentropies $Q[\tilde{\rho}(x_B)]$, giving the desired Eq. (73).

The quantity $I(O_A; X_B; T) \equiv I(O_A; X_B|T) - I(O_A; X_B)$ is known as the *interaction information*, which is a multivariate analog of the mutual information [124]. It is a measure of the correlations that exist due to interaction among all variables, is symmetric under exchange of all variables, and is also the classical analog of the (quantum)

tripartite mutual information studied in many-body physics [125,126]. In particular, our result predicts that the interaction information is independent of basis O_A and has expression $\sum_{x_B} p_d(x_B) Q[\bar{\rho}(x_B)]$, which we can bound as $\sum_{x_B} p_d(x_B) Q[\bar{\rho}(x_B)] \leq Q[\rho_A]$ using the concavity of the subentropy [105]. The interaction information quantifies the amount of information unique to a time-evolved state, i.e., with time-averaged correlations subtracted. Similar to our results in the temporal and Scrooge ensembles, this information is universal and is minimized subject to constraints.

We demonstrate this result with numerical data in Fig. 5(d), in which we plot the interaction information for three measurement bases $O_A \in \{X^{\otimes 4}, Y^{\otimes 4}, Z^{\otimes 4}\}$. While the mutual information $I(O_A; X_B|T)$ depends on the choice of O_A (inset), the interaction information $I(O_A; X_B; T)$ is independent and is equal to the weighted average subentropy [Eq. (76)].

D. Recovering the Scrooge ensemble from the generalized Scrooge ensemble

The generalized Scrooge ensemble in Theorem 4 specializes to the original Scrooge ensemble when the states $\bar{\rho}(x_B)$ are independent of measurement outcome x_B , i.e., $\bar{\rho}(x_B) \propto \rho_A$. As presented in Sec. V, this is approximately true when states evolved by the MFIM are measured in the $\{|z_B\rangle\}$ basis. The crucial property of the Z_B measurement basis is that the energy expectation values $E(|z\rangle) \equiv \langle z|H|z\rangle$ and energy variances $\sigma_H^2(|z\rangle) \equiv \langle z|(H - E_z)^2|z\rangle$ are independent of the bit string z . This is an example of “minimal correlation” between measurement basis and energy eigenbasis.

We argue that these two properties are sufficient for the Scrooge ensemble to be recovered approximately for z_B measurements in the MFIM. In particular, we show that, for any state $|\phi_A\rangle \in \mathcal{H}_A$, $\langle \phi_A | \rho_d(z_B) | \phi_A \rangle \simeq \langle \phi_A | \rho_d | \phi_A \rangle$ up to subleading corrections in the size N of the system, and, therefore, $\rho_d(z_B) \simeq \rho_d$.

From the definition of $\rho_d(z_B)$, the expectation value $\langle \phi_A | \rho_d(z_B) | \phi_A \rangle$ can be expressed in terms of the eigenstates $|E\rangle$ of H as

$$\langle \phi_A | \rho_d(z_B) | \phi_A \rangle = \sum_E |c_E|^2 |\langle \phi_A, z_B | E \rangle|^2. \quad (77)$$

We next use the theorem proved in Ref. [127], which states that, for large N , the energy distribution of the product state $|\phi_A, z_B\rangle$ converges to a Gaussian centered about $E(|\phi_A, z_B\rangle) = \langle \phi_A, z_B | H | \phi_A, z_B \rangle$ with variance $\sigma_H^2(|\phi_A, z_B\rangle) = \langle \phi_A, z_B | [H - E(|\phi_A, z_B\rangle)]^2 | \phi_A, z_B \rangle$. Furthermore, the mean $E(|\phi_A, z_B\rangle)$ and variance $\sigma_H^2(|\phi_A, z_B\rangle)$ are independent of z_B up to subleading corrections. To show this, we write the Hamiltonian as $H = H_A + H_B + H_{AB}$, where H_A and H_B act only on subsystems A and B , respectively, and H_{AB} is the remaining

term, acting on the boundary between A and B . As mentioned above, the expectation value of H_B and variance are independent of z_B :

$$E_B(|z_B\rangle) \equiv \langle z_B | H_B | z_B \rangle = 0, \quad (78)$$

$$\begin{aligned} \sigma_{H_B}^2(|z_B\rangle) &\equiv \langle z_B | [H_B - E_B(|z_B\rangle)]^2 | z_B \rangle \\ &= N_B(h_x^2 + h_y^2) + (N_B - 1)J^2. \end{aligned} \quad (79)$$

We also have $\langle \phi_A, z_B | H_{AB} | \phi_A, z_B \rangle = 0$, which gives

$$E(|\phi_A, z_B\rangle) = E_A(|\phi_A\rangle), \quad (80)$$

$$\sigma_H^2(|\phi_A, z_B\rangle) = \sigma_{H_A}^2(|\phi_A\rangle) + \sigma_{H_B}^2 + O(1), \quad (81)$$

where the terms $O(1)$ come from the boundary contributions and, hence, do not scale with system size. From Eqs. (80) and (81), we find that the energy distribution of $|\phi_A, z_B\rangle$ and, consequently, the quantity $\langle \phi_A | \rho_d(z_B) | \phi_A \rangle$ in Eq. (77) are independent of the measurement outcome z_B (up to subleading corrections in system size), recovering the Scrooge ensemble from the generalized Scrooge ensemble in the limit of large system sizes.

E. Projected ensembles of energy eigenstates

Thus far, we have investigated the projected ensembles of time-evolved states $|\Psi(t)\rangle$ and provided analytical arguments based on the temporal ensemble. In fact, the generalized Scrooge ensemble also seems to describe projected ensembles obtained from energy eigenstates, as verified by numerical simulation results illustrated in Fig. 6.

Here, we consider the projected ensembles arising from eigenstates of the MFIM, with additional terms which remove time-reversal symmetry [Eq. (82) and Appendix L]. We first test our simpler prediction: the emergence of the Scrooge ensemble by measuring in the Z_B basis. In Fig. 6(b), the projected ensemble agrees with its respective Scrooge ensemble for energy eigenstates at all energies but agrees with the Haar ensemble only at infinite temperature, when $E/N \approx 0$. Furthermore, the distance between the projected and Scrooge ensemble decreases exponentially with total system size [Fig. 6(c)].

We next test our generalized Scrooge ensemble prediction, by generating the projected ensemble with measurements in the X_B basis. In this setting, it is difficult to make quantitatively accurate predictions for the generalized Scrooge ensemble, because we do not have a prescription of the average state $\bar{\rho}(x_B)$. The expression $\bar{\rho}(x_B) \propto \mathbb{E}_t[|\tilde{\Psi}(x_B, t)\rangle \langle \tilde{\Psi}(x_B, t)|]$ in time-evolved states does not apply to energy eigenstates. Nevertheless, motivated by the ETH, we expect that the eigenstates with nearby energy values will generate highly similar projected ensembles: In particular, the average states $\bar{\rho}(x_B, E)$ should be smooth functions of the total energy. Therefore, we can

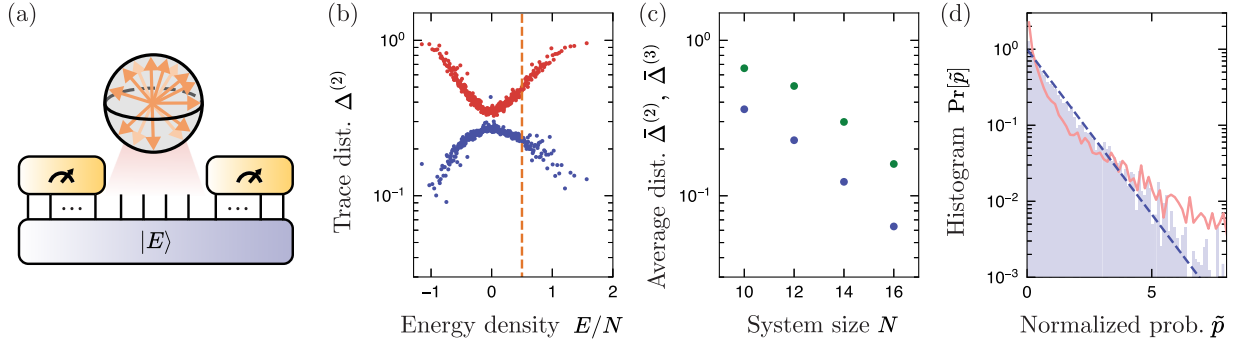


FIG. 6. Scrooge and generalized Scrooge ensembles from energy eigenstates. (a) We study the projected ensembles arising from energy eigenstates of the MFIM with broken time-reversal symmetry [Eq. (82)], with total size $N = 12$ and a subsystem of size $N_A = 4$ in the middle of the chain. (b) For each eigenstate $|E\rangle$, we construct its reduced density matrix $\rho_A^{(E)}$ and corresponding Scrooge ensemble $\text{Scrooge}[\rho_A^{(E)}]$. The projected ensemble is always close to the Scrooge ensemble prediction, as measured by the trace distance $\Delta^{(2)}$ of their second moments (blue) and is close to the Haar ensemble only at infinite temperature, when $E/N \approx 0$ (red). Curiously, and unexpectedly, the generalized Scrooge ensemble appears to be better descriptions at the edges of the spectrum. (c) The trace distances $\Delta^{(2)}$ (blue) and $\Delta^{(3)}$ (green), averaged over the 100 eigenstates closest to the target energy $E = 0.5N$ [orange line in (b)], decrease exponentially with system size. (d) For a fixed eigenstate $|E\rangle$, the histogram of normalized measurement probabilities $\tilde{p} \equiv p_{x_A}^{(E)}(x_B)/\mathbb{E}_E[p_{x_A}^{(E)}(x_B)]$ follows a PT distribution, while the raw probabilities $Dp_{x_A}^{(E)}(x_B)$ (red line) do not, supporting the presence of the generalized Scrooge ensemble. Here, we plot $x_A = 0010$ as a representative.

estimate $\bar{p}(x_B, E)$ by averaging the projected states $|\tilde{\Psi}_E(x_B)\rangle\langle\tilde{\Psi}_E(x_B)|$ over energy eigenstates of a sufficiently narrow energy window. While this method is not quantitatively accurate enough to give exponentially decreasing trace distances [the equivalent of Fig. 6(c) but for the generalized Scrooge ensemble], we can observe the PT distribution in normalized outcome probabilities [Fig. 6(d)]. We first average the probability outcomes $p_{x_A}^{(E)}(x_B) \equiv |\langle E|x_A, x_B\rangle|^2$ over a small energy window to obtain $\mathbb{E}_E[p_{x_A}^{(E)}(x_B)]$ for each value of x_A and x_B . The histogram of normalized probabilities $\tilde{p} \equiv p_{x_A}^{(E)}(x_B)/\mathbb{E}_E[p_{x_A}^{(E)}(x_B)]$ over all x_B 's (with fixed x_A) shows a PT distribution, while the raw probabilities $p_{x_A}^{(E)}(x_B)$ do not.

1. Time-reversal symmetry and the real Scrooge ensemble

Finally, we remark on the role of symmetry. Unlike in time-evolved states, the presence of time-reversal symmetry in a Hamiltonian H qualitatively affects the projected ensembles from energy eigenstates. Time-reversal symmetry—defined as the existence of a unitary operator U that maps H to its complex conjugate H^* : $UHU^\dagger = H^*$ [128]—ensures that all eigenstates of H have purely real coefficients in a certain basis. If this coincides with the measurement basis (e.g., the X_B basis in the MFIM), then the projected states are guaranteed to be real vectors. In such cases, the projected ensemble instead follows a *real Scrooge ensemble*, which is an analogous “ ρ distortion” of the ensemble of real random vectors. The real Scrooge ensemble has corresponding signatures such as a “real PT distribution” [98] which we demonstrate in Appendix L.

Here, we simply break the time-reversal symmetry by adding appropriate terms:

$$H'_{\text{MFIM}} = \sum_{j=1}^N h_x X_j + h_y Y_j + h_z Z_j + \sum_{j=1}^{N-1} J X_j X_{j+1} + J' Y_j Y_{j+1}, \quad (82)$$

with $J' = 0.4$, $h_z = 0.5$, and the values of h_x , h_y , and J kept the same as the other numerical examples [Eq. (36)].

The presence of time-reversal symmetry was not relevant to the projected ensembles of time-evolved states because the time evolution naturally breaks time-reversal symmetry. However, time-evolved states can be engineered to be purely real, requiring both time-reversal and particle-hole symmetry and suitably symmetric initial states and measurement bases (Appendix L).

VII. PORTER-THOMAS DISTRIBUTIONS AS SIGNATURES OF RANDOMNESS: HIGHER-ORDER ETH

In this work, we demonstrate instances of the PT distribution emerging from the maximum entropy principle in many-body dynamics: In the temporal and Scrooge ensembles, the PT distribution describes the *probabilities of outcome probabilities* $\text{Pr}[p]$. We found that, in the generalized Scrooge ensemble, it was necessary to appropriately rescale the outcome probabilities to obtain the PT distribution.

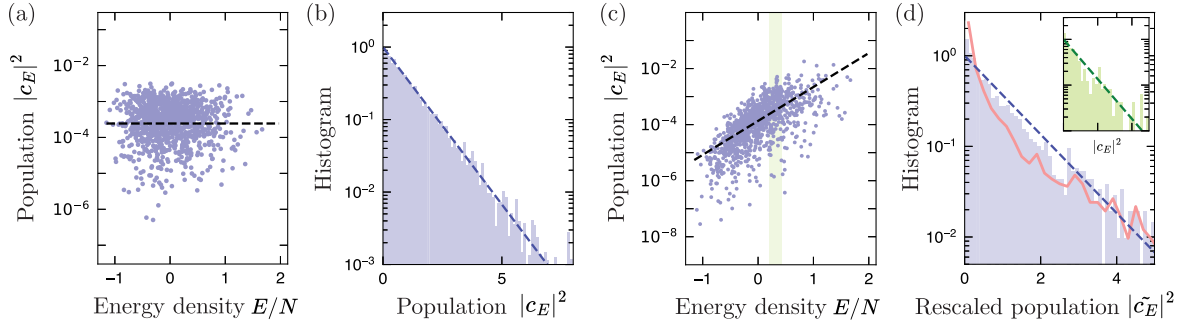


FIG. 7. PT distribution in energy populations. We observe the presence of the PT distribution in the energy populations $|c_E|^2 \equiv |\langle E|\Psi_0\rangle|^2$ as a consequence of the “higher-order ETH” [112]. In a narrow energy window, the populations are distributed according to a Porter-Thomas distribution, with a smoothly varying mean $f(E)$. (a) A state at infinite effective temperature $|\Psi\rangle = |0\rangle^{\otimes N}$ has constant mean $f(E)$ (black dashed line). (b) The histogram of $|c_E|^2$ over *all eigenstates* follows a PT distribution. (c) In contrast, for a state $|\Psi\rangle = [\exp(i\theta X/2)|0\rangle]^{\otimes N}$ at finite temperature (details below), $|c_E|^2$ monotonically increases about the smooth function $f(E) \propto \exp(\beta E)$, with β fitted (dashed line). (d) The histogram of $|c_E|^2/f(E)$ shows a PT distribution, both over the entire spectrum as well as in a narrow energy window [inset, energy window denoted by green band in (c)], while the raw populations $D|c_E|^2$ at finite temperature (red line) do not.

Here, we demonstrate another instance of the PT distribution, hinting at the wider applicability of the maximum entropy principle in many-body physics. Specifically, we find the PT distribution in the distribution of overlaps $|c_E|^2 = |\langle \Psi|E\rangle|^2$ between energy eigenstates $|E\rangle$ and a simple physical state $|\Psi\rangle$. This is consistent with a “higher-order” or k ETH [112], in which energy eigenstates in a narrow energy window are conjectured to form a state k -design in an appropriate microcanonical subspace. This is a generalization of the conventional ETH, which is a statement about the behavior of properties such as the overlaps $\langle E_j|O|E_k\rangle$ between energy eigenstates $|E_k\rangle$ and operators O . Specifically, Ref. [112] defines the k ETH to hold for a given k -copy operator O if

$$\mathbb{E}_t[U_t^{\otimes k} \Pi_E O \Pi_E U_t^{\otimes k}] = \mathbb{E}_U[U^{\otimes k} \Pi_E O \Pi_E U^{\dagger, \otimes k}], \quad (83)$$

where Π_E is a projector onto a microcanonical subspace $\mathcal{H}_{\text{MC}}$, spanned by the energy eigenstates in a narrow energy window, $U_t \equiv \exp(-iHt)$, and the average over U is over Haar-random unitaries acting on the microcanonical space $\mathcal{H}_{\text{MC}}$. They numerically verify that Eq. (83) approximately holds when $O = o^{\otimes k}$ (for local operators o) and when $O = \mathcal{S}_A$ is the two-copy swap operator on a subsystem A , which they used to derive the Page correction in the 2-Rényi entropy.

