Quantum mechanical evolution towards thermal equilibrium

Noah Linden,[1] Sandu Popescu,[2,3] Anthony J. Short,[4] and Andreas Winter[1,5]

[1]*Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, United Kingdom*
[2]*H.H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, United Kingdom*
[3]*Hewlett-Packard Laboratories, Stoke Gifford, Bristol BS12 6QZ, United Kingdom*
[4]*Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences,*
University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom
[5]*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, Singapore 117542, Singapore*

The circumstances under which a system reaches thermal equilibrium, and how to derive this from basic dynamical laws, has been a major question from the very beginning of thermodynamics and statistical mechanics. Despite considerable progress, it remains an open problem. Motivated by this issue, we address the more general question of equilibration. We prove, with virtually full generality, that reaching equilibrium is a universal property of quantum systems: almost any subsystem in interaction with a large enough bath will reach an equilibrium state and remain close to it for almost all times. We also prove several general results about other aspects of thermalization besides equilibration, for example, that the equilibrium state does not depend on the detailed microstate of the bath.

I. INTRODUCTION

Leave your hot cup of coffee or cold beer alone for a while and they soon lose their appeal—the coffee cools down and the beer warms up and they both reach room temperature. It is not only coffee and beer reaching thermal equilibrium that are ubiquitous phenomena: everything does it. Thermalization is one of the most fundamental facts of nature.

But how exactly does thermalization occur? How can one derive the existence of this phenomenon from the basic dynamical laws of nature (such as Newton's or Schrödinger's equations)? These have been open questions since the very beginning of statistical mechanics more than a century and a half ago.

One—but by no means the only—stumbling block has been the fact that the basic postulates of statistical mechanics rely on subjective lack of knowledge and ensemble averages, which is very controversial as a physical principle. Recently however there has been significant progress: it was realized that ensemble averages and subjective ignorance are not needed because *individual* quantum states of systems can exhibit statistical properties. This is a purely quantum phenomenon, and the key is entanglement, which leads to *objective* lack of knowledge. Namely, in quantum mechanics even when we have complete knowledge of the state of a system, i.e., it is in a pure state and has zero entropy, the state of a subsystem may be mixed and have nonzero entropy. In this situation we cannot describe the subsystem by any particular pure state, and for all purposes it behaves as if we have a lack of knowledge about what its pure state is (i.e., it behaves as a probability distribution over pure states). This is in stark contrast to classical physics where complete knowledge of the state of the whole system implies complete knowledge of the state of any subsystem, and hence probabilities can only arise as purely subjective lack of knowledge (i.e., the subsystem has a well-defined state only we do not know what that is).

This approach has become a very fruitful direction of research in recent years [1–3]; see also important earlier work [4–7], numerical studies [8], and analytic results for specific models [9].

Most notably, it was shown in [1–3] that almost all (pure) states of a large system are such that any small subsystem is in a canonical state, as if it were the result of averaging (with equal probability) over all possible states respecting the desired macroscopic constraints. However, the above result is limited in that it refers to the state at a given time and deals only with the case of generic states. Although one might suspect that evolution would take any initial state into a generic state, the results of [1–3] are nondynamical and cannot be used to prove this. Furthermore, the states of most interest, namely, those initially far from equilibrium, are nongeneric. In this paper we seek to address the dynamical aspects of thermalization. We are interested not in the situation at one given moment of time, but in the time evolution—in particular, under what circumstances systems reach equilibrium and how much they fluctuate.

Thermalization seems a very straightforward process—put a system in contact with a large enough thermal bath and the system will reach equilibrium at the same temperature. Closer analysis reveals however that the process of thermalization actually contains many different aspects and we can decompose it into the following elements. They constitute a rough roadmap of what one has to show in order to demonstrate that a particular physical system thermalizes.

(1) Equilibration: we say that a system equilibrates if its state evolves toward some particular state (in general mixed) and remains in that state (or close to it) for almost all times. As far as equilibration is concerned, it is irrelevant what the equilibrium state actually is. In particular, it need not be a thermal state (such as a Boltzmannian distribution) and indeed may depend on the initial state of the subsystem and/or on the initial state of the rest of the system (i.e., bath) in an arbitrary way.

(2) Bath state independence: the equilibrium state of the system should not depend on the precise initial state of the bath. That is, when defining a bath, we should only specify its macroscopic parameters, e.g., its temperature; when the system reaches equilibrium, the equilibrium state should depend only on the temperature of the bath.

(3) Subsystem state independence: if the subsystem is small compared to the bath, the equilibrium state of the subsystem should be independent of its initial state.

(4) Boltzmann form of the equilibrium state: under certain additional conditions on the Hamiltonian (especially the interaction term) and on the initial state, the equilibrium state of the subsystem can be written in the familiar Boltzmannian form $\rho_S = \frac{1}{Z} \exp(-\frac{H_S}{k_B T})$.

Realizing that thermalization can be decomposed in this way has major consequences. First, it allows us to address each aspect separately. Second, and more important, it allows us to greatly expand the scope of our study. Indeed, we will consider equilibration as a general quantum phenomenon that may occur in situations other than those usually associated with thermalization. In particular we need not restrict ourselves to standard thermal baths (that are described by a given temperature or restricted energy range), weak or short range interactions between the system and the bath, Boltzmannian distributions, situations in which energy is an extensive quantity, etc. Furthermore, we can consider situations in which the systems do not reach equilibrium and prove results about the bath or subsystem independence properties of the time-averaged state.

In this paper, with very weak assumptions, we prove the first two elements above—equilibration and bath state independence. That is, with virtually full generality, we prove that reaching equilibrium is a universal property of quantum systems and that the equilibrium state does not depend on the precise details of the bath state. After introducing the setup and basic definitions in the next section, we go through the above program step by step.

II. SETUP AND DEFINITIONS

In this section, we describe our general setup and introduce the basic notations and definitions that will be used throughout the paper.

A. System

We consider a large quantum system described by a Hilbert space $\mathcal{H}$. We decompose this system into two parts, a small subsystem S and the rest of the system that we refer to as the bath B. Correspondingly, we decompose the total Hilbert space as $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B$, where $\mathcal{H}_S$ and $\mathcal{H}_B$ (of dimensions d_S and d_B) are the Hilbert spaces of the subsystem and bath, respectively. If either part is infinite dimensional, we introduce a high-energy cutoff to render its dimension finite (and eliminate interaction terms from the Hamiltonian that would take the state outside the allowed subspace). Note, however, that at this stage we do not ascribe the subsystem or bath any special properties. The subsystem S could be a single particle, a cluster of particles next to each other, or a number of

isolated particles spread throughout the bath. It may even be something far more abstract, not necessarily related to particles at all: any arbitrary decomposition of the Hilbert space into a tensor product ($\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B$) actually defines a subsystem and a bath.

B. Hamiltonian

The evolution of the total system is governed by a Hamiltonian

$$H = \sum_k E_k |E_k\rangle\langle E_k|, \tag{1}$$

where $|E_k\rangle$ is the energy eigenstate with energy E_k. Throughout this paper, we consider the Hamiltonian to be completely general, except for the following constraint: that it has *nondegenerate energy gaps*.

We say that a Hamiltonian has nondegenerate energy gaps if any nonzero difference of eigenenergies determines the two energy values involved; i.e., for any four eigenstates with energies E_k, E_ℓ, E_m, and E_n, $E_k - E_\ell = E_m - E_n$ implies $k = \ell$ and $m = n$, or $k = m$ and $\ell = n$. Note that this also implies that the energy levels are nondegenerate.

An important physical implication of this assumption is that the Hamiltonian is fully interactive, in the sense that no matter how we partition the system into a subsystem and bath they interact. Indeed any noninteracting Hamiltonian $H = H_S + H_B$ has multiple degenerate energy gaps. Note in particular that for noninteracting Hamiltonians the energy is a sum of the system energy and bath energy. Any four energies satisfying $E_1 = E_1^S + E_1^B$, $E_2 = E_1^S + E_2^B$, $E_3 = E_2^S + E_1^B$, and $E_4 = E_2^S + E_2^B$ will lead to a degenerate gap.

We emphasize that the restriction to Hamiltonians that have no degenerate energy gaps is an extremely natural and weak restriction. Indeed, adding an arbitrarily small random perturbation to any Hamiltonian will remove all degeneracies (although such changes may take a long time to significantly influence the evolution of a state, here we are not concerned with the time scales).

Note that except for the Hamiltonian having nondegenerate energy gaps, we allow it to be completely general. In particular, in a system composed of particles, it need not describe nearest-neighbor or bipartite interactions, but could contain interactions between all particles simultaneously. Hence energy need not be an (even approximately) extensive quantity, as normally considered in statistical mechanics.

