



HAL
open science

Out-of-equilibrium phase diagram of the quantum random energy model

Giulio Biroli, Davide Facoetti, Marco Schiró, Marco Tarzia, Pierpaolo Vivo

► **To cite this version:**

Giulio Biroli, Davide Facoetti, Marco Schiró, Marco Tarzia, Pierpaolo Vivo. Out-of-equilibrium phase diagram of the quantum random energy model. 2021. hal-03425654v1

HAL Id: hal-03425654

<https://hal.science/hal-03425654v1>

Preprint submitted on 10 Nov 2021 (v1), last revised 23 Nov 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Out-of-equilibrium phase diagram of the quantum random energy modelGiulio Biroli,¹ Davide Facoetti ,^{1,2} Marco Schiró,³ Marco Tarzia ,^{4,5} and Pierpaolo Vivo ²¹*Laboratoire de Physique, Ecole Normale Supérieure, Université PSL, CNRS, Sorbonne Université, Université de Paris, 75005 Paris, France*²*Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom*³*JEIP, USR No. 3573 CNRS, Collège de France, PSL Research University, 75321 Paris, France*⁴*LPTMC, UMR No. 7600, CNRS, Sorbonne Université, 4 Place Jussieu, 75005 Paris, France*⁵*Institut Universitaire de France, 1 Rue Descartes, 75231 Paris Cedex 05, France*

(Received 14 October 2020; revised 21 December 2020; accepted 22 December 2020; published 29 January 2021)

In this paper we study the out-of-equilibrium phase diagram of the quantum version of Derrida's random energy model, which is the simplest model of mean-field spin glasses. We interpret its corresponding quantum dynamics in Fock space as a one-particle problem in very high dimension to which we apply different theoretical methods tailored for high-dimensional lattices: the forward-scattering approximation, a mapping to the Rosenzweig-Porter model, and the cavity method. Our results indicate the existence of two transition lines and three distinct dynamical phases: a completely many-body localized phase at low energy, a fully ergodic phase at high energy, and a multifractal “bad metal” phase at intermediate energy. In the latter, eigenfunctions occupy a diverging volume yet an exponentially vanishing fraction of the total Hilbert space. We discuss the limitations of our approximations and the relationship with previous studies.

DOI: [10.1103/PhysRevB.103.014204](https://doi.org/10.1103/PhysRevB.103.014204)**I. INTRODUCTION**

As discovered by the seminal work of Basko *et al.* [1], isolated disordered interacting many-body systems can show the absence of transport and thermalization even at finite energy density if the disorder is strong enough. This is known as many-body localization (MBL), which is a purely quantum phenomenon which occurs due to Anderson localization in the Fock space as the result of the interplay of disorder, quantum fluctuations, and interactions [1–3] and gives rise to a completely new mechanism for ergodicity breaking, which produces a robust dynamical phase of matter which is stable within a range of interaction and other Hamiltonian parameters. This remarkable phenomenon has attracted considerable interest recently (see Refs. [4–8] for recent reviews), as it implies that the long-time properties of MBL systems cannot be described by the conventional ensembles of quantum statistical mechanics: They can remember, forever and locally, information about their initial conditions.

Although significant and exciting progress has been made in understanding these phenomena in recent years, in both theory [4–8] and experiment [9–11], there still remain many open issues. One set of open questions is about the nature (the universality class) of the MBL phase transition between the thermal and localized phases as the randomness is increased. This transition is an eigenstate phase transition, marked by a sharp change in properties of the many-body wave functions and thus in the dynamics of the system. However, the behavior of many-body eigenstates in the Hilbert space is not firmly established [12–14]. For instance, it is still debated whether there is only one phase transition or could there possibly be some sort of intermediate phase that is neither fully localized nor fully thermal, where eigenstates are delocalized

but nonergodic [2,15–17], called the “bad metal” phase [1]. Investigations of the MBL transition so far have mostly been numerical studies based on exact diagonalization (ED) of relatively small one-dimensional systems [12,18,19], and how to do a proper finite-size scaling analysis of these numerical data remains unclear [20,21]. Also in experiments it is challenging to access the very long timescales, and possibly also long length scales, on which the critical behavior develops. Another frontier of the field is directed towards understanding the existence of MBL in higher dimensions [22–24] and its relationship with other form of ergodicity breaking such as quantum glassiness [25–28].

In this context, exactly solvable, mean-field-like, and simplified toy models might naturally play an important role in making some progress, at least partially, in these directions and in improving our understanding of MBL. Also developing new techniques and tools to tackle analytically or semianalytically the transition and its properties might provide an important step forward to provide further insight into some of the problems mentioned above. With this in mind, in this paper we study the out-of-equilibrium phase diagram of the quantum version of Derrida's random energy model [29], which is the simplest toy model of mean-field spin glasses. The quantum model's equilibrium phase diagram has been studied before [30,31] and a glassy phase was identified at low temperature and small transverse magnetic field. The MBL transition of the quantum random energy model (QREM) was also investigated previously [25,26,28,32–34]. In Refs. [25,26] the presence of a mobility edge separating ergodic eigenstates from many-body localized ones was established, based on EDs of small systems and on a perturbative expansion built on the forward-scattering approximation (FSA) [35]. A later study identified three distinct dynamical

phases, referred to as trapped, tunneling, and excited phases in the context of quantum optimization problems [32]. The interpretation of those phases, and in particular of the phase right above the MBL localized phase, has been recently called into question. In particular, in Ref. [28] the MBL phase was identified with a hyperglass where the dynamics is practically absent while the entire phase above T_{MBL} with a dynamical glass phase (or bad metal) is characterized by nonergodic extended (NEE) eigenstates. In Ref. [34] the authors derive an estimate of the transition to NEE eigenstates in agreement with Ref. [28] and argue that the NEE phase is layered in an alternating sequence of two distinct subphases. The dynamical population transfer protocol on the QREM was further analyzed in Ref. [33], yielding a numerical estimation of the dynamical phase diagram and of the fractal dimensions of the eigenstates in the NEE regime.

In this work we revisit the problem of the out-of-equilibrium phase diagram of the QREM using two complementary techniques, the first based on the FSA and on a mapping onto a paradigmatic random matrix model, the Rosenzweig-Porter model [36,37], and the second based on a generalization of the self-consistent theory of localization [38–41] (hereafter called the cavity approach) designed to take into account the local structure of the Hilbert space of the QREM. In agreement with Ref. [28], we find a multifractal bad metal phase in a broad range of intermediate energies, where eigenfunctions are delocalized but nonergodic, and out-of-equilibrium relaxation to thermal equilibrium is expected to be very slow [28,33] (exponential in the system size). We also obtain a second transition into a fully delocalized ergodic phase at higher temperatures.

The paper is organized as follows. In Sec. II we introduce the QREM and recall basic properties of its equilibrium phase diagram. Section III describes the mapping to a single-particle tight-binding Anderson problem in Hilbert space. Section IV provides qualitative arguments for the phase diagram within the FSA. Section V contains the cavity approach and the numerical results found with this method. In Sec. VI we summarize the results found with our approximations and discuss their relationship with previous results. Section VII puts forward an interpretation of the results based on a family of auxiliary Anderson toy models. Finally, in Sec. VIII we provide concluding remarks and perspective for future studies. Several technical details are reported in Appendixes A–C.

II. MODEL

The quantum random energy model for N spins $\frac{1}{2}$ is defined by the Hamiltonian

$$\mathcal{H}_{\text{QREM}} = E(\{\hat{\sigma}_a^z\}) - \Gamma \sum_a \hat{\sigma}_a^x, \quad (1)$$

where Γ is the transverse field and $E(\{\hat{\sigma}_a^z\})$ is a random operator diagonal in the $\{\hat{\sigma}_a^z\}$ basis, which takes 2^N different values for the 2^N configurations of the N spins in the z basis, identically and independently distributed according to

$$P(E) = \frac{e^{-E^2/N}}{\sqrt{\pi N}}. \quad (2)$$

Such a natural choice of the scaling of the random many-body energies ensures that they are with high probability contained in the interval $[-N\sqrt{\ln 2}, +N\sqrt{\ln 2}]$ in the thermodynamic limit. Throughout the paper we will denote by ε ($=E/N$) the intensive energy per spin corresponding to the extensive energy E . A concrete implementation of the $E(\{\hat{\sigma}_a^z\})$ is given by the $p \rightarrow \infty$ limit of the fully connected p -spin model, $E(\{\hat{\sigma}_a^z\}) = \lim_{p \rightarrow \infty} \sum_{a_1, \dots, a_p} J_{a_1, \dots, a_p} \hat{\sigma}_{a_1}^z \cdots \hat{\sigma}_{a_p}^z$, where the J_{a_1, \dots, a_p} are independent and identically distributed Gaussian random variables. In consequence, the diagonal part of the Hamiltonian corresponds to a mean-field spin-glass model and exhibits all the fundamental features of the so-called random first-order theory [42] with a one-step replica symmetry-breaking glass transition [29].

The equilibrium properties (in the canonical ensemble) of the QREM are well established [30,31]. At low transverse field, it displays the same transition as the classical model between the paramagnetic and the glass phase at the Kauzmann temperature $T_K = 1/(2\sqrt{\ln 2})$. All thermodynamic quantities are identical to the $\Gamma = 0$ classical random energy model (REM). For $T < T_K$ the system freezes in its ground state at $\varepsilon_{\text{GS}} = -\sqrt{\ln 2}$. The classical model has also a dynamical ergodicity-breaking transition below which the time to reach thermal equilibrium is exponentially large in the system size [43]. However, differently from the p -spin models with finite p , for which the dynamical transition temperature T_d is finite, in the $p \rightarrow \infty$ limit $T_d \rightarrow \infty$, due to the fact that the random energies and spin configurations are totally uncorrelated and flipping a single spin can change the energy by a large amount. Within the semiclassical approximation of Ref. [30], T_d stays infinite even when quantum fluctuations are turned on ($\Gamma > 0$).

At large magnetic field $\Gamma > \Gamma_c(T)$ the system is a standard quantum paramagnet and the REM term in the Hamiltonian does not influence the equilibrium physics of this phase. The first-order transition between these two regions takes place at $\Gamma_c(T)$, which is equal to $\sqrt{\ln 2} \approx 0.833$ for $T \rightarrow 0$ and to $\sqrt{2}/2 \approx 0.707$ for $T \rightarrow \infty$. In this paper we will only focus on the small- Γ region of the phase diagram (i.e., $\Gamma < \sqrt{2}/2$).

III. MAPPING TO ANDERSON LOCALIZATION ON THE HYPERCUBE

The QREM defined by Eq. (1) can be viewed as the simplest many-body model that displays Anderson localization in its Hilbert space: If one chooses as a basis the tensor product of the simultaneous eigenstates of the operators σ_a^z , the Hilbert space of the many-body Hamiltonian is an N -dimensional hypercube of $\mathcal{V} = 2^N$ sites. One can map a configuration of N spins to a corner of the N -dimensional hypercube by considering $\sigma_a^z = \pm 1$ as the top/bottom face of the cube's a th dimension. The random part of the Hamiltonian is by definition diagonal on this basis and gives *uncorrelated* random energies on each site orbital of the hypercube: At $\Gamma = 0$ the many-body eigenstates of Eq. (1) are simply product states of the form $|\sigma_1^z\rangle \otimes |\sigma_2^z\rangle \otimes \cdots \otimes |\sigma_N^z\rangle$ and the system is fully localized. The interacting part of the Hamiltonian acts as single spin flips on the configurations $\{\sigma_a^z\}$ and plays the role of the hopping rates connecting neighboring sites in the configuration space. The many-body quantum dynamics is

then recast as a single-particle noninteracting tight-binding Anderson model for spinless electrons in a disordered potential existing on the 2^N corners of a hypercube in N dimensions, with the spin configurations being lattice sites, $|i\rangle \equiv |\{\sigma_a^z\}\rangle$, and the transverse field playing the role of the hopping amplitude between neighboring sites

$$\mathcal{H} = -\Gamma \sum_{\langle i,j \rangle} (|i\rangle\langle j| + |j\rangle\langle i|) + \sum_{i=1}^{\mathcal{V}} E_i |i\rangle\langle i|, \quad (3)$$

where $\langle i, j \rangle$ denotes nearest neighbors on the hypercube, $\mathcal{V} = 2^N$ is the total number of sites, and Γ is the hopping kinetic energy scale. This mapping is *exact*, in the sense that the Hamiltonians (1) and (3) have *the same* eigenvalues and the eigenvectors (when the simultaneous eigenstates of the operators σ_a^z are chosen as a basis). However, for a generic interacting many-body Hamiltonian in finite dimensions the random energies defined on neighboring corners of the hypercube are strongly correlated, as they correspond to many-body configurations which only differ by a single spin flip, while for the QREM E_i are independent and identically distributed random variables distributed according to Eq. (2), since its distinguishing feature is precisely the absence of such correlations.

IV. ESTIMATE OF THE OUT-OF-EQUILIBRIUM PHASE DIAGRAM WITHIN THE FORWARD-SCATTERING APPROXIMATION

As discussed in Sec. III, the QREM can be mapped to an Anderson model on a hypercube with $\mathcal{V} = 2^N$ sites, labeled by σ^z configurations. The typical tunneling rate between two configurations depends on their energy and on the Hamming distance Nx between them (the Hamming distance Nx is defined as the minimum number of spin flips which separate the two configurations, with $0 < x < 1$). Since the energy levels are independent, the typical number of configurations of energy $|E| = N\varepsilon$ and at distance Nx from a given configuration is

$$\mathcal{N}_\varepsilon(x) = \binom{N}{Nx} \frac{e^{-N\varepsilon^2}}{\sqrt{\pi N}}. \quad (4)$$

Here we estimate the matrix elements $\mathcal{M}(\varepsilon, x)$ between these two configurations by perturbation theory in Γ , using the FSA [25,44,45]. This consists in assuming that the matrix element between two configurations at distance x is given by the product of the matrix elements obtained along the Nx spin flips that connect the two configurations, ignoring loopy contributions in which spins are flipped twice since they contribute at higher order in perturbation theory. Almost all states have energy $O(\sqrt{N})$, while $E \approx O(N)$; therefore, we take the energy differences appearing in the denominators of the perturbative expansion to be E . Since there are $(Nx)!$ such contributions, corresponding to the $(Nx)!$ to connect two configurations Nx spin flips away, the resulting matrix element reads

$$\mathcal{M}(\varepsilon, x) \approx \left(\frac{\Gamma}{N\varepsilon} \right)^{Nx} (Nx)!. \quad (5)$$

In the case of noninteracting particles in a disordered potential, the Mott argument for hybridization states that the

metal-insulator transition occurs when the number of sites in resonance with a given site i stays finite in the thermodynamic limit [46–48]. Based on the analogy with single-particle Anderson localization, one can thus characterize the localized phase by the condition (Mott criterion)

$$\lim_{N \rightarrow \infty} \mathcal{N}_\varepsilon(x) |\mathcal{M}(\varepsilon, x)| = 0 \quad (6)$$

and from this estimate the point at which many-body localization of the QREM takes place.

Analogously, the ergodicity-breaking transition can be estimated from the Fermi golden rule [28,46–48]. In fact, the spreading amplitude $\mathcal{N}_\varepsilon(x) |\mathcal{M}(\varepsilon, x)|^2$ quantifies the escape rate of a particle created at a given site i . When this amplitude is much smaller than the spreading of energy levels due to disorder, then ergodicity is broken:

$$\lim_{N \rightarrow \infty} \mathcal{N}_\varepsilon(x) |\mathcal{M}(\varepsilon, x)|^2 = 0. \quad (7)$$

The nonergodic extended phase is thus realized when

$$\mathcal{N}_\varepsilon(x) |\mathcal{M}(\varepsilon, x)|^2 \rightarrow 0, \quad \mathcal{N}_\varepsilon(x) |\mathcal{M}(\varepsilon, x)| \rightarrow \infty. \quad (8)$$

This means that although a given state is in resonance with many other states of energy ε and at distance Nx from it, the number of those resonances is not enough for the quantum dynamics to decorrelate *in a finite time* from the initial condition.