Here, we take $O = |\Psi_0\rangle\langle\Psi_0|^{\otimes k}$. Assuming that the k ETH holds for such operators, Eq. (83) predicts that the eigenstate overlaps $|c_E|^2$ in a microcanonical window follow a Porter-Thomas distribution, with a mean value $f(E)$:

$$|c_E|^2 \sim \text{PT}[f(E)]. \quad (84)$$

We further expect $f(E)$ to be a smooth function of the energy. By dividing $|c_E|^2$ by a fitted smooth function $f(E)$, we expect the *rescaled overlap* to be distributed according to a PT distribution. We empirically find that $f(E)$ is often well described by a Boltzmann distribution $\exp(-\beta E)$. In Figs. 7(a)–7(d), we demonstrate the emergence of the PT distribution in the populations over the eigenstates of the time-reversal-broken MFIM H' [Eq. (82)] for $N = 12$ for an infinite-temperature state $|\Psi_0\rangle = |0\rangle^{\otimes N}$ [Figs. 7(a) and 7(b)] and a finite-temperature state $|\Psi_0\rangle = [\exp(i\theta X/2)|0\rangle]^{\otimes N}$ [Figs. 7(c) and 7(d)], with $\theta = 0.3557$ chosen such that $\langle\Psi_0|H|\Psi_0\rangle = 0.5N$.

The presence of the PT distribution in the energy populations not only serves as an easily visualized consequence of the k ETH, it also accounts for observations in the literature, such as universal ratios of 2 and 3 between the spectral form factor (SFF) and survival probability at late times [129], which follow from the second moment of the complex and real PT distributions. Specifically, the SFF is expressed as $|\text{tr}[\exp(-iHt)]|^2/D^2 = \sum_{E,E'} \exp[i(E' - E)t]/D^2$. At late times, only the $E = E'$ contribution remains and the SFF has a late-time value of 1. In contrast, the survival probability of an initial state $|\Psi_0\rangle$ has expression $|\langle\Psi_0|\exp(-iHt)|\Psi_0\rangle|^2 = \sum_{E,E'} |c_E|^2 |c_{E'}|^2 \exp[i(E' - E)t]$. Again, at late times, only the $E = E'$ term survives, and we obtain $\sum_E |c_E|^4$. From the discussion above, at infinite effective temperature, this sum is approximately $2/D$ or $3/D$, when H , respectively, does not and does have time-reversal symmetry, and $|\Psi_0\rangle$ respects these symmetries.

VIII. OUTLOOK

In this work, we have demonstrated that a novel type of maximum entropy principle arises in natural chaotic

many-body systems. We substantiate our findings by studying two different ensembles of states, each underlying the phenomena of Hilbert-space ergodicity and deep thermalization, and find them well described by ensembles with maximum entropy. Furthermore, we present the operational meaning, information-theoretic properties, and simple, experimentally accessible signatures, i.e., Porter-Thomas distribution in certain rescaled observables, of our claimed universality.

Our work extends our understanding of deep thermalization, most notably away from infinite temperature. We leave it to future work to generalize more features of infinite-temperature projected ensembles, such as the existence of distinct thermalization times for k th moments [30,31,36] and the dynamics at short times [84]. Our maximum entropy principle may also serve as a simple theoretical prediction to complex systems such as projected ensembles in more general settings. For example, Ref. [130] studies the projected ensemble in an extremely different system—generated by Gaussian states in continuous variable systems—and discovers an ensemble with minimum accessible information. This suggests that our results may have broader applicability to the projected ensembles in various settings, such as under constraints from symmetries [37].

The universality we establish in this work may be exploited for practical purposes. For example, in Ref. [110], the presence of the PT distribution is leveraged to design a protocol to estimate fidelities in natural many-body systems. In companion experimental work, we explore deviations of the PT distribution in the presence of a bath. By understanding these deviations from universality, we design protocols to learn properties of the bath [41]. Further applications may include generalizing schemes that use aspects of quantum randomness to natural systems, such as randomized measurement protocols (including classical shadow tomography) [103,104,111,131] (with applications to characterizing thermal states [132]), as well as random number generation [133,134].

Finally, our results should be seen as complementary to and distinct from conventional results such as the ETH. The ETH models some aspects of ergodic many-body systems with random matrix theory, which has been successful in describing striking features such as the level repulsion in eigenvalues. In recent years, there has been a growing body of work scrutinizing as well as generalizing the ETH, motivated by recent interest in quantities such as multitime correlators. See, e.g., Refs. [112,135–139]. Our findings in the temporal and projected ensembles as well as our analysis of the k ETH contribute to such work, providing a guiding principle to understand the emergence of effective randomness in many-body states. We look forward to

such work and anticipate a broader understanding of complex quantum many-body systems.

ACKNOWLEDGMENTS

We thank David Huse, Wen Wei Ho, Daniel Ranard, and Saúl Pilatowsky-Cameo for insightful discussions on this work. We acknowledge support by the NSF QLCI Grant No. OMA-2016245, the DOE (DE-SC0021951), the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant No. PHY-1733907), the Center for Ultracold Atoms, an NSF Physics Frontiers Center (NSF Grant No. PHY-1734011), the DARPA ONISQ program (W911NF2010021), the AFOSR YIP (FA9550-19-1-0044), Army Research Office MURI program (W911NF2010136), NSF CAREER Grant No. 2237244, and NSF CAREER Grant No. 1753386. Support is also acknowledged from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. A. E. acknowledges funding by the German National Academy of Sciences Leopoldina under Grant No. LPDS 2021-02 and by the Walter Burke Institute for Theoretical Physics at Caltech. F. S. acknowledges support provided by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (DE-SC0020290), by Amazon Web Services, AWS Quantum Program, and by the DOE QuantISED program through the theory consortium “Intersections of QIS and Theoretical Particle Physics” at Fermilab. J. C. acknowledges support from the Terman Faculty Fellowship at Stanford University.

APPENDIX A: PROOF OF THEOREM 1 (THE TEMPORAL ENSEMBLE IS THE RANDOM PHASE ENSEMBLE)

Theorem 1. Given an initial state $|\Psi_0\rangle$ and a Hamiltonian H , the infinite-time temporal ensemble is equal to the random phase ensemble:

$$\lim_{\tau \rightarrow \infty} \mathcal{E}_{\text{Temp}}(\tau) = \mathcal{E}_{\text{Rand Phase}}, \quad (\text{A1})$$

with fixed energy populations $|c_E|^2 = |\langle \Psi_0 | E \rangle|^2$, if and only if H satisfies all k th *no-resonance conditions* modulo degeneracies.

Proof. We prove Theorem 1 by matching the k th moments of the temporal and random phase ensembles and showing that they agree for all orders k . We enumerate the possible terms with *multisets*. ■

Definition 4 (multisets). We define the *ordered multiset* $\{\alpha_j\}_{j=1}^k$ as the ordered list of elements $\alpha_j \in \{1, \dots, D\}$, with possibly repeated elements. We denote $M_k(D)$ the set of all such multisets of length k with symbols in $\{1, \dots, D\}$.

We now compute the k th moment of the temporal ensemble:

$$\rho_{\text{Temp}}^{(k)} \equiv \mathbb{E}_t \left[U_t^{\otimes k} |\Psi_0\rangle \langle \Psi_0| U_t^{\otimes k} \right] \quad (\text{A2})$$

$$= \sum_{\alpha, \beta \in M_k(D)} \left[\bigotimes_{j=1}^k |E_{\alpha_j}\rangle \langle E_{\alpha_j}| \right] |\Psi_0\rangle \langle \Psi_0| \left[\bigotimes_{j=1}^k |E_{\beta_j}\rangle \langle E_{\beta_j}| \right] \delta \left(\sum_{j=1}^k E_{\alpha_j} - \sum_{j=1}^k E_{\beta_j} \right) \quad (\text{A3})$$

$$= \sum_{\alpha \in M_k(D)} \left[\sum_{\sigma \in \pi(\alpha)} |E_{\alpha_1}, \dots, E_{\alpha_k}\rangle \langle E_{\alpha_1}, \dots, E_{\alpha_k}| |\Psi_0\rangle \langle \Psi_0| \bigotimes_{j=1}^k |E_{\sigma(\alpha_j)}\rangle \langle E_{\sigma(\alpha_j)}| \right] \quad (\text{A4})$$

$$= \sum_{\alpha \in M_k(D)} \sum_{\sigma \in \pi(\alpha)} \bigotimes_{j=1}^k |\langle E_{\alpha_j} | \Psi_0 \rangle|^2 |E_{\alpha_j}\rangle \langle E_{\sigma(\alpha_j)}| \quad (\text{A5})$$

$$= \rho_d^{\otimes k} \sum_{\alpha \in M_k(D)} \sum_{\sigma \in \pi(\alpha)} \bigotimes_{j=1}^k |E_{\alpha_j}\rangle \langle E_{\sigma(\alpha_j)}|, \quad (\text{A6})$$

where $U_t \equiv \exp(-iHt)$, $\mathbb{E}_t[\cdot] = \lim_{T \rightarrow \infty} (1/T) \int_0^T \cdot dt$ is the infinite-time average, $|E\rangle$ are energy eigenstates of the Hamiltonian, and $\rho_d \equiv \sum_E |E\rangle \langle E| |\Psi_0|^2 |E\rangle \langle E|$ is the diagonal ensemble, prominently featured in the literature on many-body thermalization [15]. In going from Eq. (A2) to Eq. (A3), we have performed an infinite-time integration to obtain a delta function, and in going to Eq. (A4), we have used the k th no-resonance condition.

We then compute the k th moment of the random phase ensemble:

$$\rho_{\text{Rand phase}}^{(k)} = \sum_{\alpha, \beta \in M_k(D)} \bigotimes_{j=1}^k c_{\alpha_j} c_{\beta_j}^* |\alpha_j\rangle \langle \beta_j| \underbrace{\prod_j \left(\frac{1}{2\pi} \int_0^{2\pi} d\theta_j \right) \exp[i(\theta_{\alpha_j} - \theta_{\beta_j})]}_{\delta_{(\alpha_j), (\beta_j)}} \quad (\text{A7})$$

$$= \sum_{\alpha \in M_k(D)} \sum_{\sigma \in \pi(M)} \bigotimes_{j=1}^k |c_{\alpha_j}|^2 |\alpha_j\rangle \langle \sigma(\alpha_j)|. \quad (\text{A8})$$

Therefore, the k th moments of the temporal and random phase ensembles are equal. Identifying a distribution from its moments is known as the “moment problem.” For complex multivariate (vector-valued) distributions $P(z_1, \dots, z_D) \equiv P(\mathbf{z})$, the complex Carleman’s condition gives a sufficient condition to the moment problem. That is, define the moments $m_j^{(k)} \equiv \int d\mathbf{z} P(\mathbf{z}) |z_j|^{2k}$. If the following sum diverges:

$$\forall j, \quad \sum_{k=1}^{\infty} \left(m_j^{(k)} \right)^{-1/2k} = \infty, \quad (\text{A9})$$

then the sequence of moments $(m_j^{(k)})$ uniquely defines the probability distribution (Theorem 15.11 in Ref. [116]). In the random phase ensemble, the complex Carleman condition is satisfied. This allows us to identify the temporal

and random phase ensembles, if the k th no-resonance condition is satisfied.

In the other direction, if the k th no-resonance condition is *not* satisfied, the delta function in Eq. (A3) will give additional contributions. Therefore, the k th moments of the temporal and random phase ensembles do not agree, giving the desired statement. ■

APPENDIX B: k TH TWIRLING IDENTITY

The above computation of $\rho_{\text{Temp}}^{(k)}$ in Appendix A can be generalized into a statement about the *twirling channel* on any operator A that acts on k copies of the Hilbert space. Note, however, that unlike Theorem 1, we now require the full k th no-resonance conditions and cannot disregard degeneracies. As an illustration, we state and explicitly prove the result for $k = 2$. This identity originally appeared in the Supplemental Material of Ref. [110].

Lemma 2 (second Hamiltonian twirling identity). If a Hamiltonian H satisfies the second no-resonance condition, then, for any operator A acting on $\mathcal{H}^{\otimes 2}$, the following identity holds:

$$\mathbb{E}_t[U_t^{\otimes 2} A U_t^{\otimes 2}] = \mathcal{D}^{\otimes 2}[A] \mathbb{I} + \mathcal{D}^{\otimes 2}[A \mathcal{S}] \mathcal{S} - \sum_E |E\rangle \langle E|^{\otimes 2} A |E\rangle \langle E|^{\otimes 2}, \quad (\text{B1})$$

where $U_t \equiv \exp(-iHt)$, $\mathbb{E}_t[\cdot] = \lim_{T \rightarrow \infty} (1/T) \int_0^T \cdot dt$ is the infinite-time average, and $|E\rangle$ are energy eigenstates of the Hamiltonian. $\mathbb{I}$ and $\mathcal{S}$ are the identity and swap operators, respectively, on $\mathcal{H}^{\otimes 2}$, defined as $\mathbb{I}(|\Psi_1\rangle \otimes |\Psi_2\rangle) = |\Psi_1\rangle \otimes |\Psi_2\rangle$ and $\mathcal{S}(|\Psi_1\rangle \otimes |\Psi_2\rangle) = |\Psi_2\rangle \otimes |\Psi_1\rangle$ for any $|\Psi_1\rangle, |\Psi_2\rangle$. Finally, $\mathcal{D}^{\otimes 2}$ is the dephasing channel in the energy eigenbasis: $\mathcal{D}^{\otimes 2}[A] \equiv \sum_{E, E'} |E, E'\rangle \langle E, E'| A |E, E'\rangle \langle E, E'|$.

Proof.

$$\mathbb{E}_t[U_t^{\otimes 2} A U_t^{\otimes 2}] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt U_t^{\otimes 2} A U_t^{\otimes 2} \quad (\text{B2})$$

$$= \sum_{E_1, E_2, E_3, E_4} \left[\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt e^{-i(E_1 - E_2 + E_3 - E_4)t} \right] |E_1, E_3\rangle \langle E_1, E_3| A |E_2, E_4\rangle \langle E_2, E_4| \quad (\text{B3})$$

$$= \sum_{E_1, E_2, E_3, E_4} \delta(E_1 - E_2 + E_3 - E_4) |E_1, E_3\rangle \langle E_1, E_3| A |E_2, E_4\rangle \langle E_2, E_4| \quad (\text{B4})$$

$$\stackrel{2\text{-NR}}{=} \sum_{E, E'} [|E, E'\rangle \langle E, E'| A |E, E'\rangle \langle E, E'| + |E, E'\rangle \langle E, E'| A |E', E\rangle \langle E', E|] - \sum_E |E, E\rangle \langle E, E| A |E, E\rangle \langle E, E|, \quad (\text{B5})$$

where “2-NR” indicates the use of the second no-resonance condition. The second no-resonance condition states that the delta function $\delta(E_1 - E_2 + E_3 - E_4)$ is nonzero in only two cases: $E_1 = E_2, E_3 = E_4$ or $E_1 = E_4, E_3 = E_2$, which give the two terms above. The final term in Eq. (B5) corresponds to $E_1 = E_2 = E_3 = E_4$. This special case belongs to the

above two terms simultaneously. It is double counted in the first two terms and, hence, is subtracted. ■

Lemma 3 (k th Hamiltonian twirling identity). If the Hamiltonian satisfies the k th no-resonance condition, its k th twirling channel acting on any operator A on k copies of the Hilbert space $\mathcal{H}^{\otimes k}$ is

$$\mathbb{E}_t[U_t^{\otimes k} A U_t^{\otimes k}] = \sum_{M=(E_1, \dots, E_k)} \left[\sum_{\sigma \in \pi(M)} |E_1, \dots, E_k\rangle \langle E_1, \dots, E_k| A |E_{\sigma(1)}, \dots, E_{\sigma(k)}\rangle \langle E_{\sigma(1)}, \dots, E_{\sigma(k)}| \right] \quad (\text{B6})$$

$$= \sum_{\sigma \in S_k} \mathcal{D}^{\otimes k}[A \text{Perm}(\sigma^{-1})] \text{Perm}(\sigma) + \text{E.C.}, \quad (\text{B7})$$

where $\text{Perm}(\sigma)$ —with $\sigma \in S_k$ —is a permutation operator acting on the k -copy Hilbert space:

$$\text{Perm}(\sigma) \equiv \sum_{\{E_i\}} |E_1, E_2, \dots, E_k\rangle \langle E_{\sigma(1)}, E_{\sigma(2)}, \dots, E_{\sigma(k)}| \quad (\text{B8})$$

and $\mathcal{D}^{\otimes k}$ is the dephasing channel in the energy basis, $\mathcal{D}^{\otimes k}[A] \equiv \sum_{\{E_i\}} |\{E_i\}\rangle \langle \{E_i\}| A |\{E_i\}\rangle \langle \{E_i\}|$.