C. Notation

We denote by $|\Psi(t)\rangle$ the global pure state of the system (i.e., of the subsystem and bath together) at time t. It is also convenient to write this state as a density matrix $\rho(t) = |\Psi(t)\rangle\langle\Psi(t)|$. The state of the subsystem $\rho_S(t)$ is obtained by tracing the global state $\rho(t)$ over the bath, i.e., $\rho_S(t) = \mathrm{Tr}_B \rho(t)$. Similarly the state of the bath at time t is described by $\rho_B(t) = \mathrm{Tr}_S \rho(t)$.

It is useful to define the time-averaged state of the system ω, which is given by

$$\omega = \langle \rho(t) \rangle_t = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau \rho(t) dt. \tag{2}$$

Similarly we define ω_S and ω_B as the time-averaged states of the system and bath, respectively.

It is also convenient to introduce the notion of the *effective dimension* of a (mixed) state ρ by

$$d^{\mathrm{eff}}(\rho) = \frac{1}{\mathrm{Tr}(\rho^2)}. \tag{3}$$

This tells us, in a certain sense, how many pure states contribute appreciably to the mixture. In particular a mixture of n orthogonal states with equal probability has effective dimension n. Unlike the support of the density matrix, this notion captures the probabilistic weight of different states in the mixture and is continuous.

It will also be important in what follows to consider the distance between two density matrices. There are many possible ways to define such a distance; here we use a strong and very natural distance, the trace distance $D(\rho_1, \rho_2) = \frac{1}{2} \mathrm{Tr} \sqrt{(\rho_1 - \rho_2)^2}$. The trace distance characterizes how hard it is to distinguish two states experimentally (even given perfect measurements). When it is small, the two states are effectively indistinguishable. More precisely, it is equal to the maximum difference in probability for any outcome of any measurement performed on the two states. Furthermore, the maximum difference in the expectation values of any operator A on the two states is the range of the eigenvalues of the operator times the trace-distance $(a_{max} - a_{min})D$.

Finally, by $\mathrm{Pr}_\Psi\{\cdot\}$ we denote the proportion of states in a Hilbert space having a particular property, according to the natural Haar measure.

III. EQUILIBRATION

We now come to the central result of our paper: every pure state of a large quantum system that is composed of a large number of energy eigenstates and which evolves under any arbitrary Hamiltonian (with nondegenerate energy gaps) is such that every small subsystem will equilibrate. That is, every small subsystem will evolve toward some particular state (in general mixed) and remain in that state (or close to it) for almost all times.

To understand the requirement that the state contains a large number of energy eigenstates we first note that this means that there will be a lot of change during the time evolution. Indeed, a single energy eigenstate does not change at all. (In this trivial sense, every energy eigenstate is already equilibrated. We however are looking for systems that do not start in equilibrium but evolve toward it.) But why do we need significant change during the time evolution? In order for equilibration to occur, we need that some information about the initial state of the subsystem leaves the subsystem. Indeed, suppose we start with a subsystem far from equilibrium. Then it will evolve through a number of distinct states on its path toward equilibrium. This implies that the state of the whole system (i.e., subsystem and bath) will also evolve through a number of distinct states. However, suppose now

that the subsystem reaches equilibrium. By definition, this means that its state does not change (significantly) anymore. Nevertheless, due to unitarity the state of the whole system cannot simply stop evolving, and it must continue to go through distinct states at the same rate as initially. In order for recurrences of the nonequilibrium state of the subsystem to occur very infrequently, the states of the whole system in which the subsystem is far from equilibrium must be only a small fraction of the total number of distinct states through which the whole system evolves. Hence the whole system must evolve through very many states. In order for this to be possible, the state of the total system must be composed of many energy eigenstates.

We have argued above that for a subsystem initially out of equilibrium to reach equilibrium, the state of the whole system must go through many distinct states. What we prove in the present paper is that this is also sufficient. That is: whenever the state of the whole system goes through many distinct states any small subsystem reaches equilibrium.

Another way of looking at the time evolution of the state is to investigate what happens to the state of the bath. As mentioned above, due to unitarity, the state of the whole system must continue evolving at the same rate even when the state of the subsystem is at equilibrium and does not change. The total state can then change in two ways: through changes in the correlations between the subsystem and the bath and through changes in the state of the bath itself. What we prove is: whenever the state of the bath goes through many distinct states, any small subsystem reaches equilibrium. Furthermore, we use our proof of this (and the earlier indented statement) to show that equilibration occurs in initial product states of the subsystem and bath, for almost all initial states of the bath.

The notion of evolving through many different distinct states is mathematically encapsulated by the effective dimension of the time-averaged state, $d^{\mathrm{eff}}(\omega)$ where $\omega = \langle \rho(t) \rangle_t$. The connection between this and the number of energy eigenstates is easily seen by expanding $|\Psi(t)\rangle$ (setting $\hbar = 1$ for convenience) as

$$|\Psi(t)\rangle = \sum_k c_k e^{-iE_k t} |E_k\rangle, \tag{4}$$

where $\Sigma_k |c_k|^2 = 1$, and hence

$$\rho(t) = \sum_{k,l} c_k c_l^* e^{-i(E_k - E_l)t} |E_k\rangle\langle E_l|. \tag{5}$$

From the nondegeneracy of the energy levels (implied by nondegenerate energy gaps),

$$\omega = \langle \rho(t) \rangle_t = \sum_k |c_k|^2 |E_k\rangle\langle E_k|, \tag{6}$$

giving

$$d^{\mathrm{eff}}(\omega) = \frac{1}{\mathrm{Tr}(\omega^2)} = \frac{1}{\sum_k |c_k|^4}. \tag{7}$$

Similarly, the notion of evolving through many different distinct states of the bath is mathematically encapsulated by the effective dimension of the time-averaged state of the bath, $d^{\text{eff}}(\omega_B)$ where $\omega_B = \langle \rho_B(t) \rangle_t$. Note that in general, we expect the time-evolved state to explore a significant portion of the bath state space, which is much larger than the subsystem's state space. Hence we anticipate that $d^{\text{eff}}(\omega_B)$ will be much larger than d_S.

We are now in the position to formulate our first theorem in rigorous mathematical terms. A quantity central to this result is $D[\rho_S(t), \omega_S]$, the distance between $\rho_S(t)$, the state of the subsystem at time t, and its time average, $\omega_S = \langle \rho_S(t) \rangle_t$. In general $\rho_S(t)$ fluctuates around its time average ω_S and the distance between them changes over time. To characterize these fluctuations we look at the time average of the distance $\langle D[\rho_S(t), \omega_S] \rangle_t$. When this average is small, the subsystem must spend almost all of its time very close to ω_S. In other words, when $\langle D[\rho_S(t), \omega_S] \rangle_t$ is small, the system equilibrates to ω_S.

Theorem 1. Consider any state $|\Psi(t)\rangle \in \mathcal{H}$ evolving under a Hamiltonian with nondegenerate energy gaps. Then the average distance between $\rho_S(t)$ and its time average ω_S is bounded by

$$\langle D[\rho_S(t), \omega_S] \rangle_t \leq \frac{1}{2} \sqrt{\frac{d_S}{d^{\text{eff}}(\omega_B)}} \leq \frac{1}{2} \sqrt{\frac{d_S^2}{d^{\text{eff}}(\omega)}}. \qquad (8)$$

The proof of this result is given in Appendix A.

By bounding $\langle D[\rho_S(t), \omega_S] \rangle_t$, our theorem tells us that the subsystem will equilibrate whenever the effective dimension explored by the bath $d^{\text{eff}}(\omega_B)$ is much larger than the subsystem dimension d_S, or whenever the effective dimension explored by the total state $d^{\text{eff}}(\omega)$ is much larger than two copies of the subsystem (dimension d_S^2). In other words, if $\langle D[\rho_S(t), \omega_S] \rangle_t$ is small, the system spends almost all its time close to the equilibrium state. Indeed, as distances are always positive, it is easy to see that the proportion of the time for which $D[\rho_S(t), \omega_S]$ is more than K times $\langle D[\rho_S(t), \omega_S] \rangle_t$ (with K an arbitrary positive constant) must be less than $1/K$. If we assume that the energy eigenvalues of H have no rational dependencies (which is much stronger than the nondegenerate energy gaps condition, but still one that holds for generic perturbations of the Hamiltonian) we can improve the bound on the proportion of time the subsystem spends away from equilibrium to one exponential in $d^{\text{eff}}(\omega)$ (see Appendix C, theorem 4).

