

## Quantum algorithm for energy matching in hard optimization problems

C. L. Baldwin<sup>1,2</sup> and C. R. Laumann<sup>1</sup><sup>1</sup>*Department of Physics, Boston University, Boston, Massachusetts 02215, USA*<sup>2</sup>*Department of Physics, University of Washington, Seattle, Washington 98195, USA*

(Received 8 April 2018; revised manuscript received 30 May 2018; published 11 June 2018)

We consider the ability of local quantum dynamics to solve the “energy-matching” problem: given an instance of a classical optimization problem and a low-energy state, find another macroscopically distinct low-energy state. Energy matching is difficult in rugged optimization landscapes, as the given state provides little information about the distant topography. Here, we show that the introduction of quantum dynamics can provide a speedup over classical algorithms in a large class of hard optimization problems. Tunneling allows the system to explore the optimization landscape while approximately conserving the classical energy, even in the presence of large barriers. Specifically, we study energy matching in the random  $p$ -spin model of spin-glass theory. Using perturbation theory and exact diagonalization, we show that introducing a transverse field leads to three sharp dynamical phases, only one of which solves the matching problem: (1) a small-field “trapped” phase, in which tunneling is too weak for the system to escape the vicinity of the initial state; (2) a large-field “excited” phase, in which the field excites the system into high-energy states, effectively forgetting the initial energy; and (3) the intermediate “tunneling” phase, in which the system succeeds at energy matching. The rate at which distant states are found in the tunneling phase, although exponentially slow in system size, is exponentially *faster* than classical search algorithms.

DOI: [10.1103/PhysRevB.97.224201](https://doi.org/10.1103/PhysRevB.97.224201)

### I. INTRODUCTION

Suppose that the perennial traveling saleswoman, having many responsibilities herself, delegates the menial task of constructing an efficient sales route to the company intern. The intern wants to impress his boss, and decides to produce not one but many routes, each sufficiently distinct from the others so that the saleswoman may pick the one which suits her best. After laboring for days on the notoriously hard traveling saleswoman’s problem [1], the intern succeeds at last in identifying one efficient route, but he shudders at the thought of having to repeat the process multiple times. Can he make use of this first route to construct others faster than he would from scratch, keeping in mind that the additional routes must be sufficiently different? This is the matching problem: given an optimization problem and one optimal or near-optimal solution, find others that are sufficiently distinct. The starting solution serves as a hint for finding the others.

The goal in solving problems is often to obtain a representative collection of all possible solutions, rather than a single one, for multiple reasons: one may desire solutions with additional properties (such as approval from the saleswoman in the example above), a larger set of solutions is more robust to parameter perturbations, and a representative set gives insight into underlying structure in the problem. Efficient matching is thus transformative, both when the original optimization problem is difficult [2–4] and when well-designed algorithms are only capable of finding a small set of special solutions [5,6]. The prospect of using a first solution to efficiently generate others is now spurring active research into the area [7,8] after being largely unexplored in the past.

The matching problem is often difficult for the same reason as the original search problem: ruggedness in the cost function to be minimized. A one-dimensional example of a rugged

cost function (or, equivalently, potential energy) is sketched in the top panel of Fig. 1, and provides intuition for the high-dimensional configuration spaces of real optimization problems. The energy landscape has deep local minima called “clusters.” A starting configuration in one cluster can be used to efficiently locate others within the same cluster (see the recent Ref. [7]), but does not provide benefit for finding distant clusters of solutions. Local search algorithms such as Metropolis Monte Carlo must excite the system out of clusters in order to fully explore the configuration space. Furthermore, it is straightforward to show that certain matching problems are NP-complete, and we provide a short proof in Appendix A.

In this paper, we assess whether *quantum Hamiltonian* dynamics may be faster than classical algorithms at energy matching in rugged landscapes. We take as a nontrivial test bed the classical random  $p$ -spin model of spin-glass theory. This model has sharply defined clusters of low-energy states (as reviewed in more detail in Sec. II). We denote the  $p$ -spin Hamiltonian by  $H_p$ , which is diagonal in the  $\hat{\sigma}^z$  basis of  $N$  spin- $\frac{1}{2}$ s. Quantum dynamics is produced by applying a uniform transverse field:

$$H = H_p - \Gamma \sum_{i=1}^N \hat{\sigma}_i^x. \quad (1)$$

Starting in a classical (i.e.,  $\hat{\sigma}^z$ ) state  $|\sigma\rangle$ , we study the probability for observing at time  $t$  a classical state  $|\sigma'\rangle$  belonging to a different cluster, i.e.,  $|\langle\sigma'|e^{-iHt}|\sigma\rangle|^2$ . Hamiltonians of the form in Eq. (1) have long been used in the context of quantum computation, particularly with time-dependent coefficients to study the quantum adiabatic algorithm for finding ground states [9–14]. Here, we instead use a static Hamiltonian to study energy-matching dynamics, and do not restrict ourselves to

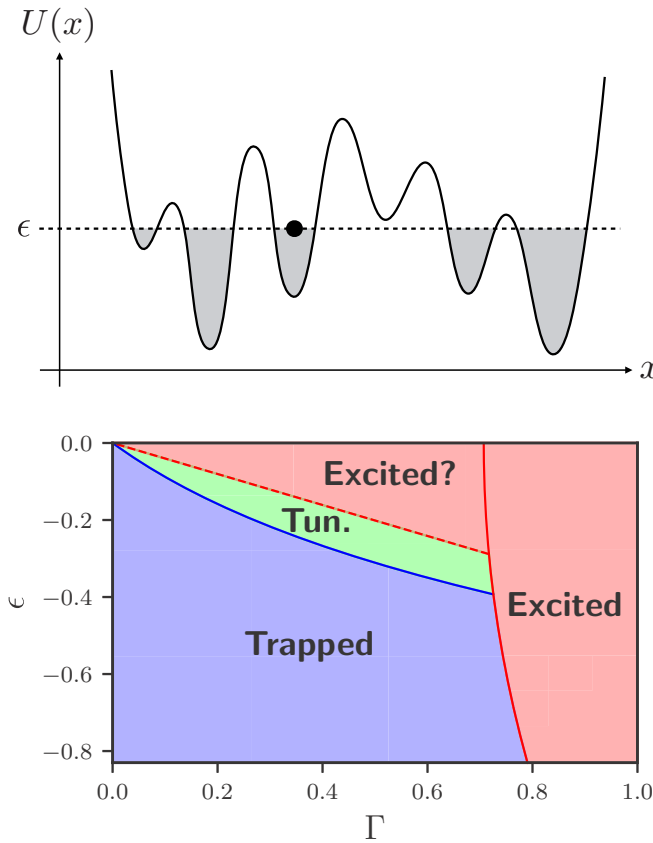


FIG. 1. (Top) A one-dimensional example of a rugged energy landscape  $U(x)$ . States with energy below the dashed line form disconnected clusters (shaded). A particle (black dot) remains in a cluster until it either surmounts or tunnels through the energy barriers. (Bottom) The dynamical phase diagram of the random energy model ( $p \rightarrow \infty$  limit). The system tunnels between clusters in the “tunneling” phase (green), remains trapped in a cluster in the “trapped” phase, and is excited out of clusters in the “excited” phase. The trapped, tunneling, and large- $\Gamma$  excited phases have been confirmed numerically. It is unclear whether the small- $\Gamma$  excited phase is truly an excited phase or a portion of the tunneling phase in which perturbation theory does not apply.

ground states. Our analysis applies to approximate as well as perfect optimization.

One might expect quantum dynamics to be efficient for two reasons: conservation of energy biases the dynamics towards classical states having the same energy as the starting configuration, and quantum fluctuations can tunnel through the energy barriers that separate those states. However, Hamiltonian dynamics conserves the full quantum mechanical energy  $\langle H \rangle$ , whereas the goal of energy matching is to find a state with the same classical energy  $\langle H_p \rangle$ . Furthermore, recent work has shown that the tunneling amplitudes between clusters can be exponentially suppressed in many-body systems [15–17]. Thus, the performance of quantum dynamics in energy matching, and its comparison to classical search algorithms, is nontrivial.

We find three sharp dynamical phases for the transverse field  $p$ -spin model, each with distinct implications for energy matching. The phase depends on the target energy per spin  $\epsilon$

and the strength of the transverse field  $\Gamma$ . The large- $p$  phase diagram is shown in the bottom panel of Fig. 1. At low  $\epsilon$  and small  $\Gamma$ , the probability of the system tunneling between clusters vanishes in the thermodynamic limit even at arbitrarily late times. The system cannot exit its initial cluster and energy matching fails in this “trapped” phase. At large  $\Gamma$ , the system moves freely out of the initial cluster but is excited to higher classical energies in return for magnetizing along the transverse field. Energy matching fails in this “excited” phase as well since the system does not locate states at the desired classical energy. Only at intermediate  $\epsilon$  and  $\Gamma$ , in the “tunneling” phase, does energy matching succeed by tunneling between clusters while roughly preserving the classical energy density [18].

The timescale for tunneling is exponential in system size, i.e., quantum dynamics cannot solve the matching problem in polynomial time. On the other hand, classical algorithms also require exponential runtime in these models, and quantum dynamics runs exponentially *faster* than classical spin-flip Monte Carlo at equilibrium. We also show that it can be exponentially faster than a nonlocal search algorithm in which new configurations are chosen independently at random (the “unstructured search”). Since the problem Hamiltonians which we consider have little to no special structure, it would be difficult to construct a classical algorithm that could perform similarly well. Furthermore, the quantum algorithm presented here is directly implementable on current experimental platforms [19,20].

Note that the tunneling phase does not exist at sufficiently low  $\epsilon$  for the  $p$ -spin model. A uniform transverse field cannot solve the matching problem near the classical ground state regardless of the field strength and regardless of runtime.

We derive these results using both perturbation theory in  $\Gamma$  and numerical exact diagonalization. Here, we present the underlying intuition. We work in the  $\hat{\sigma}^z$  basis, whose eigenstates are referred to as classical states and have definite classical energy  $H_p$ . Since matching problems start from a given classical state  $|\sigma\rangle$  with specified energy density  $\epsilon$ , we distinguish states  $|\sigma'\rangle$  not only by their classical energy densities  $\epsilon'$  but also their fractional Hamming distances  $x$  relative to  $|\sigma\rangle$ :

$$x \equiv \frac{1}{2} \left( 1 - \frac{1}{N} \sum_{i=1}^N \sigma_i \sigma'_i \right). \quad (2)$$

States having the same  $\epsilon$ , assuming  $\epsilon$  is not too close to the center of the spectrum, are disconnected: some states lie at distances less than a certain  $x^*(\epsilon)$ , and others lie at distances greater than a certain  $x^{**}(\epsilon)$ , but none lie in-between. Those states at  $x < x^*(\epsilon)$  belong to the same cluster as  $|\sigma\rangle$ , whereas those at  $x > x^{**}(\epsilon)$  belong to different clusters. We show this using the distance-resolved density of states, i.e., the number of states at distance  $x$  with energy density  $\epsilon'$ . It is exponential and written  $e^{Ng(x, \epsilon'|\epsilon)}$ . Using perturbation theory, we argue that the effective coupling between states is similarly exponential and written  $e^{-N\gamma(x, \epsilon'|\epsilon)}$ , which increases monotonically with  $\Gamma$ . The two exponents  $g(x, \epsilon'|\epsilon)$  and  $\gamma(x, \epsilon'|\epsilon)$  are the central objects in our analysis.

Tunneling between clusters occurs only if there are states in different clusters which are resonant, i.e., states whose classical energies differ by less than the effective coupling between

them. The level spacing between states at distance  $x$  and energy density  $\epsilon$  is of order  $e^{-Ng(x,\epsilon)}$ . These resonate with the initial state  $|\sigma\rangle$  if the effective coupling is larger than the level spacing at distance  $x$ . The requirement for tunneling to occur is thus

$$\max_{x \in [x^{**}(\epsilon), 1-x^{**}(\epsilon)]} [g(x, \epsilon | \epsilon) - \gamma(x, \epsilon | \epsilon)] > 0. \quad (3)$$

(Tunneling condition)

The curve  $\Gamma_{\text{tun}}(\epsilon)$  on which the left-hand side equals 0 defines the boundary between the trapped and tunneling phases.

As the initial state evolves in time, the amplitude on any nonresonant state remains exponentially small, of order  $e^{-2N\gamma(x, \epsilon' | \epsilon)}$ . However, there are exponentially many such states. If the *total* amplitude on states with  $\epsilon' \neq \epsilon$  is large, then those states will be observed at later times with high probability and the system excites into classical energy density  $\epsilon'$ . The requirement for the classical energy density to be preserved is

$$\max_{x \in [0, 1]} \left[ \max_{\epsilon'} [g(x, \epsilon' | \epsilon) - 2\gamma(x, \epsilon' | \epsilon)] \right] < 0. \quad (4)$$

(Nonexcitation condition)

The curve  $\Gamma_{\text{exc}}(\epsilon)$  on which the left-hand side equals 0 defines the boundary between the tunneling and excited phases. There is an additional transition into the excited phase that coincides with the well-known thermodynamic transition into a quantum paramagnetic phase at large  $\Gamma$ . We cannot detect this transition within perturbation theory, but we find clear evidence for it numerically.

Equations (3) and (4) are necessary conditions for the quantum dynamics to succeed at energy matching. If both are satisfied, then the time required for energy matching is simply the inverse tunneling rate between clusters. We estimate the timescale  $\tau_q$  using Fermi's golden rule with the effective coupling and distance-resolved density of states. We find  $\tau_q$  scales exponentially with exponent

$$\frac{1}{N} \ln \tau_q = \min_{x \in [x^{**}(\epsilon), 1-x^{**}(\epsilon)]} [2\gamma(x, \epsilon | \epsilon) - g(x, \epsilon | \epsilon)]. \quad (5)$$

(Tunneling timescale)

Note that the use of transverse-field quantum dynamics for energy matching is also investigated in Ref. [8], independent of and developed concurrently with this work. The analysis in Ref. [8] is complementary to that presented here: we focus on the dynamical phases of the system and the requirements for energy matching to occur, whereas they use a different approach to study the dynamics particularly at large transverse field.

The remainder of the paper is devoted to proving these results. In Sec. II A, we introduce the  $p$ -spin model  $H_p$  which we use for our analysis. We demonstrate the clustering of its low-energy states in Sec. II B, and calculate the function  $g(x, \epsilon' | \epsilon)$  in Sec. II C. In Sec. III, we consider the quantum dynamics of the full Hamiltonian  $H$  for the simple and well-controlled random energy model (the  $p \rightarrow \infty$  limit of the  $p$ -spin model). We develop the perturbation theory and calculate  $\gamma(x, \epsilon' | \epsilon)$  in Sec. III A, show that the resulting tunneling rate is superior to the Arrhenius rate from a classical spin-flip Monte Carlo simulation in Sec. III B, and use exact diagonalization of small systems to validate the perturbative results in Sec. III C. In Sec. IV, we show that these results

are robust in the more realistic mean-field models at large but finite  $p$ , and show that the quantum dynamics can now outpace the unstructured search in addition to spin-flip Monte Carlo. Finally, we conclude in Sec. V.

## II. THE $p$ -SPIN MODEL

### A. Discussion of the model

We consider energy matching in the classical  $p$ -spin model, which consists of  $N$  spin- $\frac{1}{2}$ 's with random all-to-all  $p$ -body interactions. Although originally introduced as a mean-field model for spin glasses [21–23], the  $p$ -spin model has since received attention for its connections to structural glasses [24,25] and combinatorial optimization problems [26,27]. Its versatility is due to the particularly simple Gaussian correlations between the energy levels. It serves as an analytically tractable model of high-dimensional rugged energy landscapes.

The  $p$ -spin Hamiltonian is

$$H_p = \sum_{(i_1 \dots i_p)} J_{i_1 \dots i_p} \hat{\sigma}_{i_1}^z \dots \hat{\sigma}_{i_p}^z. \quad (6)$$

The sum is over all  $p$ -tuples of the  $N$  spins. Each coupling  $J_{i_1 \dots i_p}$  is an independent Gaussian random variable of mean 0 and variance  $\frac{p!}{2N^{p-1}}$ . Denote a configuration of the spins by  $\sigma$ , where the  $i$ th spin has value  $\sigma_i$ . From Eq. (6), it follows that the energies are Gaussian distributed with mean 0 and covariance matrix

$$\mathbb{E}[H_p(\sigma)H_p(\sigma')] = \frac{N}{2}(1 - 2x(\sigma, \sigma'))^p. \quad (7)$$

Here and for the entirety of the paper,  $\mathbb{E}$  denotes averages over the disordered couplings  $J_{i_1 \dots i_p}$ .  $x(\sigma, \sigma')$  denotes the fractional Hamming distance between  $\sigma$  and  $\sigma'$  [Eq. (2)]. Equation (7) summarizes why the  $p$ -spin model is a useful test bed for disordered systems: the energy landscape is Gaussian correlated with correlations that depend only on the distance in configuration space. The parameter  $p$  sets the strength of the correlations and is a useful parameter to vary. In particular, as  $p \rightarrow \infty$  the energy levels become independent [21]. When we consider quantum dynamics below, we use the  $p \rightarrow \infty$  limit as a starting point and then argue that leading-in- $1/p$  corrections do not affect the qualitative picture.