In Eq. (B6), the sum is taken over multisets M . If M has r unique indices with multiplicities $n_1, \dots, n_r$ (such that $\sum_{i=1}^r n_i = k$), its group of permutations $\pi(M)$ is a subgroup of the symmetric group S_k with $k! / \prod_{i=1}^r n_i!$

elements. In Eq. (B7), we rewrite Eq. (B6) in terms of permutations and additional terms E.C. associated with multiple counting (similar to the $k = 2$ case). At this level, it is difficult to provide a useful bound on the size of E.C. However, bounds may be proven in specific contexts, such as Theorem 3. All terms denoted “E.C.” can be explicitly derived from Eq. (B6).

The proof of the k th twirling identity is analogous to that of the second twirling identity: In this case, there are $k!$ permutations relevant for the k th no-resonance theorem, along with more terms than the simplest one in the $k = 2$ case. As an example, for $k = 3$, in addition to the terms associated with cases where two energies are the same

$E_i = E_j$, there is a higher-order term arising when all three energies are the same: $E_1 = E_2 = E_3$.

APPENDIX C: ENTROPY OF TEMPORAL ENSEMBLE

In Sec. IV, we state that the ensemble entropy of the temporal ensemble, a negative quantity which is the negative of its relative entropy to the Haar ensemble, is given by [Eq. (30)]

$$\text{Ent}(\mathcal{E}_{\text{Temp}}) = \sum_E \log(D|c_E|^2) + \lim_{\delta \rightarrow 0} D(\log(\delta/2) - 1). \quad (\text{C1})$$

Here, we derive this result. Note that any ensemble in the constrained space can be described by the joint distribution $P(\{\theta_E\})$. We emphasize that the ensemble entropy is the differential entropy of the *distribution of states* over Hilbert space. The ensemble entropy for any ensemble of states with constrained magnitudes $|c_E|^2$ is formally negative infinite, since they lie on a subdimensional (measure zero) manifold of the Hilbert space. However, we can still extract meaningful information from the ensemble entropy as follows. We may take the probability distribution $P(\Psi)$ on the constrained space as the limit of probability distributions with independent phase and magnitude components: $P(\Psi) = P(\{|\Psi_E|\})P(\{\theta_E\})$. This separates the entropy into contributions from the phase and magnitude degrees of freedom:

$$\begin{aligned} -\mathcal{D}(\mathcal{E}||\text{Haar}) &= - \int d\Psi P(\Psi) \log\left(\frac{P(\Psi)}{P_{\text{Haar}}(\Psi)}\right) \\ &= S_\theta + S_{|\Psi_E|^2}. \end{aligned} \quad (\text{C2})$$

The phase contribution is

$$S_\theta = - \prod_E \int_0^{2\pi} d\theta_E P(\{\theta_E\}) \log[(2\pi)^D P(\{\theta_E\})], \quad (\text{C3})$$

which is uniquely maximized when the phases are uniformly distributed, taking value 0.

The magnitude component is divergent: However, if we replace the delta function with an integral over the thin shell $|\Psi_E|^2 \in [|c_E|^2(1-\delta), |c_E|^2(1+\delta)]$, we obtain a constant contribution in addition to the divergent one:

$$\begin{aligned} S_{|\Psi_E|^2} &= - \lim_{\delta \rightarrow 0} \prod_E \int_{|c_E|^2(1-\delta)}^{|c_E|^2(1+\delta)} \frac{d|\Psi_E|^2}{2|c_E|^2\delta} \log\left(\frac{1/(|c_E|^2\delta)}{2D\exp(-D|\Psi_E|^2)}\right) \\ &= \lim_{\delta \rightarrow 0} \left[\sum_E \log(|c_E|^2) - D|c_E|^2 + D\log(\delta D/2) \right] \\ &= \sum_E \log(|c_E|^2) + \lim_{\delta \rightarrow 0} D(\log(\delta D/2) - 1). \end{aligned} \quad (\text{C4})$$

We interpret the constant contribution to the entropy as the volume of Hilbert (sub)space spanned by the temporal ensemble. This subspace is, in fact, a hypertorus, with radii given by $|c_E|$. In the limit of large Hilbert-space dimension D , the volume of this hypertorus is approximately $\prod_E |c_E|$. Therefore, our measure of entropy is quantitatively equivalent to the volume of the Hilbert space.

APPENDIX D: PROOF OF THEOREM 2 (ASYMPTOTIC PRODUCT FORMULA)

Theorem 2. The random phase ensemble asymptotically approaches the *product form*:

$$\rho_{\text{Temp}}^{(k)} = \rho_d^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma) + O[\text{tr}(\rho_d^2)]. \quad (\text{D1})$$

Here, $\rho_d \equiv \sum_E |c_E|^2 |E\rangle\langle E|$ is the *diagonal ensemble*, a common construction in the thermalization literature [51,52], and its purity $\text{tr}(\rho_d^2)$ is typically exponentially small in total system size.

Proof. We want to show that

$$\rho_{\text{Temp}}^{(k)} = \rho_d^{\otimes k} \sum_{\alpha \in M_k(D)} \sum_{\sigma \in \pi(\alpha)} \bigotimes_{j=1}^k |E_{\alpha_j}\rangle\langle E_{\sigma(\alpha_j)}| \xrightarrow{D \rightarrow \infty} \rho_d^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma) = \rho_d^{\otimes k} \sum_{\alpha \in M_k(D)} \sum_{\sigma \in S_k} \bigotimes_{j=1}^k |E_{\alpha_j}\rangle\langle E_{\sigma(\alpha_j)}|. \quad (\text{D2})$$

Here, we derive the bound on the error

$$\left\| \rho_{\text{Temp}}^{(k)} - \rho_d^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma) \right\|_* \leq \frac{k!}{2} \text{tr}(\rho_d^2), \quad (\text{D3})$$

where $\|\cdot\|_*$ is the trace norm or the sum of absolute values of the eigenvalues. To do so, we note that

$$\rho_d^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma) - \rho_{\text{Temp}}^{(k)} = \rho_d^{\otimes k} \left[\sum_{\alpha \in M_k(D)} \left(\sum_{\sigma \in S_k} - \sum_{\sigma \in \pi(\alpha)} \right) \bigotimes_{j=1}^k |E_{\alpha_j}\rangle\langle E_{\sigma(\alpha_j)}| \right]. \quad (\text{D4})$$

We then divide the Hilbert space $\mathcal{H}^{\otimes k}$ into classes s of multisets α , grouped by their degeneracy structure: the number of degenerate indices α_j and their multiplicity. For example, $s = ()$ indicates the class of multisets with no degeneracies, $s = (2)$ indicates the presence of 1 degeneracy of multiplicity 2, and $s = (2, 2)$ indicates two degeneracies of multiplicity 2, respectively. The size of the permutation group $\pi(\alpha)$ is uniquely determined by the class $s(\alpha)$ and is a divisor of $|S_k| = k!$. Therefore,

$$\rho_d^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma) - \rho_{\text{Temp}}^{(k)} = \rho_d^{\otimes k} \left[\sum_{s \in M_k(D)} \sum_{\alpha \in s} \left(\sum_{\sigma \in S_k} - \sum_{\sigma \in \pi(\alpha)} \right) \bigotimes_{j=1}^k |E_{\alpha_j}\rangle \langle E_{\sigma(\alpha_j)}| \right] \quad (\text{D5})$$

$$= \sum_{s \in M_k(D)} \rho_d^{\otimes k} P_s \sum_{\sigma \in S_k} \text{Perm}(\sigma), \quad (\text{D6})$$

where P_s is a projector onto the space of multisets in the class s . Each term in the sum over s , $\rho_d^{\otimes k} P_s \sum_{\sigma \in S_k} \text{Perm}(\sigma)$, is positive semidefinite. We then use the triangle inequality to bound the trace distance, noting that the term from the class $s = ()$ gives 0 contribution. In our sum (D6) over multisets with at least one degeneracy, we can replace $\rho_d^{\otimes k}$ with $\rho_d^{(2)} \otimes \rho_d^{\otimes(k-2)}$, where $\rho_d^{(2)} \equiv \sum_E |c_E|^4 |E, E\rangle \langle E, E|$. Finally, the projector P_s only decreases the trace distance and can be omitted:

$$\left\| \rho_d^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma) - \rho_{\text{Temp}}^{(k)} \right\|_* \leq \sum_{s \neq ()} \left\| \left(\rho_d^{(2)} \otimes \rho_d^{\otimes(k-2)} \right) \sum_{\sigma \in S_k} \text{Perm}(\sigma) \right\|_* \quad (\text{D7})$$

$$= \sum_{s \neq ()} \text{tr} \left[\left(\rho_d^{(2)} \otimes \rho_d^{\otimes(k-2)} \right) \sum_{\sigma \in S_k} \text{Perm}(\sigma) \right] \quad (\text{D8})$$

$$\leq \sum_{s \neq ()} k! \text{tr}(\rho_d^2) = [p(k) - 1] k! \text{tr}(\rho_d^2) \leq k! \exp \left(\pi \sqrt{\frac{2k}{3}} \right) \text{tr}(\rho_d^2), \quad (\text{D9})$$

where $p(k)$ is the *partition number* of k (the number of possible classes s), which is upper bounded by $\exp[\pi\sqrt{(2k/3)}]$ [140]. In going from Eq. (D8) to Eq. (D9), we use the fact that

$$\text{tr} \left[\left(\rho_d^{(2)} \otimes \rho_d^{\otimes(k-2)} \right) \sum_{\sigma \in S_k} \text{Perm}(\sigma) \right] = \text{tr}(\rho_d^2) + (k-1) \text{tr}(\rho_d^3) + \binom{k-1}{2} \text{tr}(\rho_d^2)^2 + \dots \leq k! \text{tr}(\rho_d^2); \quad (\text{D10})$$

this bound is weak since we expect all terms to be exponentially smaller (in system size) than the first. This implies that the $O(k!)$ prefactor is likely not tight in terms of scaling with k . However, we typically restrict our attention to small values of k and are more interested in the factor of $\text{tr}(\rho_d^2)$. In typical systems, we expect $\text{tr}(\rho_d^2) \sim D^{-1}$ to exponentially decrease as the size of the many-body system increases. ■

APPENDIX E: CONVERGENCE OF FINITE-TIME TEMPORAL ENSEMBLES: FURTHER NUMERICAL RESULTS

Here, we present additional numerical results investigating the convergence of finite-time temporal ensembles discussed in Sec. IV B.

In Fig. 8(a), we compare convergence of first and second moments of the finite-time temporal ensemble in trace and

Frobenius distance. We consider a single Hamiltonian drawn at random from the GUE ensemble in a system with Hilbert-space dimension $d = 2^6$. For both first and second moments, we observe the same behavior at long times: Trace and Frobenius distance decay as τ^{-1} , where τ is the interval length. In contrast, the intermediate $\tau^{-1/2}$ decay of the Frobenius distance and its pronounced cross-over to the late time τ^{-1} appear to be absent when considering the trace distance. This suggests that, while the early and intermediate time behavior of trace and Frobenius distance can differ, the Frobenius distance is a quantitative proxy for the trace distance at long times.

In Fig. 8(b), we complement the results of the main text by comparing convergence of the k th moments of the finite-time temporal for different $k = 1, 2, 3$. We consider a single Hamiltonian drawn at random from the GUE ensemble in a system with Hilbert-space dimension $D = 2^6$ and the

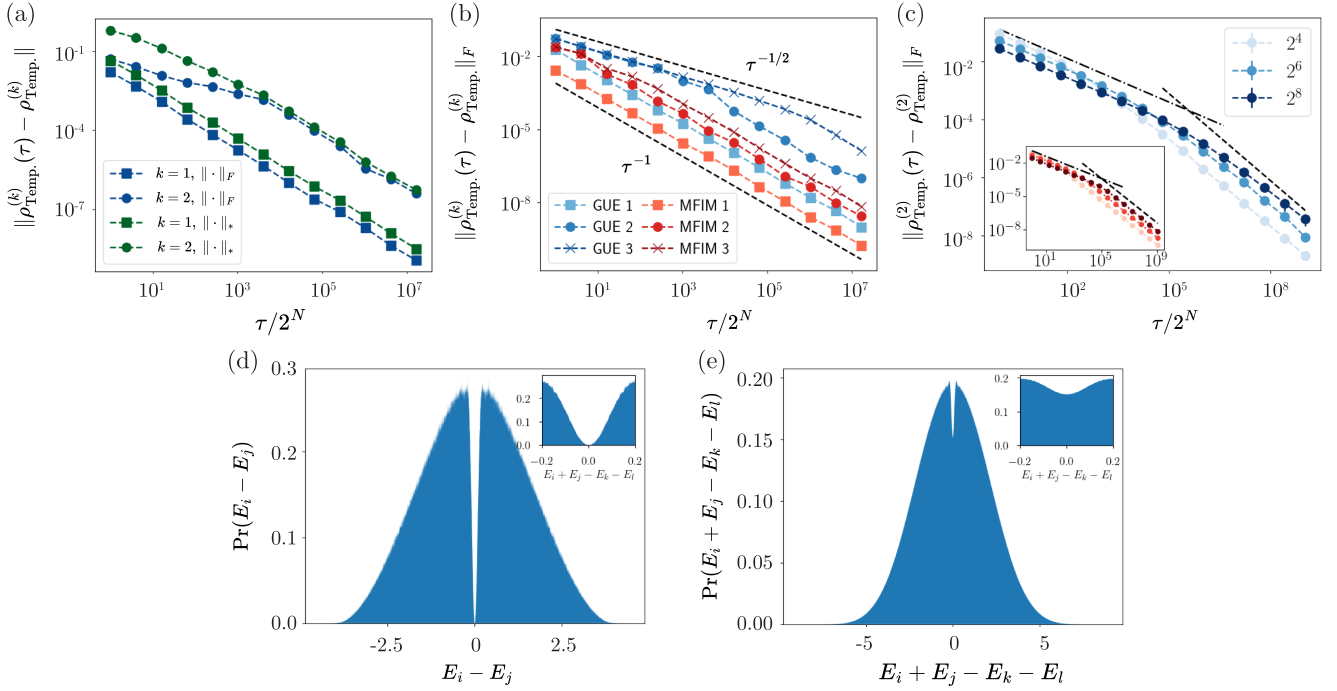


FIG. 8. Convergence of finite-time ensembles. (a) Comparison of convergence of first ($k = 1$, squares) and second moment ($k = 2$, circles) in trace (green) and Frobenius distance (blue) for a temporal ensemble generated by a single, randomly drawn GUE Hamiltonian with Hilbert-space dimension $D = 2^6$. (b) Convergence in Frobenius distance of the first three moments $k = 1, 2, 3$ (squares, circles, and crosses) for single, randomly drawn GUE Hamiltonian with Hilbert-space dimension $D = 2^6$ (blue) and the MFIM with $N = 6$ spins (red). (c) Ensemble averaged Frobenius distance of the temporal ensembles generated by a randomly drawn GUE Hamiltonians in various Hilbert-space dimension $D = 2^4, 2^6, 2^8$. The mean is taken over 10 000 ($D = 2^4, 2^6$) and 2000 random instances ($D = 2^8$). Error bars denote two standard errors of the mean. The inset shows the median of the Frobenius distances of the same data. (d), (e) Histogram of gaps and differences of gaps accumulated over 50 000 randomly chosen GUE Hamiltonian with Hilbert-space dimension $D = 2^4$. Indices i, j, k , and l take all values in $1, \dots, D$, and we exclude the (trivially vanishing) diagonal elements $i = j$ in (d) and $i = k$, $j = l$ and $i = l$, $j = k$ in (e). The insets enlarge the regions close to zero.

MFIM with $N = 6$ spins. For computational simplicity, we focus on the Frobenius norm. Independent of the value of k , we observe a late-time decay $\propto \tau^{-1}$. As discussed in the main text [cf. Eq. (34)], we attribute this to the presence of a finite minimal gap $E_i - E_j$ (difference of gaps for $k \leq 2$) for an individual Hamiltonian and the late-time decay of the sinc function.

In Fig. 8(c), we investigate the convergence of ensemble-averaged distances. Notably, we find that for the first moment, $k = 1$, the GUE ensemble mean closely resembles the behavior of individual instances and decays as τ^{-1} (not shown). For higher moments $k = 2$, we find that, at early times, ensemble average distances decay as $\tau^{-1/2}$ whereas at very late times we observe a decay consistent with τ^{-1} . The crossover between these two regimes is, however, less sharply pronounced as for single fixed Hamiltonians [as visualized by the ensemble median, representing a typical instance, in the inset in Fig. 8(c)]. We relate these findings to statistical properties of the energy eigenlevels of GUE Hamiltonians. It is well known that the statistics of

eigenvalues exhibits (strong) level repulsion; i.e., the probability to find two distinct eigenvalues close to each other $E_i - E_j \approx 0$ is vanishing. In contrast, the numerical results presented Figs. 8(d) and 8(e) suggest the gaps $E_i - E_j + E_k - E_l$ of eigenvalues exhibit only weak repulsion; i.e., the probability of finding two distinct gaps close to each other $E_i - E_j + E_k - E_l \approx 0$ is suppressed but finite. We leave a detailed investigation of such higher-order gap repulsion and its consequences for the late time decay of the ensemble-averaged Frobenius (and trace) distance to future study (see also Appendix F 1).