We want to emphasize that this result about equilibration is completely general. That is, we did not assume anything special about the interaction (apart from not having degenerate energy gaps—which rules out only a set of Hamiltonians of measure zero), neither have we assumed any special properties of the bath. In particular, we did not assume that the bath is characterized by some temperature T; in fact we did not assume that the bath is in any kind of equilibrium at all. Furthermore, we did not make the rather standard assumption that the system has a limited spread of energies ΔE. Finally, we also did not make any assumptions about the form of the equilibrium state ω_S of the subsystem; in particular there is no need for this state to have any of the usual

thermal properties, e.g., to be of Boltzmannian form or similar. Indeed, since the bath is completely arbitrary, so is the equilibrium state of the system. In other words, the equilibration phenomenon that we describe is a general phenomenon and need not have any "thermal" aspects at all.

We also note that the bounds are completely independent from the energy eigenvalues. Indeed, the energy eigenvalues did not play any role in establishing the bounds—they just gave rise to phases which averaged to zero when taking the time average. This is a very important feature of our results. In fact, apart from the discussion in the second half of Sec. IV energy does not appear anymore. Energy obviously plays a central role in establishing the rate of time evolution and its exact details, but, as we have shown here, the long-time behavior is characterized by bounds that are independent of energy.

Furthermore, it is not only the energy eigenvalues that play no role in our bounds—the form of the energy eigenstates is also irrelevant. Indeed, the quantity $\frac{1}{2}\sqrt{\frac{d_S^2}{d^{\text{eff}}(\omega)}}$ which appears in our theorem is independent of the form of the energy eigenstates.

Finally, in the discussions above we referred to a specific partition of our total system into a "subsystem" and its corresponding "bath." However, we note that the greater upper bound in theorem 1 depends only on the dimension of the subsystem, not on what that particular subsystem is, and on the effective dimension of ω which is independent of the way we partitioned our total system. Hence when one particular subsystem equilibrates then *all* other subsystems of the same dimension equilibrate as well.

A. Equilibration of typical states

Our theorem described above puts a bound on the fluctuation of the state of a subsystem around the time average. The next question is to find the cases in which the fluctuations are small so that the subsystem equilibrates. As we will now show, almost all quantum states have this property. Intuitively the idea is the following.

Consider a Hilbert subspace $\mathcal{H}_R$ of the total Hilbert space of large dimension d_R. Consider now an arbitrary orthogonal basis in this subspace. First, since the dimension of the subspace is large, it is impossible to construct all basis vectors from only a few energy eigenstates. Thus, generically, each basis vector is a superposition of many energy eigenstates. Second a typical state $|\Psi\rangle$ from this subspace will have roughly equal overlap with each of the basis vectors. Hence a typical state $|\Psi\rangle$ will have a significant overlap with many energy eigenstates, making $d^{\text{eff}}(\omega)$ very large. Consequently, when we chose a typical state from a large dimensional subspace $\mathcal{H}_R$ we find that the state is such that all the small-dimension subsystems equilibrate.

More precisely, we can prove the following.

Theorem 2. (i) The average effective dimension $\langle d^{\text{eff}}(\omega) \rangle_\Psi$ where the average is computed over uniformly random pure states $|\Psi\rangle \in \mathcal{H}_R \subset \mathcal{H}$ is such that

$$\langle d^{\text{eff}}(\omega) \rangle_\Psi \geq \frac{d_R}{2}. \tag{9}$$

(ii) For a random state $|\Psi\rangle \in \mathcal{H}_R \subset \mathcal{H}$, the probability $\text{Pr}_\Psi\{d^{\text{eff}}(\omega) < \frac{d_R}{4}\}$ that $d^{\text{eff}}(\omega)$ is smaller than $\frac{d_R}{4}$ is exponentially small, namely,

$$\text{Pr}_\Psi\left\{d^{\text{eff}}(\omega) < \frac{d_R}{4}\right\} \leq 2 \exp(-c\sqrt{d_R}), \tag{10}$$

with a constant $c = \frac{(\ln 2)^2}{72\pi^3} \approx 10^{-4}$.

The proof of (i) is given in Appendix A, and the proof of (ii), which makes use of Levy's lemma, is given in Appendix B.

Point (i) essentially tells us that the average effective dimension is larger than half the dimension of the Hilbert subspace, so when we draw states from a subspace of large dimension, the effective dimension $d^{\text{eff}}(\omega)$ of a typical state is large. Point (ii) makes the result even sharper, telling us that the probability of having a small effective dimension is exponentially small. We now apply these results to two specific cases.

1. Equilibration of generic states

We first address the question of what happens to a generic state of the whole system. In other words, what happens to a state chosen at random from the total Hilbert space $\mathcal{H}$. The answer to this question can be found immediately from the above results by taking the subspace $\mathcal{H}_R$ to be the entire space $\mathcal{H}$, meaning that in Eqs. (9) and (10) we replace d_R by d. Hence, for systems living in high dimensional Hilbert spaces, from Eqs. (9) and (10) we expect that the effective dimension $d^{\text{eff}}(\omega)$ is of the order of d. Indeed, Eq. (10) tells us that the effective dimension will be smaller than $\frac{d}{4}$ only in an exponential small number of cases. Consequently, since $d = d_S d_B$, the factor $d_S^2 / d^{\text{eff}}(\omega)$ that governs the amount of fluctuations will be approximately equal to $d_S^2 / d = d_S / d_B$. In the case of systems composed of large numbers of particles, the dimension of the Hilbert spaces of the system and bath grow exponentially with their number of particles. Hence this ratio will drop off exponentially with the total number of particles whenever the number of particles in the subsystem is no more than a constant fraction of the number in the bath, and thus each such subsystem equilibrates.

2. Equilibration of systems far from equilibrium

In the previous subsection we have shown that a typical state $|\Psi\rangle$ is such that any small enough subsystem equilibrates. This however is not the end of the story. Indeed, one can raise the following legitimate question. Is it the case that states in which a subsystem is initially far from equilibrium will equilibrate? The point is that states in which a subsystem is far from equilibrium are quite rare in the Hilbert space so they are by no means "typical." Indeed, it can be shown that the overwhelming majority of the states in the total Hilbert space are such that every small subsystem is already in a canonical state [3]. In this section we focus specifically on

what happens when the initial state of the subsystem is far from equilibrium and show that it too equilibrates.

We now consider the usual situation in which the question of equilibration is discussed. There is a bath, consisting of a large number of particles, about which we know only some macroscopic parameters (such as temperature), into which we place a much smaller subsystem. The initial state of the subsystem is arbitrary, but uncorrelated with that of the bath. The question is, does the subsystem equilibrate?

In this section we prove that for any initial state of the subsystem, and almost all initial states of the bath, the subsystem equilibrates. Note that this includes cases in which the subsystem is initially far from equilibrium.

In our formalism, we describe the above situation as follows. We consider initial state of the full system (i.e., of the subsystem and bath) to be a product state $|\Psi\rangle_{SB} = |\psi\rangle_S |\phi\rangle_B$. The initial state of the system $|\psi\rangle_S$ is an arbitrary state in $\mathcal{H}_S$. We model the macroscopic parameters of the bath by restricting its initial state $|\phi\rangle_B$ to lie within a particular subspace $\mathcal{H}_B^R \subseteq \mathcal{H}_B$ (of dimension d_B^R). Note, however, that although the motivation for imposing a restriction on the bath's Hilbert space was to model macroscopic parameters, our results apply to any restricted Hilbert space $\mathcal{H}_B^R$. In particular it need not have any thermodynamic or macroscopic meaning. Furthermore, $\mathcal{H}_B^R$ is only a restriction at the initial time, and the state of the bath may evolve outside this space over time.

Given this setup, we can apply our second theorem, Eqs. (9) and (10) with $\mathcal{H}_R = |\psi\rangle_S \otimes \mathcal{H}_B^R$, and hence $d_R = d_B^R$. This gives $d^{\text{eff}}(\omega) \geq d_B^R / 4$ for almost all initial states of the bath and any initial state of the system. Hence given an initial product state, the subsystem will equilibrate for any initial state of the system and almost any initial state of the bath (within $\mathcal{H}_B^R$), as long as $d_B^R \gg d_S^2$.

The actual mechanism by which the subsystem equilibrates when it is put in contact with the bath is however highly nontrivial. Indeed, as we show below, generically the state of the bath evolves through many distinct states and does not reach equilibrium. Moreover it may move out of the initial restricted subspace $\mathcal{H}_B^R$. At the same time, surprisingly, the subsystem that is in contact with it does equilibrate. This is even more surprising given the fact that we did not make any assumptions as to the form of the interaction between the subsystem and the bath (apart from the Hamiltonian not having degenerate energy gaps). Thus, in principle, it could be the case that the evolution of the subsystem is sensitive to the precise state of the bath.

