Replica approach to the generalized Rosenzweig-Porter model

Davide Venturelli[1*], Leticia F. Cugliandolo[2,3], Grégory Schehr[2] and Marco Tarzia[3,4]

1 SISSA – International School for Advanced Studies and INFN, via Bonomea 265, 34136 Trieste, Italy
2 Sorbonne Université, Laboratoire de Physique Théorique et Hautes Energies, CNRS UMR 7589, 4 Place Jussieu, Tour 13, 5ème étage, 75252 Paris 05, France
3 Institut Universitaire de France, 1 rue Descartes, 75231 Paris Cedex 05, France
4 Sorbonne Université, Laboratoire de Physique Théorique de la Matière Condensée, CNRS UMR 7600, 4 Place Jussieu, Tour 13, 5ème étage, 75252 Paris Cedex 05, France

* davide.venturelli@sissa.it

September 26, 2022

Abstract

The generalized Rosenzweig–Porter model arguably constitutes the simplest random matrix ensemble displaying a non-ergodic delocalized phase, which we characterize here by using replica methods. We first derive analytical expressions for the average spectral density in the limit in which the size N of the matrix is large but finite. We then focus on the number of eigenvalues in a finite interval and compute its cumulant generating function as well as the level compressibility, i.e., the ratio of the first two cumulants: these are useful tools to describe the local level statistics. In particular, the level compressibility is shown to be described by a universal scaling function, which we compute explicitly, when the system is probed over scales of the order of the Thouless energy. We confirm our results with numerical tests.

Contents

1 Introduction

Quantum non-interacting particles in a disordered potential undergo the Anderson localization transition as the disorder is increased [1]. In one and two dimensions an infinitesimal amount of disorder is sufficient to localize all eigenstates of the Hamiltonian, while in dimension larger than two a critical value of the disorder separates a metallic phase, where the eigenstates are similar to plane waves and spread over the whole volume uniformly, from an insulating phase, where the eigenstates are instead exponentially localized around specific points in space and thereby occupy a finite $\mathcal{O}(1)$ portion of the total volume. It is well established that exactly at the Anderson localization critical point the wave-functions are *multifractal* [2, 3]. This means that they are neither fully delocalized (as in the metallic regime), nor fully localized (as in the insulating phase), since their support set grows with the system size but remains a vanishing fraction of the total volume.

In the last decade, the Hilbert space localization properties of quantum disordered many-body

systems have attracted much interest. In this context, the emergence of this kind of multifractal states has been discussed as a key and robust feature of their phase diagram and has been invoked to explain some of their unconventional properties beyond the single-particle limit. In the many-body setting, multifractal eigenstates that do not cover the whole accessible Hilbert space generically lead to the violation of the eigenstate thermalization hypothesis (ETH) [4,5], and are therefore often called partially delocalized but non-ergodic (in contrast with the ergodic fully delocalized eigenstates which satisfy ETH).

In the context of Many Body Localization (MBL), recent studies indicate that the many-body eigenstates are in fact multifractal in the whole insulating phase [6–9]. Furthermore, the seminal work by Basko, Aleiner and Altshuler [10] predicted the existence of a novel unconventional "bad metal" regime in between the fully ergodic metallic phase at low disorder and the insulating one at strong disorder. Following the pioneering ideas of [11], the unusual properties of the bad metal regime have been put in relation with the possible multifractal nature of the many-body eigenstates. Recent investigations of the out-of-equilibrium phase diagram of the quantum random energy model [12–16] and Josephson junction arrays [17,18] seem to support these ideas. The existence of partially extended but non-ergodic wave-functions is also believed to have relevant practical and conceptual implications in the efficient population transfer in the context of quantum computing [15,16,19]. Moreover, recent studies of the Sachdev-Ye-Kitaev model in high-energy physics and quantum gravity have reported evidence for the emergence of a non-ergodic extended phase when the model is perturbed by a single-body term [20,21].

Matrix models have been an invaluable tool to describe and help understanding complex physical systems, in particular those with quenched randomness. The physical mechanism at the origin of the above-mentioned partially extended but non-ergodic eigenstates is one such problem, and specific matrix models have recently been used as proxies which capture the peculiar spectral properties associated with them. In this respect, the Rosenzweig-Porter (RP) random matrix ensemble [22], originally introduced to reproduce the spectral properties of complex atomic spectra, provides an archetypal illustration of a system in which a non-ergodic extended phase featuring fractal eigenstates (along with other unconventional spectral properties that will be extensively discussed below) appears in an intermediate region of the phase diagram between a fully ergodic delocalized phase and a fully Anderson localized phase. For this reason the RP model has been the focus of a strong resurgence of attention over the last few years [23–31]. Although one cannot expect that simple random matrix models could capture all the properties of interacting quantum systems, they provide natural and powerful tools to understand the deep physical mechanisms behind some of their features, which are often elusive to analytical treatments in more realistic settings.

The Hamiltonian of the RP model $\mathcal{H} = A + c(N)B$ can be written as the sum of an $N \times N$ diagonal matrix A, whose entries a_i's are independent and identically distributed (i.i.d.) random variables drawn from a Gaussian distribution $p_a(a_i)$, and another $N \times N$ random matrix B belonging to the Gaussian orthogonal (or unitary) ensemble (GOE or GUE, respectively). If the variances of the matrix elements b_{ij} are chosen of $\mathcal{O}(1)$, then the width of the spectrum of B is of $\mathcal{O}(\sqrt{N})$: thus the matrix A (whose spectral width is of $\mathcal{O}(1)$) can produce significant deviations from the GOE/GUE behavior only if $c(N)$ decays sufficiently fast for large N. The properties of this model have been extensively studied by using different techniques, such as a mapping to the Dyson Brownian motion [32], supersymmetry [33,34], resolvent methods [35,36], and first order perturbation theory [37].

Due to the strong surge of attention towards non-ergodic extended states in quantum many-body disordered systems, a generalized version of the RP model in which the distribution $p_a(a_i)$ is
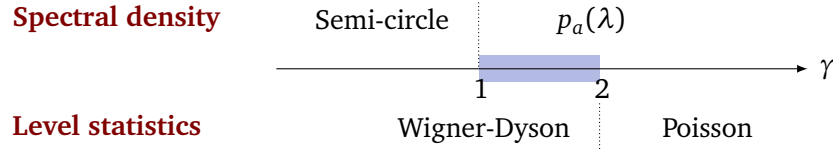
Figure 1: Sketch of the different phases of the RP model, depending on the value of the parameter γ in Eq. (1). Considering the average spectral density $\rho(\lambda)$, a transition is first observed at $\gamma = 1$ which separates a fully delocalized phase where $\rho(\lambda) = \rho_{\text{GOE}}(\lambda)$ from a non-ergodic delocalized phase where $\rho(\lambda) = p_a(\lambda)$ in the thermodynamic limit (see the main text). Focusing instead on the local level statistics, another transition is found at $\gamma = 2$ from the delocalized (non-ergodic) phase characterized by the Wigner-Dyson statistics, to an Anderson-localized phase characterized by the Poisson statistics. The shaded region $1 < \gamma < 2$ indicates the non-ergodic extended phase studied in this work.

not necessarily Gaussian has then been introduced in Ref. [23], and thoroughly investigated by using the techniques recalled above [24–31] – we will refer to this as the GRP model. In addition, new connections and applications have been pointed out in disordered elastic systems [38], many body localization [9,12–14], quantum gravity [20,21], quantum information [15,16,19], or models of theoretical ecology [39].

In this work we revisit the generalized RP model by analyzing some properties of the energy levels and their correlations which have not been investigated in the literature yet. In particular, we perform a thorough study of the finite-N corrections to the average spectral density and compute the level compressibility in the intermediate phase, thereby providing a deeper understanding of the properties of the non-ergodic extended regime. Our analysis uses the replica method largely exploited in the analysis of spin glass models [40], but not only – for example, this tool has been recently applied to study the properties of the ground states in a deformed GOE ensemble [41]. In our study, and for simplicity, we will restrict to the case in which the entries of $\mathcal{H}$ are real numbers, so that $\mathcal{H}$ belongs to the deformed GOE ensemble (the GUE case can be analyzed with only minor changes of the calculations that we develop).

In the rest of this introductory section we will present the RP model and some of its salient properties (Section 1.1), and we will outline our study and main results (Section 1.2).

1.1 The generalized Rosenzweig-Porter model

We consider the Hamiltonian represented by the $N \times N$ matrix

$$\mathcal{H} = A + \frac{\nu}{N^{\gamma/2}} B, \tag{1}$$

where the matrix B belongs to the GOE ensemble: its elements are Gaussian random variables with zero mean and unit variance (i.e., $\langle b_{ij}^2 \rangle = 1$ for $i \neq j$ and $\langle b_{ii}^2 \rangle = 2$). With this choice, the spectrum of B in the limit $N \to \infty$ converges to a Wigner semicircle supported within $\lambda \in [-\sqrt{2N}, \sqrt{2N}]$, where we denote hereafter with λ the eigenvalues of $\mathcal{H}$. The parameter ν is of $\mathcal{O}(1)$ and does not scale with N. The deformation matrix A is instead diagonal, with independent entries a_i's identically distributed according to a generic distribution $p_a(a)$, hence the name Generalized Rosenzweig-Porter model (GRP).

Following the analogy with disordered quantum many-body systems, each matrix index can be thought of as a site of the reference Hilbert space, which is connected to every other site with

the transition rates distributed according to the Gaussian law. Different phenomenologies are expected depending on the value of the parameter γ, which renders one of the two matrices A or B subleading with respect to the other in the limit of large N. As summarized in Fig. 1, the model features three distinct phases (and two transition points between them): a fully delocalized phase for $\gamma < 1$, a fully Anderson localized phase for $\gamma > 2$, and an intermediate delocalized but non-ergodic phase for $1 < \gamma < 2$.

The ergodic transition at $\gamma = 1$ manifests itself as a transition for the average spectral density $\rho(\lambda)$, which reproduces the Wigner semicircle law for $\gamma < 1$ in the $N \to \infty$ limit, while it reduces to $\rho(\lambda) = p_a(\lambda)$ if $\gamma > 1$ and the same $N \to \infty$ limit is taken. At $\gamma = 1$, $\rho(\lambda)$ interpolates between $p_a(\lambda)$ and the Wigner semicircle as the value of the parameter ν is increased. This transition becomes sharp (i.e., it occurs at a particular value of $\nu = \nu_c$) provided that $p_a(a)$ has a compact support and vanishes sufficiently fast at its upper edge [38, 42].

The value $\gamma = 2$ corresponds instead to a genuine Anderson localization transition. Indeed, the region $\gamma < 1$ is characterized by the Wigner-Dyson statistics, meaning that the eigenvectors of $\mathcal{H}$ are uniformly delocalized on N sites in the large N limit, and the average level spacing follows the Wigner surmise [43], signaling level repulsion. Conversely, in the region $\gamma > 2$ the eigenvectors are completely localized over $\mathcal{O}(1)$ sites and the mean level spacing exhibits Poisson statistics.

The intermediate region with $1 < \gamma < 2$ is particularly interesting, because the average spectral density tends to $p_a(\lambda)$, but the *local* level statistics remains of the Wigner-Dyson type. Here the eigenvectors are known to be delocalized over a large number of sites N^{D_γ}, which represent, however, a vanishing fraction of the total number of sites N in the thermodynamic limit, their fractal dimension being $D_\gamma = 2 - \gamma < 1$ [23].