### B. Clustering

The static and dynamical behavior of the  $p$ -spin model derives from the organization of its energy levels in configuration space. Central to this organization is whether configurations at the same energy density are “connected.” We say that  $\sigma$  and  $\sigma'$  are connected (as  $N \rightarrow \infty$ ) if there is a sequence of spin flips transforming one into the other which incurs only  $O(1)$  changes in energy throughout the process. For a given  $\sigma$ , the set of  $\sigma'$  to which it is connected defines a “cluster.” The motivation for these definitions is that, heuristically, stochastic dynamics such as Glauber or Metropolis quickly explores within a cluster but requires much longer times to transition between them. In many physical and computational problems, including the  $p$ -spin model, the number and geometry of clusters transition sharply at certain energy densities.

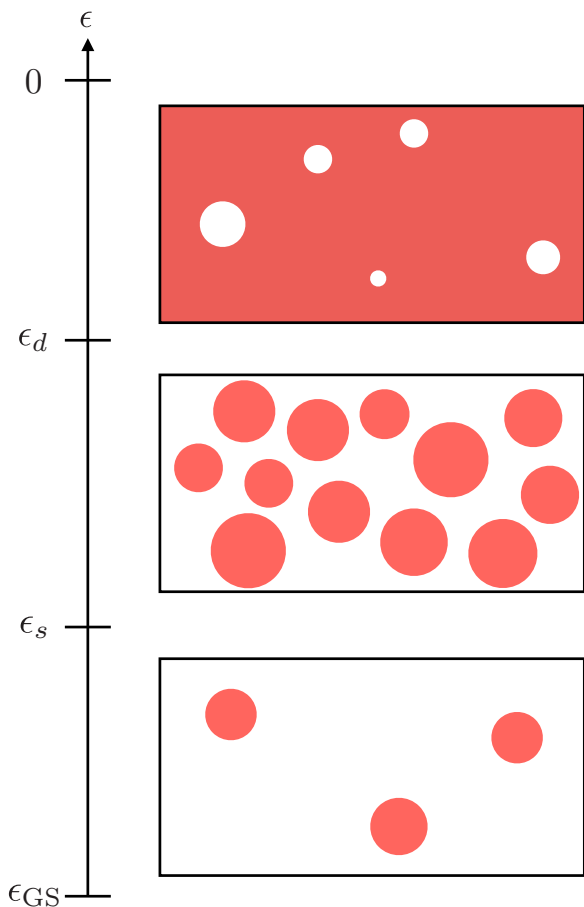


FIG. 2. Important transitions in the classical  $p$ -spin model. Each box represents the configuration space of  $N$  spin- $\frac{1}{2}$ 's, and the red areas represent the regions of configuration space that contain states of energy density  $\epsilon$ . Top:  $\epsilon_d < \epsilon < 0$ , middle:  $\epsilon_s < \epsilon < \epsilon_d$ , bottom:  $\epsilon_{GS} < \epsilon < \epsilon_s$ .

The relevant transitions for this paper are sketched in Fig. 2. The center of the spectrum, corresponding to infinite temperature, is at energy density  $\epsilon = 0$ , and the bottom is at a finite energy density  $\epsilon_{GS} (< 0)$ . In-between, transitions occur at  $\epsilon_d$  and  $\epsilon_s$ . They are best understood in terms of two order parameters. The Edwards-Anderson order parameter quantifies *dynamical* ergodicity breaking, e.g., in a Monte Carlo simulation run for time  $t$ :

$$q_{EA} \equiv \lim_{t \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{E} \left[ \left\langle N^{-1} \sum_i \sigma_i(t) \sigma_i(0) \right\rangle \right]. \quad (8)$$

The angular brackets denote a thermal average over  $\sigma(0)$  and an average over the randomness of the dynamics. The order parameter that quantifies *equilibrium* ergodicity breaking is in terms of “replicas,” i.e., copies of the system that are uncoupled from each other but have the same disorder realization:

$$q \equiv \lim_{N \rightarrow \infty} \mathbb{E} \left[ \left\langle \left| N^{-1} \sum_i \sigma_i^\alpha \sigma_i^\beta \right| \right\rangle \right]. \quad (9)$$

The superscripts  $\alpha$  and  $\beta$  denote different replicas and the angular brackets denote independent thermal averages over  $\sigma^\alpha$  and  $\sigma^\beta$ . Note that  $N^{-1} \sum_i \sigma_i^\alpha \sigma_i^\beta$ , the “overlap” between

$\alpha$  and  $\beta$ , is simply  $1 - 2x(\sigma^\alpha, \sigma^\beta)$ . Two configurations chosen *uniformly* out of all possible  $2^N$  will have  $q = 0$  ( $x = \frac{1}{2}$ ) with probability 1. In a finite-temperature paramagnetic phase,  $q$  remains at 0. In any ordered phase, whether ferromagnetic or spin glass,  $q \neq 0$ .

Now, we turn to the relevant phases of the  $p$ -spin model:

(i)  $\epsilon_d < \epsilon < 0$ :  $q_{EA} = 0, q = 0$ . A randomly selected pair of states at such  $\epsilon$  is connected with probability 1 (in the thermodynamic limit). The corresponding cluster spans the configuration space, in the sense that the overlap between a randomly selected pair is 0 with probability 1 and stochastic dynamics equilibrates throughout the space.

(ii)  $\epsilon_s < \epsilon < \epsilon_d$ :  $q_{EA} \neq 0, q = 0$ . Typical states are no longer connected, and instead the number of clusters scales exponentially with  $N$ . The timescale for transitioning between clusters is exponential in  $N$ . In particular, it diverges in the thermodynamic limit, hence,  $q_{EA}$  is nonzero. Nonetheless, the clusters are distributed throughout the configuration space. A randomly selected pair of states belong to different clusters and the overlap is still 0.

(iii)  $\epsilon_{GS} < \epsilon < \epsilon_s$ :  $q_{EA} \neq 0, q \neq 0$ . The number of clusters is  $O(1)$  with respect to  $N$ . A randomly selected pair of states has finite probability of belonging to the same cluster, which occupies only a small region of the configuration space. This finite-probability event produces a nonzero average overlap, i.e.,  $q \neq 0$ .

The transition at  $\epsilon_d$  is called the “dynamical” or “clustering” transition, as it marks the energy density (or corresponding temperature) below which stochastic dynamics fails to equilibrate the system. The transition at  $\epsilon_s$  is called the “static” transition, as it is where the equilibrium order parameter becomes nontrivial.

We focus on the range  $\epsilon_s < \epsilon < \epsilon_d$  in this paper, for which the configurations are organized into exponentially many clusters. The matching problem is solved if the system, initially prepared in one cluster, is found in a different cluster at a later time.

### C. Franz-Parisi potential

We study the geometry of clusters by computing the quantity defined in Eq. (10). It counts, for a fixed configuration  $\sigma$  having energy density  $\epsilon$ , the number of configurations  $\sigma'$  having energy density  $\epsilon'$  which are separated by a distance  $x$ . Precisely,

$$g(x, \epsilon' | \sigma, \epsilon) \equiv \frac{1}{N} \mathbb{E} [\ln \text{Tr}_{\sigma'} [\delta_{x, x(\sigma, \sigma')} \delta(\epsilon' - \epsilon(\sigma'))]]_{\epsilon(\sigma) = \epsilon}. \quad (10)$$

The subscript to the disorder average indicates that we condition on having  $\epsilon(\sigma) = \epsilon$ . The argument of the logarithm is the number of configurations at distance  $x$  from  $\sigma$  with energy density  $\epsilon'$ . Thus,  $g(x, \epsilon' | \sigma, \epsilon)$  is the conditioned average entropy density at distance  $x$ . Since the correlations between  $\sigma$  and  $\sigma'$  depend only on their separation,  $g(x, \epsilon' | \sigma, \epsilon)$  depends on  $\sigma$  only through  $x$  and  $\epsilon$ . Thus, we will write  $g(x, \epsilon' | \epsilon)$  throughout the paper.

$g(x, \epsilon' | \epsilon)$  is closely related to the Franz-Parisi potential (FPP), which is an important tool in the analysis of mean-field disordered systems [28–30]. Equation (10) is essentially the



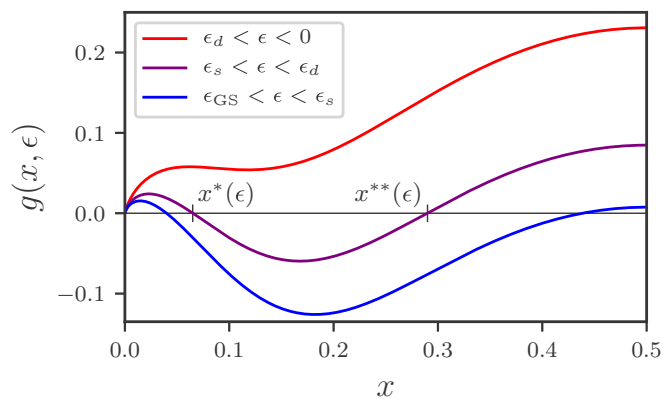


FIG. 3. Transitions in the  $p$ -spin model exhibited through the FPP,  $g(x, \epsilon | \epsilon)$ . Compare to Fig. 2. Only  $x < \frac{1}{2}$  is shown since (for even  $p$ )  $g(x, \epsilon | \epsilon)$  is symmetric between  $x \leftrightarrow 1 - x$ . For the curves shown, we made the annealed approximation using  $p = 6$  and  $\epsilon = -0.68$  (red),  $-0.78$  (purple),  $-0.828$  (blue).

FPP written in the microcanonical ensemble, as we describe in Appendix B. We shall refer to  $g(x, \epsilon' | \epsilon)$  as the FPP throughout the paper, but keep in mind that Eq. (10) is not the standard form in which it is presented.

Strictly speaking, one would need to use the replica trick to evaluate the FPP, as done in Ref. [29]. However, the essential physics remains intact if we instead take an “annealed” average by switching the order of the logarithm and disorder average:

$$\begin{aligned} g(x, \epsilon' | \epsilon) &\approx \frac{1}{N} \ln \mathbb{E}[\text{Tr}_{\sigma'}[\delta_{x, x(\sigma, \sigma')} \delta(\epsilon' - \epsilon(\sigma'))]]_{\epsilon(\sigma) = \epsilon} \\ &= -x \ln x - (1 - x) \ln(1 - x) \\ &\quad + \frac{1}{N} \ln \mathbb{E}[\delta(\epsilon' - \epsilon(\sigma'))]_{\epsilon(\sigma) = \epsilon}. \end{aligned} \quad (11)$$

Note that  $\mathbb{E}[\delta(\epsilon' - \epsilon(\sigma'))]_{\epsilon(\sigma) = \epsilon}$  is the probability of  $\epsilon(\sigma') = \epsilon'$  conditioned on  $\epsilon(\sigma) = \epsilon$ . We present the calculation of this conditional distribution in Appendix C. The end result for the FPP is

$$\begin{aligned} g(x, \epsilon' | \epsilon) &\sim -x \ln x - (1 - x) \ln(1 - x) \\ &\quad - \frac{(\epsilon' - (1 - 2x)^p \epsilon)^2}{1 - (1 - 2x)^{2p}}. \end{aligned} \quad (12)$$

Although Eq. (12) is only an annealed average, the inequality  $\mathbb{E}[\ln(\cdot)] \leq \ln \mathbb{E}[\cdot]$  shows that it is a rigorous upper bound to the exact FPP. In particular, if  $g(x, \epsilon' | \epsilon) < 0$  then there are *no* states having  $\epsilon'$  at distance  $x$  (with probability 1 in the thermodynamic limit) [26].

Define  $g(x, \epsilon) \equiv g(x, \epsilon | \epsilon)$ . As a function of  $x$ ,  $g(x, \epsilon)$  demonstrates that low-lying energy levels are organized into clusters. Figure 3 gives representative examples. Compare the shapes of  $g(x, \epsilon)$  in Fig. 3 to the sketches in Fig. 2.

(i)  $\epsilon_d < \epsilon < 0$ :  $g(x, \epsilon) > 0$  for all  $x$ . There are configurations that have the same energy density at all distances from the reference state  $\sigma$ . This suggests that each configuration is connected to all others, forming a single cluster that spans the configuration space. While it is not a proof, as the FPP distinguishes only the radial coordinate  $x$  of configurations and not angular coordinates, dynamical calculations of the classical stochastic dynamics confirm that  $q_{\text{EA}} = 0$  above  $\epsilon_d$  [24,31].

(ii)  $\epsilon_s < \epsilon < \epsilon_d$ :  $g(x, \epsilon)$  is positive for  $x$  less than a certain  $x^*(\epsilon)$  or greater than a certain  $x^{**}(\epsilon)$  (see Fig. 3), but is negative in-between. This *proves* that configurations below  $\epsilon_d$  are organized into disjoint clusters. No configurations at distances  $x \in (x^*(\epsilon), x^{**}(\epsilon))$  have energy density  $\epsilon$ , thus those at  $x > x^{**}(\epsilon)$  cannot be connected to  $\sigma$ . Furthermore, the maximum of  $g(x, \epsilon)$  over  $x > x^{**}(\epsilon)$  is greater than that over  $x < x^*(\epsilon)$ . There are exponentially more configurations disconnected to  $\sigma$  than connected, i.e., exponentially many clusters. A randomly selected  $\sigma'$  lies at distance  $\frac{1}{2}$  from  $\sigma$ .

(iii)  $\epsilon_{\text{GS}} < \epsilon < \epsilon_s$ : The maximum of  $g(x, \epsilon)$  over  $x > x^{**}(\epsilon)$  is now *less* than that over  $x < x^*(\epsilon)$ . Interpreting this result literally, one would say that most configurations belong to a single cluster of linear size  $x^*(\epsilon)$ . A randomly selected  $\sigma'$  lies within that distance.

Keep in mind that since we estimated  $g(x, \epsilon)$  through an annealed average, Eq. (12) gives only approximate values for the quantities defined above [ $\epsilon_d$ ,  $\epsilon_s$ ,  $x^*(\epsilon)$ , etc.]. In particular, the interpretation that below  $\epsilon_s$  most states belong to a single cluster is too naive: the number of clusters is  $O(1)$  but larger than 1, and it depends on  $\epsilon$  [26]. However,  $\epsilon_d$  as estimated from Eq. (12) is an exact lower bound on the location of the clustering transition. Below we shall need to compute the energy barriers between clusters, and here as well the annealed estimate gives exact lower bounds.

### III. QUANTUM DYNAMICS IN THE LARGE- $p$ LIMIT

Here, we describe the performance of quantum dynamics in tunneling between the clusters of the  $p$ -spin model, specifically in the  $p \rightarrow \infty$  limit. Our main results are the tunneling and nonexcitation conditions (3) and (4), respectively, both of which are necessary conditions for the dynamics to succeed in energy matching. They express that the tunneling amplitudes must be large enough to hybridize states between clusters, but not so large that the system is excited out of clusters. Even if both requirements are satisfied, the tunneling rate is exponentially slow in system size, with the exponent given by Eq. (5).

The Hamiltonian that we consider is

$$H = H_p - \Gamma \sum_i \hat{\sigma}_i^x \equiv H_p + V, \quad (13)$$

with  $H_p$  (the “classical” energy) as in Eq. (6). The second term, a uniform transverse field, causes spin flips. In the  $(\sigma^z)$  configuration space, it acts as a hopping term while  $H_p$  acts as a potential. Thus, Eq. (13) can be interpreted as an Anderson problem [32,33] in the many-body configuration space. The  $p \rightarrow \infty$  limit corresponds to an uncorrelated potential and small clusters, and is simplest to study for reasons which we describe below. The bottom panel of Fig. 1 presents the dynamical phase diagram in this limit. In the “trapped” phase, the system does not tunnel between clusters, even on exponentially long timescales. In the “excited” phase, the system is excited to higher classical energy densities. Only in the “tunneling” phase does the system tunnel between clusters and succeed in energy matching.