APPENDIX F: CONVERGENCE OF FINITE-TIME TEMPORAL ENSEMBLES: ANALYTICAL RESULTS FROM RANDOM MATRIX THEORY

Here, we prove the τ^{-1} scaling of the convergence of the finite-time temporal ensemble for Hamiltonians from the GUE, discussed in Sec. IV B. We focus on the (ensemble-averaged) squared Frobenius norm $\|\cdot\|_F$ for

its analytic tractability. Note that, since we average over random Hamiltonians H , the initial state can be fixed to be an arbitrary state:

$$\mathbb{E}_{H \sim \text{GUE}} \left[\|\rho_{\text{Temp}}^{(k)}(\tau) - \rho_{\text{Temp}}^{(k)}\|_F^2 \right] = \mathbb{E}_{H \sim \text{GUE}} \left[\text{tr} \left[\left(\rho_{\text{Temp}}^{(k)}(\tau) \right)^2 \right] - 2\text{tr} \left[\rho_{\text{Temp}}^{(k)}(\tau) \rho_{\text{Temp}}^{(k)} \right] + \text{tr} \left[\left(\rho_{\text{Temp}}^{(k)} \right)^2 \right] \right] \quad (\text{F1})$$

$$= \mathbb{E}_{H \sim \text{GUE}} \left[\text{tr} \left[\left(\rho_{\text{Temp}}^{(k)}(\tau) \right)^2 \right] - \text{tr} \left[\left(\rho_{\text{Temp}}^{(k)} \right)^2 \right] \right], \quad (\text{F2})$$

where we use that the long time limit average moment does not evolve in time:

$$\text{tr}[\rho_{\text{Temp}}^{(k)}(\tau) \rho_{\text{Temp}}^{(k)}] = \text{tr} \left[\int_{-\tau}^{\tau} \frac{dt}{2\tau} (e^{-iHt} |\Psi_0\rangle \langle \Psi_0| e^{iHt})^{\otimes k} \rho_{\text{Temp}}^{(k)} \right] = \langle \Psi_0 |^{\otimes k} \rho_{\text{Temp}}^{(k)} | \Psi_0 \rangle^{\otimes k} = \text{tr}[\rho_{\text{Temp}}^{(k),2}]. \quad (\text{F3})$$

To carry out the average over the GUE, we decompose $H = U\Lambda U^\dagger$ with Λ a diagonal matrix of eigenvalues of H and U the unitary matrix formed by its eigenvectors. Importantly, for H from GUE, Λ and U are statistically independent, and U is a random unitary matrix distributed according to the Haar measure on the unitary group. We can, thus, evaluate the average over the eigenvectors U separately using Weingarten calculus. We find for the first term

$$\begin{aligned} \mathbb{E}_{H \sim \text{GUE}} \left[\text{tr}[\rho_{\text{Temp}}^{(k),2}(\tau)] \right] &= \frac{1}{4\tau^2} \int_0^{2\tau} d\bar{t} \int_0^{2\bar{t}} dt' \int d\Lambda P(\Lambda) \int dU \text{tr} \left[|\Psi_0\rangle \langle \Psi_0|^{\otimes 2k} U^{\otimes 2k} \left(e^{-i\Lambda t'} \otimes e^{+i\Lambda t'} \right)^{\otimes k} U^{\dagger, \otimes 2k} \right] \\ &= \frac{1}{4\tau^2} \int_0^{2\tau} d\bar{t} \int_0^{2\bar{t}} dt' \frac{1}{D(D+1) \cdots (D+k-1)} \int d\Lambda P(\Lambda) \sum_{\pi \in S_{2k}} \text{tr} \left[W_\pi \left(e^{-i\Lambda t'} \otimes e^{+i\Lambda t'} \right)^{\otimes k} \right], \end{aligned} \quad (\text{F4})$$

where $P(\Lambda)$ denotes the joint probability distribution of the eigenvalues of Hamiltonians H from GUE, $\int dU$ denotes the Haar integral of the unitary group, W_π with $\pi \in S_{2k}$ are $2k$ permutation operators acting on the $2k$ -copy space, and S_{2k} denotes the symmetric group of the order of $2k$. In the case $k = 1$, we can simplify the expressions further. We find

$$\begin{aligned} &\int P(\Lambda) d\Lambda \sum_{\pi \in S_2} \text{tr} [W_\pi (e^{-i\Lambda\tau} \otimes e^{+i\Lambda\tau})] \\ &= \frac{1}{D(D+1)} [K(\tau) + D], \end{aligned}$$

where

$$K(\tau) = \mathbb{E}_{H \sim \text{GUE}} \left[|\text{tr}(e^{-iH\tau})|^2 \right] = \int P(\Lambda) d\Lambda |\text{tr}(e^{-i\Lambda\tau})|^2$$

is the *spectral form factor* averaged over Hamiltonians H .

Furthermore, $\rho_H^{(1)} = \sum_E |\langle E | \Psi_0 \rangle|^2 |E\rangle \langle E|$ with $\{|E\rangle\}$ the eigenvectors of H . Thus, we have $\text{tr}[\rho_H^{(1),2}] = \sum_E |\langle E | \Psi_0 \rangle|^4$, which averaged over H from GUE,

$$\mathbb{E}_{H \sim \text{GUE}} \left[\sum_E |\langle E | \Psi_0 \rangle|^4 \right] = \int dU \left[\sum_{i=1}^D |\langle i | U | \Psi_0 \rangle|^4 \right] = \frac{2}{D+1}. \quad (\text{F5})$$

Finally, we obtain

$$\begin{aligned} &\mathbb{E}_{H \sim \text{GUE}} \left[\|\rho_H^{(1)}(\tau) - \rho_H^{(1)}\|_F^2 \right] \\ &= \frac{1}{4\tau^2} \int_0^{2\tau} d\bar{t} \int_0^{2\bar{t}} dt' \frac{K(t') - D}{D(D+1)} \sim \frac{D}{\tau^2} + \mathcal{O}(\tau^{-4}). \end{aligned} \quad (\text{F6})$$

Using the improved approximation [141] for the spectral form factor $K(\tau)$ and an asymptotic expansion of the occurring Bessel function [142] gives a τ^{-2} scaling of the squared Frobenius norm. We note that, if the standard box approximation is used, Eq. (F6) does not converge to 0. As we detail below, an alternative approach is to compute the convergence for the GUE ensemble for a finite dimension D . The finite D expressions, in fact, do not require any regularization, and the large D limit can be safely taken on the final result.

1. Convergence for GUE ensemble and finite dimension

Here, we show how to study the convergence to the temporal ensemble for Hamiltonians from GUE ensembles of finite dimension D . These finite D expressions do not suffer from the problems of the infinite- D spectral form factor, which has to be properly regularized. Therefore, we can compute the distance from the temporal ensemble for finite D and take the $D \rightarrow \infty$ limit at the end.

We are interested in the average of the following distance for H belonging to the GUE ensemble:

$$\begin{aligned} & \text{Tr} \left\{ \left[\rho_{\text{Temp}}^{(k)}(\tau) - \rho_{\text{Temp}}^{(k)}(\infty) \right]^2 \right\} \\ &= \sum_{\{\alpha_j\}} \sum_{\{\beta_j\} \notin \text{Perm}(\{\alpha_j\})} \text{sinc}^2 \left(\sum_j (E_{\alpha_j} - E_{\beta_j}) \tau \right) \prod_j |c_{\alpha_j}|^2 |c_{\beta_j}|^2. \end{aligned} \quad (\text{F7})$$

Note that the sum is over the sequences of indices $\{\alpha_j\}, \{\beta_j\}$ under the condition that the two sequences are not related by a permutation. Let us consider, as a first example, the case $k = 1$. We have

$$\begin{aligned} & \mathbb{E}_{H \sim \text{GUE}} \left[\text{Tr} \left\{ \left[\rho_{\text{Temp}}^{(1)}(\tau) - \rho_{\text{Temp}}^{(1)}(\infty) \right]^2 \right\} \right] \\ &= \sum_i \sum_{j \neq i} \mathbb{E}_{H \sim \text{GUE}} [\text{sinc}^2((E_i - E_j)\tau)] \mathbb{E}_{H \sim \text{GUE}} [|c_i|^2 |c_j|^2], \end{aligned} \quad (\text{F8})$$

where we use that the eigenvalues and the eigenvectors of H are independently distributed.

For the eigenvectors, we get

$$\mathbb{E}_{H \sim \text{GUE}} [|c_i|^2 |c_j|^2] = \frac{1}{D(D+1)}. \quad (\text{F9})$$

The convergence in time depends on the following average over the distribution of the eigenvalues:

$$\begin{aligned} & \mathbb{E}_{H \sim \text{GUE}} [\text{sinc}^2((E_i - E_j)\tau)] \\ &= \int dE_1 \dots dE_D \text{sinc}^2[(E_1 - E_2)\tau] p^D(E_1, \dots, E_D) \\ &= \int du \text{sinc}^2(u\tau) q_{1|1}^D(u). \end{aligned} \quad (\text{F10})$$

Here, $p^D(E_1, \dots, E_D)$ is the joint probability distribution of the (unordered) eigenvalues $E_1, \dots, E_D$, and we define $q_{1|1}^D(u)$ as the probability distribution of the energy difference $E_1 - E_2$:

$$q_{1|1}^D(u) = \int \int dE_1 \dots dE_D \delta(E_1 - E_2 - u) p^D(E_1, \dots, E_D). \quad (\text{F11})$$

Hence, we get

$$\begin{aligned} & \mathbb{E}_{H \sim \text{GUE}} \left[\text{Tr} \left\{ \left[\rho_{\text{Temp}}^{(1)}(\tau) - \rho_{\text{Temp}}^{(1)}(\infty) \right]^2 \right\} \right] \\ &= \frac{D-1}{D+1} \int du \text{sinc}^2(u\tau) q_{1|1}^D(u). \end{aligned} \quad (\text{F12})$$

Before evaluating this integral, let us consider the case $k = 2$. By enumerating the possible sequences of indices, we get

$$\begin{aligned} \text{Tr} \left\{ \left[\rho_{\text{Temp}}^{(2)}(\tau) - \rho_{\text{Temp}}^{(2)}(\infty) \right]^2 \right\} &= \frac{D!}{(D-4)!} \mathbb{E}_{H \sim \text{GUE}} [\text{sinc}^2((E_1 + E_2 - E_3 - E_4)\tau)] \mathbb{E}_{H \sim \text{GUE}} [|c_1|^2 |c_2|^2 |c_3|^2 |c_4|^2] \\ &+ 2 \frac{D!}{(D-3)!} \mathbb{E}_{H \sim \text{GUE}} [\text{sinc}^2((E_1 + E_2 - 2E_3)\tau)] \mathbb{E}_{H \sim \text{GUE}} [|c_1|^2 |c_2|^2 |c_3|^4] \\ &+ 4 \frac{D!}{(D-3)!} \mathbb{E}_{H \sim \text{GUE}} [\text{sinc}^2((E_1 - E_2)\tau)] \mathbb{E}_{H \sim \text{GUE}} [|c_1|^4 |c_2|^4] \\ &+ D(D-1) \mathbb{E}_{H \sim \text{GUE}} [\text{sinc}^2((2E_1 - 2E_2)\tau)] \mathbb{E}_{H \sim \text{GUE}} [|c_1|^4 |c_2|^4] \\ &+ 4D(D-1) \mathbb{E}_{H \sim \text{GUE}} [\text{sinc}^2((E_1 - E_2)\tau)] \mathbb{E}_{H \sim \text{GUE}} [|c_1|^2 |c_2|^6]. \end{aligned} \quad (\text{F13})$$

Similar expressions can be obtained for generic k th moments: In general, to study the time dependence, we need to evaluate the probability distribution of linear combinations of energy eigenstates, i.e., quantities of the form

$$q_{m_1 \dots m_s | n_1 \dots n_t}^D(u) \equiv \int dE_1 \dots dE_D \delta \left[\left(\sum_{i=1}^s m_i E_i \right) - \left(\sum_{j=1}^t n_j E_{s+j} \right) - u \right] p^D(E_1, \dots, E_D). \quad (\text{F14})$$

The probability distribution $q_{m_1 \dots m_s | n_1 \dots n_t}^D(u)$ is then used to compute the following integral:

$$F_{m_1 \dots m_s | n_1 \dots n_t}^D(\tau) = \int dE_1 \dots dE_D \text{sinc}^2 \left[\left(\sum_{i=1}^s m_i E_i \right) \tau - \left(\sum_{j=1}^t n_j E_{s+j} \right) \tau \right] p^D(E_1, \dots, E_D) \quad (\text{F15})$$

$$= \int du \text{sinc}^2(u\tau) q_{m_1 \dots m_s | n_1 \dots n_t}^D(u). \quad (\text{F16})$$

a. Evaluation of the integral

To evaluate the integral in Eq. (F16), we use the following property: The probability distribution $q_{m_1 \dots m_s | n_1 \dots n_t}^D(u)$ has, in general, the expression

$$q_{m_1 \dots m_s | n_1 \dots n_t}^D(u) = e^{-\gamma u^2} P_{m_1 \dots m_s | n_1 \dots n_t}^D(\gamma u^2), \quad (\text{F17})$$

where

$$\gamma = \frac{D}{2} \left[\sum_{i=1}^s m_i^2 + \sum_{j=1}^t n_j^2 \right]^{-1} \quad (\text{F18})$$

and $P_{m_1 \dots m_s | n_1 \dots n_t}^D(x)$ is a polynomial in x . This expression can be obtained from the joint probability distribution of N levels, which can be computed with standard random matrix theory techniques [143]:

$$\int \text{dusinc}^2(u\tau) e^{-\gamma u^2} (\gamma u^2)^n = \begin{cases} \frac{\gamma^{1/2}}{\tau^2} \left(\frac{\sqrt{\pi}}{2^{2n-1}} (2n-3)!! + e^{-\tau^2/\gamma} O[\gamma^{-1} \tau^2 (n-1)] \right), & \text{for } n \geq 1, \\ \frac{\sqrt{\pi}}{\tau} \left[\sqrt{\pi} \text{Erf}(\gamma^{-1/2} \tau) - \frac{\gamma^{1/2}}{\tau} (1 - e^{-\tau^2/\gamma}) \right], & \text{for } n = 0. \end{cases} \quad (\text{F21})$$

The leading term at long time, hence, has a power-law dependence $1/\tau^2$ if $P_{m_1 \dots m_s | n_1 \dots n_t}^D(0) = 0$ and $1/\tau$ if $P_{m_1 \dots m_s | n_1 \dots n_t}^D(0) \neq 0$. For the case $k = 1$, it is possible to prove (e.g., from an explicit calculation) that $P_{1|1}^D(0) = 0$: This is the effect of level repulsion, as observed in Fig. 8(d). For generic $P_{m_1 \dots m_s | n_1 \dots n_t}^D(x)$, we do not expect similar vanishing for $x = 0$ [Fig. 8(e)]. Therefore, we get the following scaling at long times τ :

$$\mathbb{E}_{H \sim \text{GUE}} \left[\|\rho_H^{(k)}(\tau) - \rho_H^{(k)}\|_F^2 \right] \propto \begin{cases} 1/\tau^2 & \text{for } k = 1, \\ 1/\tau & \text{for } k > 1. \end{cases} \quad (\text{F22})$$

This result shows tension with our numerical results, which exhibit a crossover between $1/\tau^{1/2}$ to $1/\tau$ scaling of the Frobenius distance $\|\rho_{\text{Temp}}^{(2)}(\tau) - \rho_{\text{Temp}}^{(2)}(\infty)\|_F^2$. It suggests that, when the (squared) Frobenius norm is ensemble averaged, the crossover disappears. We believe that this is due to the fact that the crossover time changes between samples of GUE Hamiltonians, and the mean is dominated by the samples with large crossover times. As an illustration, if we numerically evaluate the mean Frobenius norm, the crossover behavior is considerably weaker [Fig. 8(c)]. If, instead, we numerically evaluate the median of the Frobenius norm, which more closely reflects the

$$p_N^D(E_1, \dots, E_N) \equiv \int \text{d}E_{N+1} \dots \text{d}E_D p^D(E_1, \dots, E_D) \\ = \frac{(D-N)!}{D!} \det(K_D(E_i, E_j))|_{i,j=1}^N. \quad (\text{F19})$$

Here, the kernel $K_D(x, y)$ is defined as

$$K_D(x, y) = e^{-(D/4)(x^2+y^2)} \sum_{j=0}^{N-1} \frac{H_j(x\sqrt{D/2}) H_j(y\sqrt{D/2})}{\sqrt{2\pi D} 2^j j!}, \quad (\text{F20})$$

where $H_j(x)$ is the j th Hermite polynomial. Using this expression, through a change of variables we can derive, in general, an expression of the form in Eq. (F17).