To show that the bath does not equilibrate generically, we will show that $d^{\text{eff}}(\omega_B)$ the effective dimension of the average state of the bath is much larger than $d^{\text{eff}}[\rho_B(t)]$, the effective dimension of the bath state at any particular time, which essentially means that the bath state continues to evolve and does not equilibrate to any particular state. Indeed, note that because the two systems are in a pure entangled state, $\text{rank}[\rho_B(t)] = \text{rank}[\rho_S(t)] \leq d_S$. Since the effective dimension of a state is always less than its rank, we obtain that

$$d^{\text{eff}}[\rho_B(t)] \leq d_S. \tag{11}$$

In contrast, Eq. (A6) implies that

$$d^{\text{eff}}(\omega_B) \geq d^{\text{eff}}(\omega)/d_S. \qquad (12)$$

Hence for a generic state which, as discussed in theorem 2 (ii) obeys $d^{\text{eff}}(\omega) > \frac{d_R}{4}$, we have

$$d^{\text{eff}}(\omega_B) \geq \frac{d^R}{4d_S} = \frac{d_B^R}{4d_S} \gg d_S \geq d^{\text{eff}}[\rho_B(t)]. \qquad (13)$$

IV. INITIAL STATE INDEPENDENCE

So far, we have addressed one issue associated with thermalization, namely, equilibration. In particular we have shown that a small subsystem will equilibrate for almost all states drawn from a large restricted subspace. We now address a second issue—how the equilibrium state reached by a subsystem depends on the initial state. So far, each initial state obeying the restriction could cause the subsystem to equilibrate to a different state. To represent this dependency we now explicitly denote the equilibrium state of the subsystem by ω_S^Ψ. However, when discussing thermalization, we would usually expect the equilibrium state to depend only on macroscopic parameters (such as the temperature of the bath), and not on the precise microscopic initial state. Our next result shows that this intuition carries over to our more general setting, by proving that almost all states within a large restricted subspace lead to the same equilibrium state of a subsystem.

Theorem 3. (i) Almost all initial states chosen from a large restricted subspace will yield the same equilibrium state of a small subsystem. In particular, with $\langle \cdot \rangle_\Psi$ referring to the average over uniformly random pure states $|\Psi(0)\rangle \in \mathcal{H}_R \subset \mathcal{H}_S \otimes \mathcal{H}_B$, and $\Omega_S = \langle \omega_S^\Psi \rangle_\Psi$:

$$\langle D(\omega_S^\Psi, \Omega_S) \rangle_\Psi \leq \sqrt{\frac{d_S \delta}{4d_R}} \leq \sqrt{\frac{d_S}{4d_R}}.$$

The first inequality gives a tighter but more complicated bound, with

$$\delta = \sum_k \langle E_k | \frac{\Pi_R}{d_R} | E_k \rangle \text{Tr}_S[\text{Tr}_B(|E_k\rangle\langle E_k|)]^2 \leq 1, \qquad (14)$$

where Π_R is the projector onto $\mathcal{H}_R$.

(ii) For a random state $|\Psi\rangle \in \mathcal{H}_R \subset \mathcal{H}$, the probability that $D(\omega_S^\Psi, \Omega_S) > \frac{1}{2}\sqrt{\frac{d_S \delta}{d_R}} + \epsilon$ drops off exponentially with $\epsilon^2 d_R$:

$$\text{Pr}_\Psi \left\{ D(\omega_S^\Psi, \Omega_S) > \frac{1}{2}\sqrt{\frac{d_S \delta}{d_R}} + \epsilon \right\} \leq 2\exp(-c'\epsilon^2 d_R), \qquad (15)$$

with a constant $c' = \frac{2}{9\pi^3}$. Note that setting $\epsilon = d_R^{-1/3}$ yields a small average distance with very high probability when $d_R \gg d_S$.

The proof is given in Appendix A [part (i)] and Appendix B [part (ii)]. The quantity δ reflects the average purity of the energy eigenstates on the subsystem (with probability weights $\langle E_k | \frac{\Pi_R}{d_R} | E_k \rangle$ determined by their relevance to $\mathcal{H}_R$). The quantity δ lies in the range $(\frac{1}{d_S}, 1)$, with highly entangled energy eigenstates yielding smaller values.

A. Bath state independence

We now apply the result above to the situation described in the previous section, in which the subsystem and bath are initially in the product state $|\Psi\rangle_{SB} = |\psi\rangle_S |\phi\rangle_B$ in the restricted subspace $\mathcal{H}_R = |\psi\rangle \otimes \mathcal{H}_R^B$. Previously, we have shown that generic states of the bath cause the subsystem to equilibrate. However, the particular equilibrium state ω_S^Ψ of the subsystem could in principle depend on the precise initial state of the bath $|\phi\rangle_B$. Here we show that this is not the case, and that almost all states of the bath in $\mathcal{H}_R^B$ lead to the same equilibrium state of the subsystem.

The proof follows simply by applying theorem 3 with $\mathcal{H}_R = |\psi\rangle \otimes \mathcal{H}_R^B$ and hence $d_R = d_R^B$. From the weaker bound (not involving δ) we see that almost all initial states of the bath yield approximately the same equilibrium state ω_S of the subsystem, as long as $d_R^B \gg d_S$. Note that since we do not use δ, this result does not depend on any properties of the energy eigenstates.

B. Subsystem state independence

We now turn to the third aspect of thermalization—subsystem state independence. A cup of coffee left untouched in a closed room eventually reaches room temperature regardless of its own initial temperature. In other words, a subsystem in contact with a bath reaches an equilibrium state that only depends on the bath but not on its particular initial state. Could we derive this behavior in our formalism? The issue of subsystem state independence turns out to be more complicated than the issues of equilibration and bath independence and we have not been able to completely solve this problem; however, we present here a number of preliminary results and observations that we believe are crucial in understanding this question.

There are many reasons why the issue of subsystem state independence is complicated. The first reason is that in fact it is not always the case that the equilibrium state is independent of the initial state of the subsystem. For a drastic example, consider the subsystem to be an atomic bomb. Then, depending on whether the bomb is functional or not, the room temperature itself will be modified, and therefore also the final equilibrium state. In other words, although the subsystem appears to be small, it may actually have a big impact on the bath, in which case the equilibrium state depends on the initial state of the subsystem. Here it seems that not only the dimension of the Hilbert spaces of the subsystem and bath are relevant (as in the previous sections), but the value of energy itself. The energy however turns out not to be the end of the story, and perhaps not even the key aspect, as we show below.

To start with, one can easily give an explicit example of a situation in which the equilibrium state of a subsystem depends on its initial state regardless on how big the bath is, or on the energy scales involved. In all our previous discussions the Hamiltonian was taken to be very general. The only condition we imposed on it was that it has no degenerate energy gaps. This condition is sufficient to ensure that the subsystem interacts with the bath (i.e., $H \neq H_S + H_B$).

However, this condition is not strong enough to ensure that there are no conserved quantities of the subsystem. When there are such quantities, then clearly initial states of the subsystem with different values of these quantities cannot equilibrate to the same state. An example would be a Hamiltonian of the form

$$H = \sum_{nm} E_{nm} |n\rangle\langle n|_S \otimes |m\rangle\langle m|_B. \quad (16)$$

As long as the energy eigenvalues E_{nm} have no degenerate energy gaps (which could be achieved by choosing them independently at random from some range), this is an interacting Hamiltonian. However, any operator $A = \sum_n a_n |n\rangle\langle n|_S$ on the subsystem will commute with H, and therefore is conserved. Evolution under this Hamiltonian will dephase the state of the subsystem in the $|n\rangle$ basis, but cannot flip an $|n\rangle$ into an $|n'\rangle$.

It is tempting to think that we could rule out such cases by only considering Hamiltonians that have no conserved quantities on the subsystem. However, this is still not a sufficient condition, as illustrated by the following counterexample (in which the system is a single spin and the bath is composed of many spins):

$$H = E\sigma_S^z + H_{\text{int}} + H_B, \quad (17)$$

where the eigenvalues of H_{int} and H_B lie between -1 and 1, and $E \gg 1$. It is easy to choose an H_{int} such that there are no conserved quantities on the subsystem. In such cases, one might imagine that in the long-time limit it would be possible for all initial states of the subsystem to equilibrate to the same state. However, it is easy to see that is not the case. Consider two initial states $|\psi_+\rangle_S|\phi\rangle_B$ and $|\psi_-\rangle_S|\phi\rangle_B$ where $\sigma_z|\psi_\pm\rangle = \pm|\psi_\pm\rangle$. The difference in expected total energy between these two states is close to $2E$ (the contribution of the terms H_{int} and H_B to the expected energy of any state lies between $+2$ and -2) and remains constant in time. If the subsystem were to equilibrate to the same state in both cases, this difference in energy would have to be accounted for by H_{int} and H_B. However these terms can generate an energy difference of at most 4. Hence such an equilibration process for the subsystem is impossible.