We demonstrate these results using perturbation theory in Sec. III A. In Sec. III B, we show that the tunneling rates thus obtained, although exponentially slow in system size, provide

exponential speedup over classical Monte Carlo simulations. We numerically validate the dynamical phase diagram in Sec. III C.

### A. Perturbative analysis

In the  $p \rightarrow \infty$  limit, Eq. (7) for the correlation between the classical energies becomes

$$\mathbb{E}[H_p(\sigma)H_p(\sigma')] \rightarrow \frac{N}{2}\delta_{0,x(\sigma,\sigma')} = \frac{N}{2}\delta_{\sigma,\sigma'}. \quad (14)$$

The classical energies are independent and distributed as

$$P_1(\epsilon) = \sqrt{\frac{N}{\pi}}e^{-N\epsilon^2}, \quad (15)$$

and the FPP is

$$g(x,\epsilon'|\epsilon) = -x \ln x - (1-x) \ln(1-x) - (\epsilon')^2. \quad (16)$$

The  $p \rightarrow \infty$  model is referred to as the ‘‘random energy model’’ (REM) [21]. It is the simplest context in which to study tunneling between clusters because the clusters have no internal structure:  $g(x,\epsilon) < 0$  for *all*  $x$  less than a certain  $x^{**}(\epsilon)$ , i.e., each ‘‘cluster’’ has size 0 and in fact consists of a single configuration.  $x^{**}(\epsilon)$  is the minimum distance between any configurations having energy density  $\epsilon$ .

To study tunneling between clusters, we formally use the Schrieffer-Wolff transformation [34] together with the forward-scattering approximation [16,35]. First, we discuss the Schrieffer-Wolff transformation. Let  $\mathcal{P}_0$  denote the subspace spanned by  $\sigma^z$  configurations having classical energy density  $\epsilon$ , and let  $\mathcal{Q}_0$  denote the orthogonal subspace. We take  $\epsilon < \epsilon_d$ , so that the configurations within  $\mathcal{P}_0$  are organized into clusters. Note that  $H$  couples  $\mathcal{P}_0$  and  $\mathcal{Q}_0$  through the transverse field, yet it does not directly couple configurations within different clusters, as multiple spin flips would be required. The Schrieffer-Wolff transformation is performed by a unitary operator  $e^{iS}$  such that  $e^{iS}He^{-iS} \equiv H_{\text{eff}}$  does not couple  $\mathcal{P}_0$  and  $\mathcal{Q}_0$ . In return,  $H_{\text{eff}}$  does have a direct coupling between configurations within different clusters, denoted  $V_{\text{eff}}$ . The situation is illustrated in Fig. 4. We denote by  $H^{(\alpha)}$  the projection of  $H$  into cluster  $\alpha$ , and similarly for  $H_{\text{eff}}^{(\alpha)}$  and  $V_{\text{eff}}^{(\alpha\beta)}$ . Since

$$\langle \sigma' | e^{-iHt} | \sigma \rangle = \langle \sigma' | e^{-iS} e^{-iH_{\text{eff}}t} e^{iS} | \sigma \rangle, \quad (17)$$

the time evolution of  $|\sigma\rangle$  into  $|\sigma'\rangle$  under  $H$  is equivalent to the evolution of  $e^{iS}|\sigma\rangle$  into  $e^{iS}|\sigma'\rangle$  under  $H_{\text{eff}}$ .

The method by which one calculates the generator  $S$  is known in the literature [34]. We quote it in Appendix D, where we also detail the forward-scattering approximation (FSA). The FSA approximates  $\langle \sigma' | V_{\text{eff}} | \sigma \rangle$  [where  $\epsilon(\sigma) = \epsilon(\sigma') = \epsilon$ ] by its lowest-order terms in the transverse field  $\Gamma$ . As reviewed in Appendix D, they correspond to directed ‘‘paths’’ in the configuration space, i.e., sequences of spin flips that transform  $\sigma$  into  $\sigma'$ :

$$\langle \sigma' | V_{\text{eff}} | \sigma \rangle \sim \sum_P \Gamma \prod_{\sigma'' \in P} \frac{\Gamma}{N(\epsilon - \epsilon(\sigma''))}. \quad (18)$$

The sum is over the  $(Nx(\sigma,\sigma'))!$  sequences of spin flips, and the product is over the intermediate configurations along each sequence. Figure 5 gives an example of such a path.

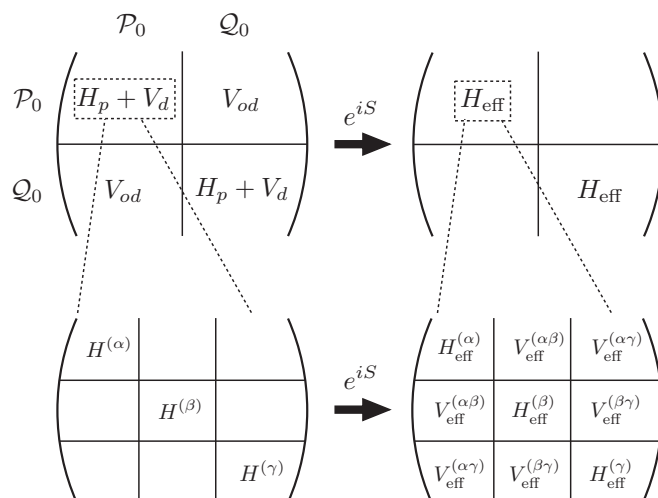


FIG. 4. The effect of the Schrieffer-Wolff transformation. The left side shows  $H$ , the right side shows  $H_{\text{eff}}$ . The top shows the full Hamiltonian, broken into  $\mathcal{P}_0$  and  $\mathcal{Q}_0$  subspaces. The transverse-field operator  $V$  is broken into a block-diagonal part  $V_d$  and a block-off-diagonal part  $V_{od}$ . The bottom is a schematic of the structure within  $\mathcal{P}_0$ . The superscripts refer to different clusters.

Each  $\epsilon(\sigma'')$  in Eq. (18) is an independent random variable of mean 0. If we replace every  $\epsilon(\sigma'')$  by  $\mathbb{E}[\epsilon(\sigma'')]$ , the effective coupling takes a simple form

$$|\langle \sigma' | V_{\text{eff}} | \sigma \rangle| \sim (Nx)! \left( \frac{\Gamma}{N|\epsilon|} \right)^{Nx} \sim e^{-N\gamma(x,\epsilon)}, \quad (19)$$

where

$$\gamma(x,\epsilon) \equiv -x \ln \frac{x\Gamma}{e|\epsilon|}. \quad (20)$$

In fact, the distribution of  $|\langle \sigma' | V_{\text{eff}} | \sigma \rangle|$  over realizations concentrates around this value. Write

$$\prod_{\sigma'' \in P} \frac{1}{|\epsilon - \epsilon(\sigma'')|} = e^{-\sum \ln |\epsilon - \epsilon(\sigma'')|}. \quad (21)$$

In order for Eq. (21) to scale as anything other than  $\exp(-Nx \ln |\epsilon|)$ , it must be that  $O(N)$  of the  $\sigma''$  along path  $P$  have  $\epsilon(\sigma'') \neq 0$ . The probability of such an event scales

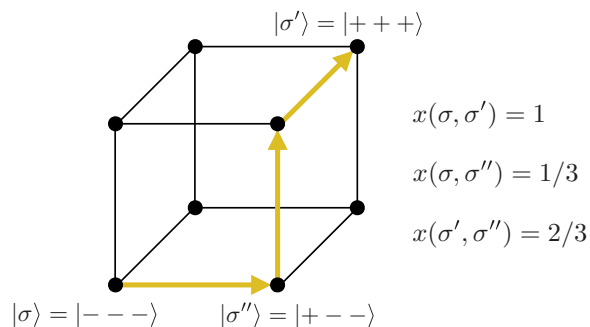


FIG. 5. The configuration space (black vertices) for  $N = 3$ . The gold arrows form a directed path from  $|\sigma\rangle$  to  $|\sigma'\rangle$ .  $|\sigma''\rangle$  is an intermediate state. Distances between configurations are shown on the right.

as  $e^{-cN^2}$ , with  $c \sim O(1)$ . Thus, the expected number of paths which contribute anything other than  $\exp(-Nx \ln |\epsilon|)$ , denoted  $\mathcal{N}_{\text{fluc}}$ , is

$$\mathbb{E}[\mathcal{N}_{\text{fluc}}] = (Nx)! e^{-cN^2} \rightarrow 0, \quad (22)$$

i.e., the expected number of atypical paths vanishes as  $N \rightarrow \infty$ . Equation (19) thus gives the effective coupling between clusters in the  $p \rightarrow \infty$  limit.

We next use time-dependent perturbation theory to calculate the transition amplitude from  $|\sigma\rangle$  to  $|\sigma'\rangle$  given in Eq. (17). One can show that taking  $e^{iS}|\sigma\rangle \sim |\sigma\rangle$  and  $e^{iS}|\sigma'\rangle \sim |\sigma'\rangle$  does not affect the lowest-order terms (see Appendix D). Furthermore, we neglect ‘‘self-energy’’ corrections that would modify the diagonal elements of  $H_{\text{eff}}$ : the corrections to the energy *densities* are only  $O(1/N)$  [36], and the classical energies are distributed randomly regardless. Then, the standard derivation [37] gives

$$P_{\sigma'}(t) \equiv |\langle \sigma' | e^{-iHt} | \sigma \rangle|^2 = \frac{4|\langle \sigma' | V_{\text{eff}} | \sigma \rangle|^2}{N^2(\epsilon(\sigma') - \epsilon(\sigma))^2} \sin^2 \left( \frac{N(\epsilon(\sigma') - \epsilon(\sigma))t}{2} \right). \quad (23)$$

We are interested in  $\epsilon(\sigma') \sim \epsilon(\sigma) \sim \epsilon$ , but it is important that we do not set  $\epsilon(\sigma')$  exactly equal to  $\epsilon(\sigma)$ . The smallest energy difference among all  $\sigma'$  is typically of the order of level spacing, which may be either smaller or larger than  $\langle \sigma' | V_{\text{eff}} | \sigma \rangle$  since both scale exponentially with  $N$ . More precisely, the level spacing among  $|\sigma'\rangle$  at distance  $x$  scales as  $e^{-Ng(x,\epsilon)}$ , with  $g(x,\epsilon)$  given by Eq. (16) (setting  $\epsilon' = \epsilon$ ). If  $g(x,\epsilon) < \gamma(x,\epsilon)$ , then Eq. (23) is exponentially small for every final state at distance  $x$ . Furthermore, the total weight on all states at distance  $x$ , obtained by summing  $P_{\sigma'}(t)$  over all  $|\sigma'\rangle$  having the specified energy density and distance, is exponentially small: those  $|\sigma'\rangle$  for which  $\epsilon(\sigma') - \epsilon(\sigma) \sim e^{-Nd}$  give a total contribution  $e^{-2N\gamma+2Nd} e^{Ng-Nd} = e^{-2N\gamma+Ng+Nd}$ , which increases with  $d$  until the smallest possible spacing at  $d = g$ . Thus, the probability of transitioning to any state at distance  $x$  vanishes in the thermodynamic limit, for all times  $t$ . Since clusters lie at distances  $x \in [x^{**}(\epsilon), 1 - x^{**}(\epsilon)]$ , if  $g(x,\epsilon) < \gamma(x,\epsilon)$  for all such  $x$  then the system never transitions into a different cluster, even on exponentially long timescales. This gives the tunneling condition presented in the Introduction as a necessary condition for quantum dynamics to succeed in energy matching:

$$\max_{x \in [x^{**}(\epsilon), 1 - x^{**}(\epsilon)]} [g(x,\epsilon) - \gamma(x,\epsilon)] > 0. \quad (3)$$

(Tunneling condition)

Using Eqs. (16) and (20), the tunneling condition becomes

$$\max_{x \in [x^{**}(\epsilon), 1 - x^{**}(\epsilon)]} \left[ x \ln \frac{\Gamma}{e|\epsilon|} - (1-x) \ln(1-x) - \epsilon^2 \right] > 0, \quad (24)$$

where  $x^{**}(\epsilon)$  solves  $-x \ln x - (1-x) \ln(1-x) = \epsilon^2$ . The curve  $\Gamma_{\text{tun}}(\epsilon)$  on which the left-hand side is 0 constitutes a sharp dynamical phase boundary, separating phases in which the system does and does not tunnel between clusters.

Note that the tunneling transition also manifests in properties of the many-body *eigenstates*, as discussed in Ref. [38]. At  $\Gamma < \Gamma_{\text{tun}}(\epsilon)$ , the eigenstates are only slightly perturbed

from the classical states  $|\sigma\rangle$ , whereas at  $\Gamma > \Gamma_{\text{tun}}(\epsilon)$ , the eigenstates are superposed from all classical states at energy density  $\epsilon$ . This picture was confirmed by extensive numerics in Ref. [16]. The use of perturbation theory and the FSA were also validated, by evaluating Eq. (18) *numerically* for instances of finite-size systems. It was found that the location of the tunneling transition predicted by the numerical FSA agrees very well with the location found by exact diagonalization.

If Eq. (3) is satisfied, then Fermi’s golden rule gives the *rate* at which the system tunnels between clusters. To ensure that we handle the exponentially small scales in the level spacing and effective coupling correctly, we present a derivation here. Consider  $|\sigma'\rangle$  at distance  $x$  and fixed time  $t$ . If  $N|\epsilon(\sigma') - \epsilon(\sigma)| \ll t^{-1}$ , then  $P_{\sigma'}(t)$  behaves as  $e^{-2N\gamma(x,\epsilon)} t^2$ , and if  $N|\epsilon(\sigma') - \epsilon(\sigma)| \gg t^{-1}$ , then  $P_{\sigma'}(t)$  has already reached its maximum value and begun oscillating. Thus, the portion of the weight at distance  $x$  which is growing with time, denoted  $P_{\text{tun}}(x,t)$ , is obtained by summing  $P_{\sigma'}(t)$  over  $|\sigma'\rangle$  with  $|\epsilon(\sigma') - \epsilon(\sigma)| \lesssim (Nt)^{-1}$ . To exponential order,

$$P_{\text{tun}}(x,t) \sim (e^{-2N\gamma(x,\epsilon)} t^2) (e^{Ng(x,\epsilon)} t^{-1}) = e^{-N(2\gamma(x,\epsilon) - g(x,\epsilon))t}, \quad (25)$$

from which the tunneling rate is apparent. The rate  $\tau_q^{-1}$  at which the system exits its initial cluster is given by integrating over  $x$ , which since  $P_{\text{tun}}(x,t)$  scales exponentially gives the result stated in the Introduction:

$$\frac{1}{N} \ln \tau_q = \min_{x \in [x^{**}(\epsilon), 1 - x^{**}(\epsilon)]} [2\gamma(x,\epsilon) - g(x,\epsilon)]. \quad (5)$$

(Tunneling timescale)

A sufficiently strong transverse field is required to satisfy Eq. (3), yet the perturbation theory breaks down if the transverse field is too strong.  $e^{iS}|\sigma\rangle$  has weight not only in the  $\mathcal{P}_0$  subspace but also in  $\mathcal{Q}_0$ , and these two components of  $e^{iS}|\sigma\rangle$  evolve separately over time (see Fig. 4). In order for perturbation theory to be valid, the total weight in  $\mathcal{Q}_0$  should be small. Let  $|\sigma''\rangle$  now be a state in  $\mathcal{Q}_0$ , so that  $\epsilon(\sigma'') \equiv \epsilon' \neq \epsilon$ . Within the same approximations as above, one finds (see Appendix D)

$$\langle \sigma' | e^{iS} | \sigma \rangle \sim \langle \sigma' | iS | \sigma \rangle \sim \sum_P \prod_{\sigma'' \in P} \frac{\Gamma}{N(\epsilon - \epsilon(\sigma''))}. \quad (26)$$

The sum is again over the direct paths from  $|\sigma\rangle$  to  $|\sigma'\rangle$ , and the product is again over intermediate states  $|\sigma''\rangle$ .