We can then evaluate the integral for each term of the polynomial $P_{m_1 \dots m_s | n_1 \dots n_t}^D(x)$ as follows:

behavior of typical samples, the crossover becomes more pronounced [Fig. 8(c), inset].

APPENDIX G: FINITE-TIME TEMPORAL MUTUAL INFORMATION

In Sec. IV F 1, we derived an expression for the mutual information of temporal ensembles of finite intervals τ : $I(Z; T) \approx 1 - \gamma - (\sqrt{\pi}/2\sigma_H\tau)$, which we provide a full derivation of here.

1. Bit string correlation time

We first introduce the “bit string correlation time” t_c , which characterizes the timescale over which the bit string distribution $p(z, t)$ becomes independent [after rescaling their time-averaged values $p_{\text{avg}}(z)$]. That is, how long does it take to form a new, independent speckle pattern? We show that $t_c = \sigma_H^{-1}$, where $\sigma_H^2 = \text{tr}(H^2) - \text{tr}(H)^2$. To do this, we compute the correlator:

$$A_z(t) = \frac{\int \text{d}t' p(z, t') p(z, t' + t)}{[\int \text{d}t' p(z, t')]^2} - 1 \\ = \frac{\left| \sum_E |\langle z|E\rangle|^2 |\langle E|\Psi_0\rangle|^2 \exp(-iEt) \right|^2}{\left| \sum_E |\langle z|E\rangle|^2 |\langle E|\Psi_0\rangle|^2 \right|^2} + \dots, \quad (\text{G1})$$

where we use “...” to denote subleading terms. In order to make this calculation tractable, we make the following approximations:

$$\sum_E \rightarrow \frac{1}{\sqrt{2\pi\sigma_H^2}} \int_{-\infty}^{\infty} dE \exp\left(-\frac{(E-E_0)^2}{2\sigma_H^2}\right), \quad (\text{G2})$$

$$|\langle z|E\rangle|^2 \rightarrow \exp(-\beta_z E)/Z_z, \quad (\text{G3})$$

$$|\langle E|\Psi_0\rangle|^2 \rightarrow \exp(-\beta_\Psi E)/Z_\Psi, \quad (\text{G4})$$

where $E_0 = \text{tr}(H)$, β_z and β_E are appropriate inverse temperatures of the state $|z\rangle$ and $|\Psi_0\rangle$, respectively, and Z_z and Z_Ψ are their corresponding partition functions. The first approximation takes the density of states of the spectrum to be a Gaussian [144], while the second and third approximations treat the energy populations of the states $|z\rangle$ and $|\Psi_0\rangle$ as Boltzmann distributions (Sec. VII). We obtain

$$\sum_E |\langle z|E\rangle|^2 |\langle E|\Psi_0\rangle|^2 \exp(-iEt) \rightarrow \frac{1}{\sqrt{2\pi\sigma_H^2}} \int_{-\infty}^{\infty} dE \exp\left(-\frac{(E-E_0)^2}{2\sigma_H^2}\right) \exp[-(\beta_z + \beta_E + it)E] \quad (\text{G5})$$

$$= \exp[(\beta_z + \beta_E + it)^2 \sigma_H^2 / 2]. \quad (\text{G6})$$

Taking the ratio of the absolute values of Eq. (G6) at nonzero t and $t = 0$ gives the desired result.

2. Temporal mutual information

We can then derive the observed scaling of the mutual information

$$I(Z; T) = H(Z) - H(Z|T) = -\sum_z \bar{p}_{[t, t+\tau]}(z) \log_2[\bar{p}_{[t, t+\tau]}(z)] + \int_t^{t+\tau} dt' \frac{1}{\tau} \sum_z p_{t'}(z) \log_2[p_{t'}(z)], \quad (\text{G7})$$

where $\bar{p}_{[t, t+\tau]}(z) = \tau^{-1} \int_t^{t+\tau} dt' p_{t'}(z)$ is the outcome probability averaged over an interval. Using the facts that (i) $p_{t'}(z) \approx p_{\text{avg}}(z) \tilde{p}$, where $\tilde{p} \sim \text{PT}(1)$, and (ii) the bit string correlation time is $t_c = \sigma_H^{-1}$, we approximate $\bar{p}/p_{\text{avg}}$ to be the average of $n = \tau/t_c$ independent PT variables. The average of n Porter-Thomas (exponential) variables follows an Erlang distribution with mean 1 [41]. Following convention, we denote this distribution as $\bar{p}/p_{\text{avg}} \sim \text{Erlang}(1/n, n)$, with probability density function

$$P_{\text{Erlang}(1/n, n)}(x) dx = \frac{(nx)^{n-1} \exp(-nx)}{(n-1)!} n dx. \quad (\text{G8})$$

This approximation allows us to calculate the entropy of $\bar{p}$:

$$-\sum_z \bar{p}_{[t, t+\tau]}(z) \log_2[\bar{p}_{[t, t+\tau]}(z)] \approx -\mathbb{E}[\bar{p}] \sum_z p_{\text{avg}}(z) \log_2[p_{\text{avg}}(z)] - \mathbb{E}[\bar{p} \log_2(\bar{p})] \sum_z p_{\text{avg}}(z) \quad (\text{G9})$$

$$= H_{p_{\text{avg}}} + \log_2(n) - \Psi(n+1) \xrightarrow{n \rightarrow \infty} \left(H_{p_{\text{avg}}} - \frac{1}{2n}\right) / \ln 2 = \left(H_{p_{\text{avg}}} - \frac{1}{2\sigma_H \tau}\right) / \ln 2, \quad (\text{G10})$$

where $\Psi(x)$ is the digamma function, related to the Gamma function $\Gamma(x)$ by $\Psi(x) = \Gamma'(x)/\Gamma(x)$. A more careful analysis modifies the effective number n of the Erlang distribution by considering the second moment of $\bar{p}$:

$$\mathbb{E}_t[\bar{p}(z, t)^2] = \frac{1}{\tau^2} \int_0^\tau dt_1 \int_0^\tau dt_2 p(z, t_1) p(z, t_2) \quad (\text{G11})$$

$$\approx 1 + \frac{1}{\tau} \int_{-\infty}^{\infty} dt' \exp(-\sigma_H^2 t'^2) = 1 + \frac{\sqrt{\pi}}{\sigma_H \tau} \equiv 1 + \frac{1}{n_{\text{eff}}}; \quad (\text{G12})$$

i.e., the effective parameter n_{eff} of the Erlang distribution is $n_{\text{eff}} = \sigma_H \tau / \sqrt{\pi}$. This gives the overall temporal mutual information for a finite-time interval:

$$I(Z; T) \approx \left(1 - \gamma - \frac{\sqrt{\pi}}{2\sigma_H \tau}\right) / \ln 2, \quad (\text{G13})$$

which agrees with numerical results [inset in Fig. 3(d)].

APPENDIX H: k TH MOMENTS OF THE SCROOGE ENSEMBLE

Here, we derive our expression for the k th moments of the Scrooge ensemble [Eq. (50)] $\text{Scrooge}[\rho]$:

$$\rho_{\text{Scr},m}^{(k)} = \left(\prod_i \lambda_i^{-1} \right) \frac{\partial^k \Lambda_k}{\partial \mu_{m_1} \dots \partial \mu_{m_k}} \Big|_{\mu_i = \lambda_i^{-1}}, \quad (\text{H1})$$

where

$$\Lambda_k \equiv \sum_{j=1}^D \frac{\mu_j^{k-2} \ln \mu_j}{\prod_{k \neq j} (\mu_j - \mu_k)} \quad (\text{H2})$$

and $\{\lambda_i\}$ are the eigenvalues of the given state ρ .

As an explicit example, we consider the simplest non-trivial parameters, $k = 2$ and $D = 2$. Let the eigenstates of ρ be $|\uparrow\rangle$ and $|\downarrow\rangle$, and the eigenvalues be λ and $1 - \lambda$. Here, the second moment $\rho^{(2)}$ is a 4×4 matrix. However, in the (doubled) eigenbasis of ρ , only six entries are nonzero:

$$\langle \uparrow \uparrow | \rho^{(2)} | \uparrow \uparrow \rangle \equiv \rho_{\text{Scr},11}^{(2)}, \quad \langle \uparrow \downarrow | \rho^{(2)} | \uparrow \downarrow \rangle = \langle \uparrow \downarrow | \rho^{(2)} | \downarrow \uparrow \rangle =$$

$$\langle \downarrow \uparrow | \rho^{(2)} | \uparrow \downarrow \rangle = \langle \downarrow \uparrow | \rho^{(2)} | \downarrow \uparrow \rangle \equiv \rho_{\text{Scr},12}^{(2)}, \text{ and } \langle \downarrow \downarrow | \rho^{(2)} | \downarrow \downarrow \rangle \equiv \rho_{\text{Scr},22}^{(2)}, \text{ where}$$

$$\rho_{\text{Scr},11}^{(2)} = \frac{\lambda^2 (2\lambda(4\lambda - 5) - 2(1 - \lambda)^2 \log(\frac{1-\lambda}{\lambda}) + 3)}{(2\lambda - 1)^3}, \quad (\text{H3})$$

$$\rho_{\text{Scr},12}^{(2)} = \frac{(1 - \lambda)\lambda(2\lambda + 2(1 - \lambda)\lambda \log(\frac{1-\lambda}{\lambda}) - 1)}{(2\lambda - 1)^3}, \quad (\text{H4})$$

$$\rho_{\text{Scr},22}^{(2)} = -\frac{(1 - \lambda)^2 (8\lambda^2 + 2\lambda^2 \log(\frac{1-\lambda}{\lambda}) - 6\lambda + 1)}{(2\lambda - 1)^3}. \quad (\text{H5})$$

1. Second moment of the Scrooge ensemble

First, we explicitly compute the second moment of the Scrooge ensemble for a D -dimensional density matrix ρ . We write ρ using its eigenvalues and eigenvectors $\rho = \sum_{m=1}^D \lambda_m |s_m\rangle \langle s_m|$. We next use the definition of the Scrooge ensemble as the ρ distortion of the Haar ensemble to write the second moment of the Scrooge ensemble as

$$\rho^{(2)} = \int \frac{(\sqrt{D\rho}|\Psi\rangle \langle \Psi| \sqrt{D\rho})^{\otimes 2}}{\|\sqrt{D\rho}|\Psi\rangle\|^2} P_{\text{Haar}}(\Psi) d\Psi \quad (\text{H6})$$

$$= \sum_{m,n,m',n'=1}^D \int \frac{D\sqrt{\lambda_m \lambda_n \lambda_{m'} \lambda_{n'}} \langle s_n | \Psi \rangle \langle s_m | \Psi \rangle \langle s_{n'} | \Psi \rangle \langle s_{m'} | \Psi \rangle}{\sum_{k=1}^D \lambda_k^2 |\langle s_k | \Psi \rangle|^2} P_{\text{Haar}}(\Psi) d\Psi |s_n\rangle \langle s_m| |s_{n'}\rangle \langle s_{m'}|. \quad (\text{H7})$$

From this equation, we clearly see that the only nonzero elements in the basis of the eigenvectors of ρ are the ones with $n = n'$, $m = m'$ and/or $n = m'$, $m = n'$. All the other terms are zero because of the average over the complex phases. We can then focus, without loss of generality, on the following matrix element of $\rho^{(2)}$:

$$\rho_{\text{Scr},nm}^{(2)} \equiv \langle s_n | \langle s_m | \rho^{(2)} | s_n \rangle | s_m \rangle = \langle s_n | \langle s_m | \rho^{(2)} | s_m \rangle | s_n \rangle \quad (\text{H8})$$

$$= \int \frac{D\lambda_m \lambda_n |\Psi_n|^2 |\Psi_m|^2 (D-1)!}{\sum_{k=1}^D \lambda_k^2 |\Psi_k|^2} \frac{\pi^D}{\pi^D} \delta\left(1 - \sum_{j=1}^D |\Psi_j|^2\right) d^2\Psi_1 \dots d^2\Psi_D, \quad (\text{H9})$$

where we use the short notation $\Psi_j = \langle s_j | \Psi \rangle$ and the explicit expression of P_{Haar} . Since the integrand depends on only the absolute values of the complex variables $\Psi_1 \dots \Psi_D$, we can perform the integral over the phases and then change variable to $x_j = \lambda_j |\Psi_j|^2$, obtaining

$$\rho_{\text{Scr},nm}^{(2)} = \left(\prod_{j=1}^D \lambda_j^{-1} \right) D! \int_0^\infty dx_1 \dots \int_0^\infty dx_D \frac{x_n x_m}{\sum_{k=1}^D x_k} \delta\left(1 - \sum_{j=1}^D \lambda_j^{-1} x_j\right). \quad (\text{H10})$$

To remove the delta function, it is useful to define the following quantity:

$$f_{nm}^{(2)}(t) = \left(\prod_{j=1}^D \lambda_j^{-1} \right) D! \int_0^\infty dx_1 \dots \int_0^\infty dx_D \frac{x_n x_m}{\sum_{k=1}^D x_k} \delta\left(t - \sum_{j=1}^D \lambda_j^{-1} x_j\right) \quad (\text{H11})$$

such that $\rho_{\text{Scr},nm}^{(2)} = f_{nm}^{(2)}(1)$, and the Laplace transform

$$\tilde{f}_{nm}^{(2)}(s) = \int_0^\infty f_{nm}^{(2)}(t) e^{-st} dt = \left(\prod_{j=1}^D \lambda_j^{-1} \right) D! \int_0^\infty dx_1 \cdots \int_0^\infty dx_D \frac{x_n x_m}{\sum_{k=1}^D x_k} \exp\left(-s \sum_{j=1}^D \lambda_j^{-1} x_j\right). \quad (\text{H12})$$

The change of variables $x_j \rightarrow x_j/s$ yields

$$\tilde{f}_{nm}^{(2)}(s) = \left(\prod_{j=1}^D \lambda_j^{-1} \right) D! s^{-D-1} \int_0^\infty dx_1 \cdots \int_0^\infty dx_D \frac{x_n x_m}{\sum_{k=1}^D x_k} \exp\left(-\sum_{j=1}^D \lambda_j^{-1} x_j\right). \quad (\text{H13})$$

Since the inverse Laplace transform of $D!s^{-D-1}$ is t^D , we find

$$\rho_{\text{Scr},nm}^{(2)} = f_{nm}^{(2)}(1) = \left(\prod_{j=1}^D \lambda_j^{-1} \right) \int_0^\infty dx_1 \cdots \int_0^\infty dx_D \frac{x_n x_m}{\sum_{k=1}^D x_k} \exp\left(-\sum_{j=1}^D \lambda_j^{-1} x_j\right). \quad (\text{H14})$$

Note that these steps allow us to get from the integral over the Haar ensemble of normalized states in Eq. (H10) to an integral over the unnormalized Gaussian ensemble.

We now define

$$\Lambda_2 = \int_0^\infty dx_1 \cdots \int_0^\infty dx_D \frac{1}{\sum_{k=1}^D x_k} \exp\left(-\sum_{j=1}^D \lambda_j^{-1} x_j\right) \quad (\text{H15})$$

such that we can now write all the $\rho_{\text{Scr},nm}^{(2)}$ from this single “partition function”:

$$\rho_{\text{Scr},nm}^{(2)} = \left(\prod_{j=1}^D \lambda_j^{-1} \right) \frac{\partial}{\partial \lambda_n^{-1}} \frac{\partial}{\partial \lambda_n^{-1}} \Lambda_2. \quad (\text{H16})$$

To find the expression for Λ_2 in a closed form, we introduce an additional integration variable z :

$$\Lambda_2 = \int_0^\infty dx_1 \cdots \int_0^\infty dx_D \int_0^\infty dz \exp\left[-\sum_{j=1}^D (\lambda_j^{-1} + z) x_j\right]. \quad (\text{H17})$$

The equivalence with Eq. (H15) is evident upon performing the integral in z . Since the integral is now factorized, we can easily compute the integral in the variables $x_1, \dots, x_D$, obtaining

$$\Lambda_2 = \int_0^\infty dz \prod_{j=1}^D (\lambda_j^{-1} + z)^{-1} = \int_0^\infty dz \sum_{j=1}^D \frac{A_j}{\lambda_j^{-1} + z}, \quad (\text{H18})$$

where

$$A_j = \prod_{k \neq j} (\lambda_k^{-1} - \lambda_j^{-1})^{-1}. \quad (\text{H19})$$

We can now perform the integral in z , and we get

$$\Lambda_2 = \lim_{M \rightarrow \infty} \sum_{j=1}^D A_j [\ln(\lambda_j^{-1} + M) - \ln(\lambda_j^{-1})]. \quad (\text{H20})$$

Using the property $\sum_j A_j = 0$, we find

$$\begin{aligned} & \lim_{M \rightarrow \infty} \sum_j A_j \ln(\lambda_j^{-1} + M) \\ &= \lim_{M \rightarrow \infty} \left[\sum_j A_j \ln(M) + \sum_j A_j \ln\left(1 + \frac{\lambda_j^{-1}}{M}\right) \right] = 0. \end{aligned} \quad (\text{H21})$$

We then get the final expression for Λ_2 :

$$\Lambda_2 = \sum_{j=1}^D A_j \ln(\lambda_j) = \sum_{j=1}^D \left[\prod_{k \neq j} (\lambda_k^{-1} - \lambda_j^{-1})^{-1} \right] \ln(\lambda_j). \quad (\text{H22})$$

This expression can be straightforwardly used to evaluate the matrix elements of the second moment of the Scrooge ensemble using Eq. (H16).