These two criteria for localization and ergodicity breaking can be illustrated using the Rosenzweig-Porter (RP) model [36,37,46,49] as a benchmark. For the sake of clarity, here we briefly recall the definition of the RP model, whose Hamiltonian is a matrix of size $\mathcal{V} \times \mathcal{V}$ given by the sum of two terms

$$H_{\text{RP}} = \mathcal{E} + \frac{\mu}{\gamma^{1/2}} \mathcal{G}, \quad (9)$$

where $\mathcal{E}_{ij} = E_i \delta_{ij}$ is diagonal with independent and identically distributed entries (the distribution does not matter as long as it has finite variance), μ is a constant of order one (whose value is unimportant), and \mathcal{G} is a random matrix drawn from the Gaussian unitary ensemble (GUE) with unit variance. The latter mimics an ergodic system (e.g., the clean lattice), while \mathcal{E} represents the on-site disorder. The parameter γ acts in the RP model as a proxy of the disorder strength: At large $\gamma > 2$, the GUE contribution is suppressed and the systems is localized; at small $\gamma < 1$, the system is ergodic, while the regime $1 < \gamma < 2$ is special, with delocalized but nonergodic wave functions which typically occupy $\mathcal{V}^{2-\gamma}$ sites close in energy. The criteria for these two transitions are exactly the ones we introduced above for Anderson localization (6) and ergodicity breaking (7).

Specializing now to the $\mathcal{N}_\varepsilon(x)$ levels at energy ε and Hamming distance Nx between them of the QREM, we can map the effective (typical) transition rates (5) onto an effective $\mathcal{N}_\varepsilon(x) \times \mathcal{N}_\varepsilon(x)$ RP random matrix with off-diagonal matrix elements roughly scaling as $\mathcal{M}(\varepsilon, x) \propto \mathcal{N}_\varepsilon(x)^{-\gamma}$. The effective exponent γ associated with a given set of configurations (ε, x) defined as $\gamma = -2 \ln \mathcal{M}(\varepsilon, x) / \ln \mathcal{N}_\varepsilon(x)$. We therefore expect that the transition from Anderson localization to nonergodic extended states for the QREM happens when at least one x sector becomes delocalized, i.e.,

$$\max_x \mathcal{N}_\varepsilon(x) \mathcal{M}(\varepsilon, x) \approx \max_x e^{N f_1(x, \varepsilon, \gamma)} \gtrsim 1, \quad (10)$$

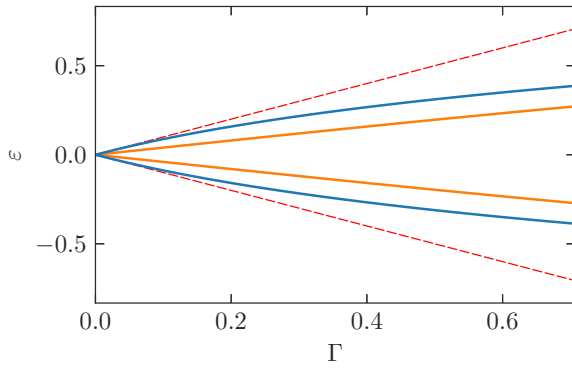


FIG. 1. Localization (blue) and ergodicity (orange) transition lines, obtained from the mapping to the RP model (11) and (14). The limits of the y axis coincide with the edges of the many-body spectrum ($|\varepsilon| < \sqrt{\ln 2}$). The red dashed lines correspond to $\varepsilon = \pm\Gamma$ (see Refs. [25,26]).

where, using Stirling's approximation for the factorials,

$$f_1(x, \varepsilon, \Gamma) = x \ln \left(\frac{\Gamma}{e\varepsilon} \right) - (1-x) \ln(1-x) - \varepsilon^2. \quad (11)$$

If $\Gamma < \varepsilon$, f_1 is always negative. Otherwise, it has a non-negative maximum at $x_1^* = 1 - \varepsilon/\Gamma$, which determines the mobility edge $\Gamma_{\text{MBL}}(\varepsilon)$ through the implicit equation

$$f_1(x_1^*, \varepsilon, \Gamma_{\text{MBL}}) = \frac{\varepsilon}{\Gamma_{\text{MBL}}} - \varepsilon^2 + \ln \frac{\Gamma_{\text{MBL}}}{e\varepsilon} = 0. \quad (12)$$

This is the same result obtained through a similar argument in Ref. [25].

Similarly, full ergodicity is recovered by requiring that at least one x sector becomes ergodic,

$$\max_x \mathcal{N}(\varepsilon, x) |\mathcal{M}(\varepsilon, x)|^2 \approx \max_x e^{Nf_2(x, \varepsilon, \gamma)} \gtrsim 1, \quad (13)$$

with

$$f_2(x, \varepsilon, \Gamma) = x \ln x - (1-x) \ln(1-x) - \varepsilon^2 + 2x \ln \left(\frac{\Gamma}{e\varepsilon} \right). \quad (14)$$

If $\Gamma < 2\varepsilon$, f_2 is always negative. Otherwise, it has a non-negative maximum at $x_2^* = 1/2[1 + \sqrt{1 - 4(\varepsilon/\Gamma)^2}]$, which gives a different implicit equation for the ergodic transition $\Gamma_{\text{erg}}(\varepsilon)$.

Expanding the solutions to the two implicit equations around $\varepsilon = 0$, we find $\Gamma_{\text{MBL}} \approx \varepsilon$ [25,26,28,32] and $\Gamma_{\text{erg}} \approx \varepsilon/2$. The analogy with the RP model therefore indicates that the QREM also undergoes two separate localization and ergodicity transitions. The estimates for the transition lines obtained in this way are shown in Figs. 1 and 2.

Note that the phase diagram of Fig. 1 implies that the nontrivial localized and nonergodic behaviors are found at nonzero energy density. Hence the fractions of localized and NEE states is exponentially small in N and vanish as $N \rightarrow \infty$, while the vast majority of states around zero energy density are fully ergodic, in agreement with Ref. [40].

Qualitatively, these results are in agreement with the recent analysis of Refs. [28,34], which are based on a different (and more thorough) strategy to estimate the off-diagonal tunneling rates between different spin configurations in the Hilbert

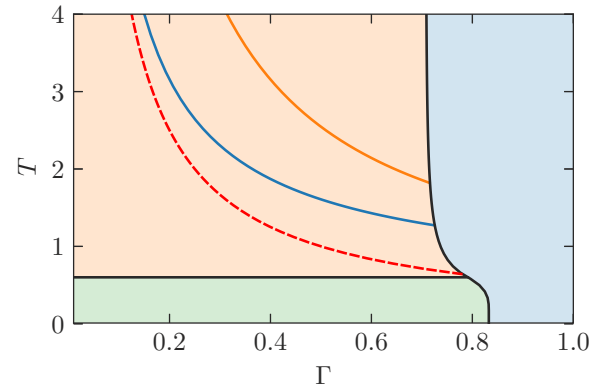


FIG. 2. Ergodicity and localization transition lines (Fig. 1), transposed on the canonical phase diagram, with $T = 1/2\varepsilon$.

space, and are also in agreement with the numerical estimations of Ref. [33]. However, while Eq. (14) predicts that the transition line from the fully ergodic to the dynamical glass (bad metal) regime is at finite energy density ε_{erg} , within the approach of Refs. [28,34] the transition line is squeezed to zero energy density, i.e., infinite temperature. We will return to this point in Sec. VI.

V. CAVITY APPROACH

A. Cluster approximation on the hypercube

In large spatial dimensions the neighbors of a given site are organized in a hierarchical way (i.e., the fraction of short loops is suppressed) and their number grows exponentially with the distance. These are distinctive features of treelike structures. In fact, it was argued originally in [2] that the (noninteracting) Anderson model on the Bethe lattice, first introduced and studied in Ref. [38], can be thought of as a toy model for MBL (see also Refs. [3,50–53] for a similar analysis and Ref. [54] for a quantitative investigation of these ideas).

Since on treelike structures the model (3) allows, in principle, for an exact solution [38], which yields the diagonal elements of the resolvent matrix, assuming that for large enough N the hypercube is well approximated by a Bethe lattice provides a very simple way to investigate analytically, although approximately, the spectral properties of the eigenvectors of the QREM (see Refs. [39–41,55] for similar approaches in the context of MBL). The simplest approximation consists in taking random regular graphs (RRGs) of $\mathcal{V} = 2^N$ sites and fixed connectivity N as the underlying lattices mimicking the Hilbert space [54], i.e., random lattices which have locally a treelike structure but have loops whose typical length scales as $\ln \mathcal{V} \propto N$ and no boundary, and which are statistically translationally invariant.¹ In practice, this corresponds to shuffling the position of the sites and/or rewiring the connections of the hypercube in a random way, keeping the total number of sites and the local connectivity of the lattice fixed.

¹The properties of random regular graphs have been extensively studied. For a review see [56].

There is however a potentially important issue related to the Bethe approximation, which we explain below. The spectrum of the kinetic term of (3) is given by the density of states (DOS) of the adjacency matrix of the N -dimensional hypercube, which coincides with the distribution of the eigenvalues of the second term of the Hamiltonian (1), i.e., a simple paramagnet, with energies contained in the interval $E \in [-N\Gamma, N\Gamma]$,

$$\rho_I^{\text{HC}}(E) = \frac{\Omega(E)}{\Gamma 2^{N+1}}, \quad (15)$$

where the term $\Gamma 2^{N-1}$ in the denominator is a normalization factor that ensures that $\int_{-N\Gamma}^{+N\Gamma} \rho_I^{\text{HC}}(E) dE = 1$ and $\Omega(E)$ is the number of spin configurations at energy E :

$$\Omega(E) = \binom{N}{(N + E/\Gamma)/2} \sim \Omega_0 e^{N s(\varepsilon/\Gamma)}. \quad (16)$$

Here Ω_0 is a normalization factor and $s(\varepsilon/\Gamma)$ is the entropy per spin at large N (apart from logarithmic corrections) for a polarization $m = \varepsilon/\Gamma = \langle \sigma^x \rangle$ of the spins in the x direction:

$$s(m) = \ln(2) - \frac{1+m}{2} \ln(1+m) - \frac{1-m}{2} \ln(1-m). \quad (17)$$

Approximating the hypercube as a treelike lattice amounts to replacing Eq. (15) with the spectrum of the adjacency matrix of a Bethe lattice of connectivity N which for large enough N tends asymptotically to a semicircle law

$$\rho_I^{\text{RRG}}(E) \approx \frac{\sqrt{4\Gamma^2 N - E^2}}{8\pi \Gamma^2 N}, \quad (18)$$

with support in the interval $E \in [-2\Gamma\sqrt{N}, 2\Gamma\sqrt{N}]$. For energies of $O(\sqrt{N})$, where the vast majority of the eigenvalues are, ρ_I^{RRG} provides in fact a reasonably good estimation for the true DOS, ρ_I^{HC} (see Fig. 10 of Appendix A for a quantitative comparison). However, for *extensive* energies of $O(N)$ Eq. (18) completely neglects the exponentially small fraction of eigenvalues in the tails of the DOS, corresponding to strongly polarized spins in the x direction. Since at finite energy density ε and N sufficiently large the energy $N\varepsilon$ will fall outside the edges of the semicircle (18) which describes the spectrum of the delocalizing kinetic term within the Bethe approximation, the Anderson localization of the eigenfunctions of the Hamiltonian (3) will occur in the far Lifshitz tails of the DOS [57] and this might affect its properties. In other words, the system might appear as localized within the Bethe approximation due to the fact that some of the matrix elements associated with the kinetic term are artificially suppressed by approximating the hypercube as a RRG.

In order to overcome, at least partially, these limitations in this paper we put forward a cluster expansion, which takes into account, at least locally, the specific structure of the hypercube up to a certain distance (in particular, including all the shortest loops of length 4, 6, 8, etc.) and improves systematically the simplest, single-site, Bethe approximation. In practice, we consider clusters of $s = 2^n$ neighboring corners on the hypercube (corresponding to spin configurations which differ by few spin flips only) and obtain self-consistent recursion equations for the $s \times s$ elements of the local resolvent matrix on each cluster by assuming that the clusters are on a treelike structure.

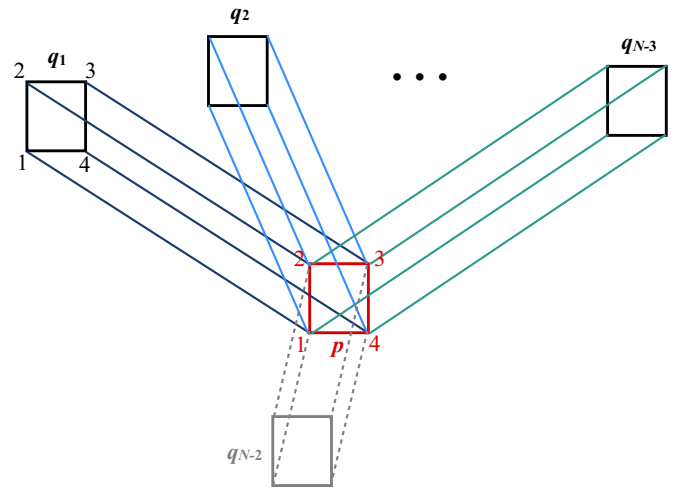


FIG. 3. Schematic representation of the recursion step which yields the self-consistent equations for the $s \times s$ elements of the (cavity) resolvent matrix for clusters of four sites. All loops up to length 12 of the hypercube are treated exactly.

The standard single-site Bethe approximation corresponds to $n = 0$, while the case $n = 2$ is schematically represented in Fig. 3 ($n = N$ corresponds, of course, to the exact solution of the problem). We will take $n = 2$ throughout.

For $n = 2$ a plaquette of four corners corresponds to four spin configurations such as

$$\begin{aligned} |1\rangle &= |\uparrow, \uparrow, \sigma_3^z, \dots, \sigma_N^z\rangle, \\ |2\rangle &= |\uparrow, \downarrow, \sigma_3^z, \dots, \sigma_N^z\rangle, \\ |3\rangle &= |\downarrow, \downarrow, \sigma_3^z, \dots, \sigma_N^z\rangle, \\ |4\rangle &= |\downarrow, \uparrow, \sigma_3^z, \dots, \sigma_N^z\rangle \end{aligned}$$

for any configuration $\{\sigma_a^z\}_{a=3, \dots, N}$. The $N - 2$ neighbors of such a plaquette on the hypercube are found by flipping the $N - 2$ spins $\sigma_3^z, \dots, \sigma_N^z$ one by one. Two neighboring plaquettes of four sites are connected by four edges. The Hilbert space will then be approximated as a RRG of $2^N/4$ square plaquettes of connectivity $N - 2$. More details are given in Appendix A.