As before, we can take  $\epsilon(\sigma'') \rightarrow \mathbb{E}[\epsilon(\sigma'')] = 0$ . Equation (26) becomes

$$|\langle \sigma' | e^{iS} | \sigma \rangle| \sim \left( \frac{x\Gamma}{e|\epsilon|} \right)^{Nx} \equiv e^{-N\gamma(x,\epsilon'|\epsilon)}. \quad (27)$$

Even though  $\gamma(x,\epsilon'|\epsilon)$  does not depend on  $\epsilon'$  [and in fact  $\gamma(x,\epsilon'|\epsilon) = \gamma(x,\epsilon)$  from Eq. (20)], we keep mention of it in our notation. The lack of  $\epsilon'$  dependence is due to the lack of correlations in the  $p \rightarrow \infty$  limit. As soon as one includes finite- $p$  corrections, as in Sec. IV,  $\gamma(x,\epsilon'|\epsilon)$  depends on  $\epsilon'$ .

The total weight in the  $\mathcal{Q}_0$  subspace is

$$\begin{aligned} & \int_0^1 dx \int d\epsilon' |\langle \sigma' | e^{iS} | \sigma \rangle|^2 e^{Ng(x, \epsilon' | \epsilon)} \\ & \sim \int_0^1 dx \int d\epsilon' e^{N(g(x, \epsilon' | \epsilon) - 2\gamma(x, \epsilon' | \epsilon))}. \end{aligned} \quad (28)$$

The result must be exponentially small if perturbation theory is to be valid. This is the nonexcitation condition given in the Introduction:

$$\max_{x \in [0, 1]} \left[ \max_{\epsilon'} [g(x, \epsilon' | \epsilon) - 2\gamma(x, \epsilon' | \epsilon)] \right] < 0. \quad (4)$$

(Nonexcitation condition)

For the REM, the maximization over  $\epsilon'$  is trivial and we have the requirement

$$\max_{x \in [0, 1]} \left[ 2x \ln \frac{\Gamma}{e|\epsilon|} + x \ln x - (1-x) \ln(1-x) \right] < 0. \quad (29)$$

Strictly speaking, violation of Eq. (4) only means that the perturbation theory is inconsistent. However, it has an immediate physical interpretation: the system excites out of its initial cluster and into higher classical energy densities. This corresponds to another failure mechanism for energy matching, as a measurement of the system will not yield the desired energy density.

Thus, we have obtained two necessary conditions for quantum dynamics to succeed in energy matching. If Eq. (3) is violated, the system never escapes its initial cluster. If Eq. (4) is violated, the system excites out of clusters and does not return. Generically, satisfying both conditions requires that the transverse field be neither too strong nor too weak, and this regime may be narrow or nonexistent depending on  $\epsilon$  and the model in question. Indeed, Fig. 1 shows that one cannot satisfy both conditions at low  $\epsilon$  in the REM.

Ultimately, the Schrieffer-Wolff transformation and the FSA are uncontrolled approximations in this context. In Sec. III C, we therefore study the random energy model numerically via exact diagonalization. We find clear evidence for the tunneling transition, although finite-size effects prevent us from confirming its specific functional form. Finite-size effects also prevent us from unambiguously identifying the excitation transition predicted by Eq. (4). However, we do find that the first-order *thermodynamic* transition into a quantum paramagnetic phase [39] is relevant for the Hamiltonian dynamics. It is itself an excitation transition in which the system excites to classical energy density 0, and is another breakdown of energy matching. As the transition is first order, the perturbative expansion in  $\Gamma$  does not locate it. Instead, the phase boundary was calculated using the replica method in Ref. [39]. We include it in Fig. 1, which highlights the limitations of quantum dynamics in energy matching.

### B. Comparison to thermal activation rate

The rate of tunneling between clusters, when it occurs at all, is always exponentially slow in system size. This is clear from the expressions for the rate [Eq. (5)] and the nonexcitation condition [Eq. (4)]: if Eq. (4) is satisfied, then  $N^{-1} \ln \tau_q$  in Eq. (5) is necessarily positive since it optimizes over fewer

variables. Thus, quantum dynamics cannot succeed at energy matching in polynomial time. However, it may nevertheless be exponentially *faster* than simple classical algorithms. We now show this for the REM, by comparing the tunneling rate found above to the equilibration timescale for stochastic (e.g., Monte Carlo) dynamics. We then additionally compare to an unstructured search algorithm in which new configurations are chosen randomly.

Stochastic dynamics can be used for energy matching as follows: from a starting configuration  $\sigma$ , run a Monte Carlo simulation at the temperature  $T$  corresponding to  $\epsilon(\sigma)$  by Legendre transform. Since the temperature is properly chosen, one will observe a configuration having the same energy density at later times with high probability. However, the later configuration will belong to the same cluster unless one waits long enough for the system to be thermally activated over the energy barriers that separate clusters. This activation timescale is the time required for Monte Carlo dynamics to succeed in energy matching.

It is straightforward to calculate the activation timescale in the REM. The  $N$  neighboring configurations to a given  $\sigma$  all have energy density 0 with high probability, as follows from Eq. (15). Thus, in a Monte Carlo simulation with single-spin update rules (e.g., Metropolis), the simulation time required to leave state  $\sigma$  is  $e^{N\beta|\epsilon(\sigma)|}$ . The thermodynamic entropy density of the REM is

$$s(\epsilon) = \ln 2 - \epsilon^2 \quad (30)$$

(see Ref. [26]), from which it follows that  $\beta(\epsilon) = -2\epsilon$ . The activation timescale in the REM starting from energy density  $\epsilon$  is therefore

$$\tau_{\text{th}} \sim e^{2N\epsilon^2}. \quad (31)$$

The tunneling timescale  $\tau_q$  also scales exponentially with  $N$  and is given by Eq. (5). Using the explicit expressions for  $\gamma(x, \epsilon)$  and  $g(x, \epsilon)$ ,

$$\begin{aligned} \frac{1}{N} \ln \tau_q = & \min_{x \in [x^{**}(\epsilon), 1-x^{**}(\epsilon)]} \left[ -2x \ln \frac{\Gamma}{e|\epsilon|} - x \ln x \right. \\ & \left. + (1-x) \ln(1-x) + \epsilon^2 \right]. \end{aligned} \quad (32)$$

It is straightforward to evaluate Eq. (32) numerically. Figure 6 shows the tunneling timescale as a function of  $\Gamma$  at representative values of  $\epsilon$ , alongside the activation timescale. Tunneling is exponentially faster than thermal activation at energy matching. One can check that this is true for all  $\epsilon$  and  $\Gamma$  in the tunneling phase of the REM. However, in general the tunneling phase may have both a regime in which tunneling is slower than activation (small  $\Gamma$ ) and a regime in which tunneling is faster (large  $\Gamma$ ). Just as in the single-particle setting, activation timescales depend on the height of energy barriers whereas tunneling timescales depend on a combination of the height and width. Either can be faster depending on the details of the energy landscape.

Note that there is another method of performing energy matching which is particularly trivial: ignore the given  $\sigma$  and pick a new configuration  $\sigma'$  uniformly out of all possible  $2^N$ , check if  $\epsilon(\sigma') = \epsilon$ , and repeat until true. The probability of selecting a  $\sigma'$  with the correct energy density is  $e^{Ns(\epsilon)}/2^N$  (and



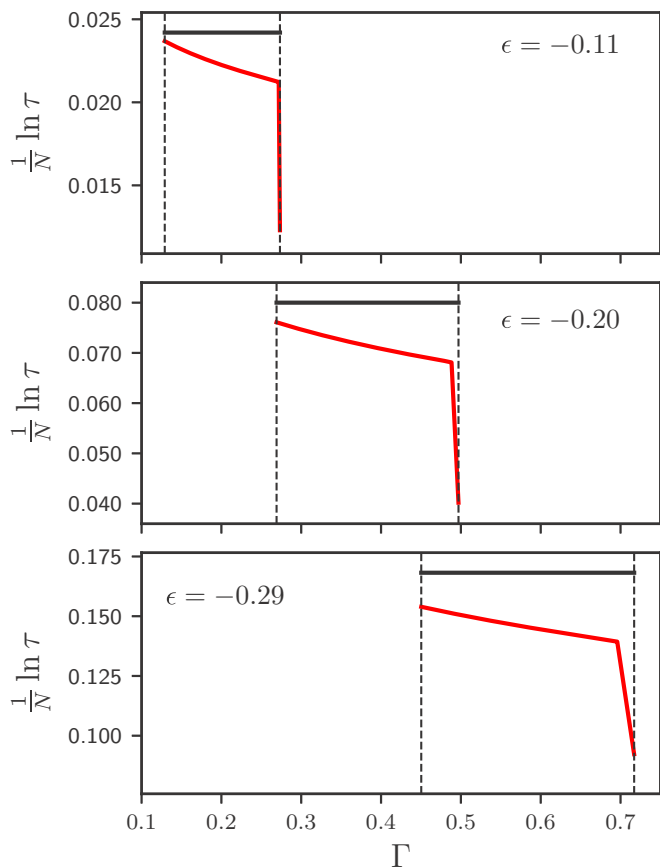


FIG. 6. Timescales for energy matching in the REM, as a function of  $\Gamma$  at the indicated  $\epsilon$ . Red is  $\tau_q$ , black is  $\tau_{th}$ . In each panel, the left dashed line is the location of the tunneling transition and the right dashed line is the location of the excitation transition.  $\tau_q$  is meaningful only in-between the two. The nonanalyticity in  $\tau_q$  comes from a change in the argmin of Eq. (32).

the event that  $\sigma' = \sigma$  is unlikely), thus the timescale  $\tau_{unstr}$  for this unstructured search is

$$\tau_{unstr} \sim e^{N\epsilon^2}. \quad (33)$$

Clearly,  $\tau_{unstr} \ll \tau_{th}$ , and indeed, the unstructured search is the most efficient *classical* algorithm for the REM since the energy levels are completely independent. Unfortunately, we find that  $\tau_{unstr} \ll \tau_q$  throughout the tunneling phase as well. However, this observation is specific to the REM and not representative of real optimization problems, which do have distance-dependent correlations between energy levels. The correlations lead to energy barriers with smaller (but still macroscopic) heights and widths, improving the runtime of both activated dynamics and quantum tunneling relative to the unstructured search. Indeed, we show in Sec. IV that tunneling can be superior to both unstructured search and activated dynamics in the more realistic *finite-p*-spin model.

### C. Numerical results

As noted above, we made uncontrolled approximations in deriving Eqs. (3) and (4), namely, using the Schrieffer-Wolff

transformation and truncating the perturbation series via the FSA. Thus, we use exact diagonalization (ED) studies of the REM to confirm that the dynamical phases in Fig. 1 exist.

Closely related work was performed in Ref. [16], in the context of eigenstate phases of the quantum REM. The tunneling transition is also a transition between localized and extended eigenstates in the configuration space, and the authors used multiple measures of localization to detect the transition numerically using ED. They also compared the location of the transition obtained by ED to the location obtained by evaluating the forward-scattering expression for  $\langle \sigma' | V_{\text{eff}} | \sigma \rangle$  numerically. They found very good agreement, supporting the validity of Schrieffer-Wolff and the FSA. Here, we supplement these results by studying explicitly *dynamical* properties of the quantum REM.

We construct instances of  $H \equiv H_{\text{REM}} - \Gamma \sum_i \hat{\sigma}_i^x$ , where  $H_{\text{REM}}$  is diagonal in the  $\hat{\sigma}^z$  basis with entries independently distributed according to Eq. (15). We then perform a “quench” simulation: find the  $\hat{\sigma}^z$  eigenstate  $|\sigma\rangle$  with classical energy density closest to a specified  $\epsilon$ , then compute  $|\Psi(t)\rangle \equiv e^{-iHt} |\sigma\rangle$ . We evaluate two observables at time  $t$ , denoted  $\hat{x}$  and  $\hat{\epsilon}$ , which are the operators corresponding, respectively, to the distance from  $|\sigma\rangle$  and the classical energy density (both are diagonal in the  $\hat{\sigma}^z$  basis). Since we evaluate the full  $2^N \times 2^N$  matrix  $e^{-iHt}$  via exact diagonalization, we are limited to very small system sizes, namely,  $N \leq 14$ .

Results for  $\langle \Psi(t) | \hat{x} | \Psi(t) \rangle$  and  $\langle \Psi(t) | \hat{\epsilon} | \Psi(t) \rangle$  are shown in Fig. 7. Both quantities reach saturated values, denoted  $x_\infty$  and  $\epsilon_\infty$ . For all  $\Gamma$  and  $\epsilon$ ,  $x_\infty$  lies between 0 and  $\frac{1}{2}$  and  $\epsilon_\infty$  lies between  $\epsilon$  and 0. Since we study finite-size systems, the dependence on the parameters is smooth. The difference between the three phases is in the flow of  $x_\infty$  and  $\epsilon_\infty$  as  $N$  increases:

- (i) Trapped phase:  $x_\infty \rightarrow 0$  and  $\epsilon_\infty \rightarrow \epsilon$ .
- (ii) Tunneling phase:  $x_\infty \rightarrow \frac{1}{2}$  and  $\epsilon_\infty \rightarrow \epsilon$ .
- (iii) Excited phase:  $x_\infty \rightarrow \frac{1}{2}$  and  $\epsilon_\infty \not\rightarrow \epsilon$ .

In Fig. 7, the behavior of the system is consistent with the trapped phase at small  $\Gamma$  (left panel), the tunneling phase at intermediate  $\Gamma$  (middle panel), and the excited phase at large  $\Gamma$  (right panel). Furthermore, it appears that the timescale on which the system approaches  $x_\infty$  is indeed exponential with  $N$  in the tunneling phase.

The middle panel, which shows behavior consistent with the tunneling phase, is at a  $(\Gamma, \epsilon)$  point which Fig. 1 predicts is in the trapped phase. There are two possible explanations for the discrepancy. First, we cannot rule out that the system becomes trapped at larger system sizes. Our analytical arguments focus only on the leading exponential order of the quantities relevant for computing the phase boundaries, and polynomial prefactors may lead to significant finite-size effects for  $N \leq 14$ . Second, since the analysis of Sec. III A is based on perturbation theory in  $\Gamma$ , it may only be quantitatively correct at small  $\Gamma$ . Regardless, previous numerical work on the eigenstate phases provides separate finite-size estimates of the phase boundaries, next to which our results are quantitatively consistent [16,38].

We are unable to conclusively identify the small- $\Gamma$  portion of the excited phase shown in Fig. 1. Figure 8 plots  $\langle \Psi(t) | \hat{x} | \Psi(t) \rangle$  and  $\langle \Psi(t) | \hat{\epsilon} | \Psi(t) \rangle$  for a  $(\Gamma, \epsilon)$  point well within that region of the phase diagram.  $x_\infty = \frac{1}{2}$  as expected, signifying that the system is not trapped within a cluster.

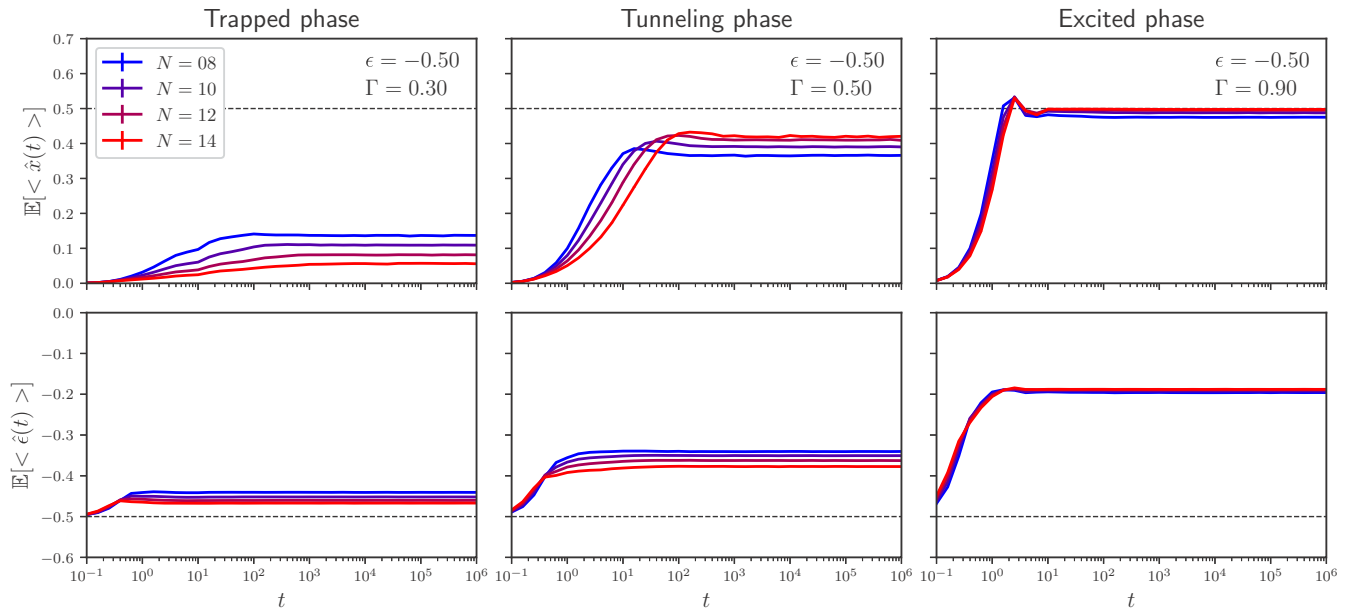


FIG. 7. Quantum dynamics in the REM, starting from a classical state at energy density  $\epsilon$ . (Top) Average distance relative to the initial configuration  $\mathbb{E}[\langle \hat{x}(t) \rangle] \equiv \mathbb{E}[\langle \Psi(t) | \hat{x} | \Psi(t) \rangle]$ . Three representative  $\Gamma$  at fixed  $\epsilon$  are shown. The dashed line is  $x = \frac{1}{2}$ . (Bottom) Average classical energy density  $\mathbb{E}[\langle \hat{\epsilon}(t) \rangle] \equiv \mathbb{E}[\langle \Psi(t) | \hat{\epsilon} | \Psi(t) \rangle]$ . The same  $\Gamma$  and  $\epsilon$  as in the top panels are used. The dashed line is the initial energy density  $\epsilon$ . Statistical error bars are smaller than the linewidths in all panels. The “tunneling phase” label refers to the observed behavior, not the location relative to the perturbatively obtained phase boundaries in Fig. 1 (see the discussion in Sec. III C).