The simplest and most intuitive way to understand the spectral properties in the intermediate region is provided by the Fermi golden rule: in the limit in which the off-diagonal matrix B is absent and all the eigenvectors are trivially localized on a single site, one has $|\psi_i\rangle = |i\rangle$, with corresponding eigenenergies $\lambda_i = a_i$. When the GOE perturbation $\nu N^{-\gamma/2} B$ is turned on, the transition probability per unit time from a state i to another state j can be evaluated perturbatively as

$$\Gamma_{i \to j} = \frac{2\pi\rho}{\hbar} \frac{4\eta}{N} |b_{ij}|^2,$$

where $\rho = \rho(\lambda)$ and, for future convenience, we have introduced the combination

$$\eta \equiv N^{1-\gamma} \nu^2 / 4. \tag{2}$$

Hence, the average escape rate per unit time for a "particle" created in site i at time $t = 0$ reads

$$\Gamma = \sum_{j \neq i} \langle \Gamma_{i \to j} \rangle = \frac{2\pi\rho}{\hbar} 4\eta. \tag{3}$$

The quantity $\hbar\Gamma$ can thus be interpreted as the bandwidth ΔE that can be reached in a time of $\mathcal{O}(1)$ from a given site i. This implies that the eigenvectors within this energy window are hybridized by the GOE perturbation. For $1 < \gamma < 2$ such energy band grows with the system size as $\Delta E \propto N^{1-\gamma}$ and is much larger than the mean level spacing

$$\delta_N \simeq [N\rho]^{-1}, \tag{4}$$

entailing that the system is not Anderson localized; still, ΔE remains much smaller than the total bandwidth, which is of $\mathcal{O}(1)$, signifying that the system is non-ergodic. The Anderson localization

transition thus occurs when ΔE becomes smaller than the mean level spacing, i.e., for $\gamma \geq 2$: this implies that the average escape time from site i grows with N, and thus the eigenfunctions remain localized on $\mathcal{O}(1)$ sites. Conversely, the transition to the fully ergodic phase takes place when ΔE becomes of the order of the total bandwidth, i.e., for $\gamma \leq 1$: this implies that, starting from site i, the wave-packet can reach any other site in a time of $\mathcal{O}(1)$.

In the intermediate phase, $1 < \gamma < 2$, the support set of the eigenvectors (i.e., the number of sites which are hybridized by the perturbation) is simply given by the spreading of the energy interval divided by the average gap between adjacent energy levels, and thus scales as $\Delta E/\delta_N \sim N^{D_\gamma}$, with $D_\gamma = 2 - \gamma$. The partially extended but non-ergodic eigenstates are therefore linear combinations of a bunch of N^{D_γ} localized states associated to nearby energy levels, i.e.,

$$|\psi_i\rangle \approx \sum_{\substack{i' \, s.t. \\ |a_i - a'_i| \leq \Delta E}} c_{i'} |i'\rangle,$$

with coefficients $c_{i'}$ of order $N^{-D_\gamma/2}$ to ensure normalization. These eigenstates give rise to the so-called *mini-bands* in the local spectrum [23]. The width of the mini-bands sets the energy scale $E_T \sim \Delta E \sim N^{D_\gamma - 1} = N^{1-\gamma}$, often called the Thouless energy [44, 45], within which GOE-like spectral correlations (and in particular level repulsion) have been established. All the moments of the wave-functions' coefficients (the so-called generalized *inverse participation ratios*, IPR) behave as $I_q = \sum_i |\langle i|\psi\rangle|^{2q} \propto N^{D_\gamma(1-q)}$: this implies that all the fractal dimensions D_q are degenerate and equal to D_γ for all q, i.e., that the intermediate phase of the GRP model is fractal but not multifractal.

As discussed above, the emergence of such non-ergodic extended phase is particularly relevant in many physical contexts. Its existence was first suggested in Ref. [23] and then rigorously proven in Ref. [24]. In recent years several generalizations of the RP model have been put forward and analysed [46–51], and many other random matrix ensembles have been shown to have an intermediate partially delocalized but non ergodic phase with similar spectral properties [52–60]. Yet, the GRP setting is still a very useful playground to analyze the properties of fractal states in a controlled framework.

1.2 Outline of this work and summary of the main results

As explained above, the GRP model has been intensively investigated over the past few years with a great variety of analytical and numerical techniques. In this paper we tackle this model by applying yet another approach, namely the replica formalism [40], which allows us to obtain new results on the average spectral density, and the statistics of the number of energy levels within a finite interval.

In Section 2 we start by analyzing the average density of states $\rho(\lambda)$. When the size N of the matrix is large, we find the leading order estimate

$$\rho(\lambda) = -\frac{1}{\pi\eta} \lim_{\varepsilon \to 0^+} \text{Re} \, C(\lambda_\varepsilon) + \mathcal{O}(1/N), \tag{5}$$

where $\lambda_\varepsilon = \lambda - i\varepsilon$, the parameter η depends on N and was introduced in Eq. (2), and the function $C(\lambda)$ is implicitly defined by the self-consistency equation

$$C(\lambda) = i\eta \, \mathcal{G}_a \left[\lambda + 2iC(\lambda)\right]. \tag{6}$$

Here $\mathcal{G}_a$ is the resolvent associated to the distribution $p_a(a)$ of the entries of A (c.f. Eq. (41)). In general, it is quite difficult to solve explicitly Eq. (6) for $C(\lambda)$ for any arbitrary distribution $p_a(a)$. However, we show in Section 2.4 that it can be solved explicitly for two special cases: (i) when $p_a(a)$ is a Wigner semicircle – which is expected, since it is stable under free-convolution [61] – and (ii) when it is a Cauchy distribution, which is more surprising.

As we show in Section 2.3, by taking the limit $N \to \infty$ with η kept finite, Eqs. (5) and (6) reduce to the free addition formula [61], sometimes called the "Zee formula" in the physics literature [62]. However, as we show in Section 2.5, these equations contain more information and allow, in the case where $\eta = \eta(N) \ll 1$ defined in Eq. (2) with $1 < \gamma < 2$, to obtain the leading $1/N$ corrections to the limiting density $\rho(\lambda) = p_a(\lambda)$ in a controlled way, even when the resolvent $\mathcal{G}_a(z)$ does not admit a closed-form analytic expression. These corrections turn out to be quite difficult to compute using the standard free addition formula, which in principle holds only in the limit $N \to \infty$.

Next, in Section 3 we analyze the behavior of the *level compressibility* $\chi(E)$, which is a simple indicator of the degree of level repulsion and is defined as follows[1]. We first introduce the empirical density of the (real) eigenvalues λ_i of $\mathcal{H}$, defined as

$$\rho_N(\lambda) = \frac{1}{N} \sum_{i=1}^{N} \delta(\lambda - \lambda_i). \tag{7}$$

Note that $\rho_N(\lambda)$ is normalised to unity. Let

$$I_N[\alpha, \beta] \equiv N \int_\alpha^\beta d\lambda \, \rho_N(\lambda) \tag{8}$$

denote the number of eigenvalues lying in the interval $[\alpha, \beta] \subseteq \mathbb{R}$, which is a random variable. Then

$$\chi(E) \equiv \frac{\kappa_2(E)}{\kappa_1(E)} = \frac{\left\langle I_N^2[-E, E] \right\rangle - \left\langle I_N[-E, E] \right\rangle^2}{\left\langle I_N[-E, E] \right\rangle}, \tag{9}$$

where κ_1 and κ_2 are the first two cumulants of I_N. For Poisson statistics, one has $\kappa_2(E) \simeq \kappa_1(E)$ for small E, and then $\chi(E) \simeq 1$ (see Appendix A). On the contrary, for a rigid spectrum like that of the GOE matrix B in Eq. (1), the mean number of eigenvalues behaves as $\langle I_N[-E, E] \rangle \propto \tilde{E}$, with $\tilde{E} \equiv N \rho_N(0) E$ and where $[N \rho_N(0)]^{-1}$ is the mean level spacing close to $E = 0$, while $\left\langle I_N^2[-E, E] \right\rangle_c \propto \ln \tilde{E}$ for large $\tilde{E}$. Hence in the GOE case one finds $\chi(E) \to 0$ for $\tilde{E} \to \infty$, i.e., for $E \gg [N \rho_N(0)]^{-1}$ (but still much smaller than $E \sim \mathcal{O}(1)$).

In Section 3.3 we provide the cumulant generating function of the variable $I_N[\alpha, \beta]$ at leading order for large N. For a symmetric interval $[-E, E]$ and a symmetric distribution $p_a(a)$, the result reads

$$\mathcal{F}_{[-E,E]}(s) \equiv \frac{1}{N} \ln \left\langle e^{-s I_N[-E,E]} \right\rangle = -ms + \ln \left\langle e^{-s f(a)} \right\rangle_a + \mathcal{O}(\eta/N), \tag{10}$$

where

$$m = -\frac{2\eta}{\pi} \text{Im} \left[\mathcal{G}_a \left(-i\Delta^{-1} \right) \right]^2, \quad f(a) \equiv \frac{1}{\pi} \arctan \left(\frac{\sin 2\theta}{a^2 r^2 + \cos 2\theta} \right) \in [0, 1], \tag{11}$$

and $\Delta(E) = r(E) e^{i\theta(E)}$ has to be determined by solving the self-consistency equation

$$\Delta^{-1} = \varepsilon - iE - 2i\eta \mathcal{G}_a \left(-i\Delta^{-1} \right). \tag{12}$$

---

[1]It is known in statistics under the name of "Fano factor" [63], and it has been studied recently in physics in the context of extremes and record statistics of time series [64, 65].

before sending $\varepsilon \to 0$. Again, we show that a closed-form solution can be found in some particular cases. Our results are supported by the comparison with the numerical diagonalization of large random matrices.

The ratio of the first two cumulants of $I_N[-E, E]$ gives the level compressibility $\chi(E)$ introduced in Eq. (9). In agreement with the non-ergodic picture presented above for the region $1 < \gamma < 2$, and having identified $E_T \propto \eta \propto N^{1-\gamma}$ as the Thouless energy of the system, we explicitly verify in Section 3.5 that $\chi(E) \sim 0$ for $E \ll E_T$ – but still much larger than the mean level spacing $\delta_N \propto N^{-1}$ – corresponding to level repulsion, while $\chi(E)$ follows the Poisson statistics for $E \gg E_T$. Finally, in the scaling limit $E = 2\pi p_a(0)\eta \cdot y$ with $N^{-1} \ll \eta \ll 1$, we show that $\chi(E)$ takes the universal form

$$\chi(E) \simeq \chi_T\left(y = \frac{E}{2\pi p_a(0)\eta}\right), \qquad \chi_T(y) = \frac{1}{\pi y}\left[2y\arctan(y) - \ln\left(1 + y^2\right)\right], \qquad (13)$$

where the scaling function $\chi_T(y)$ is independent of the specific choice of $p_a(a)$. This function is plotted in Fig. 6: it behaves as $\chi_T(y) \simeq y/\pi$ for small y, while it tends to 1 for large y. Note that the crossover energy scale $2\pi p_a(0)\eta$ coincides (apart from a factor of 4) with the width of the mini-bands identified in Eq. (3) using the Fermi golden rule, which allows us to put the intuitive arguments given above on the structure of spectral correlations on a much firmer basis.

In Section 3.6 we finally inspect, using extensive numerical diagonalization of large random matrices, the low-energy region where E is chosen on the scale of the Thouless energy. The scaling form of $\chi(E)$ presented in Eq. (13) is thus shown to represent a universal crossover between the classical GOE result $\chi(E \sim N^{-1}) \simeq \chi_{\text{GOE}}(E)$ for energies of the order of the mean level spacing (and much smaller than the Thouless energy, see Eq. (E.25)), and the model-dependent prediction of Eqs. (9) and (10), valid for energies of the order of the total spectral band-width, i.e., $E \sim \mathcal{O}(1)$.