The above situation is indeed an abstract version of the "atomic bomb" example, in the sense that the energy of the subsystem completely dominates that of the bath, despite the fact that the dimension of the Hilbert space of the subsystem is much smaller than that of the bath. It is important to note that a similar argument can be made for any globally conserved quantity in which the total value is dominated by the contribution of the subsystem. Hence the issue is not one of energy. In fact we believe that the only special role of the energy (more precisely, of the energy eigenvalues) is that of fixing the time scale of the evolution. Indeed, in all our previous arguments the energy eigenvalues drop out of all the calculations (as long as there are no degenerate energy gaps). Instead, we note that the atomic bomb example in its abstract form of the Hamiltonian in Eq. (17) is in fact an approximation of the trivially nonthermalizing Hamiltonian (16), as it

can be seen quite easily that for large E its eigenstates are arbitrarily close to product states.

The above discussion motivates us to consider the case in which the energy eigenstates are far from product. In this situation we prove that almost all initial states of the subsystem lead to the same time-averaged state. Hence when such states do equilibrate (i.e., spend almost all times close to their time average) they reach the same equilibrium state.

We emphasize that as we mentioned in the introduction, however, the issues of equilibration and subsystem independence are in fact totally separated. Indeed, the question we address here is that of subsystem independence of the *time-averaged* state of the subsystem, regardless on the fact whether or not the system has only small fluctuations around this state, in which case the time-averaged state is an equilibrium state, or not.

Consider the following situation: let the initial state of the system be a product state $|\Psi\rangle = |\psi\rangle_S|\phi\rangle_B$, where now, in contrast to the previous section, $|\phi\rangle$ is fixed but $|\psi\rangle \in \mathcal{H}_S$ is generic; i.e., $|\Psi\rangle$ comes from a subspace $\mathcal{H}_R = \mathcal{H}_S \otimes |\phi\rangle_B$. Again we apply theorem 3. However, we now note that the weaker bound (not involving δ) is essentially useless, as $d_R = d_S$. Indeed, this is to be expected: the weaker bound does not take into account any assumptions about the energy eigenstates and we have argued above that there are cases in which the time-averaged state of the subsystem depends strongly on its initial state.

This leaves us with the stronger bound involving δ. Substituting $d_R = d_S$, we see that the time-averaged state of the subsystem will be the same for almost all initial states as long as $\delta \ll 1$. However, this is precisely the case we are interested in, in which the relevant energy eigenstates are far from product. When the energy eigenstates are highly entangled, in the sense that $\delta \ll 1$, the equilibrium state of the subsystem will therefore be approximately equal for almost all initial system states.

V. CONCLUSIONS

Understanding the basic laws of statistical mechanics from first principles is a holy grail of the subject. Here we approach this question from a fundamentally different viewpoint than usual, in which the whole system is described by a pure quantum state and the probabilistic behavior of a subsystem is an objective phenomenon, due to quantum entanglement, rather than the result of subjective ignorance. We proved a general quantum result: with almost full generality all interacting large quantum systems evolve in such a way that any small subsystem equilibrates, that is, spends almost all time extremely close to a particular state. The only conditions we require are that the Hamiltonian has no degenerate energy gaps (which rules out noninteracting Hamiltonians) and that the state of the whole system contains sufficiently many energy eigenstates.

Virtually all physical situations satisfy these requirements. First, all but a measure zero set of Hamiltonians have nondegenerate energy gaps. Indeed, an infinitesimally small random perturbation will lift any such degeneracy. Second, the vast majority of states in the Hilbert space are such that they

contain (i.e., have significant overlap with) very many energy eigenstates, as required in our proof. In particular, this covers the physically interesting situation in which a subsystem, initially out of equilibrium, interacts with a large bath. In this situation, we have proved that for every state of the subsystem and almost every state of the bath, the subsystem equilibrates. Furthermore, we have also proved that the equilibrium state of the subsystem is independent of the specific initial state of the bath and only depends on its macroscopic parameters.

We would like to emphasize that the above two restrictions are the only conditions we require. We do not require the assumptions usually made in statistical mechanics. For example the interactions could be strong and long range and energy need not be an extensive quantity. Also the state of the bath could be a superposition of a very large range of energy eigenstates and hence have no well defined temperature. On the other hand, we do not currently address the important issue of the time scale on which this equilibration occurs, which could be extremely long. The specific numerical and analytic work in [8–11] may help in assessing such time scales.

Although our original motivation was to study the phenomenon of thermalization, we have found a much larger range of phenomena: we discovered that reaching equilibrium is an almost universal behavior of large quantum systems. Ordinary thermalization is just a particular example of this behavior; it occurs in specific situations and it has a number of additional characteristics that we did not address in the present paper. For example, the equilibrium state of a subsystem that is put into contact with a thermal bath is largely independent of the initial state of the subsystem, and the equilibrium state has a particular form (according to Gibbs and Boltzmann). Finally, we emphasize that we did not address the time scale for equilibration. All these are open questions for future work.

Note added. Very recently we became aware of an independent work by Reimann [12], which takes first steps toward solving the thermalization problem along similar lines as in the present paper.

ACKNOWLEDGMENTS

APPENDIX A: PROOFS OF EXPECTATION VALUES

In this appendix, we derive several expectation values that are used in the main paper.

Proof of theorem 1. We first relate the trace distance to a less natural but more mathematically tractable distance measure (the square of the Hilbert-Schmidt distance) using a standard bound [13]

$$D(\rho_1, \rho_2) = \frac{1}{2} \mathrm{Tr}_S \sqrt{(\rho_1 - \rho_2)^2} \leq \frac{1}{2} \sqrt{d_S \, \mathrm{Tr}_S(\rho_1 - \rho_2)^2}. \tag{A1}$$

Using the concavity of the square-root function, we therefore have

$$\langle D[\rho_S(t), \omega_S] \rangle_t \leq \sqrt{d_S \langle \mathrm{Tr}_S[\rho_S(t) - \omega_S]^2 \rangle_t}. \tag{A2}$$

Using the expansions in terms of the energy eigenstates given in Eqs. (5) and (6),

$$\rho_S(t) - \omega_S = \sum_{k \neq l} c_k c_l^* e^{-i(E_k - E_l)t} \, \mathrm{Tr}_B |E_k\rangle\langle E_l|, \tag{A3}$$

where $\sum_{k \neq l}$ is shorthand for a sum over k and l omitting all terms in which $k = l$. Hence,

$$\langle \mathrm{Tr}_S[\rho_S(t) - \omega_S]^2 \rangle_t = \sum_{k \neq l} \sum_{m \neq n} \mathcal{T}_{klmn} \, \mathrm{Tr}_S(\mathrm{Tr}_B |E_k\rangle$$
$$\times \langle E_l| \mathrm{Tr}_B |E_m\rangle\langle E_n|),$$

where

$$\mathcal{T}_{klmn} = c_k c_l^* c_m c_n^* \langle e^{-i(E_k - E_l + E_m - E_n)t} \rangle_t. \tag{A4}$$

Evaluating this time average, using the fact that the Hamiltonian has nondegenerate energy gaps, and that the sums only include those terms where $k \neq l$ and $m \neq n$, we find that the only nonzero terms are those where $k = n$ and $l = m$, giving

$$\langle \mathrm{Tr}_S[\rho_S(t) - \omega_S]^2 \rangle_t$$
$$= \sum_{k \neq l} |c_k|^2 |c_l|^2 \mathrm{Tr}_S(\mathrm{Tr}_B |E_k\rangle\langle E_l| \mathrm{Tr}_B |E_l\rangle\langle E_k|)$$
$$= \sum_{k \neq l} |c_k|^2 |c_l|^2 \sum_{ss'bb'} \langle sb|E_k\rangle\langle E_l|s'b\rangle\langle s'b'|E_l\rangle\langle E_k|sb'\rangle$$
$$= \sum_{k \neq l} |c_k|^2 |c_l|^2 \sum_{ss'bb'} \langle sb|E_k\rangle\langle E_k|sb'\rangle\langle s'b'|E_l\rangle\langle E_l|s'b\rangle$$
$$= \sum_{k \neq l} |c_k|^2 |c_l|^2 \mathrm{Tr}_B(\mathrm{Tr}_S|E_k\rangle\langle E_k| \mathrm{Tr}_S|E_l\rangle\langle E_l|)$$
$$= \sum_{k \neq l} \mathrm{Tr}_B[\mathrm{Tr}_S(|c_k|^2 |E_k\rangle\langle E_k|) \mathrm{Tr}_S(|c_l|^2 |E_l\rangle\langle E_l|)]$$
$$= \mathrm{Tr}_B \, \omega_B^2 - \sum_k |c_k|^4 \mathrm{Tr}_S[(\mathrm{Tr}_B|E_k\rangle\langle E_k|)^2]$$
$$\leq \mathrm{Tr}_B \, \omega_B^2, \tag{A5}$$

where $\omega_B = \mathrm{Tr}_S \omega$. To obtain a further bound, we invoke weak subadditivity of the Rényi entropy (see, e.g., [14]),

$$\mathrm{Tr}(\omega^2) \geq \frac{\mathrm{Tr}_B(\omega_B^2)}{\mathrm{rank}(\rho_S)} \geq \frac{\mathrm{Tr}_B(\omega_B^2)}{d_S}, \tag{A6}$$

and therefore

$$\langle D[\rho_S(t), \omega_S] \rangle_t \le \frac{1}{2} \sqrt{d_S \, \mathrm{Tr}_B(\omega_B^2)} \le \frac{1}{2} \sqrt{d_S^2 \, \mathrm{Tr}(\omega^2)}. \quad (A7)$$