One can show that within the cluster approximation the support of the spectrum of the kinetic term of the Hamiltonian (3) becomes indeed broader and broader as n is increased, $E_{\text{max}} \approx \Gamma(2\sqrt{N - n} + n)$ (see Appendix A). One can easily obtain, at least formally, the self-consistent recursion relations for the elements of the resolvent matrix (or Green's functions) of the Hamiltonian (3), defined as $\mathcal{G}(z) = (\mathcal{H} - z\mathcal{I})^{-1}$, at any order n of the cluster expansion. The key objects are the so-called *cavity* resolvent matrices $G_{p \rightarrow q}(z) = (\mathcal{H}_{p \leftrightarrow q} - z\mathcal{I})^{-1}$, i.e., the resolvent matrices of modified Hamiltonians $\mathcal{H}_{p \leftrightarrow q}$ where all the 2^n edges between the sites of the cluster p and the sites of the cluster q have been removed (gray dashed lines of Fig. 3). Let us assume, as explained above and as sketched in Fig. 3, that at large enough N the clusters occupy the vertices of a treelike structure (at least locally) and let us imagine taking a given cluster p and its neighbors $\{q_1, \dots, q_{N-n}\}$. If one removes such a cluster from the graph, then the clusters $\{q_1, \dots, q_{N-n}\}$ are (quasi)uncorrelated, since the lattice would

break into $N - n$ semi-infinite (quasi)disconnected branches (neglecting the large loops of length of order N). One then obtains (e.g., by Gaussian integration) the iteration relations for the elements of the cavity resolvent matrix on the s sites of the cluster [38,57,58]

$$[G_{p \rightarrow q}^{-1}(z)]_{uv} = \mathcal{H}_{uv} - z\delta_{uv} - \Gamma^2 \sum_{q_l \in \partial p/q_k} [G_{q_l \rightarrow p}(z)]_{uv}, \quad (19)$$

where $z = N\varepsilon + i\eta$, η is an infinitesimal imaginary regulator which regularizes the polelike singularities on the right-hand sides (discussed below), ε is the intensive energy density around which one chooses to study the spectral properties, \mathcal{H}_{uv} are the matrix elements of the Hamiltonian (3) between the sites u and v belonging to the cluster p , and $\partial p/q$ denotes the set of all $N - n$ neighbors of the cluster p except q . The indices $u, v = 1, \dots, s$ identifying the sites belonging to each cluster are chosen as in Fig. 3. (Note that for each cluster with $N - n$ neighbors one can define $N - n$ cavity Green's functions and $N - n$ recursion relations of this kind.)

After finding the solution of Eq. (19), one can finally obtain the resolvent matrix of the original problem on a given cluster p as a function of the cavity Green's functions on the neighboring clusters [57,58]:

$$[G_p^{-1}(z)]_{uv} = \mathcal{H}_{uv} - z\delta_{uv} - \Gamma^2 \sum_{q_k \in \partial p} [G_{q_k \rightarrow p}(z)]_{uv}. \quad (20)$$

For $n = 0$ these equations simply return the standard recursion relations for the Anderson model on the Bethe lattice (with connectivity N and Gaussian independent and identically distributed random energies) [38,57,58]. For $n = 1$ and 2 simple analytic expressions of the inverse of the local $s \times s$ resolvent matrices are known, which allows one to write simple recursion equations for its s diagonal elements and its $s(s - 1)/2$ off-diagonal elements (see Appendix A for more details). For $n \geq 3$, however, the local inversion involved in Eqs. (19) and (20) must be done numerically at each iteration step.

There are essentially two ways, which we detail in Appendix B, to solve the recursion equations for the Green's function and obtain information on the spectral statistics at finite N . The most accurate strategy, which we will refer to as the cluster belief propagation (CBP) algorithm (see Ref. [58] for a detailed explanation of this approach for the usual tight-binding Anderson model on the Bethe lattice), is to solve directly Eqs. (19) and (20) on random realizations of the hypercube of 2^N sites (i.e., N spins) (see Appendix B for details). However, due to the fact that random energies E_i of the QREM are uncorrelated, in order to access larger system sizes one can adopt another strategy, which we will refer to hereafter as the cluster population dynamics (CPD) algorithm [59], which consists in interpreting the recursion relations for the Green's functions as equations for their probability distributions once the average over the disorder is taken. In fact, since $G_{p \rightarrow q}$ and \mathcal{G}_p are random matrices, one can assume that averaging over the on-site random energies leads to functional equations on their probability distribution $Q(G)$ and $\mathcal{Q}(\mathcal{G})$ (see Appendix B for details). Figure 11 of Appendix B shows that the cluster approximation provides a quite accurate approximation of

local observables, such as the distribution of the local density of states (LDOS), as compared to exact diagonalization for small systems.

B. Spectral statistics and the $\eta \rightarrow 0^+$ limit

The statistics of the diagonal elements of the resolvent give, in the $\eta \rightarrow 0^+$ limit (discussed below), the spectral properties of \mathcal{H} . In particular, the probability distribution of the LDOS at energy $E = N\varepsilon$ is given by

$$\rho_i(\varepsilon) = \sum_{\alpha} |i|\alpha\rangle|^2 \delta(N\varepsilon - E_{\alpha}) = \lim_{\eta \rightarrow 0^+} \frac{\text{Im} \mathcal{G}_i(N\varepsilon + i\eta)}{\pi}, \quad (21)$$

from which the average DOS is simply obtained as $\rho(\varepsilon) = (1/\mathcal{V}) \sum_i \rho_i(\varepsilon) = \lim_{\eta \rightarrow 0^+} \text{Tr} \text{Im} \mathcal{G}(N\varepsilon + i\eta)/(\mathcal{V}\pi)$. (We have defined $\mathcal{G}_i = [\mathcal{G}_p]_{iuu}$ with $i = 2^n p + u$, $p = 1, \dots, 2^{N-n}$, and $u = 1, \dots, s = 2^n$.) Similarly, the spectral representation of the inverse participation ratio of the eigenstates $|\alpha\rangle$ of energy E_{α} close to $N\varepsilon$ can be obtained as

$$\begin{aligned} \Upsilon_2(\varepsilon) &= \sum_i |\langle i|\alpha\rangle|^4 \delta(N\varepsilon - E_{\alpha}) \\ &= \lim_{\eta \rightarrow 0^+} \frac{\eta \sum_{i=1}^{\mathcal{V}} |\mathcal{G}_i(N\varepsilon + i\eta)|^2}{\sum_{i=1}^{\mathcal{V}} \text{Im} \mathcal{G}_i(N\varepsilon + i\eta)}. \end{aligned} \quad (22)$$

Another useful observable is the typical DOS

$$\rho^{\text{typ}}(\varepsilon) = \lim_{\eta \rightarrow 0^+} \frac{\exp(\mathcal{V}^{-1} \sum_{i=1}^{\mathcal{V}} \ln \text{Im} \mathcal{G}_i(N\varepsilon + i\eta))}{\mathcal{V}^{-1} \sum_{i=1}^{\mathcal{V}} \text{Im} \mathcal{G}_i(N\varepsilon + i\eta)}. \quad (23)$$

However, at this point we encounter another difficulty which is due to the very unusual (and simultaneous) scaling with N of the parameters of the Hamiltonian (3). In fact, the dependence of the random energies and the connectivity of the lattice on N produces a density of states that strongly concentrates around zero energy density, as naturally expected for many-body systems: At small Γ one has that $\rho(E) \approx P(E) = e^{-E^2/N}/\sqrt{\pi N}$, while at large Γ one expects that $\rho(E) \approx \Omega(E)/\Gamma 2^{N+1}$ [see Eq. (15)]. Thus, in both cases the mean level spacing $\delta(E) = 1/\mathcal{V}\rho(E)$ is well defined locally, but depends strongly (i.e., exponentially) on the local energy density. In particular, for small Γ one has that

$$\delta(\varepsilon) \approx \sqrt{\pi N} e^{N(\varepsilon^2 - \ln 2)}. \quad (24)$$

In order for Eqs. (21)–(23) to be well defined, the limit $\eta \rightarrow 0^+$ should be taken in such a way that the imaginary regulator goes to zero on the same scale as the mean level spacing [58]. Hence, studying the asymptotic behavior of the model at large N implies varying simultaneously the following parameters [39].

- (i) The total number of sites diverges exponentially $\mathcal{V} = 2^N$.
- (ii) The connectivity of the lattice grows as N .
- (iii) The standard deviation of the random on-site energies grows as $\sqrt{N/2}$, according to Eq. (2).

(iv) The energy at which we study the system grows as $N\varepsilon$, with ε of $O(1)$.

(v) The imaginary regulator vanishes exponentially as $\delta(\varepsilon) = 1/\mathcal{V}\rho(\varepsilon)$.

Thus, the $N \rightarrow \infty$ limit of the model (3) is quite different from the usual thermodynamic limit of the standard (non-interacting) Anderson model (where one just takes the limit $\mathcal{V} \rightarrow \infty$ and $\eta \rightarrow 0$ keeping fixed the other parameters) and might give rise to unusual scaling and new properties.

In the following we will be mostly interested in studying the dependence on Γ , ε , and N of the probability distribution of the LDOS (21), from which one can compute several spectral quantities of interest, such as the inverse participation ratio (22) and the typical value of the LDOS (23). In particular, in order to assess the ergodicity of the wave functions, it is customary to introduce the fractal dimensions defined through the asymptotic behavior of these two latter quantities:

$$\begin{aligned} \Upsilon_2(\varepsilon) &\sim [\tilde{\mathcal{V}}(\varepsilon)]^{-D_2(\varepsilon)}, \\ \rho^{\text{typ}}(\varepsilon) &\sim [\tilde{\mathcal{V}}(\varepsilon)]^{D_1(\varepsilon)-1}. \end{aligned} \quad (25)$$

This definition takes into account the actual volume of the portion of the phase space accessible at finite energy density ε , $\tilde{\mathcal{V}}(\varepsilon) = 2^N \rho(\varepsilon)$, since at finite energy density ε , ergodic eigenstates are uniformly spread over the hypersurface at constant energy. In the small- Γ part of the phase diagram, as a first approximation one has that $\rho(E) \approx P(E)$.

C. Numerical results for the fractal dimensions

In Fig. 4, where we plot D_1 [Fig. 4(b)] and D_2 [Fig. 4(a)] as a function of N for $\Gamma = 0.2$, and several values of the energy density ε , mostly on the delocalized side of the MBL transition and close to the mobility edge ($\varepsilon \lesssim \Gamma$), D_1 and D_2 are obtained as numerical derivatives of ρ^{typ} and Υ_2 with respect to $\ln \tilde{\mathcal{V}}$. The figure shows three data sets, corresponding to the results obtained from EDs ($N \leq 14$, closed symbols), the CBP approach ($N \leq 25$ and $n = 2$, open symbols), and the CPD approximation ($N \leq 50$ and $n = 2$, checkered symbols).

First of all, these plots demonstrate that the CBP and CPD approximations are in reasonably good agreement with the exact results for all values of ε , at least in the range of system sizes accessible via EDs (see also Fig. 11 of Appendix B for a detailed comparison of the full probability distribution of the LDOS). At moderate energy density both fractal dimensions show a clear nonmonotonic dependence: D_1 and D_2 first rapidly increase with N and then start to decrease slowly after going through a maximum. At small enough energy density $\varepsilon \lesssim 0.14$, both D_1 and D_2 reach a finite plateau strictly smaller than one at large N (horizontal dotted lines). This behavior corresponds to genuine multifractal eigenstates, as recently predicted in Ref. [28], and is found in a broad range of energy density. The lower the energy, the higher the plateau values reached by D_1 and D_2 , i.e., the system gets closer and closer to full ergodicity as the energy density is decreased. At larger energies instead, above the mobility edge ε_{MBL} , D_1 and D_2 decay to zero in the large- N limit.

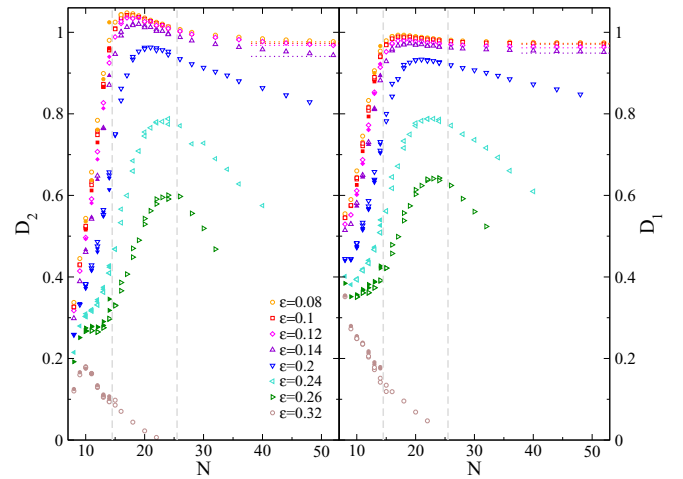


FIG. 4. Plot of (a) D_2 and (b) D_1 versus N for $\Gamma = 0.2$ and several values of the energy density as indicated in the legends. Closed symbols correspond to the results obtained from ED, open symbols give the results of the CBP approximation (with $n = 2$), and checkered symbols correspond to the result of the CPD approach (with $n = 2$). Two independent ED data sets, obtained with two different algorithms, are shown to give the idea of the typical size of the error bars. The vertical dashed lines represent the limits of the range of applicability of ED ($N \leq 14$) and the CBP approach ($N \leq 25$). The horizontal dotted lines indicate the asymptotic values of D_1 and D_2 at large N in the bad metal phase. (Note that the generalized fractal dimensions can become larger than 1 for some intermediate values of N and for ε small enough, due to logarithmic corrections to the many-body DOS.)

These results support the existence of two distinct nonergodic regions of the phase diagram: a delocalized multifractal phase ($0 < D_1, D_2 < 1$) at intermediate energy density, where eigenstates occupy a volume that diverges yet is exponentially smaller than the total Hilbert space, and an Anderson localized phase ($D_1, D_2 \rightarrow 0$), where eigenstates are exponentially localized in the Hilbert space and occupy a finite N -independent volume on the hypercube. We have repeated the same analysis for $\Gamma = 0.1$, finding similar results.

Note, however, that the cavity approach does not allow one to determine sharply the phase boundaries between the three phases because the numerical results are only available for systems of moderate size, $N \lesssim 50$, and the asymptotic values of the fractal dimensions cannot be firmly established, especially in the vicinity of the transition line between the bad metal and the MBL phases (see, e.g., the data for $\varepsilon = 0.2$ of Fig. 4). For $\Gamma = 0.2$ the MBL mobility edge within the cavity approximation is estimated within the interval $\varepsilon_{\text{MBL}} \in (0.15, 0.19)$ (see Fig. 6), which is in good agreement with the estimation of Sec. IV, Eq. (11), and Fig. 1.