2 Average spectral density

Let us begin by considering the density of states of the matrix $\mathcal{H}$ in Eq. (7). Its mean value can be obtained by means of the Edwards-Jones (E-J) formula [43, 66], which we briefly recall here. One starts from the Plemelj-Sokhotski formula: if $f(x)$ is a complex-valued function which is continuous on the real axis, and given $\alpha < 0 < \beta$, then

$$\lim_{\varepsilon \to 0^+} \int_\alpha^\beta \mathrm{d}x\, \frac{f(x)}{x \pm i\varepsilon} = \mp i\pi f(0) + \mathcal{P}\int_\alpha^\beta \mathrm{d}x\, \frac{f(x)}{x}, \qquad (14)$$

where $\mathcal{P}$ indicates the Cauchy principal value of the integral. From Eq. (7) we then have

$$\langle \rho_N(\lambda) \rangle = \frac{1}{\pi N} \lim_{\varepsilon \to 0^+} \operatorname{Im}\left\langle \sum_{i=1}^N \frac{1}{\lambda - \lambda_i - i\varepsilon} \right\rangle = \frac{1}{\pi N} \lim_{\varepsilon \to 0^+} \operatorname{Im}\frac{\partial}{\partial \lambda}\left\langle \sum_{i=1}^N \ln(\lambda_i - \lambda + i\varepsilon) \right\rangle, \qquad (15)$$

where the average is taken over the distribution of the entries of $\mathcal{H}$, and where in the last step we indicated by $\ln(z)$ the principal branch of the complex logarithm. Using the properties of Gaussian integrals, we finally obtain the E-J formula [66]

$$\rho(\lambda) \equiv \langle \rho_N(\lambda) \rangle = -\frac{2}{\pi N} \lim_{\varepsilon \to 0^+} \operatorname{Im}\frac{\mathrm{d}}{\mathrm{d}\lambda}\langle \ln \mathcal{Z}(\lambda_\varepsilon) \rangle, \qquad (16)$$

$$\mathcal{Z}(\lambda) \equiv \det(\mathcal{H} - \lambda\mathbb{1})^{-1/2} = (2\pi i)^{-N/2}\int_{\mathbb{R}^N} \mathrm{d}\mathbf{r}\, e^{-\frac{i}{2}\mathbf{r}^T(\lambda\mathbb{1} - \mathcal{H})\mathbf{r}}, \qquad (17)$$

where $\lambda_\varepsilon \equiv \lambda - i\varepsilon$ with $\varepsilon > 0$. Note that the negative imaginary part of λ_ε ensures the convergence of the integral in Eq. (17). The expectation value of the logarithm can then be handled by using the replica trick [40]

$$\langle \ln \mathcal{Z}(\lambda) \rangle = \lim_{n \to 0} \frac{1}{n} \ln \langle \mathcal{Z}^n(\lambda) \rangle . \tag{18}$$

Computing the last expectation value by standard methods (see Appendix B) leads to

$$\langle \mathcal{Z}^n(\lambda) \rangle \propto \int \mathcal{D}\mu \, \mathcal{D}\hat{\mu} \, e^{N \mathcal{S}_n[\mu, \hat{\mu}; \lambda]} , \tag{19}$$

where the proportionality holds up to an irrelevant numerical constant, and where we introduced the action

$$\mathcal{S}_n[\mu, \hat{\mu}; \lambda] \equiv -i \int d\vec{y} \, \mu(\vec{y})\hat{\mu}(\vec{y}) - \frac{\eta}{2} \int d\vec{y} \, d\vec{w} \, \mu(\vec{y})\mu(\vec{w}) \, (\vec{y} \cdot \vec{w})^2$$
$$+ \ln \int d\vec{y} \int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda - a)|\vec{y}|^2 + i\hat{\mu}(\vec{y})\right]. \tag{20}$$

The parameter η is defined in Eq. (2), while $\vec{y}, \vec{w}$ are n-dimensional vectors (one component for each of the replicas).

The strategy to obtain the finite-N averaged $\rho(\lambda)$ is the following:

1. For large N, we look for a saddle-point estimate of the path integral in Eq. (19) in the form

$$\langle \mathcal{Z}^n(\lambda) \rangle \propto e^{N \mathcal{S}_n[\mu^*, \hat{\mu}^*; \lambda] + \mathcal{O}(1)} , \tag{21}$$

where the proportionality holds up to a λ-independent (even though possibly N-dependent) prefactor.

2. Using Eq. (16), we recover the spectral density via

$$\rho(\lambda) \simeq -\frac{2}{\pi} \lim_{\varepsilon \to 0^+} \operatorname{Im} \lim_{n \to 0} \frac{1}{n} \frac{d}{d\lambda} \mathcal{S}_n[\mu^*, \hat{\mu}^*; \lambda_\varepsilon] \tag{22}$$
$$= \frac{1}{\pi} \lim_{\varepsilon \to 0^+} \operatorname{Im} \left\{ \lim_{n \to 0} \frac{i}{n} \frac{\int d\vec{y} \, |\vec{y}|^2 \int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda_\varepsilon - a)|\vec{y}|^2 + i\hat{\mu}^*(\vec{y})\right]}{\int d\vec{y} \int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda_\varepsilon - a)|\vec{y}|^2 + i\hat{\mu}^*(\vec{y})\right]} \right\} .$$

Indeed, only the third term in the action of Eq. (20) will contribute, because the dependence on λ in the first two terms is only implicit,

$$\frac{d}{d\lambda} \mathcal{S}_n[\mu, \hat{\mu}; \lambda_\varepsilon] = \partial_\lambda \mathcal{S}_n + \int d\vec{y} \left[\frac{\delta \mathcal{S}_n}{\delta \mu(\vec{y})} \frac{d\mu(\vec{y})}{d\lambda} + \frac{\delta \mathcal{S}_n}{\delta \hat{\mu}(\vec{y})} \frac{d\hat{\mu}(\vec{y})}{d\lambda} \right] , \tag{23}$$

and the term under the integral vanishes at the saddle-point (where the action is stationary by construction). In turn this implies that, to compute $\rho(\lambda)$ from Eq. (22), we do not need to determine $\mu^*(\vec{y})$, but only $\hat{\mu}^*(\vec{y})$. Finally, by the $\simeq$ symbol in Eq. (22) we mean that the corrections are *at most* $\mathcal{O}(1/N)$.

2.1 Saddle-point equations and rotationally-invariant Ansatz

The leading contribution to Eq. (19) for large N can be found by minimizing the action in Eq. (20). Omitting (to ease the notations) the superscript $*$ from $\mu^*(\vec{y})$ and $\hat{\mu}^*(\vec{y})$, and understanding the dependencies on λ as computed in correspondence of λ_ε, the saddle-point equations read

$$0 \equiv \frac{\delta \mathcal{S}_n}{\delta \mu(\vec{x})} = -i\hat{\mu}(\vec{x}) - \eta \int d\vec{w} \, \mu(\vec{w})(\vec{x} \cdot \vec{w})^2 , \tag{24}$$

$$0 \equiv \frac{\delta \mathcal{S}_n}{\delta \hat{\mu}(\vec{x})} = -i\mu(\vec{x}) + i \frac{\int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda - a)|\vec{x}|^2 + i\hat{\mu}(\vec{x})\right]}{\int d\vec{y} \int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda - a)|\vec{y}|^2 + i\hat{\mu}(\vec{y})\right]} . \tag{25}$$

Substituting the expression for $\mu(\vec{x})$ obtained from the second equation (25) in the first one (24), one obtains a closed equation for $\hat{\mu}(\vec{x})$ which reads

$$\hat{\mu}(\vec{x}) = i\eta \frac{\int d\vec{y} \int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda - a)|\vec{y}|^2 + i\hat{\mu}(\vec{y})\right](\vec{x} \cdot \vec{y})^2}{\int d\vec{y} \int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda - a)|\vec{y}|^2 + i\hat{\mu}(\vec{y})\right]} . \tag{26}$$

To make progress, we plug in the Ansatz $\hat{\mu}(\vec{x}) = \hat{\mu}(x)$, with $x \equiv |\vec{x}|$, which is rotationally symmetric in the space of replicas (i.e., it is invariant under $O(n)$ transformations). Note that requiring invariance under $O(n)$ is a stronger request than the mere replica-symmetry (RS): indeed, the exchange between any pair of components of $\vec{x}$ can be obtained by means of a $O(n)$ transformation. Stepping to spherical coordinates and using the identity

$$\int d\Omega_n (\vec{x} \cdot \vec{y})^2 = \frac{(xy)^2}{n} \int d\Omega_n , \tag{27}$$

where $d\Omega_n$ is the differential of the n-dimensional solid angle, we find

$$\hat{\mu}(x) = i\frac{\eta}{n} \frac{\int_0^\infty dy \, y^{n-1} \int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda - a)y^2 + i\hat{\mu}(y)\right](xy)^2}{\int_0^\infty dy \, y^{n-1} \int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda - a)y^2 + i\hat{\mu}(y)\right]} . \tag{28}$$

Let us now introduce the auxiliary function

$$G(y; \lambda) \equiv \int da \, p_a(a) \exp\left[-\frac{i}{2}(\lambda - a)y^2 + i\hat{\mu}(y)\right] = \exp\left[-\frac{i}{2}\lambda y^2 + i\hat{\mu}(y)\right]\psi_a(-y^2/2), \tag{29}$$

where in the second equality we recognized the characteristic function of $p_a(a)$ (i.e., its Fourier transform), namely

$$\psi_a(k) = \int da \, p_a(a)e^{-ika} . \tag{30}$$

Equation (28) can readily be expressed in terms of $G(y; \lambda)$. We then integrate by parts in the denominator of Eq. (28), finding that boundary terms disappear at least as long as $\varepsilon > 0$ in $G(y; \lambda_\varepsilon)$ – we will check *a posteriori* that the presence of $\hat{\mu}(y)$ does not spoil the convergence. We thus get

$$\hat{\mu}(x) = -i\eta x^2 \frac{\int_0^\infty dy \, y^{n+1} G(y; \lambda)}{\int_0^\infty dy \, y^n G'(y; \lambda)} , \tag{31}$$

where $G'(y;\lambda) = \partial_y G(y;\lambda)$. Under this form, we can now take the limit $n \to 0$, which yields

$$i\hat{\mu}(x) = x^2 \eta \frac{\int_0^\infty \mathrm{d}y \, y \, G(y;\lambda)}{\int_0^\infty \mathrm{d}y \, G'(y;\lambda)} \equiv x^2 C(\lambda). \tag{32}$$

The function $C(\lambda)$ must now be determined self-consistently via

$$C(\lambda) = \eta \frac{\int_0^\infty \mathrm{d}y \, y \, G(y;\lambda)}{\int_0^\infty \mathrm{d}y \, G'(y;\lambda)}, \tag{33}$$

where $G(y;\lambda)$ can be read off from Eq. (29) upon setting $i\hat{\mu}(y) = y^2 C(\lambda)$.

2.2 General result

We now repeat the same steps at the level of the saddle-point action in Eq. (22): we plug in the rotationally symmetric Ansatz, we integrate by parts in the denominator and we finally take the limit $n \to 0$. This gives

$$\rho(\lambda) = -\frac{1}{\pi} \lim_{\varepsilon \to 0^+} \mathrm{Re} \, \frac{\int_0^\infty \mathrm{d}y \, y \, G(y;\lambda_\varepsilon)}{\int_0^\infty \mathrm{d}y \, G'(y;\lambda_\varepsilon)} + \mathcal{O}(1/N) = -\frac{1}{\pi\eta} \lim_{\varepsilon \to 0^+} \mathrm{Re} \, C(\lambda_\varepsilon) + \mathcal{O}(1/N). \tag{34}$$

We then go back to Eq. (33), which contains an exact differential in its denominator. If $\mathrm{Re} \, C(\lambda_\varepsilon) \leq 0$, then $G(y \to \infty; \lambda_\varepsilon) = 0$ and $G(0;\lambda_\varepsilon) = 1$, and we obtain a self-consistency equation for $C(\lambda_\varepsilon)$,

$$C(\lambda) = -\eta \int_0^\infty \mathrm{d}z \, \psi_a(-z) e^{-i\lambda z + 2z C(\lambda)}, \tag{35}$$

where we recall that $\psi_a(x)$ is the characteristic function associated with $p_a(a)$ (see Eq. (30)). This determines $C(\lambda)$ implicitly. Equations (34) and (35) represent the main result of this section.