However,  $\epsilon_\infty$  shows a slight downward flow as  $N$  increases, and it is not clear whether the timescale on which the system approaches  $x_\infty$  is scaling exponentially or subexponentially. It is possible that the system is excited out of clusters, but it may instead be in the tunneling phase with very strong finite-size effects.

Regardless, we do find strong signatures of all three dynamical phases at lower  $\epsilon$ , as shown in Fig. 7.

#### IV. FINITE- $p$ CORRECTIONS

Here, we extend the above analysis to the finite- $p$ -spin model and demonstrate two results: first, the leading-in- $1/p$  corrections are small and thus the picture we have developed is robust to small but nonzero cluster sizes (Secs. IV A–IV C); and second, tunneling can now outperform all known classical algorithms in certain regions of the tunneling phase, including the unstructured search which is optimal for the REM (Sec. IV D) [40]. This analysis has the added complexity that clusters no longer consist of isolated configurations and the  $\epsilon(\sigma'')$  that enter into  $V_{\text{eff}}$  are no longer independent. Section IV A discusses the geometry of the clusters, Sec. IV B calculates the intercluster tunneling amplitudes, Sec. IV C describes the resulting phase diagram, and Sec. IV D calculates the tunneling rates and compares to classical algorithms.

##### A. Geometry of clusters

First, we consider the geometry of clusters at large but finite  $p$ . The radius of a cluster is no larger than  $x^*(\epsilon)$ , where  $x^*(\epsilon)$  is the smallest nonzero root of the equation  $g(x, \epsilon) = 0$  (see Fig. 3).  $g(x, \epsilon)$  is given by Eq. (12), from which we find

that

$$x^*(\epsilon) \sim e^{-O(p)}, \tag{34}$$

$$x^{**}(\epsilon) \sim x^{**}(\epsilon)_{p \rightarrow \infty} - e^{-O(p)}, \tag{35}$$

where  $x^{**}(\epsilon)_{p \rightarrow \infty}$  is the REM value of  $x^{**}(\epsilon)$ . Thus, the distance between clusters is much larger than their radii at large but finite  $p$ . The macroscopic energy barriers which separate clusters are also quantified by the FPP. Setting  $g(x, \epsilon') = 0$  at fixed  $x$  gives us a bound  $\epsilon'_-(x)$ : all configurations at distance  $x$  have energy densities greater than  $\epsilon'_-(x)$ . For  $x$  between  $x^*(\epsilon)$  and  $x^{**}(\epsilon)$ ,

$$\begin{aligned} \epsilon'_-(x) &= (1 - 2x)^p \epsilon \\ &\quad - \sqrt{[1 - (1 - 2x)^{2p}][ -x \ln x - (1 - x) \ln(1 - x) ]}, \end{aligned} \tag{36}$$

which is strictly greater than  $\epsilon$ . The shape of  $\epsilon'_-(x)$  is plotted in Fig. 9. At large  $p$ , the peak of the barrier is at

$$x_{\text{peak}} \sim \frac{\ln p}{4p}, \tag{37}$$

with a height of

$$\epsilon'_-(x_{\text{peak}}) - \epsilon \sim |\epsilon| - \frac{\ln p}{2\sqrt{p}}. \tag{38}$$

Given that each cluster has a nonzero radius, and thus that the tunneling paths between  $|\sigma\rangle$  and  $|\sigma'\rangle$  have portions lying within the clusters, the forward-scattering expansion of  $\langle \sigma' | V_{\text{eff}} | \sigma \rangle$  is more complicated than what is given in Eq. (18) for the REM. Recall that the REM expression is simply a sum

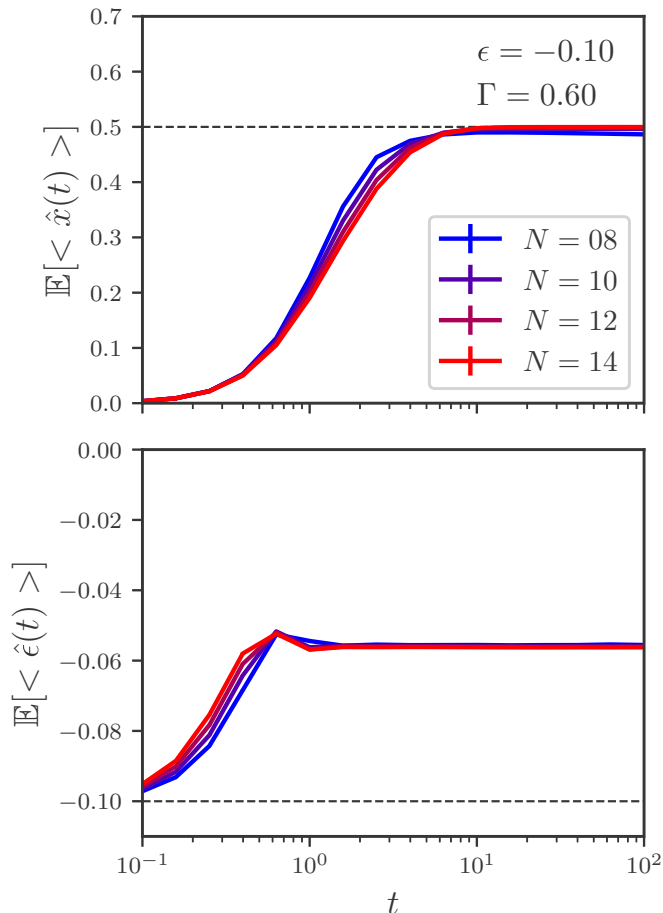


FIG. 8. Quantum dynamics in the REM starting from a high-energy classical state. (Top) Average distance relative to the initial configuration  $\mathbb{E}[\langle \hat{x}(t) \rangle] \equiv \mathbb{E}[\langle \Psi(t) | \hat{x} | \Psi(t) \rangle]$ . The dashed line is  $x = \frac{1}{2}$ . (Bottom) Average classical energy density  $\mathbb{E}[\langle \hat{\epsilon}(t) \rangle] \equiv \mathbb{E}[\langle \Psi(t) | \hat{\epsilon} | \Psi(t) \rangle]$ . The dashed line is the initial energy density  $\epsilon$ . Statistical error bars are smaller than the linewidths in both panels.

over paths, each term of which takes the same form:

$$\langle \sigma' | V_{\text{eff}} | \sigma \rangle \sim \sum_P \Gamma \prod_{\sigma'' \in P} \frac{\Gamma}{N(\epsilon - \epsilon(\sigma''))}. \quad (18)$$

At finite  $p$ , however, paths acquire different weights and different energy denominators depending on how they pass within clusters (see Appendix D). We are unable to quantitatively account for these intracluster contributions to  $\langle \sigma' | V_{\text{eff}} | \sigma \rangle$ . However, we do not expect them to modify the qualitative picture developed for the REM since the cluster size is much smaller than the separation between clusters. The majority of each tunneling path lies outside of the clusters and contributes factors of  $\Gamma/N(\epsilon - \epsilon(\sigma''))$  exactly as in Eq. (18). We expect the intracluster portions to give additive contributions to  $\gamma(x, \epsilon' | \epsilon)$  which scale as  $O(x^*(\epsilon))$ . Since  $x^*(\epsilon)$  is exponentially small in  $p$ , these effects are negligible compared to those corrections which we describe below, and we shall continue to use Eq. (18) for  $\langle \sigma' | V_{\text{eff}} | \sigma \rangle$ .

We expect the shifts in energy that arise from diagonalizing the intracluster effective Hamiltonians ( $H_{\text{eff}}^{(\alpha)}$  in Fig. 4) to be subleading as well. The number of states within a cluster

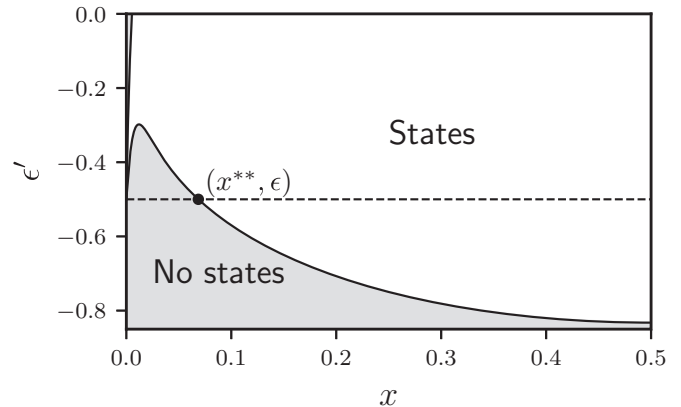


FIG. 9. Regions of the  $(x, \epsilon')$  plane in which the FPP is positive (white) and negative (gray), for  $p = 100$  and  $\epsilon = -0.5$  (dashed line). Configurations at distance  $x$  have energy densities exclusively in the white region.  $x^{**}(\epsilon)$  is marked, and  $x^*(\epsilon)$  is too close to 0 to be visible on this scale. The curve separating white and gray is  $\epsilon'_-(x)$  [Eq. (36)] [the corresponding upper root  $\epsilon'_+(x)$  is barely visible in the top-left corner].

scales as

$$\exp\left(N \max_{x \in [0, x^*(\epsilon)]} g(x, \epsilon)\right) \sim \exp(Ne^{-O(p)}). \quad (39)$$

Assuming that the hybridization energy among these states scales as the logarithm of Eq. (39), we find that the energy densities shift by amounts exponentially small in  $p$ . We neglect this effect in what follows.

## B. Intercluster tunneling

The second difference between the finite- $p$  model and the REM is that the factors of  $\Gamma/N(\epsilon - \epsilon(\sigma''))$  are no longer independent. We make the approximation

$$\epsilon(\sigma'') \approx \mathbb{E}[\epsilon(\sigma'')]_{\substack{\epsilon(\sigma) = \epsilon, \\ \epsilon(\sigma') = \epsilon'}}, \quad (40)$$

which accounts for the correlations between  $\sigma''$  and the two endpoints but neglects further correlations among the path amplitudes. The error in making this approximation is only  $O(1/p)$ , as we now show.

Consider those tunneling paths along which the energy densities are a given function  $\epsilon(y)$  ( $0 < y < x$ ). Let the number of such paths be denoted  $\mathcal{N}(\epsilon(y))$ . We can then write Eq. (18) as a path integral over  $\epsilon(y)$ :

$$|\langle \sigma' | V_{\text{eff}} | \sigma \rangle| \sim \int \mathcal{D}\epsilon(y) \mathcal{N}(\epsilon(y)) e^{N \int_0^x dy \ln \frac{\Gamma}{N|\epsilon - \epsilon(y)|}}. \quad (41)$$

We enforce the conditioning on  $\epsilon(\sigma)$  and  $\epsilon(\sigma')$  through boundary conditions:  $\epsilon(y=0) = \epsilon$  and  $\epsilon(y=x) = \epsilon'$ . Equation (41) is formally exact since now  $\mathcal{N}(\epsilon(y))$  is the random variable which depends on the disorder realization. Importantly, we integrate only over  $\epsilon(y)$  that are nowhere equal to  $\epsilon$  since the tunneling paths lie outside of the clusters. We next make another ‘‘annealed’’ approximation:

$$\mathcal{N}(\epsilon(y)) \rightarrow \mathbb{E}[\mathcal{N}(\epsilon(y))] = (Nx)! P(\epsilon(y)), \quad (42)$$

where  $P(\epsilon(y))$  is the *probability* of a given path having energy densities  $\epsilon(y)$ . Calculating  $P(\epsilon(y))$  exactly is intractable, yet we can infer the scaling with both  $N$  and  $p$  from the covariance matrix of the classical energies [Eq. (7)]:

$$\mathbb{E}[\epsilon(\sigma_1)\epsilon(\sigma_2)] = \frac{1}{2N}[1 - 2x(\sigma_1, \sigma_2)]^p. \quad (43)$$

At large  $p$ , the right-hand side is independent of  $p$  on a length scale  $x(\sigma_1, \sigma_2) \sim O(1/p)$ . Thus,  $1/p$  can be identified as the

“correlation length” for the classical energies. If  $\epsilon(y)$  deviates from its mean throughout a distance  $x$ , then since the number of correlation lengths involved is  $px$ , this roughly corresponds to a number  $px$  of independent fluctuations, each of which is exponentially rare in  $N$ . Thus,

$$P(\epsilon(y)) \sim e^{-Npc(\epsilon(y))}, \quad (44)$$

where  $c(\epsilon(y))$  is independent of both  $N$  and  $p$ , and  $c(\mathbb{E}[\epsilon(y)]) = 0$ . Thus, Eq. (41) becomes

$$|\langle \sigma' | V_{\text{eff}} | \sigma \rangle| \sim \int \mathcal{D}\epsilon(y) e^{N(x \ln \frac{x\Gamma}{e} - \int_0^x dy \ln |\epsilon - \epsilon(y)| - pc(\epsilon(y)))}. \quad (45)$$

At  $p \rightarrow \infty$ , the saddle point of the path integral is at  $\epsilon^0(y) \equiv \mathbb{E}[\epsilon(y)]$  (see Sec. III A). To compute the correction from large but finite  $p$ , write  $\epsilon(y) = \epsilon^0(y) + \delta\epsilon(y)$  and expand the exponent:

$$-\int_0^x dy \ln |\epsilon - \epsilon(y)| - pc(\epsilon(y)) \sim -\int_0^x dy \ln |\epsilon - \epsilon^0(y)| + \int_0^x dy \frac{\delta\epsilon(y)}{\epsilon - \epsilon^0(y)} - p \frac{\partial^2 c(\epsilon^0(y))}{\partial \epsilon(y) \partial \epsilon(z)} \delta\epsilon(y) \delta\epsilon(z). \quad (46)$$

There are no first derivatives of  $c(\epsilon(y))$  because  $\epsilon^0(y)$  is the location of its minimum. Competition between the second and third terms in Eq. (46) determines the location of the saddle point. They are comparable for  $\delta\epsilon(y) \sim O(1/p)$ , which changes the value of the exponent at the saddle point by  $O(1/p)$ . This is the error made to  $\langle \sigma' | V_{\text{eff}} | \sigma \rangle$ , or rather  $\gamma(x, \epsilon' | \epsilon)$ , by approximating  $\epsilon(y) \approx \mathbb{E}[\epsilon(y)]$ , i.e.,  $\delta\epsilon(y) \approx 0$ . It is indeed small at large  $p$ .