D. Level statistics

A natural question that arises concerns the statistics of the energy levels in the multifractal phase. In fact, in analogy with the RP model, it is reasonable to expect that in the mixed phase the level statistics should be described by the Gaussian orthogonal ensemble (GOE) on the scale of the mean level spacing, while it might cross over to a different, possibly

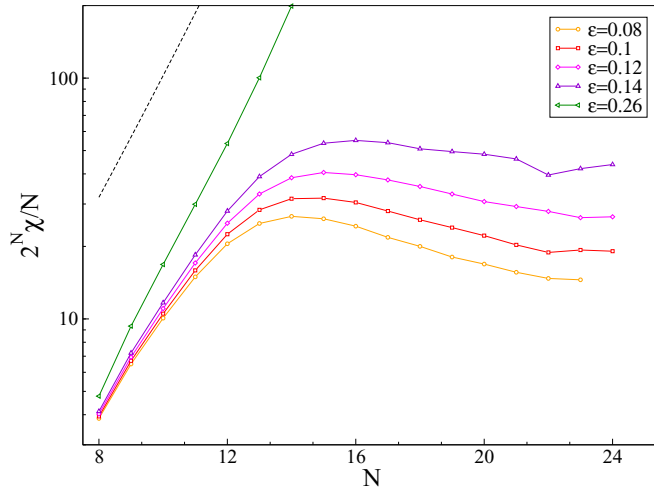


FIG. 5. Level compressibility (obtained from the CBP approach with $n = 2$) on the scale of the mean level spacing rescaled by the GOE asymptotic behavior, $2^N \chi(\varepsilon; 2c\delta)/N$, as a function of N for $\Gamma = 0.2$ and several values of ε . The black dashed line corresponds to $\chi = 1$ (Poisson statistics).

nonuniversal behavior, on a larger energy scale (proportional to $\tilde{\nu}^{D_2-1}$) which goes to zero exponentially with N but stays much larger than δ [36]. This scenario is also supported by general arguments based on the convergence of the Dyson Brownian motion to its stationary GOE distribution [37,60]. In order to check this idea we have analyzed the behavior of the level compressibility $\chi(\varepsilon; \omega)$ for the number of energy levels inside the interval $[N\varepsilon - \omega/2, N\varepsilon + \omega/2]$ [61], which display

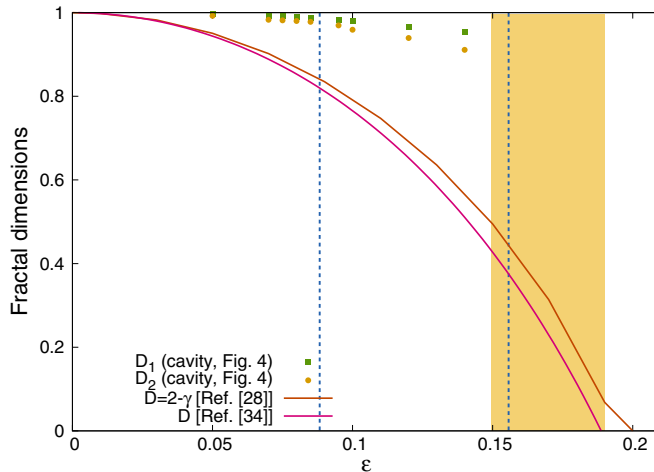


FIG. 6. Fractal dimensions D_1 (green squares) and D_2 (yellow circles) obtained from the cavity approximation as a function of ε for $\Gamma = 0.2$, determined by estimating the height of the plateaus of Fig. 4 at large N . The yellow shaded region indicates the energy interval within which the numerical results of Fig. 4 suggest that the MBL transition occurs, $0.15 \lesssim \varepsilon_{\text{MBL}} \lesssim 0.19$. The vertical blue dashed lines show the position of $\varepsilon_{\text{erg}} \approx 0.0885$ and of $\varepsilon_{\text{MBL}} \approx 0.1585$ found in Sec. IV using the FSA and the mapping onto the RP model (12) and (14). We also plot the results for the fractal dimension D obtained in Refs. [28] (orange line) and [34] (magenta line), which predict that $D \rightarrow 1$ for $\varepsilon \rightarrow 0$ (i.e., $\varepsilon_{\text{erg}} \rightarrow 0$).

different scaling behaviors for the ergodic, localized, and multifractal states [58,62–66]. The number of energy levels inside an energy interval of width ω (and centered around $N\varepsilon$) is defined as $\mathcal{L}(\varepsilon; \omega) = \int_{\varepsilon - \omega/2}^{\varepsilon + \omega/2} \sum_{\alpha=1}^{2^N} \delta(E' - E_\alpha) dE'$. The level compressibility is then the ratio between the variance of $\mathcal{L}(\varepsilon; \omega)$ and its average [61]

$$\chi(\varepsilon; \omega) = \frac{\overline{[\mathcal{L}(\varepsilon; \omega)]^2} - \overline{\mathcal{L}(\varepsilon; \omega)}^2}{\overline{\mathcal{L}(\varepsilon; \omega)}},$$

where $\overline{\dots}$ denotes the average over the disorder. In the diffusive regime of the standard ergodic metallic phase, described by the Wigner-Dyson statistics, energy levels strongly repel each other and the variance scales as $\overline{[\mathcal{L}(\varepsilon; \omega)]^2} - \overline{\mathcal{L}(\varepsilon; \omega)}^2 \propto \ln \overline{\mathcal{L}(\varepsilon; \omega)}$ [61]. The level compressibility thus vanishes as $\chi(\varepsilon; \omega) \propto N \ln 2/2^N$ for large N . Conversely, in the localized phase energy levels are thrown as random points on a line and are described by a Poisson distribution. Hence $\overline{[\mathcal{L}(\varepsilon; \omega)]^2} - \overline{\mathcal{L}(\varepsilon; \omega)}^2 = \overline{\mathcal{L}(\varepsilon; \omega)}$ and $\chi(\varepsilon; \omega) \rightarrow 1$ for $N \rightarrow \infty$. Finally, for nonergodic multifractal states the variance of the number of energy levels inside an interval should scale linearly with the average [63–66], at least in the simplest scenarios, and $\chi(\varepsilon; \omega)$ is expected to converge to a (system-dependent) constant between 0 and 1 in the large- N limit. In the following, for simplicity we will only focus on the behavior of the level compressibility when the energy interval ω is taken of the order of the mean level spacings. In particular, we will set $\omega = 2\eta = 2c\delta$, where η is given in Eq. (B1) and $c = 64$.²

As shown in Refs. [58,62], a simple spectral representation of $\mathcal{L}(\varepsilon; \omega)$ can be achieved in the framework of the CBP approach, in terms of the resolvent matrices defined on the clusters of the hypercube and of the cavity resolvent matrices defined on the edges between the clusters,

$$\mathcal{L}(\varepsilon; \omega) = \frac{1}{\pi} \lim_{\eta \rightarrow 0^+} \left\{ \sum_{p=1}^{2^{N-n}} [\Psi_p(z_+) - \Psi_p(z_-)] + \sum_{\langle p,q \rangle} [\varphi_{p \leftrightarrow q}(z_+) - \varphi_{p \leftrightarrow q}(z_-)] \right\}, \quad (26)$$

where $z_{\pm} = N\varepsilon \pm \omega/2 + i\eta$, the angle $\Psi_p(z)$ is defined as the phase of $\det \mathcal{G}_p(z)$, $\det \mathcal{G}_p(z) = |\det \mathcal{G}_p(z)| e^{i\Psi_p(z)}$, and the angle $\varphi_{p \leftrightarrow q}(z)$ is defined as the phase of $\det[\mathcal{I}_s - \Gamma^2 \mathcal{G}_{q \rightarrow p}(z) \mathcal{G}_{p \rightarrow q}(z)]$ (we have chosen here to put the branch cut in the complex plane along the negative real axis). In order to analyze the scaling properties of the level compressibility we then just need to compute the average of $\mathcal{L}(\varepsilon; \omega)$ and its fluctuations over many independent realizations of the random energies of the hypercube. The scaling behavior of χ when ω is taken on the scale of the mean level spacing $\omega = 2c\delta$ is shown in Fig. 5, where we plot the compressibility (divided by the GOE asymptotic) versus N for $\Gamma = 0.2$ and several values of the energy density in the region of multifractal eigenstates. We observe that $2^N \chi/N$ has a nonmonotonic behavior roughly on the same scale as D_1 and D_2 and seems to approach a finite

²We have checked that varying c from 16 to 128 does not modify the results.

value at large N which grows as ε is increased towards the localized phase. For $\varepsilon \gtrsim 0.26$, in the MBL phase, χ tends instead to 1 at large N , as expected for Poisson statistics. This implies that in the whole multifractal region, on energy scales proportional to δ the level compressibility goes to zero in the thermodynamic limit, as in the diffusive regime of the standard metallic phase.

VI. DISCUSSION

Above we have presented two complementary approximate strategies to determine the out-of-equilibrium phase diagram of the QREM. The first approach is based on the FSA and on the mapping to the RP model [28], while the second approach is a generalization of the self-consistent theory of Anderson localization [38–40] adapted to take into account (at least partially) the local structure of the Hilbert space of the QREM. As discussed in more detail in Appendix B, the latter possibly provides a quite accurate approximation of local observables, such as the distribution of the LDOS, while the former is expected to yield a better estimation of correlations, since it is able to capture the fact that there is a factorial number of paths connecting two points at a large distance in the configuration space.

In this section we discuss and compare in more detail the two approaches concerning the behavior of the fractal dimensions D_1 and D_2 as a function of the energy density ε , which we plot in Fig. 6 for $\Gamma = 0.2$ in the large- N limit. Within the analogy between the QREM and the RP model discussed in Sec. IV, the fractal dimension is expected to be equal to one in the ergodic phase $|\varepsilon| < \varepsilon_{\text{erg}}$ and to zero in the MBL phase $|\varepsilon| > \varepsilon_{\text{MBL}}$ and is conjectured to decrease from 1 to 0 in the intermediate NEE regime (the multifractal spectrum should be obtained as an “average” of the effective fractal dimensions over all x sectors).

We also plot the estimations of Refs. [28] (orange line) and [34], where the effective spectral dimension D is obtained using similar (although probably more accurate) methods to evaluate the amplitude of the tunneling rates $\langle \{\sigma_a^z\} | \Gamma \sigma_b^x | \{\sigma_a^z\} \rangle$ between two distant many-body configurations. The symbols are the cavity-cluster predictions, extracted from the largest size available when reasonably converged to a plateau (see Fig. 4). The shaded area indicates the energy interval within which the numerical results of Fig. 4 suggest that the MBL transition should take place. Due to the limited range of system sizes accessible via the cavity approach, we are not able to conclude whether the fractal dimension would continuously go to zero at ε_{MBL} or rather exhibit a finite jump at the transition.

All approaches agree in indicating the existence of three different phases of the QREM: a fully ergodic regime at low energy density $|\varepsilon| < \varepsilon_{\text{erg}}$; a NEE (or bad metal or dynamical glassy) one at intermediate energy density $\varepsilon_{\text{erg}} < |\varepsilon| < \varepsilon_{\text{MBL}}$, where the time to reach thermal equilibrium is exponentially large in the system size and eigenvectors are extended but multifractal, corresponding to $0 < D_{1,2} < 1$; and a fully localized one, with Anderson localized eigenstates, at high energy $|\varepsilon| > \varepsilon_{\text{MBL}}$ where $D_{1,2} \rightarrow 0$.

However, some quantitative differences also emerge between these three approximations. According to the FSA

calculation of Sec. IV, the ergodic region extends up to a finite energy (as also recently suggested by the numerical results of Ref. [33]), while the approach of Refs. [28,34] predicts instead that $\varepsilon_{\text{erg}} \rightarrow 0$. The cavity approximation indicates that if ε_{erg} is finite, it is significantly smaller than the estimation of Eqs. (13) and (14). As we are going to see in the next section, an argument based on a simplified solution of the cavity equations, equivalent to an auxiliary Anderson model in unconventional thermodynamic limit, also seems to suggest that ε_{erg} might indeed squeeze to zero energy in the thermodynamic limit (as $\sqrt{\ln N/N}$ in the large- N limit) [34]. Another difference is that within the cavity approach the fractal dimensions D_1 and D_2 might possibly exhibit an abrupt jump from a finite value smaller than one to zero at the transition between the nonergodic delocalized and fully localized regime, while the mapping to the RP model indicates that if one identifies $D = 2 - \gamma$ [36], the fractal dimensions should vanish continuously at the MBL mobility edge, as also found in Refs. [28,34]. These are still open questions for future investigations.

VII. AUXILIARY ANDERSON MODELS AND UNCONVENTIONAL THERMODYNAMIC LIMIT

In this section we further simplify the quantum cavity analysis and introduce a family of auxiliary toy Anderson tight-binding models on a Bethe lattice with connectivity $N \gg 1$ where the volume \mathcal{V} of the system is treated as an independent parameter from N and taken equal to infinity from the start (see also Refs. [39–41,55] for similar approaches in the context of MBL). The basic idea behind this procedure is that in the original Anderson model on the hypercube (see Sec. III) the scaling of the number of sites and the number of neighbors with respect to the size N of the original QREM is remarkably different, the former being exponential $\mathcal{V} = 2^N$ while the latter is linear $k = N$. As such one could hope to get some insight into the solution of the quantum cavity equations by taking the volume to infinity first, possibly providing an estimation for the transition line between the fully ergodic phase and the multifractal bad metal one.

Concretely, we consider a hybrid version of the model (3) where the total number of sites of the lattice is sent to infinity from the start keeping N fixed. Hence, for any given choice of Γ and ε , this leads to a family of tight-binding Anderson model parametrized by the connectivity N , with random on-site disorder of standard deviation $\sqrt{N/2}$, given by Eq. (2). The advantage of this procedure is that now for any choice of Γ , ε , and N , the imaginary regulator can be taken as infinitesimally small (since the mean level spacing vanishes in the thermodynamic limit) and we can study whether the system is in the localized or in the extended phase with standard techniques.

We start by determining the mobility edge $\varepsilon_{\text{loc}}(N) = E_{\text{loc}}(N)/N$ of the auxiliary models by computing the Lyapunov exponent which describes the evolution of the imaginary part Δ of the self-energy Σ , once the iteration relations (19) have been linearized [38]. At a given order n of the cluster expansion, the (cavity) self-energy on a cluster p of $s = 2^n$ sites (in the absence of the 2^n edges with one of the

neighboring clusters q) is an $s \times s$ matrix defined as

$$\Sigma_{p \rightarrow q} = S_{p \rightarrow q} + i\Delta_{p \rightarrow q} = \mathcal{H}_p - z\mathcal{I}_s - G_{p \rightarrow q}^{-1},$$

where \mathcal{H}_p is the Hamiltonian (3) acting on the sites of the cluster and \mathcal{I}_s is the $s \times s$ identity matrix. In the localized phase its imaginary part vanishes for $\eta \rightarrow 0^+$. Hence, in the thermodynamic limit and close to the localization transition, one can take the limit $\eta \rightarrow 0^+$ from the start and linearize the recursive equations (19) with respect to Δ :

$$\begin{aligned} [S_{p \rightarrow q}]_{uv} &= \Gamma^2 \sum_{q_l \in \partial p/q_k} [\mathcal{H}_{q_l} - N\varepsilon\mathcal{I}_s - S_{q_l \rightarrow p}]_{uv}^{-1}, \quad (27) \\ [\Delta_{p \rightarrow q}]_{uv} &= \Gamma^2 \sum_{q_l \in \partial p/q_k} \sum_{w,y=1}^s [\mathcal{H}_{q_l} - N\varepsilon\mathcal{I}_s - S_{q_l \rightarrow p}]_{uw}^{-1} \\ &\quad \times [\Delta_{q_l \rightarrow p}]_{wy} [\mathcal{H}_{q_l} - N\varepsilon\mathcal{I}_s - S_{q_l \rightarrow p}]_{yv}^{-1}. \quad (28) \end{aligned}$$

Since $S_{p \rightarrow q}$ and $\Delta_{p \rightarrow q}$ are random matrices, these equations naturally lead to functional self-consistent equations on their probability distribution (see also Appendix C), which can be solved with arbitrary numerical precision using a population dynamics algorithm [57,58] for each value of Γ , ε , and N .

The Lyapunov exponent Λ describes the exponential growth or the exponential decay of the imaginary part of the diagonal elements of the self-energy with the number of recursion steps r as $\Delta_{\text{typ}} \propto e^{\Lambda r}$. However, in the delocalized phase after few recursions Δ_{typ} becomes of order 1 and the exponential behavior is lost. To circumvent this problem we follow Ref. [67] and add an additional ‘‘inflationary’’ step to the recursion.

- (1) We reach to the stationary distribution of the real part of the self-energies, Eq. (27).
- (2) We initialize the imaginary parts of the self-energy to very small values, e.g., $\Delta_{\text{typ}} = \theta = 10^{-24}$.
- (3) We execute an iteration step using Eq. (28) and update the whole population.
- (4) We compute Δ_{typ} after the iteration.
- (5) We multiply all the imaginary parts by $e^{\hat{\Lambda}} = \theta/\Delta_{\text{typ}}$ such that the typical value of Δ remains equal to θ at each iteration step.
- (6) We go back to step 4.

An estimate of the Lyapunov exponent Λ can then be obtained as the average value of $\hat{\Lambda}$ at stationarity (in principle one should take $\theta \rightarrow 0$ and consider the infinite population size limit).

In Fig. 7 we report the results of an accurate numerical computation of the Lyapunov exponent for $\Gamma = 0.2$ and for several values of the energy density ε and of the connectivity N , performed with population size M ranging from 2^{25} to 2^{27} and with θ from 10^{-16} to 10^{-24} (and for $n = 2$).