Using the definition of the effective dimension (3), these are the bounds as stated in the theorem. $\square$

In order to prove theorems 2(i) and 3(i), we require a couple of additional mathematical tools, which are also used in [3]. The first is the simple and very helpful identity

$$\mathrm{Tr}(AB) = \mathrm{Tr}[(A \otimes B)\mathbb{S}], \quad (A8)$$

where $\mathbb{S}$ is the SWAP operator of the two systems. This equation is easily proved by expanding it in a basis. The second result we employ is that $\langle |\Psi\rangle\langle\Psi| \otimes |\Psi\rangle\langle\Psi| \rangle_\Psi$ is proportional to the projector onto the symmetric subspace. This is actually the simplest instance of the representation theory of the unitary group. It implies the following lemma:

Lemma. For states $|\Psi\rangle \in \mathcal{H}_R \subseteq \mathcal{H}$,

$$\langle |\Psi\rangle\langle\Psi| \otimes |\Psi\rangle\langle\Psi| \rangle_\Psi = \frac{\Pi_{RR}(\mathbb{1} + \mathbb{S})}{d_R(d_R + 1)}, \quad (A9)$$

where $\mathbb{1}$ is the identity operator on $\mathcal{H} \otimes \mathcal{H}$, and $\Pi_{RR} = \Pi_R \otimes \Pi_R$ is the projector onto $\mathcal{H}_R \otimes \mathcal{H}_R$. $\square$

For conciseness, in the proofs of theorems 2(i) and 3(i) we abbreviate energy eigenstates by their indices, such that $|E_k\rangle \equiv |k\rangle$, etc. Similarly, we write $|E_k\rangle \otimes |E_l\rangle$ as $|kl\rangle$, etc. It is also helpful to note that time averaging a state corresponds to dephasing it in the energy eigenbasis (due to the nondegeneracy of energy levels). We denote this dephasing map by $\$[\rho] := \sum_k |k\rangle\langle k|\rho|k\rangle\langle k|$, such that $\omega = \langle|\psi\rangle\langle\psi|\rangle_t = \$[|\psi\rangle\langle\psi|]$. With these tools and notation in hand, we now prove the results needed for the main paper.

Proof of theorem 2(i). We first prove a bound on the expected purity of ω,

$$\langle \mathrm{Tr}(\omega)^2 \rangle_\Psi = \langle \mathrm{Tr}[(\omega \otimes \omega)\mathbb{S}] \rangle_\Psi$$

$$= \mathrm{Tr}[\$ \otimes \$(\langle |\Psi\rangle\langle\Psi| \otimes |\Psi\rangle\langle\Psi| \rangle_\Psi)\mathbb{S}]$$

$$= \mathrm{Tr}\left\{ \$ \otimes \$\left[\frac{\Pi_{RR}(\mathbb{1} + \mathbb{S})}{d_R(d_R + 1)} \right]\mathbb{S} \right\}$$

$$= \sum_{kl} \mathrm{Tr}\left\{ |kl\rangle\langle kl| \left[\frac{\Pi_{RR}(\mathbb{1} + \mathbb{S})}{d_R(d_R + 1)} \right] |kl\rangle\langle kl|\mathbb{S} \right\}$$

$$= \sum_{kl} \mathrm{Tr}(|kl\rangle\langle lk|) \left[\frac{\langle kl|\Pi_{RR}(|kl\rangle + |lk\rangle)}{d_R(d_R + 1)} \right]$$

$$= \sum_k \frac{2\langle kk|\Pi_{RR}|kk\rangle}{d_R(d_R + 1)} \le \sum_k \frac{2\langle k|\Pi_R|k\rangle}{d_R(d_R + 1)} < \frac{2}{d_R}.$$

$$(A10)$$

It then follows straightforwardly that

$$\langle d^{\mathrm{eff}}(\omega) \rangle_\Psi = \left\langle \frac{1}{\mathrm{Tr}(\omega^2)} \right\rangle_\Psi \ge \frac{1}{\langle \mathrm{Tr}(\omega^2) \rangle_\Psi} > \frac{d_R}{2}, \quad (A11)$$

concluding the proof.

Proof of theorem 3(i). As in the proof of theorem 1, we first relate the trace distance to a less natural, but more mathematically tractable distance measure [13],

$$\langle D(\omega_S, \langle\omega_S\rangle_\Psi) \rangle_\Psi \le \left\langle \frac{1}{2} \sqrt{d_S \, \mathrm{Tr}(\omega_S - \langle\omega_S\rangle_\Psi)^2} \right\rangle_\Psi$$

$$\le \frac{1}{2} \sqrt{d_S \langle \mathrm{Tr}(\omega_S - \langle\omega_S\rangle_\Psi)^2 \rangle_\Psi}. \quad (A12)$$

We then prove bounds on the averaged term

$$\langle \mathrm{Tr}(\omega_S - \langle\omega_S\rangle_\Psi)^2 \rangle_\Psi = \langle \mathrm{Tr}(\omega_S^2) \rangle_\Psi - \mathrm{Tr}_S(\langle\omega_S\rangle_\Psi^2)$$

$$= \mathrm{Tr}[(\langle\omega_S \otimes \omega_S\rangle_\Psi - \langle\omega_S\rangle_\Psi \otimes \langle\omega_S\rangle_\Psi)\mathbb{S}]$$

$$= \mathrm{Tr}_{SS}\left\{ \mathrm{Tr}_{BB}\left[\$ \otimes \$\left(\langle |\Psi\rangle\langle\Psi| \otimes |\Psi\rangle\langle\Psi| \rangle_\Psi - \frac{\Pi_R}{d_R} \otimes \frac{\Pi_R}{d_R} \right) \right]\mathbb{S} \right\}$$

$$= \mathrm{Tr}_{SS}\left(\mathrm{Tr}_{BB}\left\{ \$ \otimes \$\left[\frac{\Pi_{RR}(\mathbb{1} + \mathbb{S})}{d_R(d_R + 1)} - \frac{\Pi_{RR}}{d_R^2} \right] \right\}\mathbb{S} \right) \le \mathrm{Tr}_{SS}\left\{ \mathrm{Tr}_{BB}\left[\$ \otimes \$\left(\frac{\Pi_{RR}\mathbb{S}}{d_R^2} \right) \right]\mathbb{S} \right\}$$

$$= \sum_{kl} \mathrm{Tr}_{SS}\left[\mathrm{Tr}_{BB}\left(|kl\rangle\langle kl| \frac{\Pi_{RR}}{d_R^2} |lk\rangle\langle kl| \right)\mathbb{S} \right]$$

$$= \sum_{kl} \frac{\langle kl|\Pi_{RR}|lk\rangle}{d_R^2} \mathrm{Tr}_S[\mathrm{Tr}_B(|k\rangle\langle k|)\mathrm{Tr}_B(|l\rangle\langle l|)] \le \sum_{kl} \frac{\langle k|\Pi_R|l\rangle\langle l|\Pi_R|k\rangle}{d_R^2} \mathrm{Tr}_S\left[\frac{(\mathrm{Tr}_B|k\rangle\langle k|)^2 + (\mathrm{Tr}_B|l\rangle\langle l|)^2}{2} \right]$$

$$= \frac{1}{d_R} \sum_k \langle k|\frac{\Pi_R}{d_R}|k\rangle \mathrm{Tr}_S[(\mathrm{Tr}_B|k\rangle\langle k|)^2] \le \frac{1}{d_R} \sum_k \langle k|\frac{\Pi_R}{d_R}|k\rangle$$

$$= \frac{1}{d_R}. \quad (A13)$$

Note that in the second inequality (sixth line) we have used the fact that $\mathrm{Tr}_S(\mathrm{Tr}_B|k\rangle\langle k| - \mathrm{Tr}_B|l\rangle\langle l|)^2 \geq 0$ (it is the trace of a positive operator).