Next, consider how  $\mathbb{E}[\epsilon(y)]$  is modified by correlations with  $\epsilon(\sigma)$  and  $\epsilon(\sigma')$ . A straightforward calculation given in Appendix C shows that

$$\mathbb{E}[\epsilon(y)] \Big|_{\substack{\epsilon(\sigma) = \epsilon \\ \epsilon(\sigma') = \epsilon'}} = \frac{\mu(y) - \mu(x)\mu(x-y)}{1 - \mu(x)^2} \epsilon + \frac{\mu(x-y) - \mu(x)\mu(y)}{1 - \mu(x)^2} \epsilon', \quad (47)$$

where  $\mu(z) \equiv (1 - 2z)^p$ . Then,  $|\langle \sigma' | V_{\text{eff}} | \sigma \rangle| \sim e^{-N\gamma(x, \epsilon' | \epsilon)}$  with

$$\gamma(x, \epsilon' | \epsilon) = -x \ln \frac{x\Gamma}{e|\epsilon|} + \int_0^x dy \ln \left| 1 - \frac{\mu(y) - \mu(x)\mu(x-y)}{1 - \mu(x)^2} - \frac{\mu(x-y) - \mu(x)\mu(y)}{1 - \mu(x)^2} \frac{\epsilon'}{\epsilon} \right|. \quad (48)$$

In the large- $p$  limit, assuming  $x \sim O(1)$  [41],

$$\gamma(x, \epsilon' | \epsilon) \sim -x \ln \frac{x\Gamma}{e|\epsilon|} - \frac{K}{p}, \quad (49)$$

where

$$K \equiv \frac{\pi^2}{12} - \int_0^\infty dy \ln \left( 1 - \frac{\epsilon'}{\epsilon} e^{-2y} \right). \quad (50)$$

The correction relative to Eqs. (20) and (27) for the REM is indeed  $O(1/p)$ .

### C. Phase boundaries

The same arguments as in Sec. III A give Eq. (3) as a necessary condition for tunneling between clusters to occur:

$$\max_{x \in [x^{**}(\epsilon), 1-x^{**}(\epsilon)]} [g(x, \epsilon) - \gamma(x, \epsilon)] > 0, \quad (3)$$

only now with Eq. (12) for  $g(x, \epsilon)$  and Eq. (49) for  $\gamma(x, \epsilon)$ . Since  $x^{**}(\epsilon) \sim O(1)$  with respect to  $p$ ,  $g(x, \epsilon)$  differs from the REM expression [Eq. (16)] by an amount exponentially small in  $p$  and  $\gamma(x, \epsilon)$  differs from Eq. (20) by  $O(1/p)$ . Thus, the same Eq. (24) determines the tunneling transition in the finite- $p$  models to within  $O(1/p)$ .

The excitation transition in the finite- $p$  models is more subtle. Within the approximations that we have made, Eq. (48) holds for all  $x$ . However, Eq. (49) only holds for  $x \sim O(1)$ . At  $x \ll 1/p$ , we instead have

$$\gamma(x, \epsilon' | \epsilon) \sim -x \ln \frac{e\Gamma}{2p^2|\epsilon|} + \left( x + \frac{\epsilon' - \epsilon}{2p^2|\epsilon|x} \right) \ln \left( x + \frac{\epsilon' - \epsilon}{2p^2|\epsilon|x} \right) - \frac{\epsilon' - \epsilon}{2p^2|\epsilon|x} \ln \frac{\epsilon' - \epsilon}{2p^2|\epsilon|x}, \quad (51)$$

$$g(x, \epsilon' | \epsilon) - 2\gamma(x, \epsilon' | \epsilon) \sim x \ln \frac{e^3\Gamma^2}{4p^4\epsilon^2x} - \frac{(\epsilon' - \epsilon - 2p|\epsilon|x)^2}{4px} - 2 \left( x + \frac{\epsilon' - \epsilon}{2p^2|\epsilon|x} \right) \ln \left( x + \frac{\epsilon' - \epsilon}{2p^2|\epsilon|x} \right) + \frac{\epsilon' - \epsilon}{p^2|\epsilon|x} \ln \frac{\epsilon' - \epsilon}{2p^2|\epsilon|x}. \quad (52)$$



The maximum over  $\epsilon'$  is at  $\epsilon' - \epsilon = 2p|\epsilon|x + O(1/p)$ , and

$$\max_{\epsilon'} [g(x, \epsilon'|\epsilon) - 2\gamma(x, \epsilon'|\epsilon)] = x \ln \frac{e^3 \Gamma^2}{4p^2 \epsilon^2 x} + O\left(\frac{1}{p^3}\right). \quad (53)$$

The maximum over  $x$  is then at

$$x_m = \frac{e^2 \Gamma^2}{4p^2 \epsilon^2} \quad (54)$$

and, importantly, the value of the maximum is positive. Equation (4) is never satisfied, and the system is excited to higher classical energy densities. However, this does not imply that the system escapes from its starting cluster: Eq. (54) indicates that the system moves through a distance  $O(1/p^2)$ , and thus the classical energy changes by  $O(1/p)$ , which is much too small to surmount the energy barrier at distance  $O(1/p)$  and height  $O(1)$  [Fig. 9 and Eq. (37)]. To check if the system is excited out of clusters, one should modify Eq. (4) to maximize over  $x \gtrsim O(1/p)$ . Then, the finite- $p$  corrections to  $\gamma(x, \epsilon'|\epsilon)$  are subleading, and Eq. (4) determines the excitation transition to within  $O(1/p)$ .

The thermodynamic transition into a quantum paramagnetic phase, constituting another excitation transition, is also present at finite  $p$  [42,43]. The phase boundary terminates at energy density  $-O(1/\sqrt{p})$  [42]. On the other hand, by setting  $\min_x [g(x, \epsilon)] = 0$ , one finds that

$$\epsilon_d \sim -\sqrt{\frac{\ln p}{p}}. \quad (55)$$

Since the energy landscape of the  $p$ -spin model is organized into clusters only below  $\epsilon_d$ , the large- $\Gamma$  excitation transition is present at all  $\epsilon$  for which energy matching is nontrivial.

#### D. Tunneling rates

Tunneling continues to be exponentially faster than activation for large but finite  $p$ . This follows from the fact that for  $x \sim O(1)$ , the density of states has exponent

$$g(x, \epsilon) = \max_{\epsilon'} [g(x, \epsilon'|\epsilon)] - \epsilon^2 + e^{-O(p)}, \quad (56)$$

and the effective coupling has exponent

$$\gamma(x, \epsilon) = \gamma(x, \epsilon' \neq \epsilon|\epsilon) - O\left(\frac{1}{p}\right). \quad (57)$$

See Eqs. (12) and (49), respectively. Thus,  $N^{-1} \ln \tau_q$  is corrected only by  $O(1/p)$  relative to the REM value. Similarly, given that the height of the energy barriers is close to the REM value [Eq. (38)] and the distance from the cluster centers is close to 0 [Eq. (37)],  $N^{-1} \ln \tau_{th}$  also has corrections small in  $1/p$ . These corrections cannot compensate for the leading-order result that  $N^{-1} \ln \tau_q < N^{-1} \ln \tau_{th}$  by  $O(1)$ .

The comparison between tunneling and unstructured search is more interesting at finite  $p$ , and indeed tunneling is now *faster* in a portion of the tunneling phase. It is known that the entropy density of the  $p$ -spin model is exactly equal to the REM expression [Eq. (30)] for all  $\epsilon > \epsilon_s$  [23], thus, the runtime for unstructured search is again given by Eq. (33). From

Eqs. (56) and (57),

$$\begin{aligned} \frac{1}{N} \ln \tau_q &= - \max_{x \in [x^{**}(\epsilon), 1-x^{**}(\epsilon)]} [g(x, \epsilon) - 2\gamma(x, \epsilon)] \\ &= - \max_{x \in [x^{**}(\epsilon), 1-x^{**}(\epsilon)]} \left[ \max_{\epsilon'} [g(x, \epsilon'|\epsilon) - 2\gamma(x, \epsilon'|\epsilon)] \right] \\ &\quad + \epsilon^2 - O\left(\frac{1}{p}\right). \end{aligned} \quad (58)$$

Precisely on the excitation boundary, the first term on the right-hand side is zero by definition [44]. Thus,  $N^{-1} \ln \tau_q$  is smaller than  $\epsilon^2 = N^{-1} \ln \tau_{unstr}$  by  $O(1/p)$  on the boundary. By continuity, tunneling is faster than unstructured search for an entire range of  $\Gamma < \Gamma_{exc}(\epsilon)$  in the tunneling phase.

As trivial as the unstructured search is, we are not aware of any classical search algorithm that can perform better at large  $p$ , where the energy landscape is similar to the uncorrelated landscape of the REM (for which unstructured search must be optimal since the energy levels are independent). Thus, to the best of our knowledge, the quantum energy-matching algorithm presented here is more efficient than any classical method for solving matching problems in this regime. The algorithm does not require exponential precision or knowledge of the level structure. While one must be able to identify the tunneling phase and tune the transverse field close to the excitation boundary, these requirements do not scale with the *size* of the problem. This approach thus has a fundamental advantage over other quantum algorithms, which require unreasonable amounts of control [45].

#### V. CONCLUSION

We have demonstrated that the application of quantum dynamics to energy-matching problems is severely constrained, but that it is capable of outperforming all known classical algorithms in the tunneling phase. Our test bed is the  $p$ -spin model at large  $p$ , for which we use perturbation theory and exact diagonalization. The goal of energy matching is to find states at a target energy given one state at the same energy. In general, finding low-energy states is difficult when the energy landscape is rugged, i.e., contains many local minima separated by large barriers. Ideally, starting at a low energy state serves as a “hint” for finding others, but the same ruggedness makes energy matching difficult: in order to find sufficiently distinct target states, any classical algorithm that flips only a few spins per step must return to and explore high-energy states at intermediate steps. This raises the possibility for quantum dynamics to provide a speedup over classical algorithms by tunneling through those energy barriers.

In many regimes of transverse-field strength and target energy, quantum dynamics *cannot* succeed in energy matching even if allowed arbitrarily long runtime. If the applied field is too weak, the system never tunnels out of its starting cluster of states. If the applied field is too strong, the system is excited to higher (classical) energy densities and thus never locates target states. Only at intermediate fields does the system successfully tunnel between clusters while roughly conserving the classical energy density. These possibilities constitute distinct dynamical phases, demarcated by sharp phase boundaries in the thermodynamic limit (Fig. 1). The combination of three

features of the energy landscape underlies the physics: the macroscopic height of energy barriers, the macroscopic width of those barriers, and the exponential number of clusters in which to tunnel (referred to in the spin-glass community as the “complexity” of the model). These features are typical in many of the well-studied classes of optimization problems [46–49], thus, we expect the results of this paper to apply more generally.

A remarkable feature of the dynamical phase diagram for the  $p$ -spin model is that there is no tunneling phase at energies close to the ground state. The barriers between such low-energy states are too large and wide, and the complexity of clusters is too small. As a result, it is not possible to tunnel between ground or near-ground states of the  $p$ -spin model for any applied transverse field. However, ground-state tunneling may be possible in problems with larger ground-state complexity, e.g., the satisfiability problem [2,27,50], and we leave this as an interesting open question.

If the system is in its tunneling phase, the timescale for intercluster tunneling is exponential in system size. Thus, quantum dynamics never solves the energy-matching problem in polynomial time. However, we find that tunneling is exponentially faster than certain simple classical algorithms, such as Metropolis Monte Carlo simulation, throughout the tunneling phase. Furthermore, tunneling outperforms even the unstructured search close to the excitation boundary at large but finite  $p$ . The unstructured search is the optimal classical algorithm in the infinite- $p$  limit, and we do not know of any classical method that could perform better at large  $p$  [40].

It would in particular be interesting to compare with the performance of quantum Monte Carlo (which is still a classical algorithm). Recent work suggests that quantum Monte Carlo algorithms can perform as well as incoherent tunneling processes [14]. However, that analysis is limited to “single-path” tunneling in integrable and/or highly symmetric models, and it is unclear whether a similar statement can be made for nonintegrable disordered models.

One drawback to the approach presented here is that the rates we find are exponentially slower than those which could be achieved by a universal quantum computer implementing a variation of Grover’s algorithm [45]. Indeed, in the REM ( $p \rightarrow \infty$  limit), there are strong arguments that *no* algorithm can outperform Grover’s since the energy landscape is unstructured [51]. However, Grover’s algorithm requires both exponential tuning precision and detailed knowledge of the adiabatic level structure. The advantage of the algorithm presented here is that it requires neither. Effective Hamiltonians of the form in Eq. (1) are realizable in near-term quantum annealing devices [19,20], and the energy-matching protocol proposed here is straightforward to implement and test on them.

#### ACKNOWLEDGMENTS

We would like to thank A. Chandran, D. Huse, F. Krzakala, A. Pal, F. Ricci-Tersenghi, O. Sattath, A. Scardicchio, and, in particular, V. Smelyanskiy for valuable discussions. We would also like to thank the Simons Center for Geometry and Physics and the Kavli Institute for Theoretical Physics for hospitality during which part of this work was completed. C.L.B. acknowledges the support of the NSF through a Graduate Research Fellowship, Grant No. DGE-1256082. C.R.L. acknowledges

support from the Sloan Foundation through a Sloan Research Fellowship and the NSF through Grant No. PHY-1656234.

#### APPENDIX A: NP-COMPLETENESS OF GROUND-STATE ENERGY MATCHING IN $K$ -CSPs

Here, we show that the matching problem for the ground states of  $K$ -body constraint satisfaction problems (CSPs) is NP-complete when  $K \geq 4$ . To be precise, we define the decision problem [52]:

##### Matching $K$ -CSP

**Input:**  $(H, \sigma, x)$ , where  $H$  is a classical energy function encoding a  $K$ -CSP,  $\sigma \in \{0, 1\}^N$  is a satisfying ground state of  $H$ , and  $x \in (0, 1)$  is a fractional Hamming distance.

**Output:** YES if there exists a satisfying state  $\sigma'$  such that  $x(\sigma, \sigma') \geq x$ .

It is clear that matching  $K$ -CSP is in NP. In order to show that it is NP-complete, we construct a reduction from  $(K - 1)$ -CSP to matching  $K$ -CSP. Since  $(K - 1)$ -CSP is NP-complete for  $K - 1 \geq 3$ , this proves that matching  $K$ -CSP is NP-complete for  $K \geq 4$ .

Let  $H_p$  be an instance of  $(K - 1)$ -CSP acting on  $N$  bits  $s_i$ . We construct an instance of matching  $K$ -CSP by introducing  $M = 2N$  auxiliary bits  $\tau_j$  and defining the  $K$ -CSP Hamiltonian:

$$\tilde{H}_p = H_p \otimes \left( \frac{1}{2M} \sum_{j=1}^M (1 + \tau_j) \right) + I \otimes \left( M - \frac{1}{M} \sum_{j,j'=1}^M \tau_j \tau_{j'} \right), \quad (\text{A1})$$

where the left factor in each tensor product refers to the original bits and the right factor to the auxiliary bits. The first term interpolates between 0 when all  $\tau_j = -1$  and  $H_p$  when  $\tau_j = +1$ . The second term contributes an energy cost when any two auxiliary bits are antialigned. Clearly,  $\tilde{H}_p$  involves no more than  $K$ -body interactions.

A state  $|s\rangle \otimes |\tau\rangle$  has zero energy under  $\tilde{H}_p$  if it satisfies *one* of the following conditions:

- (1) All  $\tau_j = -1$ .
- (2) All  $\tau_j = +1$  and  $H_p |s\rangle = 0$ .

All other states have positive energy under  $\tilde{H}_p$ . The given zero-energy state of the matching problem is  $|-\rangle \otimes |-\rangle$ , i.e., all original and auxiliary bits have value  $-1$ , which satisfies condition 1.

It is then clear that  $H_p$  has a zero-energy state if and only if  $\tilde{H}_p$  has a zero-energy state at distance greater than  $\frac{1}{3}$  from  $|-\rangle \otimes |-\rangle$ . If  $H_p$  has a zero-energy state  $|s\rangle$ , then  $\tilde{H}_p(|s\rangle \otimes |+\rangle) = 0$  and  $|s\rangle \otimes |+\rangle$  is at distance greater than  $\frac{2}{3}$  from  $|-\rangle \otimes |-\rangle$ . If  $\tilde{H}_p$  has a zero-energy state  $|s\rangle \otimes |\tau\rangle$  at sufficient distance, then  $|s\rangle \otimes |\tau\rangle$  cannot satisfy condition 1 as that would imply a distance less than  $\frac{1}{3}$  and so must satisfy condition 2, i.e.,  $|s\rangle$  is a satisfying configuration of  $H_p$ . We thus have the desired mapping from any instance  $H_p$  of  $(K - 1)$ -CSP to an instance  $(\tilde{H}_p, |-\rangle \otimes |-\rangle, 1/3)$  of matching  $K$ -CSP.