One observes that the critical energy density $\varepsilon_{\text{loc}}(N)$ at which the Lyapunov exponent vanishes slowly but continuously decreases as N is increased. In Fig. 8 we plot the dependence of $\varepsilon_{\text{loc}}(N)$ as a function of N for three different values of Γ : We find a similar behavior for all values of the transverse field, at least in the region of the phase diagram where the physical properties of the QREM are dominated by the random term $\Gamma < \sqrt{2}/2$ [30,31]. In particular, we observe that $\varepsilon_{\text{loc}}(N)$ seems to vanish in the large- N limit as $\sqrt{\ln N/N}$. Consistently, we find that the slope $|d\Lambda/d\varepsilon|_{\varepsilon_{\text{loc}}}$ around ε_{loc}

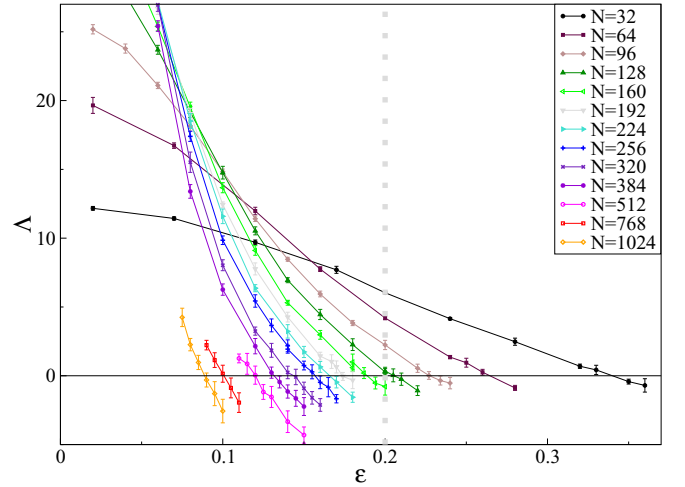


FIG. 7. Lyapunov exponent Λ as a function of the energy $\varepsilon = E/N$ of the hybrid Anderson tight-binding models when \mathcal{V} is sent to infinity at fixed N , for several values of N ranging from 32 to 1024 and for $\Gamma = 0.2$. The results are obtained within the cluster expansion with $n = 2$. The vertical gray dotted line corresponds to the prediction of the FSA for the localization threshold $\varepsilon_{\text{loc}} = \Gamma$ (see Appendix C).

grows roughly as \sqrt{N} (inset). This behavior can be understood in terms of the analytic computation, carried out in Appendix C in full detail, of the largest eigenvalue of the integral operator associated with the self-consistent equation

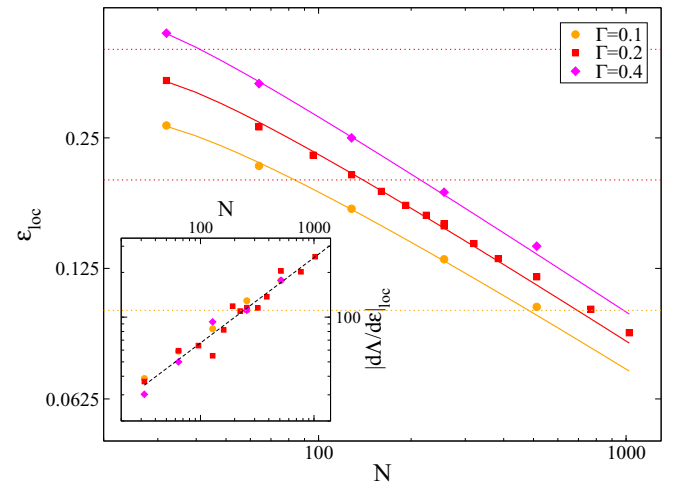


FIG. 8. Mobility edge $\varepsilon_{\text{loc}}(N)$ [such that $\Lambda(\varepsilon_{\text{loc}}) = 0$] of the family of models (3) when the thermodynamic limit is taken from the start, as a function of N for three different values of Γ (and for $n = 2$). The solid curves correspond to the analytical prediction of Eq. (29), with the fitting parameters c_1 and c_2 smoothly varying with Γ as $c_1 \approx -11.51, -9.27, \text{ and } -6.62$ and $c_2 \approx 3.25, 4.54, \text{ and } 6.5$ for $\Gamma = 0.1, 0.2, \text{ and } 0.4$ respectively. The horizontal dotted lines correspond to the prediction of the FSA, $\varepsilon_{\text{loc}} = \Gamma$. The inset shows the slope of the Lyapunov exponent computed at ε_{loc} as a function of N for the same values of Γ as in the main panel. The black dashed line is a power-law fit of the data as $|d\Lambda/d\varepsilon|_{\varepsilon_{\text{loc}}} \simeq AN^\gamma$, with $\gamma \approx 0.57$.

for $Q(S, \Delta)$ in the large-connectivity limit and for $n = 0$ (i.e., in the standard single-site Bethe approximation, when the underlying lattice is taken as a RRG of connectivity N), which yields [38,68,69]

$$\varepsilon_{\text{loc}} = \left[\frac{\ln\{\sqrt{N}\Gamma^2/\pi[\ln(N/\Gamma^4) + c_1]\} + c_2}{N} \right]^{1/2}, \quad (29)$$

where c_1 is a real constant of $O(1)$ which only depends on Γ and c_2 can be expressed in terms of the solution μ_* of the self-consistent equations (C9) for the real part of the self-energy. The predictions of this equation are plotted in Fig. 8 on top of the numerical points, showing good agreement with the numerical results.

It is interesting to compare the asymptotic behavior at large N of the localization threshold ε_{loc} found here for the family of auxiliary models with the large connectivity limit of the standard Anderson model on the Bethe lattice [38,68,69]. In fact, for Bethe lattices of connectivity $k + 1$ and on-site random energies taken from a box distribution of width W , in the large- k (and large-disorder) limit the localization transition (at zero energy) takes place at disorder W_L given by

$$4\rho \ln(W_L) = \frac{1}{k}, \quad (30)$$

where ρ is the density of state in the middle of the spectrum which, at strong disorder, is just given by $\rho(0) \simeq 1/W$. In order to translate this relation to our case, assuming that that $\rho(E) \approx P(E)$, one finds that the exponential dependence on N of the DOS is exactly canceled for $E_{\text{loc}} = N\varepsilon_{\text{loc}}$ given by Eq. (29):

$$\rho(\varepsilon_{\text{loc}}) \approx \frac{e^{-c_2}}{N\Gamma \ln(N/\Gamma^4)}.$$

(We have neglected the constant c_1 for simplicity.) The variance of the random energies of the QREM scales as $\sigma_E^2 = N/2$, which leads us to identify the effective disorder as $W \approx \sqrt{6N}$. Thus, from Eq. (30) one gets

$$\frac{4e^{-c_2} \ln \sqrt{6N}}{N\Gamma \ln(N/\Gamma^4)} \approx \frac{1}{N},$$

which is satisfied for $c_2 = \ln(2/\Gamma)$.

Returning now to the QREM and to its Hilbert space formulation (3) defined on finite Boolean hypercubes of $\mathcal{V} = 2^N$ sites, we argue that the mobility edges of the auxiliary models provide asymptotically in the large- N limit an estimation for the transition line between the fully ergodic phase and the delocalized but multifractal one. The argument goes as follows. Consider a wave function defined on the N -dimensional hypercube which decays exponentially over a finite N -independent length ξ . This corresponds to a multifractal many-body state which occupies roughly ξ^N sites of the Hilbert space, with a fractal dimension $D \sim \log_2 \xi$. In the original many-body setting, when the number of spins N is sent to infinity, the volume of the Hilbert space (2^N) and the connectivity of the hypercube (N) diverge concomitantly, and such wave function will preserve its multifractal nature at all N . If instead the volume of the Hilbert space is sent to infinity while the connectivity is kept fixed, as in the auxiliary model, then such a wave function will appear as genuinely

Anderson localized as it occupies a finite volume. In this sense, it is natural to argue that the apparent N -dependent Anderson localization transition of the family of auxiliary models may in fact capture the transition from ergodic to nonergodic eigenstates of the original problem.

On the other hand, the FSA analysis of the linearized recursion relations of the auxiliary models on the Bethe lattice, which consists in neglecting the real part of the self-energy in the denominators of Eqs. (27) and (28), simply predicts (again in the simplest $n = 0$ setting) that $\varepsilon_{\text{loc}}^{\text{FSA}} = \Gamma$, irrespectively of N (see Appendix C for a detailed calculation). This is the same result for the many-body mobility edge of the QREM at the lowest order in Γ [25]. We argue that the localization threshold predicted by the FSA does not depend on whether the $\mathcal{V} \rightarrow \infty$ limit is taken before the $N \rightarrow \infty$ one or not. In fact, the FSA only keeps the leading-order contribution to the wave-function amplitude at each site and determines the convergence of the perturbative expansion by counting the relative number of resonances found at a given distance compared to the total number of sites accessible at such a distance. In this sense, this approximation captures the transition from the Anderson-localized regime (where the perturbative expansion is convergent, the eigenstates are weakly dressed single configurations of spins and occupy a finite volume on the hypercube) to a delocalized regime (where the perturbative expansion does not converge and resonances can be found at arbitrary large distances) irrespectively of the multifractal nature of the eigenstates. These arguments thus suggest that while $\varepsilon_{\text{loc}}^{\text{FSA}}$ gives a rough estimate of the mobility edge between the MBL phase and the NEE phase [25], ε_{loc} given in Eq. (29) provides an estimation of the transition between the fully ergodic phase and the delocalized nonergodic one:

$$\varepsilon_{\text{MBL}} \approx \Gamma, \quad \varepsilon_{\text{erg}} \approx \left[\frac{\ln\sqrt{N}\Gamma^2/\pi + O(\ln(\ln N))}{N} \right]^{1/2}. \quad (31)$$

According to this interpretation, fully ergodic eigenstates of the QREM are only found in a narrow energy window around $|\varepsilon| < \varepsilon_{\text{erg}}$, which concentrates around zero in the thermodynamic limit (in agreement with Refs. [28,34]). However, the fraction of ergodic eigenfunctions at large N is approximately given by $\int_{-N\varepsilon_{\text{erg}}}^{N\varepsilon_{\text{erg}}} \rho(E)dE \approx 1 - C/\sqrt{N(\ln N)^3}$, where $\rho(E) \approx e^{-E^2/N}/\sqrt{\pi N}$ and C is a constant of order 1 which depends on Γ . Hence, although $\varepsilon_{\text{erg}} \rightarrow 0$, due to the scaling of the many-body energies with N , only a fraction of order $1/\sqrt{N}$ of the 2^N eigenstates are delocalized but nonergodic. Furthermore, a finite value of the mobility edge $\varepsilon_{\text{MBL}} \approx \Gamma$ only corresponds to an exponentially small fraction of Anderson-localized eigenstates in the tails of the DOS.

VIII. CONCLUSIONS AND PERSPECTIVES

In this work we have revisited the dynamical phase diagram of the QREM, using a complementary set of approaches, the FSA coupled to a mapping to the RP model [28] and the self-consistent theory of localization [38] extended to include the local Hilbert space structure of the QREM [39,40]. While the FSA is expected to yield a better estimation of correlations

and large-distance physics, the latter provides a quite accurate approximation of local observables. These approaches provide a qualitatively similar scenario for the phase diagram in the energy-transverse field plane, namely, the existence of three dynamical phases: a fully ergodic delocalized one, an intermediate nonergodic extended regime with multifractal behavior, and an Anderson localized one. Concerning the quantitative features of the phase diagram and the properties of the three phases, there remain however many open questions. The FSA and our RP mapping seem to suggest that ergodic delocalized states exist in an entire region around zero energy density (see also [32]), while the analysis of the quantum cavity equations suggests that if such a region exists it is much narrower in energy, in agreement with Refs. [28,34]. On the other hand, an approximate analytic solution of the cavity equations, corresponding to an auxiliary Anderson model on a Bethe lattice where the connectivity is sent to infinity after the thermodynamic limit is taken, mimicking the exponential scaling of the number of sites of the hypercube, points toward a threshold energy for full delocalization squeezing to zero in the thermodynamic limit [28,34].

It is worth stressing upon concluding that certain features of the QREM make it very peculiar and produce some specific and unique features compared to other generic interacting disordered models such as one-dimensional disordered spin chains or mean-field spin-glass models. One is the absence of correlations between the many-body energies E_i and the spin configurations $\{\sigma_a^z\}$. The other unique feature of the QREM is the fact that in the frozen glassy phase $T < T_K$, the Edwards-Anderson order parameter is equal to one, implying that essentially no spin can be flipped with respect to the initial state. This property implies that in the MBL phase of the QREM many-body wave functions are genuinely Anderson localized and occupy a finite volume in the Hilbert space, while for generic interacting models one expects that the volume occupied by many-body eigenstates in the configuration space is subexponentially large due to the presence of a finite fraction of active spins [14,45]. This makes the investigation using techniques developed here of other disordered mean-field models interesting and worth pursuing.

ACKNOWLEDGMENTS

We warmly thank L. Ioffe for helpful discussions. This work was partially supported by the Simons Foundation through Grant No. 454935 (G.B.). D.F. was partially supported by the EPSRC (CANES, Grant No. EP/L015854/1) and the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 723955, GlassUniversality).

APPENDIX A: RECURSION RELATIONS FOR THE MATRIX ELEMENTS OF THE RESOLVENT WITHIN THE CLUSTER APPROXIMATION

For $n = 1$ the clusters are simply made by two sites (corresponding to two spin configurations which differ by a single spin flip) connected by an edge (Fig. 9). The cavity resolvent matrix on such a cluster can then just be parametrized by three

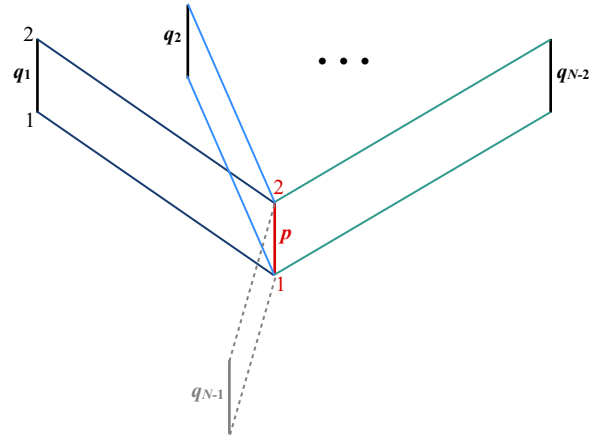


FIG. 9. Schematic representation of the recursion step which yields the self-consistent equations for the 2×2 elements of the (cavity) resolvent matrix for clusters of two sites ($n = 1$).

complex numbers

$$G_{p \rightarrow q} = \begin{pmatrix} g_1^{p \rightarrow q} & g_{12}^{p \rightarrow q} \\ g_{12}^{p \rightarrow q} & g_2^{p \rightarrow q} \end{pmatrix}.$$

The matrix elements of the Hamiltonian (3) on the sites of the cluster are

$$\mathcal{H}_c = \begin{pmatrix} -E_1 & -\Gamma \\ -\Gamma & -E_2 \end{pmatrix}.$$