2. Second moment in the limit of small purity of ρ

Here, we show that, in the limit that ρ has small purity, the second moment of the Scrooge ensemble has a simple form $\rho^{(2)} \simeq (1 + \mathbb{S})\rho^{\otimes 2}$, equivalent to the product form for the temporal ensemble (Theorem 2). To do so, we compute an approximate formula for Λ_2 in the case of a state ρ with small purity $\text{Tr}[\rho^2] = \sum_k \lambda_k^2 \ll 1$. Using that $\sum_k \lambda_k^n \leq \sum_k \lambda_k^2 \ll 1$ for every $n \geq 2$, from Eq. (H18) we get

$$\Lambda_2 = \left(\prod_{j=1}^D \lambda_j \right) \int_0^\infty dz \prod_{k=1}^D (1 + \lambda_k z)^{-1} \quad (\text{H23})$$

$$= \left(\prod_{j=1}^D \lambda_j \right) \int_0^\infty dz \left\{ 1 + \left(\sum_{k=1}^D \lambda_k \right) z + \frac{1}{2} \left[\left(\sum_{k=1}^D \lambda_k \right)^2 - \sum_{k=1}^D \lambda_k^2 \right] z^2 + \dots \right\}^{-1} \quad (\text{H24})$$

$$\approx \left(\prod_{j=1}^D \lambda_j \right) \int_0^\infty dz \left[\sum_{n=0}^\infty \frac{1}{n!} \left(\sum_{k=1}^D \lambda_k \right)^n z^n \right]^{-1} \quad (\text{H25})$$

$$= \left(\prod_{j=1}^D \lambda_j \right) \int_0^\infty dz \exp \left(-z \sum_{k=1}^D \lambda_k \right) = \left(\prod_{j=1}^D \lambda_j \right) \left(\sum_{k=1}^D \lambda_k \right)^{-1}. \quad (\text{H26})$$

From this equation, we obtain

$$\rho_{nm}^{(2)} \approx \lambda_n \lambda_m [(1 + \delta_{nm})(1 - \lambda_n - \lambda_m) + 2\lambda_m \lambda_n], \quad (\text{H27})$$

and, assuming that the Schmidt coefficients are small, we have $\rho_{\text{Scr},nm}^{(2)} \simeq \lambda_n \lambda_m (1 + \delta_{nm})$, which implies $\rho^{(2)} \simeq (1 + \mathbb{S})\rho^{\otimes 2}$.

3. Higher moments

We now generalize the above derivation to compute the matrix elements of the k th moment ($k < D$). The nonzero matrix elements in the eigenbasis of ρ are of the form

$$\rho_{m_1, \dots, m_k}^{(k)} \equiv \langle s_{m_1} | \dots \langle s_{m_k} | \rho^{(2)} | s_{m'_1} \rangle \dots | s_{m'_k} \rangle, \quad (\text{H28})$$

where $m'_1, \dots, m'_k$ is a permutation of the indices $m_1, \dots, m_k$. With the same procedure used above for the second moment, we get

$$\begin{aligned} \rho_{m_1, \dots, m_k}^{(k)} &= \left(\prod_{j=1}^D \lambda_j^{-1} \right) \int_0^\infty dx_1 \dots \\ &\times \int_0^\infty dx_D \frac{x_{m_1} \dots x_{m_k}}{(\sum_{j=1}^D x_j)^{k-1}} \exp \left(- \sum_{j=1}^D \lambda_j^{-1} x_j \right). \end{aligned} \quad (\text{H29})$$

We can compute the this matrix element for arbitrary indices using the generating function Λ_k here defined:

$$\rho_{m_1, \dots, m_k}^{(k)} = \left(\prod_j \lambda_j^{-1} \right) \frac{\partial^k \Lambda_k}{\partial \lambda_{m_1}^{-1} \dots \partial \lambda_{m_k}^{-1}}, \quad (\text{H30})$$

$$\Lambda_k = \int_0^\infty dx_1 \dots \int_0^\infty dx_D \frac{(-1)^k}{(\sum_{j=1}^D x_j)^{k-1}} \exp \left(- \sum_{j=1}^D \lambda_j^{-1} x_j \right). \quad (\text{H31})$$

Similarly to the case $k = 2$, the integral can be computed by introducing an additional integration variable:

$$\begin{aligned} \Lambda_k &= \int_0^\infty dx_1 \dots \int_0^\infty dx_D \exp \left(- \sum_{j=1}^D \lambda_j^{-1} x_j \right) \\ &\times \int_0^\infty dz \frac{(-1)^k z^{k-2}}{(k-2)!} \exp \left(-z \sum_j x_j \right) \end{aligned} \quad (\text{H32})$$

$$= \int_0^\infty dz \frac{(-1)^k z^{k-2}}{(k-2)!} \prod_{j=1}^D (\lambda_j^{-1} + z)^{-1} \quad (\text{H33})$$

$$= \int_0^\infty dz \frac{(-1)^k z^{k-2}}{(k-2)!} \sum_{j=1}^D \frac{A_j}{\lambda_j^{-1} + z} \quad (\text{H34})$$

$$= \sum_{j=1}^D \lambda_j^{-k+2} A_j \ln \lambda_j. \quad (\text{H35})$$

This derivation holds for $k > 1$, but we can define, in general, Λ_k using Eq. (H34). Then, the subentropy derived in Ref. [89] can be written as

$$\begin{aligned} \mathcal{Q}(\rho) &= - \left(\prod_{j=1}^D \lambda_j^{-1} \right) \sum_{j=1}^D A_j \lambda_j^2 \ln \lambda_j \\ &= - \left(\prod_{j=1}^D \lambda_j^{-1} \right) \Lambda_0. \end{aligned} \quad (\text{H36})$$

APPENDIX I: PROOF OF LEMMA 1: kTH MOMENTS OF TEMPORAL ENSEMBLE OF PROJECTED STATES

Here, we provide the proof for Lemma 1.

Lemma 1 (temporal kth moment of the projected state). Consider the temporal ensemble comprised of the

(unnormalized) projected states $|\tilde{\Psi}(x_B, t)\rangle \equiv (\mathbb{I}_A \otimes |x_B\rangle) |\Psi(t)\rangle$. The k th moment of this temporal ensemble is equal to the k th moment of the corresponding unnormalized Scrooge ensemble, up to small corrections:

$$\begin{aligned} \rho_{\text{Temp Proj}}^{(k)}(x_B) &\equiv \mathbb{E}_t[|\tilde{\Psi}(x_B, t)\rangle \langle \tilde{\Psi}(x_B, t)|^{\otimes k}] \\ &= p_d(x_B)^k \tilde{\rho}_{\text{Scr}}^{(k)}[\bar{\rho}(x_B)] + O[\Delta_\beta(x_B) p_d(x_B)^k], \end{aligned}$$

where $\Delta_\beta(x_B) \equiv \text{tr}[(\mathbb{I}_A \otimes |x_B\rangle \langle x_B|)^{\otimes 2} \rho_d^{(2)}] / p_d(x_B)^2$ is a correction term that is typically exponentially small in system size. To see this, note that $\Delta_\beta(x_B)$ can be interpreted as the purity of a probability distribution over D elements. Specifically, we define the quantity $q_{x_B}(E) \equiv |c_E|^2 \langle E | (\mathbb{I}_A \otimes |x_B\rangle \langle x_B|) | E \rangle / p_d(x_B)$, which is a probability distribution because it is non-negative and satisfies $\sum_E q_{x_B}(E) = 1$. $\Delta_\beta(x_B) = \sum_E q_{x_B}(E)^2$ is the purity of $q_{x_B}(E)$, which we expect to be typically $O(D^{-1})$.

Proof. This is a simple application of an intermediate step from Appendix D, which states that

$$\begin{aligned} \left\| \bar{\rho}(x_B)^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma) - \rho_{\text{Temp Proj}}^{(k)} \right\|_* &= \left\| \sum_{s \in M_k(D)} \bar{\rho}(x_B)^{\otimes k} \left(1 - \frac{|\pi(s)|}{|S_k|} \right) P_s \sum_{\sigma \in S_k} \text{Perm}(\sigma) \right\|_* \\ &\leq C_k \sum_{\substack{s \neq () \\ s \in M_k(D)}} \text{tr} \left[\bar{\rho}(x_B)^{\otimes (k-2)} \otimes \bar{\rho}^{(2)}(x_B) \sum_{\sigma \in S_k} \text{Perm}(\sigma) \right] \\ &\leq (C_k k!) \text{tr}[\bar{\rho}^{(2)}(x_B)] p_d(x_B)^{k-2} \equiv C'_k \Delta_\beta(x_B) p_d(x_B)^k. \end{aligned} \quad (11)$$

■.

APPENDIX J: PROOF OF THEOREM 4: WEIGHTED k TH MOMENTS OF PROJECTED ENSEMBLE

Here, we prove Theorem 4.

Theorem 4. At typical (late) times t , the weighted k th moment $\tilde{\rho}_{\text{Proj}}^{(k)}(t)$ of the unnormalized projected ensemble is equal to the k th moment of the generalized Scrooge ensemble:

$$\tilde{\rho}_{\text{Proj}}^{(k)}(t) \equiv \sum_{x_B} \frac{|\tilde{\Psi}(x_B, t)\rangle \langle \tilde{\Psi}(x_B, t)|^{\otimes k}}{p_d(x_B)^{k-1}} \quad (J1)$$

$$= \sum_{x_B} p_d(x_B) \tilde{\rho}_{\text{Scr}}^{(k)}[\bar{\rho}(x_B)] + O(\Delta_\beta^{1/2}), \quad (J2)$$

where $\Delta_\beta \equiv \sum_{x_B} p_d(x_B) \Delta_\beta(x_B)$ the weighted mean of $\Delta_\beta(x_B)$.

Proof. First, we compute the time average:

$$\mathbb{E}_t[\tilde{\rho}_{\text{Proj}}^{(k)}(t)] \equiv \sum_{x_B} p_d(x_B) \tilde{\rho}_{\text{Scr}}^{(k)}[\bar{\rho}(x_B)] + O(\Delta_\beta). \quad (J3)$$

This follows from Appendix I. Next, we compute the variance $\text{var}_t[\tilde{\rho}_{\text{Proj}}^{(k)}(t)]$ by computing the trace norm

$$\|\mathbb{E}_t[\tilde{\rho}_{\text{Proj}}^{(k)}(t)^{\otimes 2}] - \mathbb{E}_t[\tilde{\rho}_{\text{Proj}}^{(k)}(t)]^{\otimes 2}\|_* = O[\text{tr}(\rho_d^2)] = O(\Delta_\beta). \quad (J4)$$

The temporal averages can be in terms of permutations of $2k$ and k elements, respectively, and correction terms E.C. in Lemma 3. The dominant term is from the former: permutations of $2k$ elements that cannot be written as disjoint permutations of two groups of k elements. The largest possible contribution comes from the permutation where only two elements from each group are swapped and all others are not permuted. This has trace

$$\sum_{x_B, x'_B} \text{tr}[(\mathbb{I}_A \otimes |x_B\rangle \langle x_B|) \rho_d(\mathbb{I}_A \otimes |x'_B\rangle \langle x'_B|) \rho_d] = \text{tr}(\rho_d^2), \quad (J5)$$

and the contributions of all other relevant permutations and correction terms can be bounded by Eq. (J5). Finally, we can bound $\text{tr}(\rho_d^2) \leq \Delta_\beta$ as follows: We use Sedrakyan's inequality, a specialization of the Cauchy-Schwarz inequality. It states that, for positive u_i and v_i , $\sum_i u_i^2 / \sum_i v_i \leq \sum_i u_i^2 / v_i$. Applied to our context, we have

$$\begin{aligned} \text{tr}(\rho_d^2) &= \sum_E |c_E|^4 = \sum_E |c_E|^4 \frac{(\sum_{x_B} \langle E | (\mathbb{I}_A \otimes |x_B\rangle \langle x_B|) | E \rangle)^2}{\sum_{x_B} p_d(x_B)} \\ &\leq \sum_E |c_E|^4 \sum_{x_B} \frac{\langle E | (\mathbb{I}_A \otimes |x_B\rangle \langle x_B|) | E \rangle^2}{p_d(x_B)} \\ &\leq \sum_{x_B} p_d(x_B) \Delta_\beta(x_B) \equiv \Delta_\beta. \end{aligned} \quad (J6)$$

While the tightest general bound we can achieve is $\Delta_\beta \leq 1$ (saturated when the initial state is an energy eigenstate $|\Psi_0\rangle = |E\rangle$), we expect Δ_β to be exponentially small in

total system size. As discussed above, $\Delta_\beta(x_B) = \sum_E q_{x_B}(E)^2$ is the purity of a distribution $q_{x_B}(E)$ and, hence, is typically $O(D^{-1})$. Δ_β is the weighted average of such purities. ■

APPENDIX K: ADDITIONAL NUMERICAL EVIDENCE

1. Chaotic XXZ model

In this appendix, we support our results on the projected ensemble with another chaotic model. Specifically, we study a variant of the XXZ model [145]:

$$H = \frac{1}{4} \sum_{j=1}^{N-1} [J(X_j X_{j+1} + Y_j Y_{j+1}) + \Delta Z_j Z_{j+1}] + \frac{\Delta'}{4} \sum_{j=1}^{N-2} Z_j Z_{j+2}, \quad (\text{K1})$$

where X_j , Y_j , and Z_j are the Pauli matrices on site j . Following Ref. [145], we set $J = \sqrt{2}$, $\Delta = (\sqrt{5} + 1)/4$, and $\Delta' = 1$ and perform simulations on a one-dimensional chain with open boundary conditions.

This model serves as a further check of our theoretical conclusions but also serves to illustrate the effect of conserved quantities on the projected ensemble. This model has a $U(1)$ symmetry, leading to a conserved global magnetization $M \equiv \sum_j Z_j$.

The $U(1)$ symmetry leads to significant correlations between projected state and outcome. For simplicity, we consider Z -basis measurements on the complementary

system. If the measured bit string z_B has magnetization m , the projected state $|\Psi(z_B)\rangle$ is guaranteed to have magnetization $M - m$. As a result, even in the case where the outcomes z_B are not “energy revealing,” they reveal charge, another conserved quantity. We no longer expect the entire projected ensemble to be described by the Scrooge ensemble. Rather, following Ref. [29], we form the projected ensemble only out of measurement outcomes z_B with fixed magnetization m and conjecture that this is described by a Scrooge ensemble of the “sector reduced density matrix,” $\rho_m \propto \sum_{z_B, |z_B|=m} |\Psi(z_B)\rangle \langle \Psi(z_B)|$.

In the more general case where the outcomes z_B are energy revealing, Theorem 4 applies without modification: Each measurement outcome z_B has its own average state $\bar{\rho}(z_B)$; the projected state $|\Psi(z_B)\rangle$ is sampled from the corresponding Scrooge ensemble $\text{Scrooge}[\bar{\rho}(z_B)]$. We illustrate the validity of this result in Fig. 9(a). We plot the trace distances between the third moments of the projected and generalized Scrooge ensemble, for a time evolution of the initial state with one domain wall $|\Psi_0\rangle = |0\rangle^{\otimes(N/2-1)} |1\rangle^{\otimes(N/2+1)}$. The trace distances decay exponentially with increasing system size, consistent with Theorem 4. Meanwhile, the Scrooge ensemble $\text{Scrooge}[\rho_m]$ is a good description of the projected ensemble for smaller system sizes, but deviations become noticeable at larger system sizes, illustrated by the trace distance between the third moments of the projected and Scrooge ensemble, which saturates to a nonzero value (light green). As with the MFIM, the Porter-Thomas distribution is visible in the outcome probabilities $p(z_A, z_B)$, serving as a measurable signature of the generalized Scrooge ensemble.