2.2.1 Limiting cases

Here we briefly check the consistency of the result we just obtained in the limiting cases in which only one of the two matrices A or B in Eq. (1) are retained.

First, when $\nu = \eta = 0$ we find $\mathcal{H} = A$, and the level statistics can only be determined by $p_a(a)$. Indeed, from Eq. (24) we read $\hat{\mu}(\vec{x}) = 0$, and even without invoking a particular Ansatz we get from Eq. (22), upon sending $N \to \infty$,

$$\rho(\lambda) = -\frac{1}{\pi} \lim_{\varepsilon \to 0^+} \mathrm{Re} \, \frac{\int_0^\infty \mathrm{d}y \, y \, G(y;\lambda_\varepsilon)}{\int_0^\infty \mathrm{d}y \, G'(y;\lambda_\varepsilon)}, \tag{36}$$

where $G(y;\lambda)$ is given by Eq. (29) with $\hat{\mu} = 0$. For any $\varepsilon > 0$ the function $G(y;\lambda_\varepsilon)$ is well behaved at $y \to +\infty$, so that

$$\rho(\lambda) = \frac{1}{\pi} \lim_{\varepsilon \to 0^+} \mathrm{Re} \int_0^\infty \mathrm{d}y \, y \, G(y;\lambda_\varepsilon) = \frac{1}{\pi} \lim_{\varepsilon \to 0^+} \mathrm{Re} \int_0^\infty \mathrm{d}z \, \psi_a(-z) e^{-i\lambda_\varepsilon z} = p_a(\lambda), \tag{37}$$

where we changed variable to $z = y^2/2$ in the first step, and in the last one we used $\psi_a^*(-z) = \psi_a(z)$.

Conversely, the GOE case is formally recovered by setting $\gamma = 1$ and $a_i = 0$ identically.[2] With this choice, the limiting spectrum of $\mathcal{H}$ is supported within $\lambda \in [-\sqrt{2}\nu, \sqrt{2}\nu]$, and we set for simplicity $\nu = 1$ (corresponding to $\eta = 1/4$). Using Eq. (33) (without integrating by parts in the denominator, but really computing $G'(y; \lambda)$) one easily finds

$$C(\lambda) = \frac{1}{4}\left(i\lambda \pm \sqrt{2 - \lambda^2}\right). \tag{38}$$

We choose the branch with the minus sign, as we assumed above that $\operatorname{Re} C(\lambda_\varepsilon) \leq 0$. Using Eq. (34) then immediately renders, in the limit $N \to \infty$, the semicircle law

$$\rho(\lambda) \to \rho_{\text{GOE}}(\lambda) = \frac{\sqrt{2 - \lambda^2}}{\pi} \Theta(\sqrt{2} - |\lambda|), \tag{39}$$

where $\Theta(x)$ is the Heaviside step function, with $\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ if $x \leq 0$.

2.3 Resolvent formulation and connection with the Zee formula

One can observe that, under the same convergence hypotheses as above, the self-consistency equation (35) which determines $C(\lambda)$ can be rewritten as

$$C(\lambda) = i\eta \, \mathcal{G}_a \left[\lambda + 2iC(\lambda)\right], \tag{40}$$

where

$$\mathcal{G}_a(z) = \int \mathrm{d}a \, \frac{p_a(a)}{z - a} \tag{41}$$

is the resolvent (or Cauchy-Stieltjes transform) of the distribution $p_a(a)$. The resolvent can be inverted to give back

$$p_a(x) = \frac{1}{\pi} \lim_{\varepsilon \to 0^+} \operatorname{Im} \mathcal{G}_a(x - i\varepsilon), \tag{42}$$

which can be easily proved by using the Plemelj-Sokhotski formula in Eq. (14).

By comparing Eq. (42) with Eq. (34), one immediately realizes that our function $C(\lambda)$ is nothing but

$$C(\lambda) = i\eta \, \mathcal{G}(\lambda) + \mathcal{O}(1/N), \tag{43}$$

where we denoted by $\mathcal{G}(\lambda)$ the resolvent of the spectral density $\rho(\lambda)$ of our model. Choosing $\gamma = 1$ (so that η in Eq. (2) becomes N-independent) and taking the limit $N \to \infty$, the correction in Eq. (43) vanishes, and Eq. (40) takes the form

$$\mathcal{G}(\lambda_\varepsilon) = \mathcal{G}_a(\lambda_\varepsilon - 2\eta \mathcal{G}(\lambda_\varepsilon)). \tag{44}$$

This is analogous to Eq. (148) in Ref. [38], which was derived in the case in which the matrix B belongs to the GUE (and not the GOE) ensemble, and it is consistent with the results of Ref. [67]. Moreover, in Appendix C we show how Eq. (44) can be recovered by direct application of the Zee formula for the addition of two random matrices derived in Ref. [62].

One may legitimately wonder, at this point, whether our calculation is solely another way of obtaining Zee's result [62] for the particular case in which one of the two matrices being summed

---

[2]The way we constructed the action in Eq. (20) does not allow us to explore the region $\gamma < 1$ where A becomes subleading for large N (although this could of course be achieved with minor modifications of the calculation presented in Appendix B).

is a GOE matrix. In Section 2.4 we will analyze a few cases in which the self-consistency equation (35) (or (40)) can be solved exactly, so that from Eq. (34) one can find an expression for $\rho(\lambda)$ which is correct, up to $\mathcal{O}(1/N)$, for any value of η. These cases are in fact the very same which could be cracked by applying the Zee formula to the deformed GOE matrix. The advantage of our framework is that, whenever η is N-dependent and decays slower than $1/N$ (as it happens, for instance, in Eq. (2) for $1 < \gamma < 2$), then we are still able to keep track of the finite-N corrections. Moreover, in Section 2.5 we will provide approximate solutions which can be used whenever the resolvent $\mathcal{G}_a$ is not available in closed form, so that the Zee route is not viable.

2.4 Exactly solvable cases

In some particular cases, the self-consistency equation (35) admits an analytic solution, and we can access the limiting distribution $\rho(\lambda)$ for any value of η (i.e., not necessarily small). This happens whenever the following conditions are met:

(i) the resolvent $\mathcal{G}_a(z)$ associated with $p_a(a)$ is known analytically, and

(ii) the self-consistency equation for $C(\lambda)$ resulting from Eq. (35) or Eq. (40) is not transcendental, so that we can solve for $C(\lambda)$.

Below we present two such examples, which will also prove useful in our discussion of the level compressibility presented later in Section 3.4.

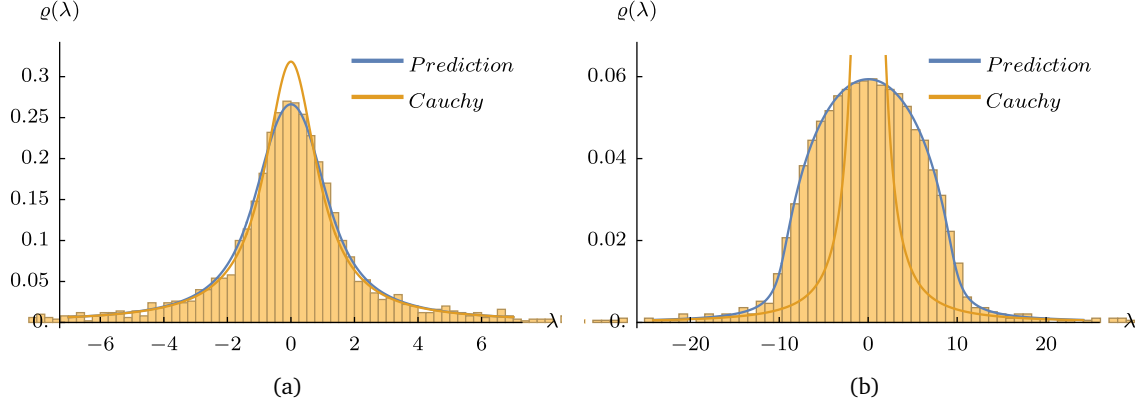(a)   (b)

Figure 2: Distribution of the eigenvalues $\rho(\lambda)$ in the case in which $p_a(a)$ is the Cauchy distribution (see Section 2.4.1). We show a numerical check of Eq. (47) in the regime of (a) small η (here $\eta = 0.12$) and (b) large η (here $\eta = 11.7$). The histogram was built using $\mu = 0$, $\omega = 1$, $\gamma = 1.1$ and $N = 2000$, with $\nu = 1$ in (a) and $\nu = 10$ in (b).

2.4.1 Cauchy distributed a_i

Let us choose $p_a(a)$ to be a Lorentzian of width ω and centered at μ,

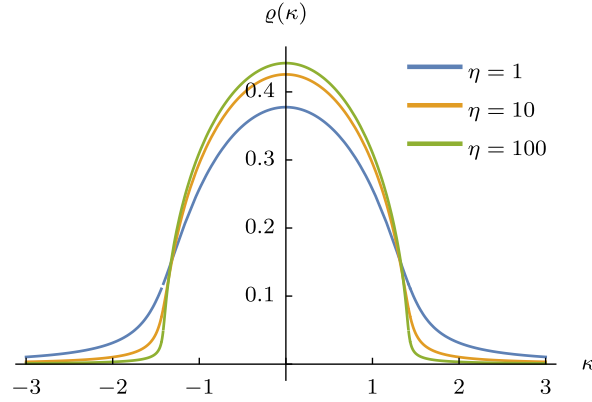$$p_a(a) = \frac{1}{\pi\omega}\left[\frac{\omega^2}{(a-\mu)^2 + \omega^2}\right]. \tag{45}$$

Figure 3: Limiting distribution of the rescaled eigenvalues $\kappa = \lambda/\sqrt{4\eta}$, with $\rho(\lambda)$ given in Eq. (47) for the Cauchy case (see Section 2.4.1). In this plot we used $\mu = 0$ and $\omega = 0.5$.

Its characteristic function is an exponential, $\psi_a(z) = \exp(-i\mu z - \omega|z|)$, and then by using Eq. (35) we can compute $C(\lambda)$ in closed form:

$$C(\lambda) = \frac{1}{4}\left\{\omega + i(\lambda - \mu) \pm \sqrt{[\omega + i(\lambda - \mu)]^2 + 8\eta}\right\}. \tag{46}$$

We choose the branch with the minus sign for which $\operatorname{Re} C(\lambda) \leq \omega/2$. Using Eq. (34) then yields

$$\rho(\lambda) = -\frac{\omega - \left\{4\omega^2(\lambda - \mu)^2 + \left[8\eta + \omega^2 - (\lambda - \mu)^2\right]^2\right\}^{1/4}\cos(\theta_\lambda/2)}{4\pi\eta} + \mathcal{O}(1/N), \tag{47}$$

with

$$\theta_\lambda \equiv \arg\left\{[\omega + i(\lambda - \mu)]^2 + 8\eta\right\}. \tag{48}$$

In Fig. 2a we plot Eq. (47) against numerical results in the small-η region: we find a good agreement with the theoretical prediction, as well as visible departures from the Cauchy distribution, especially in the bulk.