Thus, Eq. (19) becomes

$$\begin{aligned} \frac{g_2^{p \rightarrow q_k}}{\det G_{p \rightarrow q_k}} &= -E_1 - z - \Gamma^2 \sum_{q_l \in \partial p / q_k} g_1^{q_l \rightarrow p}, \\ \frac{g_1^{p \rightarrow q_k}}{\det G_{p \rightarrow q_k}} &= -E_2 - z - \Gamma^2 \sum_{q_l \in \partial p / q_k} g_2^{q_l \rightarrow p}, \\ \frac{g_{12}^{p \rightarrow q_k}}{\det G_{p \rightarrow q_k}} &= -\Gamma + \Gamma^2 \sum_{q_l \in \partial p / q_k} g_{12}^{q_l \rightarrow p}, \end{aligned}$$

where $\det G_{p \rightarrow q_k} = g_1^{p \rightarrow q_k} g_2^{p \rightarrow q_k} - (g_{12}^{p \rightarrow q_k})^2$. From Eq. (20) one can then write down the equations for the elements of the resolvent matrix on the cluster

$$\begin{aligned} \frac{\mathfrak{g}_2^p}{\det \mathcal{G}_p} &= -E_1 - z - \Gamma^2 \sum_{q_k \in \partial p} g_1^{q_k \rightarrow p}, \\ \frac{\mathfrak{g}_1^p}{\det \mathcal{G}_p} &= -E_2 - z - \Gamma^2 \sum_{q_k \in \partial p} g_2^{q_k \rightarrow p}, \\ \frac{\mathfrak{g}_{12}^p}{\det \mathcal{G}_p} &= -\Gamma + \Gamma^2 \sum_{q_k \in \partial p} g_{12}^{q_k \rightarrow p}, \end{aligned}$$

where \mathfrak{g}_1^p and \mathfrak{g}_2^p are the diagonal elements on sites i and j of the cluster p of the resolvent matrix, \mathfrak{g}_{12}^p is the off-diagonal element, and $\det \mathcal{G}_p = \mathfrak{g}_1^p \mathfrak{g}_2^p - (\mathfrak{g}_{12}^p)^2$. When $g_{12}^{p \rightarrow q_k}$ and $g_{12}^{p \rightarrow q_k}$ are set to zero, these equations return the standard (cavity) recursion equations for the single-site Anderson model on the Bethe lattice [38,57]. Moreover, since the off-diagonal elements are proportional to Γ , in the limit of small transverse

field one might expand these equations in powers of Γ to obtain the systematic corrections to the zeroth-order equations due to the small loops

$$\begin{aligned} g_{1,0}^{p \rightarrow qk} &= -E_1 - z - \Gamma^2 \sum_{q_l \in \partial p/qk} g_{1,0}^{q_l \rightarrow p}, \\ g_{2,0}^{p \rightarrow qk} &= -E_2 - z - \Gamma^2 \sum_{q_l \in \partial p/qk} g_{2,0}^{q_l \rightarrow p}, \\ \delta g_{12,1}^{p \rightarrow qk} &= -\Gamma g_{1,0}^{p \rightarrow qk} g_{2,0}^{p \rightarrow qk} + O(\Gamma^3), \\ \delta g_{1,1}^{p \rightarrow qk} &= \Gamma^2 (g_{1,0}^{p \rightarrow qk})^2 g_{2,0}^{p \rightarrow qk} + O(\Gamma^4), \\ \delta g_{2,1}^{p \rightarrow qk} &= \Gamma^2 (g_{2,0}^{p \rightarrow qk})^2 g_{1,0}^{p \rightarrow qk} + O(\Gamma^4), \end{aligned}$$

where $g_{1,0}^{p \rightarrow qk}$ and $g_{2,0}^{p \rightarrow qk}$ are the diagonal elements of the resolvent at the zeroth order of the cluster expansion (i.e., within the standard single-site Bethe approximation) and $\delta g_{12,1}^{p \rightarrow qk}$, $\delta g_{1,1}^{p \rightarrow qk}$, and $\delta g_{2,1}^{p \rightarrow qk}$ are the corrections for $n = 1$ up to the lowest order in Γ .

One can proceed in a similar way for $n = 2$ and obtain closed equations for the ten independent elements of the cavity resolvent matrices on each cluster in terms of the elements of the cavity resolvent matrices on the neighboring clusters. However, the equations are much longer and we do not write them here explicitly. It is just worth mentioning that in this case the off-diagonal elements of the resolvent matrix between pairs of sites of the cluster that are connected by an edge on the hypercube (e.g., $g_{12}^{p \rightarrow qk}$, $g_{23}^{p \rightarrow qk}$, $g_{34}^{p \rightarrow qk}$, and $g_{14}^{p \rightarrow qk}$, using the notation of Fig. 3) are proportional to Γ . Conversely, the Green's functions between pairs of sites that are not connected by an edge on the hypercube (e.g., $g_{13}^{p \rightarrow qk}$ and $g_{24}^{p \rightarrow qk}$) are (as expected) proportional to Γ^2 .

1. Spectrum of the kinetic term

It is instructive to study how the spectrum of the kinetic term of the Hamiltonian is modified by the cluster expansion. To this aim we consider the pure limit (in the absence of disorder) of the equations above for $n = 1$. In the pure case ($E_i = E_j = 0$) the hypercube is translationally invariant. One can thus look for a uniform solution of the equations in the form

$$\begin{aligned} \frac{g}{g^2 - h^2} &= -z - (N - 2)\Gamma^2 g, \\ \frac{g}{g^2 - h^2} &= -\Gamma + (N - 2)\Gamma^2 h, \end{aligned}$$

where $g = g_1^{p \rightarrow qk} = g_2^{p \rightarrow qk}$ and $h = g_{12}^{p \rightarrow qk}$ for all p and q . One can then introduce the variables $g_+ = g + h$ and $g_- = g - h$ in terms of which the equations above become

$$\begin{aligned} \frac{1}{g_+} &= -z + \Gamma - (N - 2)\Gamma^2 g_+, \\ \frac{1}{g_-} &= -z - \Gamma - (N - 2)\Gamma^2 g_-, \end{aligned}$$

which coincide with the equations that one obtains for the standard single-site Anderson model on the Bethe lattice in the uniform limit, with energies shifted by $\pm\Gamma$. The DOS is thus modified accordingly. In particular, the edges of the spectrum

are shifted as well by $+\Gamma$ on the right edge and by $-\Gamma$ on the left edge. One can indeed define $\mathbf{g} = \mathbf{g}_1^p = \mathbf{g}_2^p$ and $\mathbf{h} = \mathbf{g}_{12}^p$ for all p and introduce the variables $\mathbf{g}_\pm = \mathbf{g} \pm \mathbf{h}$, which verify the equation

$$\frac{1}{\mathbf{g}_\pm} = -z \pm \Gamma - (N - 1)\Gamma^2 \mathbf{g}_\pm,$$

in terms of which the DOS can be obtained as $\rho_I^{(n=1)} = \text{Im}(\mathbf{g}_+ + \mathbf{g}_-)/2\pi$.

For $n = 2$ a similar treatment of the equations yields

$$\begin{aligned} \frac{1}{g_\pm} &= -z \pm 2\Gamma - (N - 3)\Gamma^2 g_\pm, \\ \frac{1}{g_0} &= -z - (N - 3)\Gamma^2 g_0, \end{aligned}$$

where $g_\pm = g + f \pm h$ and $g_0 = g - f$, where $g = g_u^{p \rightarrow qk}$, $h = g_{12}^{p \rightarrow qk} = g_{23}^{p \rightarrow qk} = g_{34}^{p \rightarrow qk} = g_{14}^{p \rightarrow qk}$, and $f = g_{13}^{p \rightarrow qk} = g_{24}^{p \rightarrow qk}$ for all u , p , and q . Similarly one defines $\mathbf{g} = \mathbf{g}_u^p$, $\mathbf{h} = \mathbf{g}_{12}^p = \mathbf{g}_{23}^p = \mathbf{g}_{34}^p = \mathbf{g}_{14}^p$, and $\mathbf{f} = \mathbf{g}_{13}^p = \mathbf{g}_{24}^p$ for all u and p and introduces the variables $\mathbf{g}_\pm = \mathbf{g} + \mathbf{f} \pm \mathbf{h}$ and $\mathbf{g}_0 = \mathbf{g} - \mathbf{f}$, which verify the equations

$$\begin{aligned} \frac{1}{\mathbf{g}_\pm} &= -z \pm 2\Gamma - (N - 2)\Gamma^2 \mathbf{g}_\pm, \\ \frac{1}{\mathbf{g}_0} &= -z - (N - 2)\Gamma^2 \mathbf{g}_0, \end{aligned}$$

in terms of which the DOS reads

$$\rho_I^{(n=2)} = \text{Im} \left[\frac{\mathbf{g}_+ + \mathbf{g}_- + 2\mathbf{g}_0}{4\pi} \right].$$

A comparison between the exact spectrum of the kinetic term of the Hamiltonian (3), i.e., the adjacency matrix of the N -dimensional hypercube (15), the DOS resulting from the cluster approximation for $n = 0$, i.e., the adjacency matrix of a RRG of connectivity N [Eq. (18)], $n = 1$, and $n = 2$ is shown in Fig. 10 for $\Gamma = 0.2$ and three values of N . The imaginary regulator is set to the value used to solve the recursion equations within the CBP approximation in the presence of the disordered many-body on-site energies, Eq. (B1). At the order n of the cluster expansion the (right) edge of the spectrum is shifted by $+n\Gamma$ ($-n\Gamma$) to the right (left) to the leading order in N .

APPENDIX B: SOLUTION OF THE RECURSION EQUATIONS AND COMPARISON WITH EXACT DIAGONALIZATION

As explained in the main text, the cluster approximation allows us to derive a system of closed equations (19) and (20) for the diagonal elements of the resolvent matrix of (3). A first, and crucial, question that we want to address here is to what extent this approximation provides a good qualitative and quantitative description of the spectral statistics of the QREM. To this aim, in this Appendix we consider samples of moderate size and compare the probability distributions of the LDOS computed from the numerical solution of the recursion relations (19) and (20) with those obtained from EDs of the QREM.

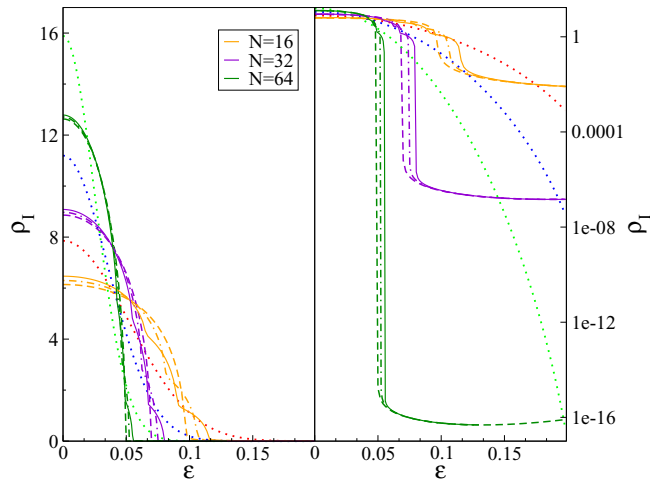


FIG. 10. Comparison between the exact DOS of the delocalizing interacting part of the QREM [i.e., Γ times the adjacency matrix of the N -dimensional hypercube (15), dotted red, blue, and green curves] and the spectra of the kinetic term of Eq. (3) within the cluster approximation for $n = 0$ [i.e., the standard single-site Bethe approximation, which yields $(\Gamma$ times) the adjacency matrix of a RRG of connectivity N , Eq. (18), dashed orange, violet, and dark green curves], $n = 1$ (dash-dotted violet and dark green curves), and $n = 2$ (solid violet and dark green curves) for $\Gamma = 0.2$ and three values of N . At each order of the cluster expansion the edge of the spectrum is shifted by Γ to the right (to the leading order in N). The imaginary regulator η is set to be $\eta = c\delta$ [Eq. (B1)], with $\delta = 1/2^N \rho(E)$.

There are essentially two ways, which we detail below, to solve the recursion equations for the Green's function and obtain information on the spectral statistics at finite N .

(i) *CBP on the hypercube.* The most accurate strategy, which we will refer to as the cluster belief propagation algorithm (see Ref. [58] for a detailed explanation of this approach for the usual tight-binding Anderson model on the Bethe lattice), is to solve directly Eqs. (19) and (20) on random realizations of the hypercube of 2^N sites (i.e., N spins). In practice, one proceeds as follows.

(a) One first generates a random instance of the hypercube drawing the 2^N on-site energies from the distribution (2).

(b) One finds a partition of the hypercube in 2^{N-n} clusters of 2^n sites each (note that the choice of the partition is *not* unique).

(c) Then one finds the fixed point of Eq. (19), which constitute a system of $(s+1)(N-n)2^{N-1}$ coupled equations for the $s(s+1)/2$ independent elements of the cavity Green's functions on each cluster [this can be done iteratively with arbitrary precision in a time which scales linearly with $(s+1)(N-n)2^{N-1}$].

(d) Using Eq. (20) one obtains the $s(s+1)/2$ independent elements of the resolvent matrix on each cluster of that specific instance (and for that specific choice of the partition of the hypercube in clusters).

(e) One then repeats this procedure several times to average over different realizations of the on-site disorder (and over different choices of the cluster partitioning).

As discussed above (see also [58]), in order for the recursive equations to converge to the physical fixed point, the broadening η must be larger than the mean level spacing. Hence, in order to implement the $\eta \rightarrow 0^+$ limit correctly, for any given choice of the parameters Γ , ε , and N , the imaginary regulator is self-consistently set to be a constant of order 1 times the mean level spacing

$$\eta = \frac{c}{2^N \rho_\eta(\varepsilon)} = \frac{c\pi}{\sum_{i=1}^{\mathcal{V}} \text{Im} \mathcal{G}_i(N\varepsilon + i\eta)}. \quad (\text{B1})$$

As shown in [58], for large enough system sizes and for c large enough, ρ_η converges to its asymptotic value obtained in the limit $\mathcal{V} \rightarrow \infty$ and $\eta \rightarrow 0^+$. We will take $c = 64$ throughout.² Within the CBP approach one can study hypercubes of sizes up to $N = 26$, which are considerably larger than the ones accessible via the most efficient ED algorithms.

(ii) *CPD algorithm on the RRG.* In order to access even larger system sizes, one can adopt another strategy, which we will hereafter refer to as the cluster population dynamics algorithm [59], which consists in interpreting the recursion relations for the Green's functions as equations for their probability distributions once the average over the disorder is taken. In fact, since $G_{p \rightarrow q}$ and \mathcal{G}_p are random matrices, one can assume that averaging over the on-site random energies leads to functional equations on their probability distribution $\mathcal{Q}(G)$ and $\mathcal{Q}(\mathcal{G})$. From Eq. (19) we naturally get

$$\mathcal{Q}(G) = \int \prod_{u=1}^{2^n} dP(E_u) \prod_{q=1}^{N-n-1} dQ(G_q) \times \delta \left(G^{-1} + \mathcal{H}_c + z\mathcal{I}_s + \Gamma^2 \sum_{q=1}^{N-n-1} G_q \right), \quad (\text{B2})$$

where $P(E)$ is given by Eq. (2), \mathcal{H}_c is the Hamiltonian (3) on the sites of the cluster (see, e.g., Appendix A for the explicit expression of \mathcal{H}_c for $n = 1$), which contains the s diagonal random energies E_1, \dots, E_s , and \mathcal{I}_s is the $s \times s$ identity matrix. [The notation $dQ(G_q)$ is just a shortcut for the integration over the $s(s+1)/2$ independent elements of G_q .] Once the fixed point of this equation is obtained, using Eq. (20), one can find an equation for the probability distribution of the elements of the resolvent

$$\mathcal{Q}(\mathcal{G}) = \int \prod_{u=1}^{2^n} dP(E_u) \prod_{i=q}^{N-n} dQ(G_q) \times \delta \left(\mathcal{G}^{-1} + \mathcal{H}_c + z\mathcal{I}_s + \Gamma^2 \sum_{q=1}^{N-n} G_q \right). \quad (\text{B3})$$

As before, $z = N\varepsilon + i\eta$ and the imaginary regulator is self-consistently set to be c times the mean level spacing, $\eta = 2^{-N} c\pi / \langle \text{Im} \mathcal{G} \rangle$, where the average is performed over the distribution $\mathcal{Q}(\mathcal{G})$. This set of functional equations can be solved numerically with an arbitrary degree of precision using a population dynamics algorithm [57–59,67]. Hereafter, we will show results obtained using populations of M fields going from $M = 2^{25}$ to $M = 2^{27}$ (and for $n = 2$). Note that, differently from the CBP approach, within the CPD approximation the specific structure of the hypercube is completely lost for

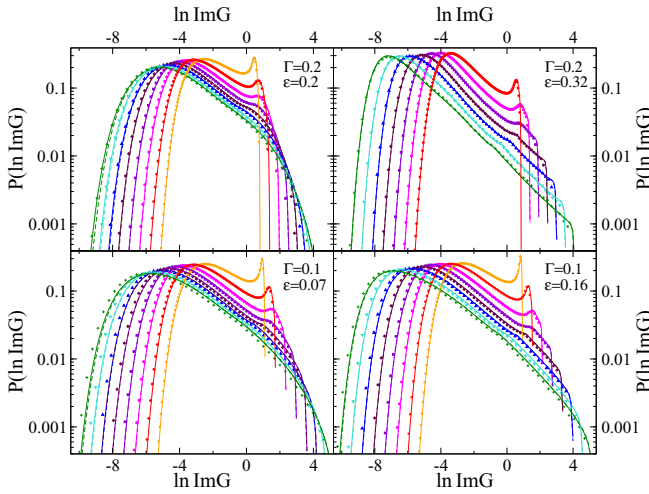


FIG. 11. Probability distributions $P(\ln \text{Im}G)$ obtained from ED (closed symbols), CBP ($n = 2$, continuous curves), and CPD ($n = 2$, dashed curves) for $N = 8$ (orange), 10 (red), 11 (violet), 12 (maroon), 13 (blue), 14 (turquoise), and 15 (green) and for (a) $\Gamma = 0.2$ and $\varepsilon = 0.2$, (b) $\Gamma = 0.2$ and $\varepsilon = 0.32$, (c) $\Gamma = 0.1$ and $\varepsilon = 0.07$, and (d) $\Gamma = 0.1$ and $\varepsilon = 0.16$.

distances larger than the size of the clusters (apart from the local connectivity of each cluster equal to $N - n$).