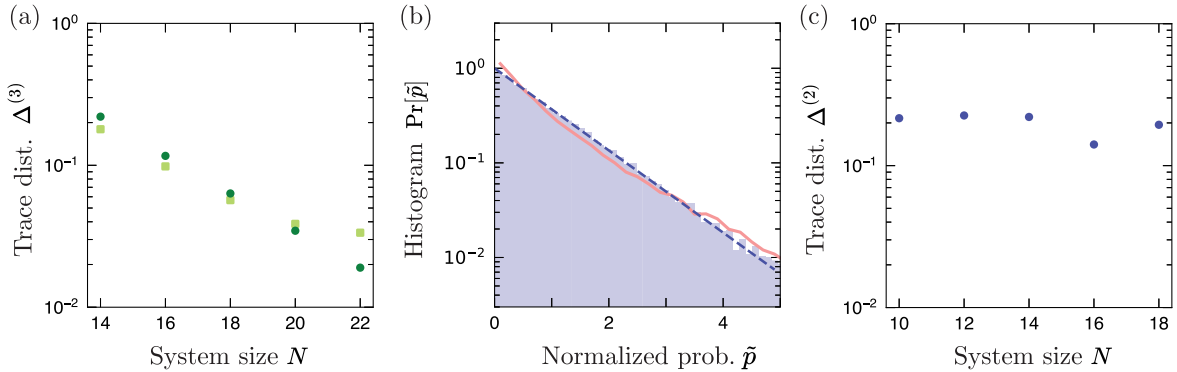


FIG. 9. Projected ensembles in a chaotic and integrable model. (a) We time evolve an initial state and plot the trace distances between the third moments of the Scrooge (light green) and generalized Scrooge (dark green) ensembles and the projected ensemble generated by a evolving a state under a variant of the XXZ model [Eq. (K1)] for a subsystem of four qubits and measurements in the Z_B basis. Like in the MFIM, the projected ensemble is close to the Scrooge ensemble but deviates from it at large system sizes. In contrast, the generalized Scrooge ensemble provides good agreement, even for the largest system sizes studied in this work. (b) The PT distribution is visible in the statistics of the normalized outcome probabilities $\tilde{p} \equiv p_{z_A}(z_B)/\mathbb{E}_t[p_{z_A}(z_B)]$ [while the raw probabilities $Dp_{z_A}(z_B)$ show small deviations] (histogram plotted over all z_A and z_B). (c) In contrast, the trace distance $\Delta^{(2)}$ between the Scrooge ensemble and the projected ensemble generated from the state $\exp(-iH_{\text{TFIM}}t)|0\rangle^{\otimes N}$ remains high, where H_{TFIM} is the integrable transverse field Ising model [Eq. (36) with $h_x = 0$], the subsystem A consists of two qubits, and the complement B is measured in the X basis.

2. Transverse field Ising model

We also demonstrate an example where the projected ensemble is not described by the generalized Scrooge ensemble, using the time-evolved states of the transverse field Ising model (TFIM), which is the mixed field Ising model [Eq. (36)] with $h_x = 0$. We time evolve the initial state $|0\rangle^{\otimes N}$ and construct the projected ensemble on two sites near the middle of the chain, with measurements in the X_B basis. Unlike the MFIM, the projected ensemble from the TFIM is not close to the generalized Scrooge ensemble [Fig. 9(c)]. This is because the TFIM is integrable, has solution by free fermions and, hence, does not satisfy the k th no-resonance conditions. We have also taken care to ensure that our projected ensemble should not be described by the real Scrooge ensemble (Appendix L below): The time-reversal symmetry is broken by time evolution, and the particle-hole symmetry is broken by the choice of initial state.

APPENDIX L: TIME-REVERSAL SYMMETRY AND REAL SCROOGE ENSEMBLES

Here, we discuss a notable deviation from the Scrooge ensemble: In the presence of relevant symmetries, states in the ensemble can be constrained to be purely real. In this case, one has to construct a maximally entropic ensemble of *real* states, which we dub the *real Scrooge ensemble*.

A simple way that this can emerge is the presence of time-reversal symmetry: the existence of a unitary operator U that maps the Hamiltonian H to its complex conjugate H^* : $UHU^\dagger = H^*$. The eigenstates of such Hamiltonians are guaranteed to be real in an appropriate basis. Under appropriate measurement bases z_B , the projected states are guaranteed to be real. The real Scrooge ensemble can be

defined as an analogous ρ distortion of the maximally entropic ensemble of *real* states, which has been studied in, e.g., Ref. [48] and we dub the *real Haar ensemble*.

As an example, in analogy to Eq. (H16), the second moment of the real Scrooge ensemble has nonzero entries:

$$\rho_{nm}^{(2)} = \left(\prod_{k=1}^{d_A} \lambda_k^{-1} \right)^{1/2} \frac{\partial}{\partial \mu_m} \frac{\partial}{\partial \mu_n} \Lambda_2^{\text{real}} \Big|_{\mu_i = \lambda_i^{-1}}, \quad (\text{L1})$$

where λ_i are the eigenvalues of the first moment ρ and

$$\Lambda_2^{\text{real}} = 2 \int_0^\infty dz \prod_{k=1}^{d_A} (\mu_k + z)^{-1/2}. \quad (\text{L2})$$

While there is no closed-form expression for Λ_2^{real} analogous to Eq. (H22), for a given first moment ρ , we can numerically evaluate Eq. (L1) to obtain the higher moments of the real Scrooge ensemble.

In Fig. 10(b), we show that the real Scrooge ensemble describes the projected ensemble of eigenstates of the MFIM. As discussed in Sec. VI E, the MFIM is time-reversal symmetric. By measuring in the x_B basis, the projected states can be written as real states. We compute the trace distance between the second moments of the real Scrooge ensemble (blue) and the real Haar ensemble (red). The real Haar ensemble describes only the projected ensemble at infinite temperature $E \approx 0$, while the real Scrooge ensemble describes the projected ensemble at all energies and temperatures, with exponentially increasing accuracy with increasing system size [Fig. 10(b)].

In addition to trace distances, we take the *real Porter-Thomas distribution* as smoking-gun evidence of the real

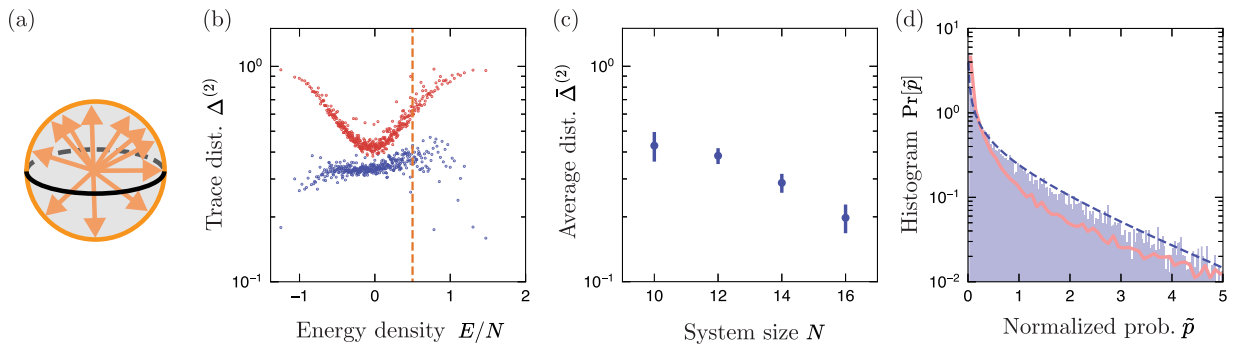


FIG. 10. Real Scrooge ensemble. (a) Under appropriate conditions, the states in the temporal and projected ensembles can be restricted to be always real, indicated here with single-qubit states lying in the X - Z plane of the Bloch sphere (orange circle). (b) We demonstrate this with the eigenstates of the MFIM, a time-reversal-symmetric Hamiltonian. We compare the second moments of the projected ensemble with those of the real Haar ensemble (red) and the real Scrooge ensemble (blue). The real Scrooge ensemble consistently has the smaller trace distance $\Delta^{(2)}$. (c) With increasing system size, the projected ensembles appear to converge to the real Scrooge ensemble. (d) We verify the generalized real Scrooge ensemble with a similar procedure as Fig. 6(e): The quantity $\tilde{p}_{z_A}(z_B) = p_{z_A}^{(E)}(z_B)/\mathbb{E}_E[p_{z_A}^{(E)}(z_B)]$ (blue) obeys a real PT distribution [Eq. (L3), dashed line; note cusp at $\tilde{p} = 0$], for $z_A = 0100$, while the raw probabilities $Dp_{z_A}^{(E)}(z_B)$ do not (red).

Scrooge ensemble. The real Porter-Thomas distribution [98] is an analogous universal distribution, given by

$$P_{\text{PT},\text{R}}(x)dx = \frac{\exp(-x/2)}{\sqrt{2\pi x}} dx. \quad (\text{L3})$$

In Fig. 10(d), we show the real PT distribution in relevant measurement outcomes $p_E(o_A, x_B)$, following the same procedure as Fig. 6. By using the same normalization procedure in Fig. 6(e), we see even better agreement with the real PT distribution (red), indicating that the *generalized* real Scrooge ensemble is likely a better description than the real Scrooge ensemble, as expected.

Finally, we remark that time-evolved states can also be made to be all real. Unlike with energy eigenstates, time-reversal symmetry is not sufficient, because time evolution breaks this symmetry. Rather, we need both time-reversal symmetry and *particle-hole symmetry*, which ensures the energy spectrum is symmetric about 0. With appropriate initial states, the time-evolved state will always be real, and as such the projected ensemble will be described by the real Scrooge and generalized real Scrooge ensembles. See Supplemental Material of Ref. [110] for more details.

-
- [1] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Elsevier, New York, 2013), Vol. 5.
 - [2] J. D. Bekenstein, *Generalized second law of thermodynamics in black-hole physics*, *Phys. Rev. D* **9**, 3292 (1974).
 - [3] T. M. Cover, *Elements of Information Theory* (John Wiley & Sons, New York, 1999).
 - [4] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University, New York, 2009).
 - [5] C. H. Bennett and R. Landauer, *The fundamental physical limits of computation*, *Sci. Am.* **253**, 48 (1985), <https://www.scientificamerican.com/article/the-fundamental-physical-limits-of-computation/>.
 - [6] E. T. Jaynes, *Information theory and statistical mechanics*, *Phys. Rev.* **106**, 620 (1957).
 - [7] W. Grandy, Jr., *Principle of maximum entropy and irreversible processes*, *Phys. Rep.* **62**, 175 (1980).
 - [8] L. Martyushev and V. Seleznev, *Maximum entropy production principle in physics, chemistry and biology*, *Phys. Rep.* **426**, 1 (2006).
 - [9] J. R. Banavar, A. Maritan, and I. Volkov, *Applications of the principle of maximum entropy: From physics to ecology*, *J. Phys. Condens. Matter* **22**, 063101 (2010).
 - [10] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, *Principles of maximum entropy and maximum caliber in statistical physics*, *Rev. Mod. Phys.* **85**, 1115 (2013).
 - [11] C. Gogolin, M. P. Müller, and J. Eisert, *Absence of thermalization in nonintegrable systems*, *Phys. Rev. Lett.* **106**, 040401 (2011).
 - [12] C. Gogolin and J. Eisert, *Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems*, *Rep. Prog. Phys.* **79**, 056001 (2016).
 - [13] A. J. Short, *Equilibration of quantum systems and subsystems*, *New J. Phys.* **13**, 053009 (2011).
 - [14] A. Riera, C. Gogolin, and J. Eisert, *Thermalization in nature and on a quantum computer*, *Phys. Rev. Lett.* **108**, 080402 (2012).
 - [15] J. M. Deutsch, *Eigenstate thermalization hypothesis*, *Rep. Prog. Phys.* **81**, 082001 (2018).
 - [16] M. Ueda, *Quantum equilibration, thermalization and prethermalization in ultracold atoms*, *Nat. Rev. Phys.* **2**, 669 (2020).
 - [17] P. Reimann, *Foundation of statistical mechanics under experimentally realistic conditions*, *Phys. Rev. Lett.* **101**, 190403 (2008).
 - [18] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner *et al.*, *Probing many-body dynamics on a 51-atom quantum simulator*, *Nature (London)* **551**, 579 (2017).
 - [19] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, *Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator*, *Nature (London)* **551**, 601 (2017).
 - [20] B. Neyenhuis, J. Zhang, P. W. Hess, J. Smith, A. C. Lee, P. Richerme, Z.-X. Gong, A. V. Gorshkov, and C. Monroe, *Observation of prethermalization in long-range interacting spin chains*, *Sci. Adv.* **3**, e1700672 (2017).
 - [21] J. Choi, H. Zhou, S. Choi, R. Landig, W. W. Ho, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, D. A. Abanin *et al.*, *Probing quantum thermalization of a disordered dipolar spin ensemble with discrete time-crystalline order*, *Phys. Rev. Lett.* **122**, 043603 (2019).
 - [22] 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 (2021).
 - [23] C. Gross and I. Bloch, *Quantum simulations with ultracold atoms in optical lattices*, *Science* **357**, 995 (2017).
 - [24] C. S. Chiu, G. Ji, A. Bohrdt, M. Xu, M. Knap, E. Demler, F. Grusdt, M. Greiner, and D. Greif, *String patterns in the doped Hubbard model*, *Science* **365**, 251 (2019).
 - [25] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, *Probing Rényi entanglement entropy via randomized measurements*, *Science* **364**, 260 (2019).
 - [26] P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. C. Lang, T. Lahaye *et al.*, *Quantum simulation of 2D antiferromagnets with hundreds of Rydberg atoms*, *Nature (London)* **595**, 233 (2021).
 - [27] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho *et al.*, *Quantum phases of matter on a 256-atom programmable quantum simulator*, *Nature (London)* **595**, 227 (2021).
 - [28] J. Choi, A. L. Shaw, I. S. Madjarov, X. Xie, R. Finkelstein, J. P. Covey, J. S. Cotler, D. K. Mark, H.-Y. Huang, A. Kale *et al.*, *Preparing random states and benchmarking with many-body quantum chaos*, *Nature (London)* **613**, 468 (2023).
 - [29] J. S. Cotler, D. K. Mark, H.-Y. Huang, F. Hernandez, J. Choi, A. L. Shaw, M. Endres, and S. Choi, *Emergent*