Another interesting limit is the one of large η, i.e., $\gamma = 1$ and ν large. It has been shown in Ref. [38] that, whenever $p_a(a)$ is rapidly decaying close to the edge of its finite support, then the spectral density $\rho(\lambda)$ interpolates between $p_a(a)$ and $\rho_{\text{GOE}}(\lambda)$ as the value of ν is increased. Note, however, that in the present case $p_a(a)$ decays algebraically and its support is not compact, so the outcome is less clear. The correct way of taking this limit is to rescale the eigenvalues as $\kappa \equiv \lambda/\sqrt{4\eta}$ and look for the distribution $\rho_\kappa(\kappa) = \sqrt{4\eta}\rho(\sqrt{4\eta}\kappa)$. From Eq. (46) we see that, if $\eta \gg \omega, \mu$, then

$$C(\sqrt{4\eta}\kappa) = \frac{\sqrt{\eta}}{2}\left[i\kappa - \sqrt{2 - \kappa^2}\right] + \mathcal{O}(\eta^0, 1/N), \tag{49}$$

and for large N we get from Eq. (34) that $\rho_\kappa(\kappa) \to \rho_{\text{GOE}}(\kappa)$ (see Eq. (39)). For large but finite η and $N \to \infty$, on the other hand, the bulk distribution of κ looks like a semicircle (as in the GOE ensemble), but with fat tails whose width grows with ω (see Fig. 3). Moreover, the whole distribution shifts rigidly by changing its center $\mu/\sqrt{4\eta}$. Numerical results in the large-η region are again nicely reproduced by Eq. (47), as shown in Fig. 2b.

Note that one can equivalently get to Eq. (46) by first computing the resolvent associated with the Lorentzian distribution in Eq. (45), i.e.,

$$\mathcal{G}_{\text{Cauchy}}(\lambda) = \frac{1}{\lambda - \mu \pm i\omega}, \tag{50}$$

where the $\pm$ branches correspond to $\text{Im}\{\lambda\} > 0$ or $\text{Im}\{\lambda\} < 0$, respectively. This can be obtained by explicitly performing the complex integral in Eq. (41), which only entails simple poles for a Cauchy distribution. One can then easily solve the self-consistency equation (40), which turns out to be quadratic, and thus recover Eq. (46).

2.4.2 Wigner distributed a_i

Another simple case is the one in which $p_a(a)$ itself is chosen as the (μ-centered) Wigner distribution

$$p_a(a) \equiv \frac{2}{\pi\sigma^2} \sqrt{\sigma^2 - (a-\mu)^2}, \tag{51}$$

whose corresponding resolvent is

$$\mathcal{G}_a(z) = \frac{2}{\sigma^2} \left[z - \mu - \sqrt{(z-\mu)^2 - \sigma^2} \right]. \tag{52}$$

The self-consistency equation (40) is again quadratic and it yields

$$C(\lambda) = 2\eta \frac{i(\lambda - \mu) \pm \sqrt{\sigma^2 + 8\eta - (\lambda - \mu)^2}}{\sigma^2 + 8\eta}, \tag{53}$$

so that by choosing the branch with the minus sign and using Eq. (34) we find

$$\rho(\lambda) = \frac{2\sqrt{\sigma^2 + 8\eta - (\lambda - \mu)^2}}{\pi(\sigma^2 + 8\eta)} \Theta\left(\sigma^2 + 8\eta - (\lambda - \mu)^2\right) + \mathcal{O}(1/N). \tag{54}$$

As expected, this is still a Wigner distribution centered in $\lambda = \mu$, but its width gets corrected as $\sigma^2 \to \sigma^2 + 8\eta$.

2.5 Approximate solutions

For most choices of the diagonal disorder distribution $p_a(a)$, Eq. (35) cannot be solved exactly. However, since η is a small parameter for large N and $\gamma > 1$ (which is the case we are mostly interested in), it may be sufficient to note that $C(\lambda) \sim \mathcal{O}(\eta) \sim \mathcal{O}(N^{1-\gamma})$: this can be taken as a starting point for constructing self-consistent approximations for large N.

By expanding recursively the exponential in Eq. (35) and using that $C(\lambda) \sim \mathcal{O}(\eta)$, one gets

$$C(\lambda) = -\eta \int_0^\infty dz\, \psi_a(-z) e^{-i\lambda z}$$
$$+ 2\eta^2 \int_0^\infty dz \int_0^\infty ds\, z \psi_a(-z) \psi_a(-s) e^{-i\lambda(z+s)} + \mathcal{O}(\eta^3). \tag{55}$$

Note that

$$\text{Re}\, C(\lambda_\varepsilon) = -\eta\, p_a(\lambda_\varepsilon) + \mathcal{O}(\eta^2) \xrightarrow[\varepsilon \to 0]{N \gg 1} -\eta\, p_a(\lambda) \leq 0, \tag{56}$$

in such a way that $G(y)$ is indeed well-behaved at $y \to +\infty$, as we had assumed in Section 2.1. Plugging Eq. (55) into Eq. (34) now gives

$$\rho(\lambda) = p_a(\lambda) - \frac{2\eta}{\pi} \operatorname{Re} \int_0^\infty \mathrm{d}z \int_0^\infty \mathrm{d}s \, z \psi_a(-z) \psi_a(-s) e^{-i\lambda(z+s)} + \mathcal{O}(\eta^2, N^{-1}). \tag{57}$$

Of course this approximation becomes meaningless if $\gamma \geq 2$, where the higher order corrections we neglected when we took the saddle-point approximation mingle with our correction in η.

Alternatively, one can expand the exponential in Eq. (35) as

$$C(\lambda) \simeq -\eta \int_0^\infty \mathrm{d}z \, \psi_a(-z) e^{-i\lambda z} - 2\eta C(\lambda) \int_0^\infty \mathrm{d}z \, z \psi_a(-z) e^{-i\lambda}, \tag{58}$$

corresponding to a partial resummation of the perturbative series in η: this is the so-called self-consistent Hartree-Fock approximation [68]. From Eq. (58) we then obtain

$$C(\lambda) \simeq \frac{-\eta \int_0^\infty \mathrm{d}z \, \psi_a(-z) e^{-i\lambda z}}{1 + 2\eta \Sigma(\lambda)}, \tag{59}$$

with the *self-energy*

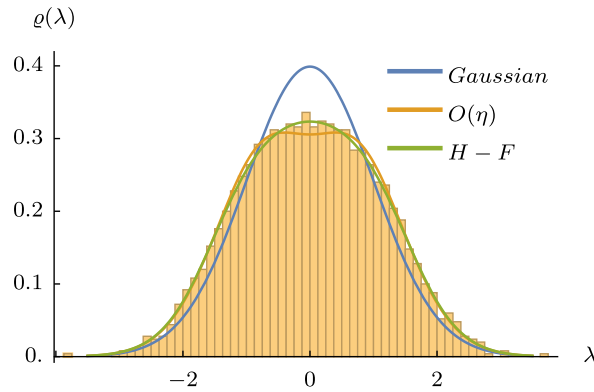$$\Sigma(\lambda) = \int_0^\infty \mathrm{d}z \, z \psi_a(-z) e^{-i\lambda z}. \tag{60}$$

Figure 4: Distribution of the eigenvalues $\rho(\lambda)$ in the case in which $p_a(a)$ is Gaussian with unit variance, and B is a GOE matrix as in Eq. (1). Numerical results are compared to the first $\mathcal{O}(\eta)$ correction given in Eq. (61), and to the Hartree-Fock approximation given in Eq. (59). The histogram was built using $N = 2000$, $\nu = 1$ and $\gamma = 1.1$, corresponding to $\eta = 0.12$.

An explicit example: the Gaussian case. We finally test both approximations in the case where $p_a(a)$ is chosen to be Gaussian (with zero mean and variance σ^2), as in the original Rosenzweig-Porter model [22]. By using $\psi_a(z) = \exp(-z^2\sigma^2/2)$ in Eq. (57), one finds

$$\rho(\lambda) = p_a(\lambda) + \sqrt{\frac{2}{\pi}} \frac{\eta}{\sigma^3} e^{-\lambda^2/(2\sigma^2)} \left[\frac{2\sqrt{2}\lambda}{\sigma} F\left(\frac{\lambda}{\sqrt{2}\sigma}\right) - 1 \right] + \mathcal{O}(\eta^2, N^{-1}), \tag{61}$$

where $F(z)$ is the Dawson function defined as $F(z) \equiv e^{-z^2} \int_0^z e^{y^2} \, \mathrm{d}y$. The corrected spectral density is plotted in Fig. 4, and comparison with the eigenvalue spectrum computed numerically shows a fair agreement. Using instead the Hartree-Fock approximation for $C(\lambda)$ given in Eq. (59) results in a better agreement with the numerical data in the central region of the spectrum.

3 Number of eigenvalues in an interval and level compressibility

In this Section we consider the number of eigenvalues $I_N[\alpha, \beta]$ in a finite interval $[\alpha, \beta]$, as given by Eq. (8), and compute its cumulant generating function. This, in turn, can be used to access the level compressibility $\chi(E)$ defined in Eq. (9). As we explained in the Introduction, $\chi(E)$ represents a simple measure of the rigidity of the spectrum, which in turn allows us to distinguish between the phases of the model.

This program can be achieved by following the replica-based procedure introduced and exploited in Refs. [69, 70], which we briefly outline here. Starting from the definition of the spectral density in Eq. (7), we first rewrite Eq. (8) as

$$I_N[\alpha, \beta] = \sum_{i=1}^{N} [\Theta(\beta - \lambda_i) - \Theta(\alpha - \lambda_i)] . \tag{62}$$

Now we recall that the Heaviside function can be represented in terms of the discontinuity of the complex logarithm,

$$\Theta(-x) = \frac{1}{2\pi i} \lim_{\varepsilon \to 0^+} [\ln(x + i\varepsilon) - \ln(x - i\varepsilon)] , \tag{63}$$

so that we interpret

$$\sum_{i=1}^{N} \Theta(\alpha - \lambda_i) = \frac{1}{2\pi i} \lim_{\varepsilon \to 0^+} \left[\ln \det(\mathcal{H} - \alpha_\varepsilon \mathbb{1}) - \ln \det(\mathcal{H} - \alpha_\varepsilon^* \mathbb{1}) \right] , \tag{64}$$

where we called as before $\alpha_\varepsilon \equiv \alpha - i\varepsilon$ with $\varepsilon > 0$. This allows us to express $I_N[\alpha, \beta]$ in terms of the partition function given in Eq. (17),

$$I_N[\alpha, \beta] = -\frac{1}{\pi i} \lim_{\varepsilon \to 0^+} \ln \left[\frac{\mathcal{Z}(\beta_\varepsilon) \mathcal{Z}(\alpha_\varepsilon^*)}{\mathcal{Z}(\beta_\varepsilon^*) \mathcal{Z}(\alpha_\varepsilon)} \right] . \tag{65}$$

In order to compute the moments of I_N, we first address its cumulant generating function

$$\mathcal{F}_{[\alpha, \beta]}(s) \equiv \frac{1}{N} \ln \left\langle e^{-s I_N[\alpha, \beta]} \right\rangle = \frac{1}{N} \ln \left\langle e^{\frac{s}{\pi i} \lim_{\varepsilon \to 0^+} \ln \{ \mathcal{Z}(\beta_\varepsilon) \mathcal{Z}(\alpha_\varepsilon^*) [\mathcal{Z}(\beta_\varepsilon^*) \mathcal{Z}(\alpha_\varepsilon)]^{-1} \}} \right\rangle . \tag{66}$$

Assuming now that one can move the limit $\varepsilon \to 0^+$ at the front of this expression, we obtain

$$\mathcal{F}_{[\alpha, \beta]}(s) = \lim_{\varepsilon \to 0^+} \frac{1}{N} \ln Q_{[\alpha, \beta]}(s), \tag{67}$$

where we introduced

$$Q_{[\alpha, \beta]}(s) \equiv \left\langle \left[\mathcal{Z}(\beta_\varepsilon^*) \mathcal{Z}(\alpha_\varepsilon) \right]^{is/\pi} \left[\mathcal{Z}(\beta_\varepsilon) \mathcal{Z}(\alpha_\varepsilon^*) \right]^{-is/\pi} \right\rangle . \tag{68}$$

The latter can be accessed by first evaluating

$$Q_{[\alpha,\beta]}(n_\pm) \equiv \left\langle \left[\mathcal{Z}(\beta_\varepsilon^*)\mathcal{Z}(\alpha_\varepsilon)\right]^{n_+} \left[\mathcal{Z}(\beta_\varepsilon)\mathcal{Z}(\alpha_\varepsilon^*)\right]^{n_-} \right\rangle \tag{69}$$

within the replica formalism with $n_\pm$ integer, and then performing its analytic continuation to

$$Q_{[\alpha,\beta]}(s) = \lim_{n_\pm \to \pm i s/\pi} Q_{[\alpha,\beta]}(n_\pm). \tag{70}$$

To obtain the level compressibility in Eq. (9), we finally compute the cumulants

$$\frac{\kappa_j[\alpha,\beta]}{N} = (-1)^j \left. \partial_s^j \mathcal{F}_{[\alpha,\beta]}(s)\right|_{s=0}, \tag{71}$$

and we evaluate them at $\alpha = -E$, $\beta = E$.