On the other hand, from exact diagonalization³ we can easily obtain the matrix elements \mathcal{G}_i for a given instance of the QREM in terms of the eigenvalues E_α and the eigenvectors $|\alpha\rangle$ of (1) as

$$\mathcal{G}_i(N\varepsilon + i\eta) = \sum_{\alpha=1}^{2^N} |\langle \alpha | i \rangle|^2 \frac{E_\alpha - N\varepsilon + i\eta}{(E_\alpha - N\varepsilon)^2 + \eta^2}. \quad (\text{B4})$$

For each choice of the parameters Γ , N , and ε , the imaginary regulator is set to the same value as the one used to solve the self-consistent recursion relations (B1).

In Fig. 11 we focus on the probability distribution of the imaginary part of the Green's functions and plot $\mathcal{Q}(\ln \text{Im}\mathcal{G})$ for $\Gamma = 0.1$ and 0.2 , for N ranging from 10 to 15, and for two values of ε which are supposed to be on the ergodic side of the MBL transition and close to the mobility edge, respectively [25,26]. In all cases, we observe good agreement between the probability distributions found from EDs and Eq. (B4), and their CBP counterpart, found from the numerical solution of the recursion relations (19) and (20) on the hypercube for $n = 2$. In the figure we also plot the distributions $\mathcal{Q}(\ln \text{Im}\mathcal{G})$ obtained from the CPD algorithm (B2) and (B3), which presents very small deviations from the CBP results only in the very far tails of the distributions at small $\text{Im}\mathcal{G}$ and are only visible for some values of Γ , ε , and N .

³In fact, the diagonal elements of the resolvent can be obtained by matrix inversion, which is slightly faster than ED.

APPENDIX C: ANALYTIC COMPUTATION OF THE LOCALIZATION THRESHOLD OF THE AUXILIARY ANDERSON MODELS IN THE LARGE- N LIMIT

In this Appendix we discuss the analytical computation of the localization threshold(s) of the family of auxiliary Anderson tight-binding models described by the Hamiltonian (3), when the thermodynamic limit $\mathcal{V} \rightarrow \infty$ is taken from the start while keeping N fixed. For simplicity, we will only consider the simplest setting $n = 0$, i.e., the standard single-site Bethe approximation in which the hypercube is approximated by a treelike structure of connectivity N .

1. Probability distribution of the real part of the self-energy

The first step is to realize that the recursion relations for the real part of the self-energies in the linearized regime are independent of the imaginary part and can be solved as explained below. It is useful to introduce the variables $X_{i \rightarrow j}$ (i.e., the real part of the diagonal elements of the resolvent matrix in the linearized regime) defined as

$$X_{i \rightarrow j} = -\frac{1}{E_i + N\varepsilon + S_{i \rightarrow j}} = G_{i \rightarrow j}^R. \quad (\text{C1})$$

In terms of these variables at the zeroth order of the cluster expansion Eqs. (27) and (28) become

$$\begin{aligned} S_{i \rightarrow j} &= \Gamma^2 \sum_{j' \in \partial i / j} X_{j' \rightarrow i}, \\ \Delta_{i \rightarrow j} &= \Gamma^2 \sum_{j' \in \partial i / j} X_{j' \rightarrow i}^2 \Delta_{j' \rightarrow i}. \end{aligned} \quad (\text{C2})$$

Hence, the probability distribution of the real part of the self-energy $R_S(S)$ can be obtained in terms of the probability distribution $R_X(X)$,

$$\begin{aligned} R_S(S) &= \int \prod_{i=1}^{N-1} dX_i R_X(X_i) \delta\left(S - \Gamma^2 \sum_i X_i\right) \\ &= \int \frac{dk}{2\pi} e^{ikS} \left[\prod_i \frac{dx_i dk_i}{2\pi} e^{i(k_i - \Gamma^2 k)x_i} \hat{R}_X(k_i) \right], \end{aligned}$$

where $\hat{R}_X(k)$ is the characteristic function of $R_X(X)$. Assuming that at small k it behaves as the characteristic function of a Cauchy distribution,

$$\hat{R}_X(k) \simeq 1 - A|k| - ik\mu, \quad (\text{C3})$$

implies that in the large- N limit also $R_S(S)$ is given by a Cauchy distribution

$$\begin{aligned} \hat{R}_S(k) &= [\hat{R}_X(\Gamma^2 k)]^{N-1} \simeq e^{-(N-1)A\Gamma^2|k| - i(N-1)\Gamma^2\mu k}, \\ R_S(S) &= \frac{A_S}{\pi[(S - \mu_S)^2 + A_S^2]}, \end{aligned} \quad (\text{C4})$$

with

$$A_S = N\Gamma^2 A, \quad \mu_S = N\Gamma^2 \mu. \quad (\text{C5})$$

(Throughout we will consider the large- N limit $N - 1 \approx N$.)

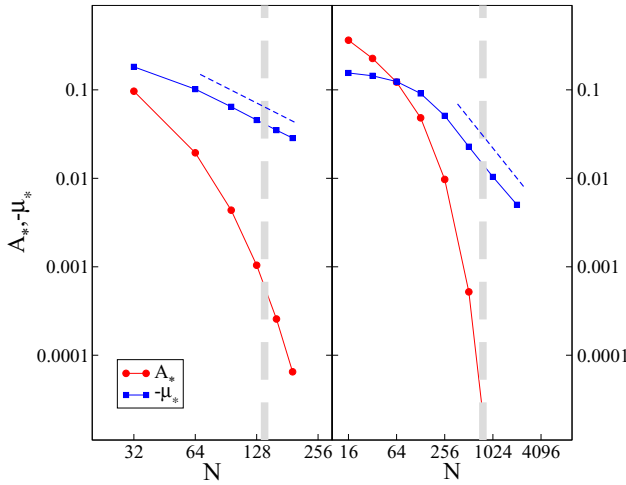


FIG. 12. Solutions A_* (red) and (minus) μ_* (blue) of Eqs. (C9) for $\Gamma = 0.2$ and (a) $\varepsilon = 0.2$ and (b) $\varepsilon = 0.1$ as a function of N . The blue dashed lines correspond to $\mu_* \sim -1/N$. The vertical thick gray dashed lines indicate the localization threshold where the Lyapunov exponent vanishes for these particular values of Γ and ε (see Fig. 7).

On the other hand, from Eq. (C1) we have that

$$\begin{aligned} R_X(X) &= \int dE dS P(E) R_S(S) \delta\left(X + \frac{1}{E + N\varepsilon + S}\right) \\ &= \frac{1}{|X|^2} \int dE P(E) R_S\left(-E - N\varepsilon - \frac{1}{X}\right). \end{aligned} \quad (\text{C6})$$

After some simple algebra we obtain that

$$R_S\left(-E - N\varepsilon - \frac{1}{X}\right) = \frac{cX^2}{\pi[(X - X_0)^2 + c^2]}, \quad (\text{C7})$$

with

$$c = \frac{A_S}{(E + N\varepsilon + \mu_S)^2 + A_S^2}, \quad X_0 = -\frac{E + N\varepsilon + \mu_S}{(E + N\varepsilon + \mu_S)^2 + A_S^2}.$$

As a result, from the second line of Eq. (C6) and from the relations above we get

$$R_X(X) = \int dE P(E) \frac{c}{\pi[(X - X_0)^2 + c^2]}. \quad (\text{C8})$$

We can now finally compute self-consistently the characteristic function of $R_X(X)$ by expanding the equation above up to first order in k :

$$\hat{R}_X(k) = \int dE P(E) e^{-c|k| - ikX_0} \simeq 1 - \int dE P(E) (c|k| + ikX_0).$$

From (C3), (C5), and the preceding equation we can thus obtain two self-consistent relations for the coefficients A and μ :

$$\begin{aligned} A_* &= \int dE P(E_i) \frac{N\Gamma^2 A_*}{(E + N\varepsilon + N\Gamma^2 \mu_*)^2 + (N\Gamma^2 A_*)^2}, \\ \mu_* &= - \int dE P(E) \frac{E + N\varepsilon + N\Gamma^2 \mu_*}{(E + N\varepsilon + N\Gamma^2 \mu_*)^2 + (N\Gamma^2 A_*)^2}. \end{aligned} \quad (\text{C9})$$

These equations can be easily solved numerically. In Fig. 12 we show the solutions A_* and (minus) μ_* of Eqs. (C9) for $\Gamma = 0.2$ and $\varepsilon = 0.2$ [Fig. 12(a)] and $\varepsilon = 0.1$ [Fig. 12(b)] as

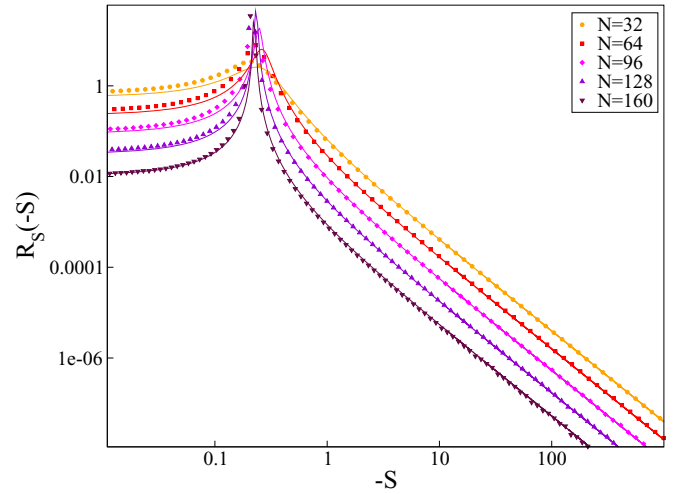


FIG. 13. Probability distributions $R_S(-S)$ for $\Gamma = 0.2$, $\varepsilon = 0.2$, and several values of N across the localization transition of the auxiliary models. Closed symbols are obtained as the numerical solution of the linearized recursion relations for the self-energy with the CPD algorithm for $n = 0$, while solid lines correspond to the analytic prediction of Eqs. (C4), (C5), and (C9).

a function of N . While A_* decays very fast (exponentially) with N , μ_* decreases much slower, roughly as $1/N$ (blue dashed lines). The vertical thick gray dashed lines indicate the localization threshold where the Lyapunov exponent vanishes for these particular values of Γ and ε (see Fig. 7).

In Fig. 13 we plot the probability distribution of the real part of the self-energy $R_S(S)$ for $\Gamma = 0.2$, $\varepsilon = 0.2$, and several values of N across the localization threshold. We focus on the negative real axis since the peak of the distribution is located in $S = \mu_S$, which turns out to be negative. Closed symbols correspond to the numerical solution found using the CPD algorithm (for $n = 0$) of the linearized recursion equations for the self-energy (28), while the solid lines correspond to the analytic prediction (C4), with A_S and μ_S given by Eqs. (C5) and (C9). The agreement between the numerical results and the analytic solution is excellent, and it improves for large N .

2. Computation of the Lyapunov exponent

Once the probability distribution of the real part of the self-energy has been obtained, we can focus on the integral equation for the joint distributions of the real and the imaginary part (for $n = 0$):

$$\begin{aligned} Q(S, \Delta) &= \int \prod_{i=1}^{N-1} [dE_i P(E) dS_i d\Delta_i Q(S_i, \Delta_i)] \\ &\times \delta\left(S + \Gamma^2 \sum_i \frac{1}{E_i + N\varepsilon + S_i}\right) \\ &\times \delta\left(\Delta - \Gamma^2 \sum_i \frac{\Delta_i}{(E_i + N\varepsilon + S_i)^2}\right). \end{aligned} \quad (\text{C10})$$

We replace the δ functions by their integral representation in the Fourier space and also write $Q(S_i, \Delta_i)$ as the inverse Fourier transform of $\hat{Q}_2(S_i, k_i)$ with respect to the second

argument, defined as

$$\hat{Q}_2(S, k) = \int_{-\infty}^{+\infty} d\Delta e^{-ik\Delta} Q(S, \Delta), \quad (\text{C11})$$

yielding

$$\hat{Q}(k_1, k_2) = \int \prod_{i=1} \left[dE_i P(E_i) \frac{dS_i d\Delta_i dk_i}{2\pi} \hat{Q}_2(S_i, k_i) \times e^{ik_1 \Gamma^2 / (E_i + N\varepsilon + S_i)} e^{\Delta_i [k_1 - k_2 \Gamma^2 / (E_i + N\varepsilon + S_i)^2]} \right].$$

We can now perform the integration over $d\Delta_i$, which gives $2\pi \delta(k_1 - k_2 \Gamma^2 / (E_i + N\varepsilon + S_i)^2)$, and then integrate over k_i :

$$\hat{Q}(k_1, k_2) = \left[\int dE P(E) dS \hat{Q}_2 \left(S, \frac{k_2 \Gamma^2}{(E + N\varepsilon + S)^2} \right) \times e^{ik_1 \Gamma^2 / (E + N\varepsilon + S)} \right]^{N-1}. \quad (\text{C12})$$

Similarly to Ref. [38], we assume that in the localized phase the asymptotic form of $Q(S, \Delta)$ holds for large enough Δ ,

$$Q(S, \Delta) \simeq \frac{A(S)}{\Delta^{1+\beta}} \quad \text{for } \Delta \rightarrow \infty, \\ \hat{Q}_2(S, k) \simeq \hat{Q}_2(S, 0) - \alpha |k|^\beta A(S) \quad \text{for } k \rightarrow 0, \quad (\text{C13})$$

where $\hat{Q}_2(S, 0)$ is by definition the marginal of $Q(S, \Delta)$ once we integrate over Δ , i.e., $Q(S, \Delta) = R_S(S)$. Plugging the asymptotic form (C13) into both sides of Eq. (C12), we obtain

$$\hat{R}_S(k_1) - \alpha |k_2|^\beta \hat{A}(k_1) \simeq \left[\int dE P(E) dS \left(R_S(S) - \alpha \left| \frac{k_2 \Gamma^2}{(E + N\varepsilon + S)^2} \right|^\beta A(S) \right) e^{ik_1 \Gamma^2 / (E + N\varepsilon + S)} \right]^{N-1}. \quad (\text{C14})$$

We can now expand the right-hand side of (C14) in powers of k_2 up to the order $|k_2|^\beta$ and define

$$I_1 = \int dE P(E) dS R_S(S) e^{ik_1 \Gamma^2 / (E + N\varepsilon + S)}, \\ I_2 = \int dE P(E) dS \frac{\Gamma^{2\beta}}{|E + N\varepsilon + S|^{2\beta}} A(S) e^{ik_1 \Gamma^2 / (E + N\varepsilon + S)}.$$