## APPENDIX B: FPP IN THE CANONICAL AND MICROCANONICAL ENSEMBLES

The Franz-Parisi potential is typically presented in the canonical ensemble [26,28–30]. It represents the free energy of a system at inverse temperature  $\beta'$  constrained to be at distance  $x$  from a system in equilibrium at inverse temperature  $\beta$ :

$$v(x, \beta' | \beta) = \mathbb{E} \left[ \text{Tr}_\sigma \left[ \frac{e^{-N\beta\epsilon(\sigma)}}{Z(\beta)} \ln \text{Tr}_{\sigma'} [\delta_{x,x(\sigma,\sigma')} e^{-N\beta'\epsilon(\sigma')}] \right] \right], \quad (\text{B1})$$

where  $Z(\beta)$  is the partition function. However, in this paper it is more natural to use  $g(x, \epsilon' | \epsilon)$  as defined in Eq. (10). Here, we show the relationship between the two.

First consider

$$\mathbb{E} \left[ \text{Tr}_\sigma \left[ \frac{e^{-N\beta\epsilon(\sigma)}}{Z(\beta)} f(\sigma) \right] \right]$$

for a function  $f(\sigma)$  growing slower than exponential with  $N$  but otherwise arbitrary. The trace is dominated by those  $\sigma$  at the energy density  $\epsilon$  which maximizes  $s(\epsilon) - \beta\epsilon$ , where  $s(\epsilon)$  is the disorder-averaged entropy density [the sample-to-sample fluctuations about  $s(\epsilon)$  are expected to vanish as  $N \rightarrow \infty$ ]. The configurations at other energy densities collectively give an exponentially small contribution. Thus, as  $N \rightarrow \infty$ ,

$$\mathbb{E} \left[ \text{Tr}_\sigma \left[ \frac{e^{-N\beta\epsilon(\sigma)}}{Z(\beta)} f(\sigma) \right] \right] \sim \mathbb{E} \left[ \text{Tr}_\sigma \left[ \frac{\delta(\epsilon - \epsilon(\sigma))}{\mathcal{N}(\epsilon)} f(\sigma) \right] \right], \quad (\text{B2})$$

where  $\epsilon = \arg \max_z [s(z) - \beta z]$  and  $\mathcal{N}(\epsilon) = e^{Ns(\epsilon)}$ .

Equation (B2) applies to  $v(x, \beta' | \beta)$  with

$$f(\sigma) = \ln \text{Tr}_{\sigma'} [\delta_{x,x(\sigma,\sigma')} e^{-N\beta'\epsilon(\sigma')}]. \quad (\text{B3})$$

In fact, since all configurations  $\sigma$  are statistically equivalent, the argument of the trace is independent of  $\sigma$  and we have that

$$v(x, \beta' | \beta) = \mathbb{E} [\ln \text{Tr}_{\sigma'} [\delta_{x,x(\sigma,\sigma')} e^{-N\beta'\epsilon(\sigma')}]]_{\epsilon(\sigma)=\epsilon}, \quad (\text{B4})$$

for an arbitrary reference configuration  $\sigma$ .

Next, write

$$\begin{aligned} & \text{Tr}_{\sigma'} [\delta_{x,x(\sigma,\sigma')} e^{-N\beta'\epsilon(\sigma')}] \\ &= \int d\epsilon' e^{-N\beta'\epsilon'} \text{Tr}_{\sigma'} [\delta_{x,x(\sigma,\sigma')} \delta(\epsilon' - \epsilon(\sigma'))]. \end{aligned} \quad (\text{B5})$$

Assuming, as is typical, that  $g(x, \epsilon' | \epsilon)$  is self-averaging, the trace on the right-hand side is precisely  $e^{Ng(x, \epsilon' | \epsilon)}$  [cf. Eq. (10)]. The integral is dominated by  $\epsilon' = \arg \max_{z'} [g(x, z' | \epsilon) - \beta' z']$ . We thus have the desired relationship between the canonical FPP  $v(x, \beta' | \beta)$  and the microcanonical FPP  $g(x, \epsilon' | \epsilon)$ ;

$$v(x, \beta' | \beta) = g(x, \epsilon' | \epsilon) - \beta' \epsilon', \quad (\text{B6})$$

$$\epsilon = \arg \max_z [s(z) - \beta z], \quad (\text{B7})$$

$$\epsilon' = \arg \max_{z'} [g(x, z' | \epsilon) - \beta' z']. \quad (\text{B8})$$

Note that the relationship between  $\epsilon'$  and  $\beta'$  is not set by the full entropy  $s(\epsilon)$  but rather by the distance-resolved entropy  $g(x, \epsilon' | \epsilon)$ .

Equations (B6)–(B8) express  $v(x, \beta' | \beta)$  in terms of  $g(x, \epsilon' | \epsilon)$  through a Legendre transform. By inverting the transform, one obtains  $g(x, \epsilon' | \epsilon)$  in terms of  $v(x, \beta' | \beta)$ .

## APPENDIX C: CORRELATIONS AMONG ENERGY LEVELS

Here, we calculate the conditional expectation values needed in the main text, namely, for obtaining  $g(x, \epsilon' | \epsilon)$  and  $\gamma(x, \epsilon' | \epsilon)$  at finite  $p$  [Eqs. (12) and (48)]. The joint distribution of energy levels in the classical  $p$ -spin model follows immediately from the fact that the energies are Gaussian distributed with covariances given by Eq. (7). For a subset of levels  $\epsilon \equiv (\epsilon(\sigma_1) \dots \epsilon(\sigma_n))^T$ , the joint distribution is

$$P_n(\epsilon) = e^{-N\epsilon^T \Theta^{-1} \epsilon} \quad (\text{C1})$$

up to normalization, where

$$\Theta \equiv \begin{pmatrix} 1 & (1-2x_{12})^p & \dots & (1-2x_{1n})^p \\ (1-2x_{12})^p & 1 & \dots & (1-2x_{2n})^p \\ \vdots & \vdots & \ddots & \vdots \\ (1-2x_{1n})^p & (1-2x_{2n})^p & \dots & 1 \end{pmatrix}, \quad (\text{C2})$$

with  $x_{ij}$  shorthand for  $x(\sigma_i, \sigma_j)$ . Thus, the distribution for a single level  $\epsilon$  is

$$P_1(\epsilon) = e^{-N\epsilon^2}, \quad (\text{C3})$$

and the distribution for a pair of levels  $\epsilon$  and  $\epsilon'$  is

$$P_2(\epsilon, \epsilon') = \exp \left( -N \frac{\epsilon^2 - 2(1-2x)^p \epsilon \epsilon' + \epsilon'^2}{1 - (1-2x)^{2p}} \right). \quad (\text{C4})$$

The conditional distribution of  $\epsilon'$  given  $\epsilon$  is

$$\frac{P_2(\epsilon', \epsilon)}{P_1(\epsilon)} = \exp \left( -N \frac{[\epsilon' - (1-2x)^p \epsilon]^2}{1 - (1-2x)^{2p}} \right), \quad (\text{C5})$$

from which Eq. (12) follows.

To obtain Eq. (47), we need to average over  $\epsilon''$  in the joint distribution  $P_3(\epsilon, \epsilon', \epsilon'')$ . Write the  $3 \times 3$  matrix  $\Theta^{-1}$  in block-diagonal form as

$$\Theta^{-1} = \begin{pmatrix} A & \mathbf{b} \\ \mathbf{b}^T & c \end{pmatrix}, \quad (\text{C6})$$

where  $A$  is  $2 \times 2$  and acts in the  $(\epsilon, \epsilon')$  subspace,  $\mathbf{b}$  is  $2 \times 1$ , and  $c$  is  $1 \times 1$ . Then,

$$\begin{aligned} & P_3(\epsilon, \epsilon', \epsilon'') \\ &= \exp(-N(c\epsilon''^2 + 2\mathbf{b}^T \epsilon \epsilon'' + \epsilon^T A \epsilon)) \\ &= \exp \left( -Nc \left( \epsilon'' + \frac{\mathbf{b}^T \epsilon}{c} \right)^2 - N\epsilon^T \left( A - \frac{\mathbf{b}\mathbf{b}^T}{c} \right) \epsilon \right), \end{aligned} \quad (\text{C7})$$

where  $\epsilon \equiv (\epsilon \ \epsilon')^T$ . We see that

$$\mathbb{E}[\epsilon'']_{\epsilon, \epsilon'} = -\frac{\mathbf{b}^T \epsilon}{c}. \quad (\text{C8})$$

A direct calculation of the inverse gives that

$$-\frac{1}{c} \mathbf{b} = \begin{pmatrix} \frac{(1-2y)^p - (1-2x)^p(1-2x+2y)^p}{1 - (1-2x)^{2p}} \\ \frac{(1-2x+2y)^p - (1-2x)^p(1-2y)^p}{1 - (1-2x)^{2p}} \end{pmatrix}, \quad (\text{C9})$$

from which Eqs. (47) and (48) follow.

## APPENDIX D: SCHRIEFFER-WOLFF AND FORWARD SCATTERING

### 1. Effective coupling

Consider a Hamiltonian  $H = H_0 + \eta V$  with projectors  $P_0$  and  $Q_0 \equiv 1 - P_0$ , such that  $H_0 = P_0 H_0 P_0 + Q_0 H_0 Q_0$  (i.e.,  $H_0$  is block diagonal with respect to  $P_0$  and  $Q_0$ ). Owing to  $V$ , one may have that  $H$  is not block diagonal. The Schrieffer-Wolff transformation is a unitary transformation  $U$  such that  $H_{\text{eff}} \equiv U H U^\dagger$  is block diagonal, i.e.,  $P_0 H_{\text{eff}} Q_0 =$

$Q_0 H_{\text{eff}} P_0 = 0$ . Here, we discuss how to compute the generator of  $U$  within the forward-scattering approximation, which amounts to retaining the lowest-order-in- $\eta$  terms for each matrix element of  $H_{\text{eff}}$  (different matrix elements may become nonzero at different orders, and we work to lowest nonzero order for each separately).

We use the review by Bravyi *et al.* [34] as our starting point. First, we present their notation.  $V$  is broken into diagonal and off-diagonal parts

$$V_d \equiv P_0 V P_0 + Q_0 V Q_0, \quad V_{od} \equiv P_0 V Q_0 + Q_0 V P_0. \quad (\text{D1})$$

Eigenstates of  $H_0$  are used as the basis, denoted  $|i\rangle, |j\rangle, \dots$  with corresponding energies  $E_i, E_j, \dots$ . The unitary transformation  $U$  is expressed as  $e^S$ , with  $S$  anti-Hermitian. For consistency with [34], we shall use this convention throughout the Appendix, even though we refer to a Hermitian generator in the main text. Finally, Bravyi *et al.* define superoperators

$$\hat{S}(\cdot) \equiv [S, \cdot], \quad (\text{D2})$$

$$\mathcal{L}(\cdot) \equiv \sum_{\substack{i \in P_0 \\ j \in Q_0}} \left( |i\rangle \frac{\langle i | \cdot | j \rangle}{E_i - E_j} \langle j | + |j\rangle \frac{\langle j | \cdot | i \rangle}{E_j - E_i} \langle i | \right), \quad (\text{D3})$$

where the center dot denotes an arbitrary operator. Note that Eq. (D3) only has off-block-diagonal matrix elements. With this notation, the condition that  $H_{\text{eff}}$  be block diagonal gives an equation for the generator  $S$ :

$$S = \mathcal{L} \hat{S}(\eta V_d) + \mathcal{L} \hat{S} \coth(\hat{S})(\eta V_{od}). \quad (\text{D4})$$

The effective Hamiltonian is

$$H_{\text{eff}} = H_0 + \eta V_d + \tanh\left(\frac{\hat{S}}{2}\right)(\eta V_{od}). \quad (\text{D5})$$

See [34] for the derivation.

Equation (D4) is naturally suited to an expansion in  $\eta$ :

$$S = \sum_{n=1}^{\infty} \eta^n S_n, \quad (\text{D6})$$

with each  $S_n$  an anti-Hermitian operator. From Eq. (D4),

$$S_1 = \mathcal{L}(V_{od}), \quad S_2 = \mathcal{L} \hat{S}_1(V_d), \quad S_n = \mathcal{L} \hat{S}_{n-1}(V_d) + \sum_{j=1}^{\infty} a_{2j} \sum_{\substack{n_1, \dots, n_{2j} \geq 1 \\ n_1 + \dots + n_{2j} = n}} \mathcal{L} \hat{S}_{n_1} \dots \hat{S}_{n_{2j}}(V_{od}), \quad (\text{D7})$$

where the last line refers to  $n \geq 3$ , and  $a_{2j}$  is the  $2j$ th Taylor coefficient of  $x \coth x$  about 0. Consider the first two orders:

$$S_1 = \sum_{\substack{i \in P_0 \\ j \in Q_0}} \left( |i\rangle \frac{\langle i | V_{od} | j \rangle}{E_i - E_j} \langle j | + |j\rangle \frac{\langle j | V_{od} | i \rangle}{E_j - E_i} \langle i | \right), \quad (\text{D8})$$

$$S_2 = \sum_{\substack{i \in P_0 \\ j \in Q_0 \\ k \in Q_0}} \left( |i\rangle \frac{\langle i | V_{od} | j \rangle \langle j | V_d | k \rangle}{(E_i - E_k)(E_i - E_j)} \langle k | - |k\rangle \frac{\langle k | V_d | j \rangle \langle j | V_{od} | i \rangle}{(E_k - E_i)(E_j - E_i)} \langle i | \right) \\ + \sum_{\substack{i \in P_0 \\ j \in P_0 \\ k \in Q_0}} \left( |i\rangle \frac{\langle i | V_d | j \rangle \langle j | V_{od} | k \rangle}{(E_i - E_k)(E_j - E_k)} \langle k | - |k\rangle \frac{\langle k | V_{od} | j \rangle \langle j | V_d | i \rangle}{(E_k - E_i)(E_k - E_j)} \langle i | \right). \quad (\text{D9})$$

It becomes very tedious to write higher-order terms, yet one already sees the structure of the expansion. Each term in  $S_n$  has a numerator which is a string of matrix elements of  $V$  and a denominator which is a string of energy differences. Note that each matrix element can be either  $V_d$  or  $V_{od}$ , and each energy denominator is between a state in  $P_0$  and a state in  $Q_0$ . The same structure



holds when we insert the expansion of  $S$  into Eq. (D5) and obtain an expansion of  $H_{\text{eff}}$  in powers of  $\eta$ . For example, two of the fourth-order terms in  $H_{\text{eff}}$  are

$$H_{\text{eff}} = \dots + \frac{\eta^4}{2} |i\rangle \frac{\langle i|V_{od}|j\rangle \langle j|V_{od}|k\rangle \langle k|V_{od}|l\rangle \langle l|V_{od}|m\rangle}{(E_i - E_l)(E_i - E_k)(E_i - E_j)} \langle m| + \frac{\eta^4}{2} |i\rangle \frac{\langle i|V_{od}|j\rangle \langle j|V_{od}|k\rangle \langle k|V_{od}|l\rangle \langle l|V_{od}|m\rangle}{(E_i - E_l)(E_i - E_k)(E_j - E_k)} \langle m| + \dots \quad (\text{D10})$$

Thus far, all that we have presented is completely general. Now, we show how the above equations, which are ultimately used to determine  $H_{\text{eff}}$ , are considerably simplified by making the forward-scattering approximation (FSA). It is best to first consider the FSA within a simple toy problem, a one-dimensional nearest-neighbor tight-binding model with open boundary conditions:

$$H = \sum_{i=0}^L E_i |i\rangle \langle i| - \eta \sum_{i=0}^{L-1} (|i\rangle \langle i+1| + |i+1\rangle \langle i|). \quad (\text{D11})$$

Suppose that  $E_0$  and  $E_L$  are much lower than all other  $E_i$ , and we want to study tunneling from site 0 to site  $L$ . We take  $P_0$  to project onto  $|0\rangle$  and  $|L\rangle$ ,  $V$  to be the hopping term, and aim to compute  $\langle 0|H_{\text{eff}}|L\rangle$  to lowest order in  $\eta$ .

The lowest-order terms are  $O(\eta^L)$  since at least  $L$  applications of the hopping term are required to couple  $|0\rangle$  and  $|L\rangle$ . However, alongside those  $L$ th-order terms which do couple  $|0\rangle$  and  $|L\rangle$ , there are many  $L$ th-order terms in  $H_{\text{eff}}$  which do not. Only terms for which the operator string is  $V_{od}V_d^{L-2}V_{od}$  contribute to  $\langle 0|H_{\text{eff}}|L\rangle$  at  $L$ th order since  $\{|0\rangle, |L\rangle\} \in P_0$  and  $\{|1\rangle, \dots, |L-1\rangle\} \in Q_0$ . For example, taking  $L = 4$ , the first term in Eq. (D10) does contribute to  $\langle 0|H_{\text{eff}}|4\rangle$  at fourth order (taking  $i = 0, j = 1$ , etc.) but the second term does not.