We note that this calculation involves the product of partition functions and not a logarithm of them, and thus there is no question of quenched vs. annealed averages here.

3.1 Replica action and saddle-point equations

The details of this derivation are reported in Appendix D.1. In analogy with Section 2, here we express

$$Q_{[\alpha,\beta]}(n_\pm) \propto \int \mathcal{D}(i\varphi) \exp\{N\mathcal{S}_{n_\pm}[\varphi;\hat{\Lambda}]\}, \tag{72}$$

$$\mathcal{S}_{n_\pm}[\varphi;\hat{\Lambda}] \equiv \frac{1}{2\eta} \int \mathrm{d}\vec{\tau}\,\mathrm{d}\vec{\tau}'\,\varphi(\vec{\tau})M(\vec{\tau},\vec{\tau}')\varphi(\vec{\tau}')$$
$$+ \ln \int \mathrm{d}\vec{\tau}\exp\left[-\frac{i}{2}\vec{\tau}\,\hat{\Lambda}\,\vec{\tau} - \int \mathrm{d}\vec{\tau}'\,M(\vec{\tau},\vec{\tau}')\varphi(\vec{\tau}')\right]\psi_a\left(-\frac{1}{2}\vec{\tau}\,\hat{L}\,\vec{\tau}\right), \tag{73}$$

where the replica vectors $\vec{\tau} \in \mathbb{R}^n$ live in dimension $n = 2(n_+ + n_-)$. Each of the four replica sets corresponds to one sector in the block matrices

$$\hat{\Lambda} \equiv \begin{pmatrix} \alpha_\varepsilon \mathbb{1}_{n_+} & & & \\ & \bar{\beta}_\varepsilon \mathbb{1}_{n_+} & & \\ & & \beta_\varepsilon \mathbb{1}_{n_-} & \\ & & & \bar{\alpha}_\varepsilon \mathbb{1}_{n_-} \end{pmatrix}, \qquad \hat{L} \equiv \begin{pmatrix} \mathbb{1}_{n_+} & & & \\ & -\mathbb{1}_{n_+} & & \\ & & \mathbb{1}_{n_-} & \\ & & & -\mathbb{1}_{n_-} \end{pmatrix}, \tag{74}$$

where we called $\bar{\alpha}_\varepsilon = -\alpha_\varepsilon^*$, and similarly for $\bar{\beta}_\varepsilon$. Finally, in Eq. (73) we introduced the function

$$M(\vec{\tau},\vec{\tau}') \equiv \left(\vec{\tau}\,\hat{L}\,\vec{\tau}'\right)^2, \tag{75}$$

while we recall that $\psi_a(z)$ was given in (30). Next, we need to compute the integration over $\mathcal{D}(i\varphi)$ in Eq. (72) in the limit of large N. From Eq. (73), the saddle-point equation follows simply as

$$\varphi_0(\vec{\tau}) = \frac{\eta}{Z_\varphi} \exp\left[-\frac{i}{2}\vec{\tau}\,\hat{\Lambda}\,\vec{\tau} - \int \mathrm{d}\vec{\tau}'\,M(\vec{\tau},\vec{\tau}')\varphi_0(\vec{\tau}')\right]\psi_a\left(-\frac{1}{2}\vec{\tau}\,\hat{L}\,\vec{\tau}\right), \tag{76}$$

with

$$Z_\varphi \equiv \int \mathrm{d}\vec{\tau}\exp\left[-\frac{i}{2}\vec{\tau}\,\hat{\Lambda}\,\vec{\tau} - \int \mathrm{d}\vec{\tau}'\,M(\vec{\tau},\vec{\tau}')\varphi_0(\vec{\tau}')\right]\psi_a\left(-\frac{1}{2}\vec{\tau}\,\hat{L}\,\vec{\tau}\right). \tag{77}$$

In order to better quantify the finite-size corrections and to make contact with the calculation performed in Ref. [70] for the pure GOE case, in Appendix D.2 we also compute the Gaussian fluctuations around the saddle-point $\varphi_0(\vec{\tau})$, leading to

$$Q_{[\alpha,\beta]}(n_\pm) = \exp\left\{ N\mathcal{S}_{n_\pm}[\varphi_0;\hat{\Lambda}] + \frac{1}{2}\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\operatorname{Tr}T^k \right\} + \mathcal{O}(1/N^2), \tag{78}$$

with

$$T(\vec{\tau}_1,\vec{\tau}_2) \equiv \varphi_0(\vec{\tau}_1)\left[M(\vec{\tau}_1,\vec{\tau}_2) - \frac{1}{\eta}\int d\vec{\tau}'\, M(\vec{\tau}_2,\vec{\tau}')\varphi_0(\vec{\tau}') \right]. \tag{79}$$

Note that $\varphi_0(\vec{\tau}) \sim \mathcal{O}(\eta)$ (see Eq. (76)), so that the function $T(\vec{\tau}_1,\vec{\tau}_2)$ itself is also of $\mathcal{O}(\eta)$.

3.2 Rotationally-invariant Ansatz

In the pure GOE case (which is recovered, for instance, by setting $\psi_a(z) = 1$ identically in the expressions above), the saddle-point equation (76) suggests to look for a replica-symmetric solution in the form of a Gaussian, i.e.,

$$\varphi_0(\vec{\tau}) = \mathcal{N}\exp\left(-\frac{1}{2}\vec{\tau}\,\hat{C}^{-1}\,\vec{\tau}\right), \tag{80}$$

where $\hat{C}$ is another block-diagonal matrix of the form

$$\hat{C} \equiv \begin{pmatrix} \Delta_\alpha \mathbb{1}_{n_+} & & & \\ & \bar{\Delta}_\beta \mathbb{1}_{n_+} & & \\ & & \Delta_\beta \mathbb{1}_{n_-} & \\ & & & \bar{\Delta}_\alpha \mathbb{1}_{n_-} \end{pmatrix}. \tag{81}$$

The four parameters $\Delta_\alpha, \bar{\Delta}_\alpha, \Delta_\beta$ and $\bar{\Delta}_\beta$ and are to be fixed, together with the prefactor $\mathcal{N}$ in Eq. (80), by substituting the Ansatz in Eq. (80) back into the saddle-point equation (76).

This strategy has proven effective in Ref. [70] and, for completeness, in Appendix E we sketch the entire calculation for the GOE case. However, in the presence of a nontrivial $\psi_a(z)$ in Eq. (76), there is no reason why the Ansatz in Eq. (80) should work. We are thus naturally led to extend it to the form

$$\varphi_0(\vec{\tau}) = \mathcal{N}\exp\left(-\frac{1}{2}\vec{\tau}\,\hat{C}^{-1}\,\vec{\tau}\right)\psi_a\left(-\frac{1}{2}\vec{\tau}\,\hat{L}\,\vec{\tau}\right). \tag{82}$$

This can be plugged into Eq. (76) to first obtain $\mathcal{N} = \eta/Z_\varphi$, where

$$Z_\varphi = \int d\vec{\tau}\,\varphi_0(\vec{\tau}). \tag{83}$$

By using Gaussian integration one can then show that

$$\int d\vec{\tau}'\, M(\vec{\tau},\vec{\tau}')\varphi_0(\vec{\tau}') = \eta\vec{\tau}\,\hat{K}\,\vec{\tau}, \tag{84}$$

where $\hat{K}$ is a diagonal matrix given by

$$\hat{K} = -i\hat{L}\,\mathcal{G}_a\left((i\hat{L}\hat{C})^{-1}\right), \tag{85}$$

$\mathcal{G}_a$ is the resolvent in Eq. (41), and $\hat{L}$ is given in Eq. (74). The remaining free parameters in Eq. (80) can be determined by solving the set of four self-consistency equations which follow from Eq. (76), i.e.,

$$\hat{C}^{-1} = 2\eta\hat{K} + i\hat{\Lambda}. \tag{86}$$

Finally, both the action and its Gaussian fluctuation matrix T (see Eq. (79)) can be computed by using the saddle-point solution in Eq. (82). First, notice that the action in Eq. (73) takes the form

$$
\begin{aligned}
\mathcal{S}_{n_\pm}[\varphi_0;\hat{\Lambda}] &= \frac{1}{2}\int d\vec{\tau}\left(\vec{\tau}\,\hat{K}\,\vec{\tau}\right)\varphi_0(\vec{\tau}) + \ln Z_\varphi \\
&= \frac{\eta}{2}\sum_i K_{ii}^2 + \ln\int_{-\infty}^{\infty} da\, p_a(a)\exp\left\{-\frac{1}{2}\operatorname{Tr}\ln[(i\hat{L}\hat{C})^{-1} - a]\right\} + \text{const.}.
\end{aligned} \tag{87}
$$

The constant term vanishes upon taking the analytic continuation $n_\pm \to \pm is/\pi$, yielding

$$
\begin{aligned}
\mathcal{S}_{\pm\frac{is}{\pi}}[\varphi_0;\hat{\Lambda}] = &\frac{is\eta}{2\pi}\left(k_\alpha^2 + \bar{k}_\beta^2 - k_\beta^2 - \bar{k}_\alpha^2\right) \\
&+ \ln\int_{-\infty}^{\infty} da\, p_a(a)\exp\left\{-\frac{is}{2\pi}\ln\left[\frac{(\bar{\Delta}_\beta^{-1} + ia)(\Delta_\alpha^{-1} - ia)}{(\bar{\Delta}_\alpha^{-1} + ia)(\Delta_\beta^{-1} - ia)}\right]\right\},
\end{aligned} \tag{88}
$$

where we introduced for brevity $\hat{K} \equiv \operatorname{diag}(k_\alpha\mathbb{1}_{n_+}, \bar{k}_\beta\mathbb{1}_{n_+}, k_\beta\mathbb{1}_{n_-}, \bar{k}_\alpha\mathbb{1}_{n_-})$. Note that this action coincides at leading order with the cumulant generating function in Eq. (67), i.e.,

$$\mathcal{F}_{[\alpha,\beta]}(s) = \lim_{\varepsilon\to 0^+}\mathcal{S}_{\pm\frac{is}{\pi}}[\varphi_0;\hat{\Lambda}] + \mathcal{O}(\eta/N), \tag{89}$$

where we used Eq. (78) (the estimate of the large-N correction will soon be justified). In the next Section we will make this result more explicit in the case of an interval which is symmetric around the origin.

The Gaussian fluctuations around the saddle-point are studied in Appendix D.2. A closed-form result is not available in this case (in contrast to the GOE case, see Appendix E), because the calculation involves increasingly complex generalizations of the resolvent $\mathcal{G}_a(z)$ which encode higher order correlations (see Eq. (D.18)). However, we can show that the Gaussian fluctuations bring to the leading order term in Eq. (89) a correction of $\mathcal{O}(\eta/N) = \mathcal{O}(N^{-\gamma})$, which is strongly suppressed for large N in the region $\gamma > 1$ which we focus on here.