Inserting these results into Eq. (A12), specifically, the bounds given by the third line from bottom and the last line, we obtain the inequalities stated in the theorem. □

APPENDIX B: USES OF LEVY'S LEMMA

In this appendix we prove theorems 2(ii) and 3(ii), in which exponential bounds are placed on the proportion of states which have values of $d^{\mathrm{eff}}(\omega)$ and $D(\omega_S^\Psi, \Omega_S)$ far from the average value, using Levy's lemma for measure concentration on a hypersphere.

Levy's lemma [15]. Let $f: S^{D-1} \to \mathbb{R}$ be a real-valued function on the $(D-1)$-dimensional Euclidean sphere (which we think of as embedded into D-dimensional Euclidean space), with Lipschitz constant $\lambda = \sup_{x_1, x_2} |f(x_1) - f(x_2)|/|x_1 - x_2|_2$. Then, for a uniformly random point $X \in S^{D-1}$,

$$\mathrm{Pr}_X\{f(X) > \langle f \rangle + \epsilon\} \leq 2\exp\left(-\frac{D\epsilon^2}{9\pi^3\lambda^2}\right). \quad \text{(B1)}$$

□

Note that the pure quantum states $|\Psi\rangle \in \mathcal{H}_R$ can be thought of as lying on a $(2d_R - 1)$-dimensional hypersphere, with coordinates given by the real and imaginary components of the state's overlap with an orthonormal basis $|\hat{n}\rangle$ of $\mathcal{H}_R$:

$$x_{2n-1}(\Psi) = \mathrm{Re}[\langle\hat{n}|\Psi\rangle], \quad x_{2n}(\Psi) = \mathrm{Im}[\langle\hat{n}|\Psi\rangle]. \quad \text{(B2)}$$

In this coordinate system Euclidean and Hilbert space norm coincide: $|\vec{x}(\Psi_1) - \vec{x}(\Psi_2)|_2 = \||\Psi_1\rangle - |\Psi_2\rangle\|_2$.

Because it involves the simplest application of Levy's lemma, we begin with the proof relating to initial state independence.

Proof of theorem 3(ii). To prove that almost all initial states $|\Psi\rangle$ yield the same equilibrium state, we apply Levy's lemma directly to the function $f(\Psi) \equiv f[\vec{x}(\Psi)] = D(\omega_S^\Psi, \Omega_S)$ on the $(2d_R - 1)$-dimensional hypersphere of quantum states. Using an analog of Eq. (C6) to give

$$|D(\omega_S^{\Psi_1}, \Omega_S) - D(\omega_S^{\Psi_2}, \Omega_S)| \leq \||\Psi_1\rangle - |\Psi_2\rangle\|_2, \quad \text{(B3)}$$

we find that the Lipschitz constant of the function satisfies $\lambda \leq 1$. Substituting this into Levy's lemma, as well as the average value obtained in theorem 3(i), we obtain the desired result. □

Proof of theorem 2(ii). Unfortunately we cannot prove the result in theorem 2(ii) directly by applying Levy's lemma to $d^{\mathrm{eff}}(\omega)$. Instead, we apply it to the function

$$f(\Psi) \equiv f[\vec{x}(\Psi)] = \ln\{\mathrm{Tr}[\$(|\Psi\rangle\langle\Psi|)^2]\}, \quad \text{(B4)}$$

where the superoperator $\$$ acts on the subspace $\mathcal{H}_T \subseteq \mathcal{H}$ spanned by energy eigenstates with nonzero projection onto $\mathcal{H}_R$ (i.e., states $|k\rangle$ satisfying $\langle k|\Pi_R|k\rangle \neq 0$). The subspace $\mathcal{H}_T$ contains all states which could arise during the evolution of an initial state in $\mathcal{H}_R$, and $\$$ maps these states back into $\mathcal{H}_R$ according to

$$\$[\rho] = \sum_k |\tilde{k}\rangle\langle k|\rho|k\rangle\langle\tilde{k}| \quad \text{and} \quad |\tilde{k}\rangle = \frac{1}{\sqrt{\langle k|\Pi_R|k\rangle}}\Pi_R|k\rangle. \quad \text{(B5)}$$

Note that when the Hamiltonian commutes with Π_R, $\$$ is identical to the normal $\$$ on $\mathcal{H}_T$. Computing the average value of our function we find

$$\langle\ln\{\mathrm{Tr}[\$(|\Psi\rangle\langle\Psi|)^2]\}\rangle_\Psi$$

$$\leq \ln\langle\mathrm{Tr}[\$(|\Psi\rangle\langle\Psi|)^2]\rangle_\Psi$$

$$= \ln\mathrm{Tr}\{\$ \otimes \$(\langle|\Psi\rangle\langle\Psi| \otimes |\Psi\rangle\langle\Psi|\rangle_\Psi)\mathbb{S}\}$$

$$= \ln\left(\mathrm{Tr}\left\{\$ \otimes \$\left[\frac{\Pi_{RR}(\mathbb{1} + \mathbb{S})}{d_R(d_R+1)}\right]\mathbb{S}\right\}\right)$$

$$= \ln\left(\sum_{kl}\mathrm{Tr}\left\{|\widetilde{kl}\rangle\langle kl|\left[\frac{\Pi_{RR}(\mathbb{1} + \mathbb{S})}{d_R(d_R+1)}\right]|kl\rangle\langle\widetilde{kl}|\right\}\right)$$

$$= \ln\left\{\sum_{kl}\langle\widetilde{lk}|\widetilde{kl}\rangle\left[\frac{\langle kl|\Pi_{RR}(|kl\rangle + |lk\rangle)}{d_R(d_R+1)}\right]\right\}$$

$$\leq \ln\left\{\sum_{kl}\langle\widetilde{lk}|\widetilde{kl}\rangle\left[\frac{\langle kl|\Pi_{RR}(|kl\rangle + |lk\rangle)}{d_R(d_R+1)}\right]\right\}$$

$$\leq \ln\left[\frac{2}{d_R(d_R+1)}\sum_{kl}\langle lk|\Pi_{RR}|kl\rangle\right]$$

$$= \ln\left[\frac{2}{d_R(d_R+1)}\sum_k\langle k|\Pi_R|k\rangle\right]$$

$$< \ln\left(\frac{2}{d_R}\right). \quad \text{(B6)}$$

To bound the Lipschitz constant of the function $f(\Psi)$, we actually employ a further function

$$g(\Psi) = \ln\mathrm{Tr}\left[\left(\sum_n |\hat{n}\rangle\langle\hat{n}|\$[|\Psi\rangle\langle\Psi|]|\hat{n}\rangle\langle\hat{n}|\right)^2\right], \quad \text{(B7)}$$

where $|\hat{n}\rangle$ is an orthonormal basis of $\mathcal{H}_R$. Writing

$$t_{nk0} = \mathrm{Re}[\langle\hat{n}|\tilde{k}\rangle\langle k|\Psi\rangle] \quad \text{and} \quad t_{nk1} = \mathrm{Im}[\langle\hat{n}|\tilde{k}\rangle\langle k|\Psi\rangle], \quad \text{(B8)}$$

it follows that

$$g(\Psi) = \ln\mathrm{Tr}\left[\left(\sum_{nkz} t_{nkz}^2|\hat{n}\rangle\langle\hat{n}|\right)^2\right] = \ln\sum_n\left(\sum_{kz} t_{nkz}^2\right)^2. \quad \text{(B9)}$$

To bound the Lipschitz constant of g, it is sufficient to find an upper bound on its gradient,

$$\frac{\partial g}{\partial t_{nkz}} = \frac{1}{\sum_{n'}\left(\sum_{k'z'} t_{n'k'z'}^2\right)^2}.2.2.t_{nkz}\sum_{k'z'} t_{nk'z'}^2. \quad \text{(B10)}$$

Introducing the notation $p_n = \sum_{kz} t_{nkz}^2$, and noting that $\sum_n p_n = 1$, we find

$$|\nabla g|^2 = \sum_{nkz} \left(\frac{\partial g}{\partial t_{nkz}} \right)^2$$

$$= \frac{16 \sum_n p_n^3}{\left(\sum_n p_n^2 \right)^2} \le \frac{16 \left(\sum_n p_n^2 \right)^{3/2}}{\left(\sum_n p_n^2 \right)^2}$$

$$= \frac{16}{\left(\sum_n p_n^2 \right)^{1/2}} \le 16 \sqrt{d_R}, \qquad (B11)$$

and hence that the Lipschitz constant of g is at most $4\sqrt[4]{d_A}$.