In fact, only two terms at  $L$ th order have the correct operator string:

$$\begin{aligned} \langle 0|H_{\text{eff}}|L\rangle &\sim \frac{\eta^L}{2} \left( \frac{\langle 0|V_{od}|1\rangle \langle 1|V_d|2\rangle \dots \langle L-2|V_d|L-1\rangle \langle L-1|V_{od}|L\rangle}{(E_0 - E_1)(E_0 - E_2) \dots (E_0 - E_{L-2})(E_0 - E_{L-1})} \right. \\ &\quad \left. + \frac{\langle 0|V_{od}|1\rangle \langle 1|V_d|2\rangle \dots \langle L-2|V_d|L-1\rangle \langle L-1|V_{od}|L\rangle}{(E_L - E_1)(E_L - E_2) \dots (E_L - E_{L-2})(E_L - E_{L-1})} \right) + O(\epsilon^{L+1}). \end{aligned} \quad (\text{D12})$$

If  $E_0 = E_L$ , the final expression is particularly simple:

$$\langle 0|H_{\text{eff}}|L\rangle \sim \eta \prod_{i=1}^{L-1} \frac{\eta}{E_0 - E_i}. \quad (\text{D13})$$

Compare to Eq. (18) in the main text (noting that the toy problem has only one path from 0 to  $L$ ). This is the FSA, in which only the lowest-order terms in  $\langle 0|H_{\text{eff}}|L\rangle$  are kept.

To add another level of complexity, let us consider the REM in a transverse field, i.e., the  $p \rightarrow \infty$  limit of Eq. (13) in which the classical energy levels become independent and a finite- $\epsilon$  ‘‘cluster’’ corresponds to a single configuration. Now,  $P_0$  is the projector onto configurations with classical energy density  $\epsilon$ ,  $V = -\Gamma \sum_i \hat{\sigma}_i^x$ , and we calculate  $\langle \sigma|H_{\text{eff}}|\sigma'\rangle$  for  $|\sigma\rangle, |\sigma'\rangle \in P_0$ . Let the distance between  $\sigma$  and  $\sigma'$  be  $x$ .  $\langle \sigma|H_{\text{eff}}|\sigma'\rangle \sim O(\Gamma^{Nx})$ , and although there are many more terms at  $Nx$ th order than in the toy problem, the same arguments hold here. Each surviving term has an operator string of the form  $V_{od}V_d^{Nx-2}V_{od}$ , with the intermediate states constituting a sequence of spin flips transforming  $\sigma$  into  $\sigma'$  (a ‘‘path’’ in configuration space). This gives us Eq. (18) in the main text. The only subtlety is that some paths may pass through other states in  $P_0$ , giving operator strings  $V_{od}V_d^{Ny-2}V_{od}^2V_d^{N(x-y)-2}V_{od}$ . However, such paths are an exponentially small fraction of the total, and thus negligible: out of the  $(Nx)!$  paths from  $\sigma$  to  $\sigma'$ , the expected number with at least one intermediate configuration having energy density  $\epsilon$  scales as  $Ne^{-N\epsilon^2}(Nx)!$  [53]. Note that these atypical paths do not have amplitudes large enough to compensate for the smaller quantity, as the Schrieffer-Wolff formalism ensures that there are no resonant denominators.

## 2. Time evolution

Since the Schrieffer-Wolff transformation is unitary, time evolution from  $|\sigma\rangle$  to  $|\sigma'\rangle$  through  $H$  is equivalent to time evolution from  $e^S|\sigma\rangle$  to  $e^S|\sigma'\rangle$  through  $H_{\text{eff}}$ . Keep in mind that  $H_{\text{eff}}$  is block diagonal with respect to  $P_0$  and  $Q_0$ . Therefore,  $e^S|\sigma\rangle$  consists of two components, one that evolves within  $P_0$  and one that evolves within  $Q_0$ . Furthermore,  $S$  is off block diagonal since every term in its expansion involves an odd number of  $V_{od}$  factors [see Eq. (D7)]. Thus,  $\cosh(S)|\sigma\rangle$  is the  $P_0$  component and  $\sinh(S)|\sigma\rangle$  is the  $Q_0$  component. In the main text, we simply take  $e^S|\sigma\rangle \sim |\sigma\rangle$ , and time-dependent perturbation theory in  $V_{\text{eff}}$  then gives the rate at which the system tunnels between clusters. However, to justify this, we must consider two effects which are not described by time-dependent perturbation theory. First,  $\cosh(S)|\sigma\rangle$  has amplitude not only in the initial cluster, but in other clusters as well. The overlap with final states vanishes at  $t = 0$  since  $\langle \sigma'|e^{-S}e^S|\sigma\rangle = 0$ , yet this is due to interference between the terms in the expansion of  $e^S$ . Such interference presumably decoheres by  $t \sim O(1)$ , meaning that the system may develop significant amplitude in other clusters on  $O(1)$  timescales. Second,  $\sinh(S)|\sigma\rangle$  has nonzero weight, corresponding to probability for the system to be excited to higher classical energy densities. The dynamics within the  $Q_0$  subspace cannot be described by our method either. In this section, we develop the conditions under which one can neglect these two effects.

Consider  $\cosh(S)|\sigma\rangle$ . We want to calculate the amplitude on  $|\sigma'\rangle \in P_0$ . We will continue to work within the FSA, meaning that we compute  $\langle \sigma'|\cosh(S)|\sigma\rangle$  to lowest nonzero order in  $\Gamma$ . By the same arguments as above, the relevant terms again have the operator string  $V_{od}V_d^{Nx-2}V_{od}$ , which all come from  $\frac{1}{2}S^2$  in the expansion of  $\cosh(S)$ . Unlike above, however,

there are many more terms: the string in the first factor of  $S$  must begin on  $|\sigma\rangle$  but can terminate on any intermediate  $|\sigma''\rangle$ , and the string in the second factor must then begin at  $|\sigma''\rangle$  and terminate on  $|\sigma'\rangle$ . Thus,

$$\langle\sigma'|\cosh(S)|\sigma\rangle\sim\frac{1}{2}\sum_{\sigma''}\sum_{P_{\sigma''}}\prod_{\sigma'''\in P_{\sigma''}}\frac{\Gamma}{N(\epsilon-\epsilon(\sigma'''))}. \quad (\text{D14})$$

The outer sum is over all  $|\sigma''\rangle$  intermediate between  $|\sigma\rangle$  and  $|\sigma'\rangle$ . The inner sum is over paths  $P_{\sigma''}$  that pass through  $|\sigma''\rangle$ .

Focus on the REM for simplicity. Then, we again take  $\epsilon(\sigma''')\rightarrow\mathbb{E}[\epsilon(\sigma''')]=0$ , and

$$\begin{aligned} &|\langle\sigma'|\cosh(S)|\sigma\rangle| \\ &\sim\int_0^x N dy \binom{Nx}{Ny}(Ny)!(N(x-y))!\left(\frac{\Gamma}{N|\epsilon|}\right)^{Nx} \\ &= Nx(Nx)!\left(\frac{\Gamma}{N|\epsilon|}\right)^{Nx}\sim Nx e^{-N\gamma(x,\epsilon)}, \end{aligned} \quad (\text{D15})$$

with  $\gamma(x,\epsilon)$  as in the main text. The extra factor of  $Nx$  does not modify the exponential scaling and can be neglected. To obtain the *total* weight on other clusters, we multiply  $e^{-2N\gamma(x,\epsilon)}$  by  $e^{Ng(x,\epsilon)}$  and integrate over all  $x\in[x^{**}(\epsilon),1-x^{**}(\epsilon)]$ . The

result is  $e^{-Nr(\epsilon)}$  with

$$r(\epsilon)=\min_{x\in[x^{**}(\epsilon),1-x^{**}(\epsilon)]}[2\gamma(x,\epsilon)-g(x,\epsilon)]. \quad (\text{D16})$$

Interestingly, the total weight is governed by the same exponent as the Fermi's golden rule rate [Eq. (5)]. The field strength required for the transformed state  $\cosh(S)|\sigma\rangle$  to have significant weight on other clusters, which could then decohere and become observable on  $O(1)$  timescales, is exactly the field strength required for the tunneling rate to become  $O(1)$  regardless. If the tunneling rate is  $O(1)$ , our use of perturbation theory is questionable anyway. Thus, as long as we focus on the portion of the phase diagram for which  $r(\epsilon)>0$ , it is justified to take  $\cosh(S)|\sigma\rangle\sim|\sigma\rangle$ .

Next, consider  $\sinh(S)|\sigma\rangle$ . We compute  $\langle\sigma'|\sinh(S)|\sigma\rangle$  with  $|\sigma'\rangle\notin P_0$ . Thus,  $\epsilon(\sigma')\equiv\epsilon'\neq\epsilon$ . The lowest-order operator strings are of the form  $V_d^{Nx-1}V_{od}$ , which come from  $S$  in the expansion of  $\sinh(S)$ . There is only one such term in Eq. (D7), giving

$$\langle\sigma'|\sinh(S)|\sigma\rangle\sim\sum_P\prod_{\sigma''\in P}\frac{\Gamma}{N(\epsilon-\epsilon(\sigma''))}. \quad (\text{D17})$$

The notation is the same as for Eqs. (18) and (26) in the main text. For the REM, we evaluate Eq. (D17) as before and obtain Eq. (27). The conditions under which the total weight of  $\sinh(S)|\sigma\rangle$  is negligible then follow as discussed in the main text.

- 
- [1] W. J. Cook, *In Pursuit of the Traveling Salesman* (Princeton University Press, Princeton, NJ, 2012).
- [2] M. Mezard, G. Parisi, and R. Zecchina, *Science* **297**, 812 (2002).
- [3] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, *Phys. Rev. Lett.* **89**, 268701 (2002).
- [4] J. Raymond, A. Sportiello, and L. Zdeborová, *Phys. Rev. E* **76**, 011101 (2007).
- [5] F. Krzakala and L. Zdeborová, *Phys. Rev. Lett.* **102**, 238701 (2009).
- [6] C. Baldassi, C. Borgs, J. T. Chayes, A. Ingrosso, C. Lucibello, L. Saglietti, and R. Zecchina, *Proc. Natl. Acad. Sci. USA* **113**, E7655 (2016).
- [7] A. J. Ochoa, D. C. Jacob, S. Mandra, and H. G. Katzgraber, [arXiv:1801.07681](https://arxiv.org/abs/1801.07681).
- [8] V. N. Smelyanskiy, K. Kechedzhi, S. Boixo, S. V. Isakov, H. Neven, and B. Altshuler, [arXiv:1802.09542](https://arxiv.org/abs/1802.09542).
- [9] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, *Science* **292**, 472 (2001).
- [10] S. Knysh and V. N. Smelyanskiy, *Phys. Rev. E* **78**, 061128 (2008).
- [11] A. P. Young, S. Knysh, and V. N. Smelyanskiy, *Phys. Rev. Lett.* **104**, 020502 (2010).
- [12] L. Foini, G. Semerjian, and F. Zamponi, *Phys. Rev. Lett.* **105**, 167204 (2010).
- [13] E. Farhi, D. Gosset, I. Hen, A. W. Sandvik, P. Shor, A. P. Young, and F. Zamponi, *Phys. Rev. A* **86**, 052334 (2012).
- [14] S. V. Isakov, G. Mazzola, V. N. Smelyanskiy, Z. Jiang, S. Boixo, H. Neven, and M. Troyer, *Phys. Rev. Lett.* **117**, 180402 (2016).
- [15] B. Altshuler, H. Krovi, and J. Roland, *Proc. Natl. Acad. Sci. USA* **107**, 12446 (2010).
- [16] C. L. Baldwin, C. R. Laumann, A. Pal, and A. Scardicchio, *Phys. Rev. B* **93**, 024202 (2016).
- [17] C. L. Baldwin, C. R. Laumann, A. Pal, and A. Scardicchio, *Phys. Rev. Lett.* **118**, 127201 (2017).
- [18] Dynamical fluctuations in the classical energy density are  $O(1/N)$  in the REM, and are  $O(1/p)$  in the  $p$ -spin model.
- [19] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk *et al.*, *Nature (London)* **473**, 194 (2011).
- [20] S. Boixo, T. F. Ronnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, *Nat. Phys.* **10**, 218 (2014).
- [21] B. Derrida, *Phys. Rev. Lett.* **45**, 79 (1980).
- [22] D. Gross and M. Mezard, *Nucl. Phys. B* **240**, 431 (1984).
- [23] E. Gardner, *Nucl. Phys. B* **257**, 747 (1985).
- [24] T. R. Kirkpatrick and D. Thirumalai, *Phys. Rev. B* **36**, 5388 (1987).
- [25] G. Biroli and J. P. Bouchaud, in *The Random First-Order Transition Theory of Glasses*, edited by P. G. Wolynes and V. Lubchenko (Wiley, Hoboken, NJ, 2012), Chap. 2.
- [26] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford Graduate Texts (Oxford University Press, Oxford, 2009).
- [27] V. Bapst, L. Foini, F. Krzakala, G. Semerjian, and F. Zamponi, *Phys. Rep.* **523**, 127 (2013).
- [28] J. Kurchan, G. Parisi, and M. A. Virasoro, *J. Phys. I (France)* **3**, 1819 (1993).

- [29] S. Franz and G. Parisi, *J. Phys. I (France)* **5**, 1401 (1995).
- [30] S. Franz and G. Parisi, *Phys. Rev. Lett.* **79**, 2486 (1997).
- [31] H. Sompolinsky and A. Zippelius, *Phys. Rev. Lett.* **47**, 359 (1981).
- [32] P. W. Anderson, *Phys. Rev.* **109**, 1492 (1958).
- [33] D. Thouless, *Phys. Rep.* **13**, 93 (1974).
- [34] S. Bravyi, D. P. DiVincenzo, and D. Loss, *Ann. Phys.* **326**, 2793 (2011).
- [35] F. Pietracaprina, V. Ros, and A. Scardicchio, *Phys. Rev. B* **93**, 054201 (2016).
- [36] T. Jörg, F. Krzakala, J. Kurchan, and A. C. Maggs, *Phys. Rev. Lett.* **101**, 147204 (2008).
- [37] J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, Boston, 2011).
- [38] C. R. Laumann, A. Pal, and A. Scardicchio, *Phys. Rev. Lett.* **113**, 200405 (2014).
- [39] Y. Y. Goldschmidt, *Phys. Rev. B* **41**, 4858 (1990).
- [40] The recent Ref. [7] presents a classical algorithm that constructs new ground states in  $O(1)$  time given two ground states as a seed. However, one can show that the states thus produced cannot lie in different clusters if both initial states lie in the same cluster. Thus, that algorithm cannot solve the matching problem as we have presented it.
- [41] To obtain Eq. (49) from Eq. (48), substitute  $y = w/p$  and take  $\mu(w/p) \sim e^{-2w}$ .
- [42] V. Dobrosavljevic and D. Thirumalai, *J. Phys. A: Math. Gen.* **23**, L767 (1990).
- [43] T. M. Nieuwenhuizen and F. Ritort, *Phys. A (Amsterdam)* **250**, 8 (1998).
- [44] Strictly speaking, on the excitation boundary the maximum over all  $x \in [0,1]$  is 0. However, one can check that the maximum in fact lies within  $[x^{**}(\epsilon), 1 - x^{**}(\epsilon)]$  for  $p \rightarrow \infty$  (and by continuity for large  $p$ ).
- [45] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
- [46] G. Biroli, R. Monasson, and M. Weigt, *Eur. Phys. J. B* **14**, 551 (2000).
- [47] M. Mézard and R. Zecchina, *Phys. Rev. E* **66**, 056126 (2002).
- [48] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina, *J. Stat. Phys.* **111**, 505 (2003).
- [49] L. Zdeborová and F. Krzakala, *Phys. Rev. E* **76**, 031131 (2007).
- [50] S. Kirkpatrick and B. Selman, *Science* **264**, 1297 (1994).
- [51] E. Farhi, J. Goldstone, S. Gutmann, and D. Nagaj, *Int. J. Quantum Inf.* **6**, 503 (2008).
- [52] It would be interesting to show that matching-finite-energy density solutions is likewise NP-complete, analogous to a hardness of approximation result, although we have not done so.
- [53] By Markov's inequality, a typical sample cannot have more than a finite factor times the expected number of paths as  $N \rightarrow \infty$ .