3.3   General result in the case of a symmetric interval

We consider now the simpler case in which $\alpha = -E$ and $\beta = E$, and we take a symmetric distribution $p_a(a)$. From Eqs. (85) and (86) one can deduce that the entries of the matrix $\hat{C}$ are related by the following symmetries:

$$\Delta_\alpha \equiv \Delta \equiv re^{i\theta}, \quad \bar{\Delta}_\alpha = \Delta_\alpha^* = \Delta_\beta, \quad \bar{\Delta}_\beta = \Delta_\beta^*. \tag{90}$$

The same holds for the entries of $\hat{K}$ (see Eq. (85)), hence we will simply call $k_\alpha \equiv k$. The problem is then reduced to computing one single unknown, namely Δ: from Eqs. (85) and (86), this amounts to solving the self-consistency equations

$$
\begin{cases}
\Delta^{-1} = \varepsilon - iE + 2\eta k, \\
k = -i\mathcal{G}_a\left(\dfrac{1}{i\Delta}\right).
\end{cases} \tag{91}
$$

The action in Eq. (88) then takes the form

$$\mathcal{S}[\varphi_0; \hat{\Lambda}] = -\frac{2\eta s}{\pi} \operatorname{Im}\{k^2\} + \ln \int_{-\infty}^{\infty} da \, p_a(a) \exp\left[-\frac{s}{\pi} \arctan\left(\frac{\sin 2\theta}{a^2 r^2 + \cos 2\theta}\right)\right], \quad (92)$$

where the branch of the arctan is chosen so that it returns an angle in $[0, \pi]$. From Eq. (67) we can then read out the leading order contribution to the rate function, namely

$$\mathcal{F}_{[-E,E]}(s) = \mathcal{S}[\varphi_0; \hat{\Lambda}] + \mathcal{O}(\eta/N) = -ms + \ln\left\langle e^{-sf(a)}\right\rangle_a + \mathcal{O}(\eta/N). \quad (93)$$

Here we used the notation $\langle \bullet \rangle_a$ to indicate an average over $p_a(a)$, and we introduced for brevity

$$m \equiv \frac{2\eta}{\pi} \operatorname{Im}\{k^2\} = -\frac{2\eta}{\pi} \operatorname{Im}\left[\mathcal{G}_a\left(\frac{1}{i\Delta}\right)\right]^2, \quad (94)$$

$$f(a) \equiv \frac{1}{\pi} \arctan\left(\frac{\sin 2\theta}{a^2 r^2 + \cos 2\theta}\right) \in [0, 1], \quad (95)$$

where in the first line we used Eq. (91). The cumulant generating function in Eq. (93), together with the self-consistency equations (91), represent our second main result. As we stressed above, when η is given by Eq. (2), the correction to Eq. (93) is of $\mathcal{O}(\eta/N) = \mathcal{O}(N^{-\gamma})$, which is strongly suppressed for large N in the region $\gamma > 1$.

Expanding Eq. (93) in powers of s we get

$$\mathcal{F}_{[-E,E]}(s) \simeq -s\left[m + \langle f(a)\rangle_a\right] + \frac{s^2}{2}\left[\left\langle f^2(a)\right\rangle_a - \langle f(a)\rangle_a^2\right] + \mathcal{O}(s^3, \eta/N), \quad (96)$$

and by comparison with Eq. (71) we can identify the first two cumulants

$$\frac{\kappa_1}{N} = m + \langle f(a)\rangle_a + \mathcal{O}(\eta/N), \quad (97)$$

$$\frac{\kappa_2}{N} = \left\langle f^2(a)\right\rangle_a - \langle f(a)\rangle_a^2 + \mathcal{O}(\eta/N). \quad (98)$$

From Eq. (9), we finally obtain the level compressibility

$$\chi(E) = \frac{\left\langle f^2(a)\right\rangle_a - \langle f(a)\rangle_a^2}{m + \langle f(a)\rangle_a} + \mathcal{O}(\eta/N). \quad (99)$$

We remark that not only the level compressibility, but actually all the moments of $I_N[-E, E]$ can be simply computed starting from Eq. (93): they read (at leading order for large N)

$$\left\langle (I_N[-E, E])^m \right\rangle \simeq N \left\langle [f(a)]^m \right\rangle_a, \quad m \geq 2. \quad (100)$$

3.3.1 Limit of a pure diagonal matrix with random i.i.d. entries

It is instructive, at this point, to consider the limit $\eta \to 0$. In this case the GOE part of Eq. (1) is neglected, and the spectral properties are completely determined by the matrix A, whose entries are independent and identically distributed according to $p_a(a)$. The self consistency equation (86) then reduces to

$$\hat{C} = (i\hat{\Lambda})^{-1}, \quad (101)$$

whence

$$\Delta = \Delta_\alpha \Big|_{\alpha = -E} = \frac{iE + \varepsilon}{E^2 + \varepsilon^2} \equiv r e^{i\theta} , \tag{102}$$

and there is no need to determine the entries of $\hat{K}$ since it does not enter the expression of the saddle point action (92) for $\eta = 0$. From Eq. (92) we obtain, for $\varepsilon \to 0^+$,

$$
\begin{aligned}
\mathcal{S}[\varphi_0; \hat{\Lambda}] &= \ln \int_{-\infty}^{\infty} da\, p_a(a) \exp\left\{ -\frac{s}{\pi} \arctan\left[\frac{0^+}{(a/E)^2 - 1} \right] \right\} \\
&= \ln\left[1 + (e^{-s} - 1) \int_{-E}^{E} da\, p_a(a) \right] ,
\end{aligned}
\tag{103}
$$

where we used that the branch of $\arctan(z) \in [0, \pi]$ has a discontinuity in $z = 0$, i.e., it jumps from π to 0 as z becomes positive. From Eqs. (8) and (93), we can then read out the cumulant generating function

$$\mathcal{F}_{[-E,E]}(s) = \ln\left[1 + (e^{-s} - 1) \frac{\langle I_N[-E, E] \rangle_a}{N} \right] . \tag{104}$$

In this way we recover the standard textbook result for the cumulant generating function in the case of i.i.d. random variables, which we sketch for completeness in Appendix A. In particular, the level compressibility in Eq. (9) reads in this case

$$\chi(E) = 1 - \frac{\langle I_N[-E, E] \rangle_a}{N} , \tag{105}$$

so that in general $\chi(E) \sim 1$ for small E, and $\chi(E) \to 0$ for large E.

Finally, we note that Eq. (104) is exact, because we have stressed above that the Gaussian fluctuations around the saddle point are at least of $\mathcal{O}(\eta)$ and so they vanish in the limit $\eta \to 0$.

3.3.2   Limit of a pure GOE matrix

The opposite limit in which A is neglected in Eq. (1) is obtained by setting $p_a(a) = \delta(a)$, whose corresponding resolvent is $\mathcal{G}_a(z) = 1/z$ (indeed, summing zeros to the matrix B in Eq. (1) does not change its spectrum). Equation (85) then implies $\hat{K} = \hat{C}$, where we used $\hat{L}^2 = \mathbb{1}$ (see Eq. (74)). The self-consistency equations (86) are then seen to coincide with Eq. (E.2), corresponding to the GOE case studied in Ref. [70] and here revisited in Appendix E. From Eq. (92) we obtain, in terms of r and θ introduced in Eq. (90),

$$\mathcal{S}[\varphi_0; \hat{\Lambda}] = -\frac{2\eta r^2 s}{\pi} \sin 2\theta - \frac{2s\theta}{\pi} , \tag{106}$$

which is linear in s: we deduce that the cumulants higher than the average $\langle I_N[-E, E] \rangle$ are subleading for large N, and they are only accessible by explicitly computing the Gaussian fluctuations around the saddle-point (see Appendix D.2). Still, the leading order term in the first cumulant (see Eq. (E.16)) is correctly reproduced by Eq. (106).

3.4   Exactly solvable cases

The cumulant generating function we found in Eq. (93) is formal, in that $\Delta = r e^{i\theta}$ must first be determined by solving the self-consistency equation (86). In the following, we will present two cases in which the result can be expressed in closed form, namely those in which $p_a(a)$ is the Cauchy or the Wigner distribution, respectively. While the former presents fat tails and is thus slowly decaying, the latter has a compact support and a well-defined edge.

3.4.1 Cauchy distributed a_i

Let us start from the case in which $p_a(a)$ is the Cauchy distribution, see Eq. (45). We set $\mu = 0$, so that the problem remains symmetric around the origin. Using the expression of $\mathcal{G}_a(z)$ in Eq. (50), $\hat{K}$ in Eq. (85) reduces to

$$\hat{K} = \frac{\hat{C}}{1 + \omega\hat{C}} \, . \tag{107}$$

Solving the self-consistency equation (86) then yields

$$\Delta = \Delta_\alpha \bigg|_{\alpha=-E} = 2\left[\varepsilon - iE - \omega + \sqrt{8\eta + (\varepsilon - iE + \omega)^2}\right]^{-1} \equiv re^{i\theta} \, , \tag{108}$$

where we chose the positive branch of the square root so that $\operatorname{Re}\Delta_\alpha > 0$. After this choice, the constant ε can be safely sent to zero. The resulting cumulant generating function is given in Eq. (93), where the averages are taken over the Cauchy distribution in Eq. (45), and the first linear contribution reads

$$m = \frac{4\eta r^2 \sin\theta \, (r\omega + \cos\theta)}{\pi\left[1 + (r\omega)^2 + 2r\omega\cos\theta\right]^2} \, . \tag{109}$$

(a)   (b)
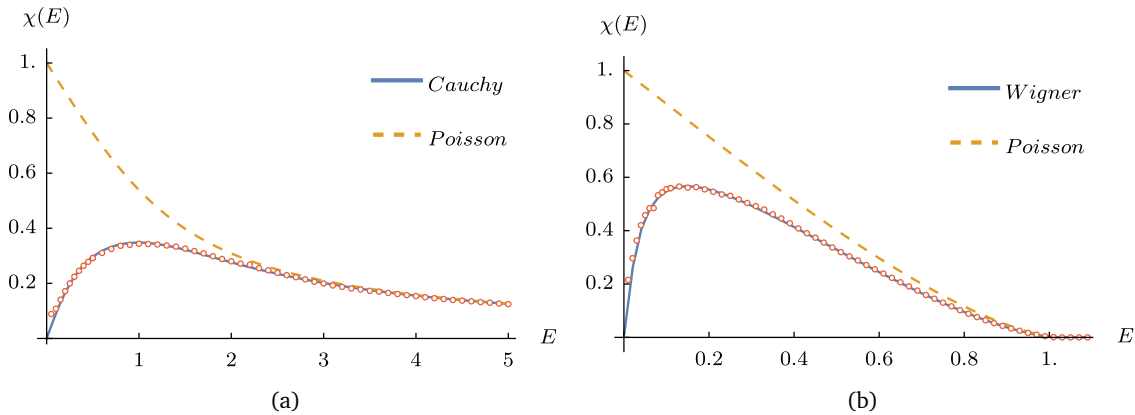
Figure 5: Numerical check of the level compressibility $\chi(E)$ predicted in Eq. (99) (solid blue line), for the case in which the elements of the diagonal matrix A belong to the (a) Cauchy, or (b) Wigner distribution. Numerical data (symbols) are obtained from the numerical diagonalization of $N_{\text{tot}} = 1000$ random matrices of size $N = 5000$. The (dashed yellow) line denoted by *Poisson* shows the level compressibility as it would be in the absence of level repulsion – see the main text. In both plots we used the parameters $\mu = 0$, $\gamma = 1.1$. In (a) we chose $\omega = 1$ and $\nu = 1$, corresponding to $\eta = 0.125$, while in (b) we set $\sigma = 1$ and $\nu = 0.2$, yielding $\eta = 0.005$.