The right-hand side of Eq. (C14) is given by $(I_1 - \alpha |k_2|^\beta I_2)^{N-1} \simeq I_1^{N-1} [1 - \alpha(N-1) |k_2|^\beta I_2 / I_1] = I_1^{N-1} - \alpha(N-1) |k_2|^\beta I_1^{N-2} I_2$. From Eq. (28) we have that by definition $I_1^{N-1} = \hat{R}_S$. Hence, in the large- N limit $I_1^{N-2} \simeq I_1^{N-1} = \hat{R}_S(k_1)$ we get

$$\hat{A}(k_1) \simeq N \Gamma^{2\beta} \hat{R}_S(k_1) \int dE P(E) dS A(S) \frac{e^{ik_1 \Gamma^2 / (E + N\varepsilon + S)}}{|E + N\varepsilon + S|^{2\beta}}.$$

Changing the variable to $w = E + N\varepsilon + S$, replacing $A(S)$ by the inverse Fourier transform of $\hat{A}(k)$, and integrating over dE , we obtain

$$\hat{A}(k_1) \simeq N \Gamma^{2\beta} \hat{R}_S(k_1) \int \frac{dw dk}{2\pi} e^{-Nk^2/4 - ikN\varepsilon} \times \hat{A}(k) \frac{e^{ikw + ik_1 \Gamma^2 / w}}{|w|^{2\beta}}. \quad (\text{C15})$$

For a given choice of the parameters Γ , ε , and N , the localization threshold at the zeroth order of the cluster expansion is thus given by the value of the energy $\varepsilon_{\text{loc}}(\Gamma, \varepsilon, N)$ such that the largest eigenvalue λ_β of the integral operator defined by the equation above becomes equal to one. As first noticed in Ref. [38] (see also Refs. [68–70]), the kernel of the integral operator is symmetric around $\beta = \frac{1}{2}$ under the transformation $\beta \rightarrow 1 - \beta$, which implies that $\lambda = 1$ if and only if $\beta = \frac{1}{2}$. Since this is the value of interest for the transition, hereafter we will focus on the case $\beta = \frac{1}{2}$ only. The integral over dw can then be performed in terms of modified Bessel functions:

$$\int dw \frac{e^{ikw + ik_1 \Gamma^2 / w}}{|w|} = -2\pi Y_0(2\Gamma \sqrt{kk_1}).$$

Following Ref. [68], we now assume that in the large connectivity limit the eigenvector of the integral operator defined by Eq. (C15) for $\beta = \frac{1}{2}$ is very well approximated by $\hat{R}(k)$. Assuming that the localization transition occurs on such energy scales (apart from logarithmic corrections)

$$\varepsilon = \frac{\tilde{\varepsilon}}{\sqrt{N}}, \quad (\text{C16})$$

with $\tilde{\varepsilon}$ of $O(1)$, the equation for the mobility edge becomes

$$1 = N \Gamma \int \frac{dk}{2\pi} e^{-Nk^2/4 - N\Gamma^2 A_* |k| - ik(\sqrt{N}\tilde{\varepsilon} + N\Gamma^2 \mu_*)} \times [-2\pi Y_0(2\Gamma \sqrt{kk_1})],$$

where we have used Eqs. (C5). Here A_* is exponentially small in N and can be neglected, while $N\mu_*$ is of order 1 at the transition (see Fig. 12) and gives a correction of order $1/\sqrt{N}$ to $\tilde{\varepsilon}$ as $\tilde{\varepsilon}' = \tilde{\varepsilon} + \Gamma^2 \sqrt{N} \mu_*$. Since the integral over k is cut off on a scale $1/\sqrt{N}$ we can then expand the Bessel function, keeping only the leading logarithmic divergence at small k , $Y_0(x) \approx 2 \ln(k)/\pi$. In the $N \rightarrow \infty$ limit we can then change the variable to $\tilde{k} = \sqrt{N}k$, yielding

$$1 = \frac{\sqrt{N}\Gamma}{\sqrt{\pi}} e^{-(\tilde{\varepsilon}')^2} [\ln N - 4 \ln \Gamma + O(1)].$$

Putting everything together, we finally obtain the equation for the mobility edge (29) given in the main text.

3. Forward-scattering approximation

The FSA consists in neglecting the real part of the self-energy in the denominators of Eqs. (27) and (28). From Eq. (C6) one can then compute the probability distribution $R_X(X)$ for $n = 0$,

$$R_X(X) = \frac{1}{|X|^2} \frac{e^{-(1/N)(1/X - N\varepsilon)^2}}{\sqrt{\pi N}},$$

which does not verify exactly the asymptotic form (C3) for its behavior at large X in the presence of the real part S . This implies that, differently from what happens for the large connectivity limit of the usual Anderson tight-binding model on the Bethe lattice, the distribution $R_S(S)$ found within the FSA does not coincide exactly with the distribution obtained in Sec. C1 in the presence of the real parts in the denominators.

Following the steps of the calculation detailed above for the largest eigenvalue of the linearized recursion relations, one

can compute the distribution of the imaginary part of the self-energy as

$$Q(\Delta) = \int \prod_{i=1}^{N-1} [dE_i P(E) \Delta_i Q(\Delta_i)] \times \delta \left(\Delta - \Gamma^2 \sum_i \frac{\Delta_i}{(E_i + N\varepsilon)^2} \right),$$

which gives the self-consistent equation for its Fourier transform

$$\hat{Q}(k) = \left[\int dE P(E) \hat{Q} \left(\frac{k\Gamma^2}{(E + N\varepsilon)^2} \right) \right]^{N-1}.$$

Assuming, as before, the asymptotic form $\hat{Q}(k) \approx 1 - \alpha|k|^\beta$ and expanding the preceding equation at large N , one gets the

equation for the localization threshold within the FSA:

$$1 \approx N\Gamma^{2\beta} \int dE P(E) \frac{1}{|E + N\varepsilon|^{2\beta}}. \quad (\text{C17})$$

The symmetry $\lambda(\beta) = \lambda(1 - \beta)$ is now lost. Hence the transition point is not achieved at $\beta = \frac{1}{2}$, but rather at a given point $\beta_* \in [0, \frac{1}{2}]$ which depends on the other parameters of the auxiliary models. Equation (C17) can be easily solved numerically for any choice of Γ , ε , N , and β and gives the localization threshold $\varepsilon_{\text{loc}}^{\text{FSA}} = \Gamma$, with $\beta_* \rightarrow \frac{1}{2}$ for $N \rightarrow \infty$. Indeed, since the random energy E is typically of order \sqrt{N} , if ε is of order 1, one can expand the denominator in powers of $E/N\varepsilon$, yielding

$$1 \approx N^{1-2\beta} \left(\frac{\Gamma}{\varepsilon} \right)^{2\beta} \left(1 + \frac{\beta(2\beta - 1)}{2N\varepsilon^2} + \dots \right),$$

which, in the $N \rightarrow \infty$ limit and $\beta \rightarrow \frac{1}{2}$, gives $\varepsilon_{\text{loc}}^{\text{FSA}} = \Gamma$.

-
- [1] D. Basko, I. Aleiner, and B. Altshuler, *Ann. Phys. (NY)* **321**, 1126 (2006).
- [2] B. L. Altshuler, Y. Gefen, A. Kamenev, and L. S. Levitov, *Phys. Rev. Lett.* **78**, 2803 (1997).
- [3] I. V. Gornyi, A. D. Mirlin, and D. G. Polyakov, *Phys. Rev. Lett.* **95**, 206603 (2005).
- [4] E. Altman and R. Vosk, *Annu. Rev. Condens. Matter Phys.* **6**, 383 (2015).
- [5] R. Nandkishore and D. A. Huse, *Annu. Rev. Condens. Matter Phys.* **6**, 15 (2015).
- [6] D. A. Abanin and Z. Papić, *Ann. Phys. (Berlin)* **529**, 1700169 (2017).
- [7] F. Alet and N. Laflorencie, *C. R. Phys.* **19**, 498 (2018).
- [8] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, *Rev. Mod. Phys.* **91**, 021001 (2019).
- [9] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, *Science* **349**, 842 (2015).
- [10] P. Bordia, H. P. Lüschen, S. S. Hodgman, M. Schreiber, I. Bloch, and U. Schneider, *Phys. Rev. Lett.* **116**, 140401 (2016).
- [11] J.-y. Choi, S. Hild, J. Zeiher, P. Schauß, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, *Science* **352**, 1547 (2016).
- [12] D. J. Luitz, N. Laflorencie, and F. Alet, *Phys. Rev. B* **91**, 081103(R) (2015).
- [13] D. J. Luitz and Y. Bar Lev, *Ann. Phys. (Berlin)* **529**, 1600350 (2017).
- [14] N. Macé, F. Alet, and N. Laflorencie, *Phys. Rev. Lett.* **123**, 180601 (2019).
- [15] M. Pino, L. B. Ioffe, and B. L. Altshuler, *Proc. Natl. Acad. Sci. USA* **113**, 536 (2016).
- [16] M. Pino, V. E. Kravtsov, B. L. Altshuler, and L. B. Ioffe, *Phys. Rev. B* **96**, 214205 (2017).
- [17] E. J. Torres-Herrera and L. F. Santos, *Ann. Phys. (Berlin)* **529**, 1600284 (2017).
- [18] V. Oganesyan and D. A. Huse, *Phys. Rev. B* **75**, 155111 (2007).
- [19] A. Pal and D. A. Huse, *Phys. Rev. B* **82**, 174411 (2010).
- [20] J. Šuntajs, J. Bonča, T. Prosen, and L. Vidmar, *Phys. Rev. E* **102**, 062144 (2020).
- [21] D. A. Abanin, J. H. Bardarson, G. D. Tomasi, S. Gopalakrishnan, V. Khemani, S. A. Parameswaran, F. Pollmann, A. C. Potter, M. Serbyn, and R. Vasseur, *arXiv:1911.04501*.
- [22] W. De Roeck and F. Huveneers, *Phys. Rev. B* **95**, 155129 (2017).
- [23] D. J. Luitz, F. Huveneers, and W. De Roeck, *Phys. Rev. Lett.* **119**, 150602 (2017).
- [24] P. Ponte, C. Laumann, D. A. Huse, and A. Chandran, *Philos. Trans. R. Soc. A* **375**, 20160428 (2017).
- [25] C. R. Laumann, A. Pal, and A. Scardicchio, *Phys. Rev. Lett.* **113**, 200405 (2014).
- [26] C. L. Baldwin, C. R. Laumann, A. Pal, and A. Scardicchio, *Phys. Rev. B* **93**, 024202 (2016).
- [27] C. L. Baldwin, C. R. Laumann, A. Pal, and A. Scardicchio, *Phys. Rev. Lett.* **118**, 127201 (2017).
- [28] L. Faoro, M. V. Feigel'man, and L. Ioffe, *Ann. Phys. (NY)* **409**, 167916 (2019).
- [29] B. Derrida, *Phys. Rev. Lett.* **45**, 79 (1980).
- [30] Y. Y. Goldschmidt, *Phys. Rev. B* **41**, 4858 (1990).
- [31] T. Jörg, F. Krzakala, J. Kurchan, and A. C. Maggs, *Phys. Rev. Lett.* **101**, 147204 (2008).
- [32] C. L. Baldwin and C. R. Laumann, *Phys. Rev. B* **97**, 224201 (2018).
- [33] T. Parolini and G. Mossi, *arXiv:2007.00315*.
- [34] V. Smelyanskiy, K. Kechedzhi, S. Boixo, H. Neven, and B. Altshuler, *arXiv:1907.01609*.
- [35] P. W. Anderson, *Phys. Rev.* **109**, 1492 (1958).
- [36] V. E. Kravtsov, I. M. Khaymovich, E. Cuevas, and M. Amini, *New J. Phys.* **17**, 122002 (2015).
- [37] D. Facoetti, P. Vivo, and G. Biroli, *Europhys. Lett.* **115**, 47003 (2016).
- [38] R. Abou-Chacra, D. J. Thouless, and P. W. Anderson, *J. Phys. C* **6**, 1734 (1973).
- [39] D. E. Logan and S. Welsh, *Phys. Rev. B* **99**, 045131 (2019).
- [40] S. Roy and D. E. Logan, *Phys. Rev. B* **101**, 134202 (2020).

- [41] S. Roy and D. E. Logan, *Phys. Rev. Lett.* **125**, 250402 (2020).
- [42] T. R. Kirkpatrick, D. Thirumalai, and P. G. Wolynes, *Phys. Rev. A* **40**, 1045 (1989).
- [43] M. Baity-Jesi, G. Biroli, and C. Cammarota, *J. Stat. Mech.* (2018) 013301.
- [44] F. Pietracaprina, V. Ros, and A. Scardicchio, *Phys. Rev. B* **93**, 054201 (2016).
- [45] M. Tarzia, *Phys. Rev. B* **102**, 014208 (2020).
- [46] E. Bogomolny and M. Sieber, *Phys. Rev. E* **98**, 042116 (2018).
- [47] P. A. Nosov, I. M. Khaymovich, and V. E. Kravtsov, *Phys. Rev. B* **99**, 104203 (2019).
- [48] V. E. Kravtsov, I. M. Khaymovich, B. L. Altshuler, and L. B. Ioffe, [arXiv:2002.02979](https://arxiv.org/abs/2002.02979).
- [49] G. De Tomasi, M. Amini, S. Bera, I. M. Khaymovich, and V. E. Kravtsov, *SciPost Phys.* **6**, 014 (2019).
- [50] P. Jacquod and D. L. Shepelyansky, *Phys. Rev. Lett.* **79**, 1837 (1997).
- [51] D. E. Logan and P. G. Wolynes, *Phys. Rev. B* **36**, 4135 (1987).
- [52] D. E. Logan and P. G. Wolynes, *J. Chem. Phys.* **93**, 4994 (1990).
- [53] R. Bigwood, M. Gruebele, D. M. Leitner, and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **95**, 5960 (1998).
- [54] A. De Luca and A. Scardicchio, *Europhys. Lett.* **101**, 37003 (2013).
- [55] S. Roy and A. Lazarides, *Phys. Rev. Research* **2**, 023159 (2020).
- [56] N. C. Wormald, in *Surveys in Combinatorics, 1999*, edited by J. D. Lamb and D. A. Preece, London Mathematical Society Lecture Note Series Vol. 267 (Cambridge University Press, Cambridge, 1999), pp. 239–298.
- [57] G. Biroli, G. Semerjian, and M. Tarzia, *Prog. Theor. Phys. Suppl.* **184**, 187 (2010).
- [58] G. Biroli and M. Tarzia, [arXiv:1810.07545](https://arxiv.org/abs/1810.07545).
- [59] M. Mézard and G. Parisi, *Eur. Phys. J. B* **20**, 217 (2001).
- [60] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Electron. J. Probab.* **18**, 1 (2013).
- [61] M. L. Mehta, *Random Matrices*, 3rd ed. (Academic, New York, 2004).
- [62] F. L. Metz and I. P. Castillo, *Phys. Rev. B* **96**, 064202 (2017).
- [63] B. Altshuler, I. K. Zharekeshev, S. Kotochigova, and B. Shklovskii, *Zh. Eksp. Teor. Fiz.* **94**, 343 (1988) [*Sov. Phys. JETP* **67**, 625 (1988)].
- [64] J. T. Chalker, V. E. Kravtsov, and I. V. Lerner, *J. Exp. Theor. Phys. Lett.* **64**, 386 (1996).
- [65] E. Bogomolny and O. Giraud, *Phys. Rev. Lett.* **106**, 044101 (2011).
- [66] A. D. Mirlin, *Phys. Rep.* **326**, 259 (2000).
- [67] B. L. Altshuler, E. Cuevas, L. B. Ioffe, and V. E. Kravtsov, *Phys. Rev. Lett.* **117**, 156601 (2016).
- [68] V. Bapst, *J. Math. Phys.* **55**, 092101 (2014).
- [69] G. Parisi, S. Pascazio, F. Pietracaprina, V. Ros, and A. Scardicchio, *J. Phys. A: Math. Theor.* **53**, 014003 (2019).
- [70] B. Altshuler and V. Prigodin, *Zh. Eksp. Teor. Fiz.* **95**, 348 (1989) [*Sov. Phys. JETP* **68**, 198 (1989)].