To obtain the Lipschitz constant of f, we note that $g(\Psi) \ge f(\Psi)$, with equality if $\{|\hat{n}\rangle\}$ is an eigenbasis of $\text{Tr}_B|\Psi\rangle\langle\Psi|$. Now, for any two vectors, we may without loss of generality assume that $f(\Psi_1) \ge f(\Psi_2)$, and take $\{|\hat{n}\rangle\}$ to be the eigenbasis of $\text{Tr}_B|\Psi_2\rangle\langle\Psi_2|$. Thus,

$$f(\Psi_1) - f(\Psi_2) \le g(\Psi_1) - g(\Psi_2)$$

$$\le 4\sqrt[4]{d_A} ||\Psi_1\rangle - |\Psi_2\rangle|_2, \qquad (B12)$$

and hence the Lipschitz constant of f is also upper bounded by $4\sqrt[4]{d_A}$.

Applying Levy's lemma to $f(\Psi)$, and using the observation that $\Pr\{x > a\} \le b$ and $x \ge y$ implies $\Pr\{y > a\} \le b$ to substitute the bound on $\langle f(\Psi) \rangle_\Psi$ obtained above, and the further bound

$$\ln\{\text{Tr}[\mathcal{S}(|\Psi\rangle\langle\Psi|)^2]\} \ge \ln\{\text{Tr}[\mathcal{S}(|\Psi\rangle\langle\Psi|)^2]\} \qquad (B13)$$

gives

$$\Pr_\Psi \left\{ \ln(\text{Tr}[\mathcal{S}(|\Psi\rangle\langle\Psi|)^2]) > \ln \frac{2e^\epsilon}{d_R} \right\}$$

$$\le 2 \exp\left(-\frac{\epsilon^2 \sqrt{d_R}}{72\pi^3} \right). \qquad (B14)$$

Negating both sides of the expression inside the braces and then taking their exponent yields

$$\Pr_\Psi \left\{ d^{\text{eff}}(\omega) < \frac{d_R}{2e^\epsilon} \right\} \le 2 \exp\left(-\frac{\epsilon^2 \sqrt{d_R}}{72\pi^3} \right). \qquad (B15)$$

Finally setting $\epsilon = \ln 2$ yields the desired result. $\square$

APPENDIX C: FLUCTUATIONS FROM THE TIME AVERAGE

Here we prove an exponential bound on the proportion of time the state spends a long way from its equilibrium state, using measure concentration results [16], and a stronger assumption on H. Making the assumption that the eigenenergies E of H have no rational dependencies (which is much stronger than our nondegenerate energy gaps condition), the trajectory

$$|\Psi(t)\rangle = \sum_k e^{-itE_k} c_k |E_k\rangle \qquad (C1)$$

over time fills the torus

$$\mathcal{T} = \left\{ \Psi(\underline{\alpha}) := \sum_k e^{i\alpha_k} c_k |E_k\rangle : 0 \le \alpha_k \le 2\pi \right\} \qquad (C2)$$

uniformly. What this means in fact is that an ergodic theorem holds, equating the time averages of all continuous functions of the state with their respective "ensemble" averages, i.e., with integration over independent phase angles α_k. The same is true for the indicator functions of open sets, as the one in the following theorem.

Theorem 4. Under the above assumption of ergodicity,

$$\Pr_t \left\{ D[\rho_S(t), \omega_S] > \sqrt{\frac{d_S}{d^{\text{eff}}(\omega_B)}} + \epsilon \right\}$$

$$= \Pr_{\underline{\alpha}} \left\{ D[\rho_S(\underline{\alpha}), \omega_S] > \sqrt{\frac{d_S}{d^{\text{eff}}(\omega_B)}} + \epsilon \right\}$$

$$\le \exp[-c'' \epsilon^4 d^{\text{eff}}(\omega)], \qquad (C3)$$

with a constant $c'' = \frac{1}{128\pi^2}$

Proof. We already argued the first equality, so only the second inequality is to be proved.

From theorem 1 in the main paper, the average of $D[\rho_S(\underline{\alpha}), \omega_S]$ is upper bounded by $\frac{1}{2}\sqrt{\frac{d_S}{d^{\text{eff}}(\omega_B)}}$, hence by Markov's inequality its median is upper bounded by $\sqrt{\frac{d_S}{d^{\text{eff}}(\omega_B)}}$.

The space on which we want to study the measure concentration is the direct product of unit circles. As the metric on that space we choose the following weighted ℓ^1 metric:

$$d(\underline{\alpha}, \underline{\beta}) = \frac{1}{2\pi} \sum_k |c_k|^2 |\alpha_k - \beta_k|. \qquad (C4)$$

The kth factor has diameter $|c_k|^2$, so using [16], we conclude that the concentration function ϑ of this space obeys

$$\vartheta(r) \le e^{-r^2/8L^2}, \quad \text{where} \quad L^2 = \sum_k |c_k|^4 = 1/d^{\text{eff}}(\omega). \qquad (C5)$$

"Concentration function" means that for any set $A \subset [0; 2\pi]^d$ of measure $\ge 1/2$, the r-neighborhood $A_r = \{\underline{\alpha} \text{ s.t. } \exists \underline{\beta} \in A d(\underline{\alpha}, \underline{\beta}) \le r\}$ has measure at least $1 - \vartheta(r)$.

Hence, all that remains is to relate the distance r to the variation of the function $D[\rho_S(\underline{\alpha}), \omega_S]$. We first obtain the general result

$$|D[\rho_S(\underline{\alpha}), \omega_S] - D[\rho_S(\underline{\beta}), \omega_S]|$$

$$\le D[\rho_S(\underline{\alpha}), \rho_S(\underline{\beta})] \le D[\rho(\underline{\alpha}), \rho(\underline{\beta})]$$

$$= \sqrt{1 - |\langle\psi(\underline{\alpha})|\psi(\underline{\beta})\rangle|^2} \le ||\psi(\underline{\alpha})\rangle - |\psi(\underline{\beta})\rangle|_2. \qquad (C6)$$

Together with

$$\||\psi(\underline{\alpha})\rangle - |\psi(\underline{\beta})\rangle\|_2^2 = \sum_k |c_k|^2 |e^{i\alpha_k} - e^{i\beta_k}|^2$$

$$\leq 2\sum_k |c_k|^2 |e^{i\alpha_k} - e^{i\beta_k}|$$

$$\leq 2\sum_k |c_k|^2 |\alpha_k - \beta_k|$$

$$= 4\pi d(\underline{\alpha},\underline{\beta}), \tag{C7}$$

this results in

$$\left|D[\rho_S(\underline{\alpha}),\omega_S] - D[\rho_S(\underline{\beta}),\omega_S]\right|$$

$$\leq \sqrt{4\pi}\sqrt{d(\underline{\alpha},\underline{\beta})}. \tag{C8}$$

So, to change the value of $D(\rho_S,\omega_S)$ by more than ϵ above the median, one has to go from an element in the submedian set a distance of at least $r = \frac{\epsilon^2}{4\pi}$ in our chosen metric d. Inserting this into the above concentration function yields the result. $\qquad\square$

---

[1] J. Gemmer, M. Michel, and G. Mahler, *Quantum Thermodynamics* (Springer-Verlag, Berlin, 2004).

[2] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghì, Phys. Rev. Lett. **96**, 050403 (2006).

[3] S. Popescu, A. J. Short, and A. Winter, Nat. Phys. **2**, 754 (2006). Full version e-print arXiv:quant-ph/0511225.

[4] P. Bocchieri and A. Loinger, Phys. Rev. **114**, 948 (1959).

[5] S. Lloyd, Ph.D thesis, Rockefeller University, 1988.

[6] H. Tasaki, Phys. Rev. Lett. **80**, 1373 (1998).

[7] J. M. Deutsch, Phys. Rev. A **43**, 2046 (1991); M. Srednicki, Phys. Rev. E **50**, 888 (1994).

[8] M. Rigol, V. Dunjko, and M. Olshanii, Nature (London) **452**, 854 (2008).

[9] M. Cramer, C. M. Dawson, J. Eisert, and T. J. Osborne, Phys.

[10] C. Kollath, A. M. Läuchli, and E. Altman, Phys. Rev. Lett. **98**, 180601 (2007).

[11] T. Barthel and U. Schollwöck, Phys. Rev. Lett. **100**, 100601 (2008).

[12] P. Reimann, Phys. Rev. Lett. **101**, 190403 (2008).

[13] C. A. Fuchs and J. van de Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[14] P. Hayden and W. van Dam, e-print arXiv:quant-ph/0204093.

[15] V. Milman and G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed Spaces*, 2nd ed. (Springer-Verlag, Berlin, 2001).

[16] M. Ledoux, *The Concentration of Measure Phenomenon* (AMS Monographs, Providence, RI, 2001), Vol. 89.

Rev. Lett. **100**, 030602 (2008).