- quantum state designs from individual many-body wave functions, *PRX Quantum* **4**, 010311 (2023).
- [30] M. Ippoliti and W. W. Ho, *Dynamical purification and the emergence of quantum state designs from the projected ensemble*, *PRX Quantum* **4**, 030322 (2023).
- [31] M. Ippoliti and W. W. Ho, *Solvable model of deep thermalization with distinct design times*, *Quantum* **6**, 886 (2022).
- [32] P. W. Claeys and A. Lamacraft, *Emergent quantum state designs and biunitarity in dual-unitary circuit dynamics*, *Quantum* **6**, 738 (2022).
- [33] M. Lucas, L. Piroli, J. De Nardis, and A. De Luca, *Generalized deep thermalization for free fermions*, *Phys. Rev. A* **107**, 032215 (2023).
- [34] H. Wilming and I. Roth, *High-temperature thermalization implies the emergence of quantum state designs*, *arXiv:2202.01669*.
- [35] T. Bhore, J.-Y. Desaulles, and Z. Papić, *Deep thermalization in constrained quantum systems*, *Phys. Rev. B* **108**, 104317 (2023).
- [36] A. Chan and A. D. Luca, *Projected state ensemble of a generic model of many-body quantum chaos*, *arXiv:2402.16939*.
- [37] N. D. Varikuti and S. Bandyopadhyay, *Unraveling the emergence of quantum state designs in systems with symmetry*, *Quantum* **8**, 1456 (2024).
- [38] C. Vairogs and B. Yan, *Extracting randomness from quantum ‘magic’*, *arXiv:2402.10181*.
- [39] S. Pilatowsky-Cameo, C. B. Dag, W. W. Ho, and S. Choi, *Complete Hilbert-space ergodicity in quantum dynamics of generalized Fibonacci drives*, *Phys. Rev. Lett.* **131**, 250401 (2023).
- [40] S. Pilatowsky-Cameo, I. Marvian, S. Choi, and W. W. Ho, *Hilbert-space ergodicity in driven quantum systems: Obstructions and designs*, *arXiv:2402.06720*.
- [41] A. L. Shaw, D. K. Mark, J. Choi, R. Finkelstein, P. Scholl, S. Choi, and M. Endres, *Universal fluctuations and noise learning from Hilbert space ergodicity*, *arXiv:2403.11971*.
- [42] T. Mori and N. Shiraishi, *Thermalization without eigenstate thermalization hypothesis after a quantum quench*, *Phys. Rev. E* **96**, 022153 (2017).
- [43] T. Mori, T. N. Ikeda, E. Kaminishi, and M. Ueda, *Thermalization and prethermalization in isolated quantum systems: A theoretical overview*, *J. Phys. B* **51**, 112001 (2018).
- [44] G. De Palma, A. Serafini, V. Giovannetti, and M. Cramer, *Necessity of eigenstate thermalization*, *Phys. Rev. Lett.* **115**, 220401 (2015).
- [45] A. W. Harrow and Y. Huang, *Thermalization without eigenstate thermalization*, *arXiv:2209.09826*.
- [46] Y. Nakata, P. S. Turner, and M. Murao, *Phase-random states: Ensembles of states with fixed amplitudes and uniformly distributed phases in a fixed basis*, *Phys. Rev. A* **86**, 012301 (2012).
- [47] A. Ambainis and J. Emerson, *Quantum t -designs: t -wise independence in the quantum world*, *arXiv:quant-ph/0701126*.
- [48] A. W. Harrow, *The church of the symmetric subspace*, *arXiv:1308.6595*.
- [49] E. Rosenberg *et al.*, *Dynamics of magnetization at infinite temperature in a Heisenberg spin chain*, *Science* **384**, 48 (2024).
- [50] J. F. Wienand, S. Karch, A. Impertro, C. Schweizer, E. McCulloch, R. Vasseur, S. Gopalakrishnan, M. Aidelsburger, and I. Bloch, *Emergence of fluctuating hydrodynamics in chaotic quantum systems*, *Nat. Phys.* **20**, 1732 (2024).
- [51] M. Rigol, V. Dunjko, and M. Olshanii, *Thermalization and its mechanism for generic isolated quantum systems*, *Nature (London)* **452**, 854 (2008).
- [52] A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore, *Colloquium: Nonequilibrium dynamics of closed interacting quantum systems*, *Rev. Mod. Phys.* **83**, 863 (2011).
- [53] D. C. Brody and L. P. Hughston, *Information content for quantum states*, *J. Math. Phys. (N.Y.)* **41**, 2586 (2000).
- [54] F. Anza and J. P. Crutchfield, *Quantum information dimension and geometric entropy*, *PRX Quantum* **3**, 020355 (2022).
- [55] D. Hahn, D. J. Luitz, and J. T. Chalker, *The statistical properties of eigenstates in chaotic many-body quantum systems*, *Phys. Rev. X* **14**, 031029 (2024).
- [56] A. S. Holevo, *The capacity of the quantum channel with general signal states*, *IEEE Trans. Inf. Theory* **44**, 269 (1998).
- [57] I. Kukuljan, S. Grozdanov, and T. Prosen, *Weak quantum chaos*, *Phys. Rev. B* **96**, 060301(R) (2017).
- [58] A. Nahum, S. Vijay, and J. Haah, *Operator spreading in random unitary circuits*, *Phys. Rev. X* **8**, 021014 (2018).
- [59] X. Mi, P. Roushan, C. Quintana, S. Mandrà, J. Marshall, C. Neill, F. Arute, K. Arya, J. Atalaya, R. Babbush *et al.*, *Information scrambling in quantum circuits*, *Science* **374**, 1479 (2021).
- [60] F. Barratt, U. Agrawal, A. C. Potter, S. Gopalakrishnan, and R. Vasseur, *Transitions in the learnability of global charges from local measurements*, *Phys. Rev. Lett.* **129**, 200602 (2022).
- [61] M. Ippoliti and V. Khemani, *Learnability transitions in monitored quantum dynamics via eavesdropper’s classical shadows*, *PRX Quantum* **5**, 020304 (2024).
- [62] A. Altland, D. A. Huse, and T. Micklitz, *Maximum entropy quantum state distributions*, *arXiv:2203.12580*.
- [63] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th anniversary edition (Cambridge University Press, Cambridge, England, 2010).
- [64] For discrete ensembles, the integral reduces to a sum over all states in the ensemble; see Sec. III.
- [65] M. P. Fisher, V. Khemani, A. Nahum, and S. Vijay, *Random quantum circuits*, *Annu. Rev. Condens. Matter Phys.* **14**, 335 (2023).
- [66] A. C. Potter and R. Vasseur, *Entanglement dynamics in hybrid quantum circuits*, in *Entanglement in Spin Chains: From Theory to Quantum Technology Applications* (Springer, New York, 2022), pp. 211–249.
- [67] Y. Li, X. Chen, and M. P. A. Fisher, *Quantum Zeno effect and the many-body entanglement transition*, *Phys. Rev. B* **98**, 205136 (2018).
- [68] B. Skinner, J. Ruhman, and A. Nahum, *Measurement-induced phase transitions in the dynamics of entanglement*, *Phys. Rev. X* **9**, 031009 (2019).
- [69] Even when the ensembles are finite, such as the projected ensemble, we treat them as finite samples from an underlying continuous ensemble.

- [70] F. Anza and J. P. Crutchfield, *Beyond density matrices: Geometric quantum states*, *Phys. Rev. A* **103**, 062218 (2021).
- [71] F. Anza and J. P. Crutchfield, *Geometric quantum thermodynamics*, *Phys. Rev. E* **106**, 054102 (2022).
- [72] F. Anza and J. P. Crutchfield, *Maximum geometric quantum entropy*, *Entropy* **26**, 225 (2024).
- [73] E. Stechel and E. Heller, *Quantum ergodicity and spectral chaos*, *Annu. Rev. Phys. Chem.* **35**, 563 (1984).
- [74] M. Srednicki, *The approach to thermal equilibrium in quantized chaotic systems*, *J. Phys. A* **32**, 1163 (1999).
- [75] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghi, *Long-time behavior of macroscopic quantum systems*, *Eur. Phys. J. H* **35**, 173 (2010).
- [76] A. Vikram and V. Galitski, *Dynamical quantum ergodicity from energy level statistics*, *Phys. Rev. Res.* **5**, 033126 (2023).
- [77] J. L. Lebowitz and O. Penrose, *Modern ergodic theory*, *Phys. Today* **26**, No. 2, 23 (1973).
- [78] N. Linden, S. Popescu, A. J. Short, and A. Winter, *Quantum mechanical evolution towards thermal equilibrium*, *Phys. Rev. E* **79**, 061103 (2009).
- [79] I. Nechita and S. Singh, *A graphical calculus for integration over random diagonal unitary matrices*, *Linear Algebra Appl.* **613**, 46 (2021).
- [80] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandão, D. A. Buell *et al.*, *Quantum supremacy using a programmable superconducting processor*, *Nature (London)* **574**, 505 (2019).
- [81] M. Brin and G. Stuck, *Introduction to Dynamical Systems* (Cambridge University Press, Cambridge, England, 2002).
- [82] S. Goldstein, J. L. Lebowitz, C. Mastrodonato, R. Tumulka, and N. Zanghi, *Universal probability distribution for the wave function of a quantum system entangled with its environment*, *Commun. Math. Phys.* **342**, 965 (2016).
- [83] W. W. Ho and S. Choi, *Exact emergent quantum state designs from quantum chaotic dynamics*, *Phys. Rev. Lett.* **128**, 060601 (2022).
- [84] H. Shrotriya and W. W. Ho, *Nonlocality of deep thermalization*, *arXiv:2305.08437*.
- [85] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, *Quantum entanglement growth under random unitary dynamics*, *Phys. Rev. X* **7**, 031016 (2017).
- [86] H. Kim and D. A. Huse, *Ballistic spreading of entanglement in a diffusive nonintegrable system*, *Phys. Rev. Lett.* **111**, 127205 (2013).
- [87] W. W. Ho and D. A. Abanin, *Entanglement dynamics in quantum many-body systems*, *Phys. Rev. B* **95**, 094302 (2017).
- [88] G. De Palma and L. Hackl, *Linear growth of the entanglement entropy for quadratic Hamiltonians and arbitrary initial states*, *SciPost Phys.* **12**, 021 (2022).
- [89] R. Jozsa, D. Robb, and W. K. Wootters, *Lower bound for accessible information in quantum mechanics*, *Phys. Rev. A* **49**, 668 (1994).
- [90] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghi, *On the distribution of the wave function for systems in thermal equilibrium*, *J. Stat. Phys.* **125**, 1193 (2006).
- [91] P. Reimann, *Typicality of pure states randomly sampled according to the Gaussian adjusted projected measure*, *J. Stat. Phys.* **132**, 921 (2008).
- [92] As a technical point, Scrooge ensembles have maximum entropy only when they are regarded as ensembles of *unnormalized states* [53,72,90]; see Sec. III.
- [93] While our proof requires times that are exponential in system size, we find from numerical experiments that we require only times that grow weakly (at most polynomially) with system size.
- [94] B. Schumacher, *Quantum coding*, *Phys. Rev. A* **51**, 2738 (1995).
- [95] R. Lewis-Swan, A. Safavi-Naini, A. Kaufman, and A. Rey, *Dynamics of quantum information*, *Nat. Rev. Phys.* **1**, 627 (2019).
- [96] S. Xu and B. Swingle, *Scrambling dynamics and out-of-time-ordered correlators in quantum many-body systems*, *PRX Quantum* **5**, 010201 (2024).
- [97] S. Choi, Y. Bao, X.-L. Qi, and E. Altman, *Quantum error correction in scrambling dynamics and measurement-induced phase transition*, *Phys. Rev. Lett.* **125**, 030505 (2020).
- [98] C. E. Porter and R. G. Thomas, *Fluctuations of nuclear reaction widths*, *Phys. Rev.* **104**, 483 (1956).
- [99] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, *Characterizing quantum supremacy in near-term devices*, *Nat. Phys.* **14**, 595 (2018).
- [100] We ignore subtleties about the global phase of the states $|\Psi_j\rangle$ and assume that the states are gauge fixed to, for example, have real positive values of the first entry $\langle u_1 | \Psi_j \rangle$ in some basis $\{|u_i\rangle\}$.
- [101] A. L. Shaw, Z. Chen, J. Choi, D. K. Mark, P. Scholl, R. Finkelstein, A. Elben, S. Choi, and M. Endres, *Benchmarking highly entangled states on a 60-atom analog quantum simulator*, *Nature (London)* **628**, 71 (2023).
- [102] M. Christandl, *The structure of bipartite quantum states—Insights from group theory and cryptography*, *arXiv:quant-ph/0604183*.
- [103] H.-Y. Huang, R. Kueng, and J. Preskill, *Predicting many properties of a quantum system from very few measurements*, *Nat. Phys.* **16**, 1050 (2020).
- [104] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, *The randomized measurement toolbox*, *Nat. Rev. Phys.* **5**, 9 (2023).
- [105] N. Datta, T. Dorlas, R. Jozsa, and F. Benatti, *Properties of subentropy*, *J. Math. Phys. (N.Y.)* **55**, 062203 (2014).
- [106] G. Valiant and P. Valiant, *The power of linear estimators*, in *Proceedings of the 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science* (IEEE, New York, 2011), pp. 403–412.
- [107] G. Valiant and P. Valiant, *Estimating the unseen: Improved estimators for entropy and other properties*, *J. ACM* **64**, 1 (2017).
- [108] C. Nation and D. Porras, *Off-diagonal observable elements from random matrix theory: Distributions, fluctuations, and eigenstate thermalization*, *New J. Phys.* **20**, 103003 (2018).

- [109] C. Nation and D. Porras, *Ergodicity probes: Using time-fluctuations to measure the Hilbert space dimension*, *Quantum* **3**, 207 (2019).
- [110] D. K. Mark, J. Choi, A. L. Shaw, M. Endres, and S. Choi, *Benchmarking quantum simulators using ergodic quantum dynamics*, *Phys. Rev. Lett.* **131**, 110601 (2023).
- [111] Z. Liu, Z. Hao, and H.-Y. Hu, *Predicting arbitrary state properties from single Hamiltonian quench dynamics*, *arXiv:2311.00695*.
- [112] K. Kaneko, E. Iyoda, and T. Sagawa, *Characterizing complexity of many-body quantum dynamics by higher-order eigenstate thermalization*, *Phys. Rev. A* **101**, 042126 (2020).
- [113] Y. Huang, *Extensive entropy from unitary evolution*, *arXiv:2104.02053*.
- [114] J. Riddell and N. Pagliaroli, *No-resonance conditions, random matrices, and quantum chaotic models*, *arXiv:2307.05417*.
- [115] If there is a symmetry and $|\Psi_0\rangle$ is supported only on every eigenstate in a symmetry sector, we may conclude that the k th no-resonance conditions modulo degeneracies hold for the eigenvalues in that sector.
- [116] K. Schmüdgen, *The Moment Problem* (Springer International, Cham, 2017), pp. 381–398.
- [117] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, *Random quantum circuits anticconcentrate in log depth*, *PRX Quantum* **3**, 010333 (2022).
- [118] The subscript z indicates that the measurement outcome z is being fixed and the distribution $p_z(t)$ is taken over time t .
- [119] More accurately, we bound the distance between the k th moments of $p_z(t)$ and the Porter-Thomas distribution. We do not currently have a means to directly bound the distance between distributions, such as the total variation distance.
- [120] H. Kim, T. N. Ikeda, and D. A. Huse, *Testing whether all eigenstates obey the eigenstate thermalization hypothesis*, *Phys. Rev. E* **90**, 052105 (2014).
- [121] G. Parfionov and R. Zapatin, *‘Lazy’ quantum ensembles*, *J. Phys. A* **39**, 10891 (2006).
- [122] C. Murthy and M. Srednicki, *Structure of chaotic eigenstates and their entanglement entropy*, *Phys. Rev. E* **100**, 022131 (2019).
- [123] More precisely, the norm is a random variable which is a weighted sum of PT variables, where the weights are the eigenvalues of $\tilde{\rho}(x_B)$. The resulting distribution is known as a *hypoexponential distribution* and is discussed extensively in companion work [41] but depends on the spectrum of $\tilde{\rho}(x_B)$.
- [124] W. McGill, *Multivariate information transmission*, *Trans. IRE Prof. Group Inf. Theory* **4**, 93 (1954).
- [125] M. Rangamani and M. Rota, *Entanglement structures in qubit systems*, *J. Phys. A* **48**, 385301 (2015).
- [126] M. Rota, *Tripartite information of highly entangled states*, *J. High Energy Phys.* **04** (2016) 075.
- [127] M. Hartmann, G. Mahler, and O. Hess, *Gaussian quantum fluctuations in interacting many particle systems*, *Lett. Math. Phys.* **68**, 103 (2004).
- [128] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. Ludwig, *Topological insulators and superconductors: Tenfold way and dimensional hierarchy*, *New J. Phys.* **12**, 065010 (2010).
- [129] M. Schiulaz, E. J. Torres-Herrera, and L. F. Santos, *Thouless and relaxation time scales in many-body quantum systems*, *Phys. Rev. B* **99**, 174313 (2019).
- [130] C. Liu, Q. C. Huang, and W. W. Ho, *Deep thermalization in continuous-variable quantum systems*, *arXiv:2405.05470*.
- [131] M. C. Tran, D. K. Mark, W. W. Ho, and S. Choi, *Measuring arbitrary physical properties in analog quantum simulation*, *Phys. Rev. X* **13**, 011049 (2023).
- [132] L. Coopmans, Y. Kikuchi, and M. Benedetti, *Predicting Gibbs-state expectation values with pure thermal shadows*, *PRX Quantum* **4**, 010305 (2023).
- [133] R. Bassirian, A. Bouland, B. Fefferman, S. Gunn, and A. Tal, *On certified randomness from quantum advantage experiments*, *arXiv:2111.14846*.
- [134] S. Aaronson and S.-H. Hung, *Certified randomness from quantum supremacy*, *arXiv:2303.01625*.
- [135] L. Foini and J. Kurchan, *Eigenstate thermalization hypothesis and out of time order correlators*, *Phys. Rev. E* **99**, 042139 (2019).
- [136] J. Wang, M. H. Lamann, J. Richter, R. Steinigeweg, A. Dymarsky, and J. Gemmer, *Eigenstate thermalization hypothesis and its deviations from random-matrix theory beyond the thermalization time*, *Phys. Rev. Lett.* **128**, 180601 (2022).
- [137] S. Pappalardi, L. Foini, and J. Kurchan, *Eigenstate thermalization hypothesis and free probability*, *Phys. Rev. Lett.* **129**, 170603 (2022).
- [138] M. Fava, J. Kurchan, and S. Pappalardi, *Designs via free probability*, *arXiv:2308.06200*.
- [139] S. Jindal and P. Hosur, *Generalized Free Cumulants for Quantum Chaotic Systems*, *arXiv:2401.13829*.
- [140] P. Erdős, *On an elementary proof of some asymptotic formulas in the theory of partitions*, *Ann. Math.* **43**, 437 (1942).
- [141] J. Liu, *Spectral form factors and late time quantum chaos*, *Phys. Rev. D* **98**, 086026 (2018).
- [142] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, tenth printing ed. (Dover, New York, 1964), p. 364.
- [143] M. L. Mehta, *Random Matrices* (Elsevier, New York, 2004).
- [144] J. P. Keating, N. Linden, and H. J. Wells, *Spectra and eigenstates of spin chain Hamiltonians*, *Commun. Math. Phys.* **338**, 81 (2015).
- [145] T. LeBlond, D. Sels, A. Polkovnikov, and M. Rigol, *Universality in the onset of quantum chaos in many-body systems*, *Phys. Rev. B* **104**, L201117 (2021).