The resulting level compressibility $\chi(E)$ (see Eq. (99)) is plotted in Fig. 5a, and it is compared to numerical results showing excellent agreement. We show in the same plot the level compressibility under the hypothesis that the level statistics is of the Poisson type, i.e., that the energy levels do not repel each other. This is given by Eq. (105) upon interpreting the average $\langle \bullet \rangle$ as taken over the average eigenvalue density $\rho(\lambda)$ for the case in which the matrix A has Cauchy-distributed entries: this has been found previously in Eq. (47). Even for very small values of η, the behavior at low energies E of the compressibility $\chi(E)$ is qualitatively very different: in the Poisson case we

have $\chi(E) \sim 1$, while in the GRP model it is $\chi(E) \sim 0$. The latter $\chi(E)$ increases up to a maximum, whose position $E_{\max}(\eta)$ grows monotonically (and sublinearly) with η.

3.4.2 Wigner distributed a_i

Let us now consider the case in which $p_a(a)$ is the Wigner distribution, see Eq. (51). Again we set $\mu = 0$, so that the problem remains symmetric around the origin. Using the expression of $\mathcal{G}_a(z)$ in Eq. (52), the quantity $\hat{K}$ in Eq. (85) becomes

$$\hat{K} = -\frac{2}{\sigma^2 \hat{C}} \left[1 - \sqrt{1 + (\sigma \hat{C})^2} \right].$$ (110)

Solving the self-consistency equation (86) yields, after choosing the positive branch of the square root so that $\operatorname{Re} \Delta_\alpha > 0$ and letting $\varepsilon \to 0$,

$$\Delta = \Delta_\alpha \bigg|_{\alpha = -E} = \frac{iE(\sigma^2 + 4\eta) + 4\eta\sqrt{\sigma^2 + 8\eta - E^2}}{16\eta^2 + \sigma^2 E^2} \equiv re^{i\theta}.$$ (111)

The resulting cumulant generating function is again given in Eq. (93), where the averages are taken over the Wigner distribution in Eq. (51). The analytical expression of the quantity m in Eq. (94) is cumbersome, but it follows readily from Eq. (110). In Fig. 5b we plot the level compressibility $\chi(E)$ as we did for the Cauchy case, finding a qualitatively similar behavior.

3.5 Scaling limit and Thouless energy

In this Section we focus on the limit in which $E = x\eta^\delta$ and $\eta \ll 1$, while $x \sim \mathcal{O}(1)$. We can envision a different behavior depending on whether the exponent $\delta > 1$, $\delta < 1$ or $\delta = 1$. The latter case turns out to be particularly interesting: we will show that the level compressibility computed in $\chi(E = x\eta^\delta)$ assumes for $\delta = 1$ a *universal* scaling form, which is independent of the particular choice of the distribution $p_a(a)$ of the entries of the diagonal matrix A.

To this end, let us go back to the self-consistency equations (91), which we can rewrite for $\delta = 1$ and $\eta \ll 1$ as

$$\Delta^{-1} = \varepsilon - ix\eta - 2i\eta \mathcal{G}_a\left(-i\Delta^{-1}\right)$$
$$\simeq \varepsilon - ix\eta - 2i\eta \mathcal{G}_a\left(-i\varepsilon + \mathcal{O}(\eta)\right),$$ (112)

where again $\Delta = re^{i\theta}$. By taking the complex conjugate of Eq. (112) and by summing and subtracting the two equations, we obtain the two conditions

$$r^{-1}\cos\theta = \varepsilon + 2\eta \operatorname{Im}\mathcal{G}_a\left(-i\varepsilon + \mathcal{O}(\eta)\right) \xrightarrow[\varepsilon \to 0^+]{\eta \ll 1} 2\pi\eta p_a(0),$$ (113)

$$r^{-1}\sin\theta = x\eta + 2\eta \operatorname{Re}\mathcal{G}_a\left(-i\varepsilon + \mathcal{O}(\eta)\right) \xrightarrow[\varepsilon \to 0^+]{\eta \ll 1} x\eta,$$ (114)

where we used the Plemelj-Sokhotski formula recalled in Eq. (14), and the fact that for a symmetric distribution $p_a(a)$ one has

$$\mathcal{P}\int da \, \frac{p_a(a)}{a} = 0.$$ (115)

One then easily obtains, at leading order for small η (and with $x = E/\eta$),

$$\tan\theta \simeq \frac{x}{2\pi p_a(0)} \equiv y,\tag{116}$$

$$r^{-1} \simeq 2\pi\eta p_a(0)\sqrt{1+y^2}.\tag{117}$$

We can now estimate the level compressibility in this limit. First, notice that

$$\mathcal{G}_a(-i\Delta^{-1}) = i\pi p_a(0) + \mathcal{O}(\eta),\tag{118}$$

so that from Eq. (94) we read

$$m = -\frac{2\eta}{\pi}\,\mathrm{Im}\big[\mathcal{G}_a(-i\Delta^{-1})\big]^2 = \mathcal{O}(\eta^2).\tag{119}$$

From Eq. (97), the first cumulant κ_1 thus reduces to

$$\begin{aligned}\frac{\kappa_1}{N} &\simeq \langle f(a)\rangle_a = \int \mathrm{d}a\, p_a(a)f(a) \simeq \frac{p_a(0)}{r}\int_{-\infty}^{\infty}\mathrm{d}u\, f(u/r)\\ &= \frac{p_a(0)}{r}\frac{2y}{\sqrt{1+y^2}} = 2p_a(0)x\eta,\end{aligned}\tag{120}$$

where in the first line we changed variable to $u = ra$ and we used the fact that $r^{-1} \sim \mathcal{O}(\eta)$ (see Eq. (117)), while in the second line we explicitly computed the integral

$$\frac{1}{\pi}\int_0^{\infty}\mathrm{d}u\,\arctan\left(\frac{\sin 2\theta}{u^2+\cos 2\theta}\right) = \frac{\tan\theta}{\sqrt{1+\tan^2\theta}} = \sin\theta,\qquad \theta \in \left[0,\frac{\pi}{2}\right],\tag{121}$$

and we inserted the expression for r found in Eq. (117) (note that $f(u/r)$ is actually r-independent – see Eq. (95)). One could alternatively compute κ_1 by taking the average of Eq. (8): this leads to the same result upon expanding for small E and η, since

$$\rho(\lambda) = p_a(\lambda) + \mathcal{O}(\eta).\tag{122}$$

The same steps can be repeated for the second (and possibly any other) cumulant κ_2, yielding

$$\frac{\kappa_2}{N} \simeq \langle f^2(a)\rangle_a \simeq \frac{p_a(0)}{r}\int_{-\infty}^{\infty}\mathrm{d}u\, f^2(u/r).\tag{123}$$

From Eq. (99) we thus obtain the leading order estimate for the level compressibility when $\eta \ll 1$, which takes the universal scaling form

$$\chi\left(y = \frac{E}{2\pi p_a(0)\eta}\right) \simeq \chi_T(y) \equiv \frac{\sqrt{1+y^2}}{\pi^2 y}\int_0^{\infty}\mathrm{d}u\left\{\arctan\left[\frac{2y}{u^2(1+y^2)+1-y^2}\right]\right\}^2,\tag{124}$$

where we stress that we have chosen the branch $\arctan(z) \in [0,\pi]$. Upon integrating by parts and performing some algebra [71, 72], the integral over u can be computed explicitly to give

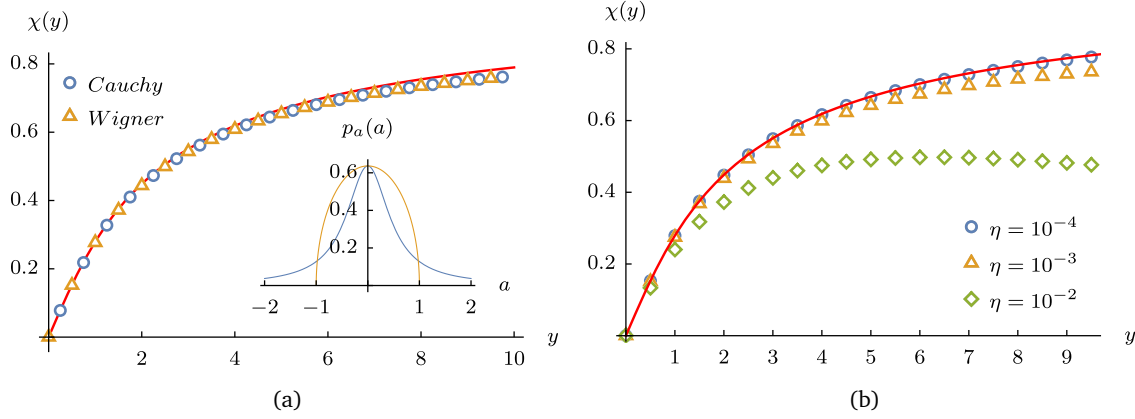$$\chi_T(y) = \frac{1}{\pi y}\left[2y\arctan(y) - \ln\left(1+y^2\right)\right].\tag{125}$$

Figure 6: Scaling form of the level compressibility, in the limit in which $E \propto y\eta$ and $\eta \ll 1$. In (a) we compare the universal prediction in Eq. (125) (solid red line) with the two exactly solvable cases studied in Section 3.4 (symbols), showing a good agreement at low energies $y = E/[2\pi p_a(0)\eta]$ (i.e., the condition under which Eq. (125) was derived). We chose $\eta = 5 \cdot 10^{-4}$ and $\omega = 0.5$, $\sigma = 1$, so that $p_a(0)$ assumes the same value for the two distributions (see inset). In (b) we exemplify in the Cauchy case how the curves corresponding to different values of η collapse onto the same master curve, when plotted as a function of $y \propto E/\eta$ for $\eta \ll 1$.

The function $\chi_T(y)$ grows monotonically from 0 to 1 as we increase y, and it is plotted in Fig. 6a together with the level compressibility for the two cases explicitly solved above, i.e., Cauchy and Wigner. We find a good agreement at low energies y, while we observe a departure at large energies: here the scaling prediction keeps growing, while the actual compressibility must hit a maximum and start decreasing – see Fig. 5. The same trend can be observed in Fig. 6b, where we evaluate the level compressibility for different values of η, and show that they collapse on a common master curve when they are plotted as a function of $y \propto E/\eta$.

The other two cases ($\delta > 1$ or $\delta < 1$) can be easily addressed by the same token. When $\delta > 1$, by studying the self-consistency equations as in Eq. (112) we obtain at leading order

$$\tan\theta \simeq y\eta^{\delta-1}, \tag{126}$$

$$r^{-1} \simeq 2\pi\eta p_a(0). \tag{127}$$

It can be readily seen that $\kappa_1/N \simeq 2p_a(0)x\eta^\delta$ and $\kappa_2/N \sim \mathcal{O}(\eta^{2\delta-1})$, so that in this limit

$$\chi(E = x\eta^\delta) \sim \mathcal{O}(\eta^{\delta-1}), \qquad \delta > 1, \, \eta \ll 1. \tag{128}$$

This resembles the behavior of $\chi(E)$ in the case of a pure GOE matrix, see Section 3.3.2 and Appendix E. Conversely, for $\delta < 1$ the self-consistency equations (91) reduce to

$$\Delta^{-1} \simeq \varepsilon - ix\eta^\delta, \tag{129}$$

and, by comparison with Eq. (101), we identify this limit as that in which the eigenvalues behave as i.i.d. random variables. In particular,

$$\chi(E = x\eta^\delta) \simeq 1 - \frac{\langle I_N[-E, E]\rangle_a}{N}, \qquad \delta < 1, \, \eta \ll 1. \tag{130}$$

The above analysis suggests to interpret the quantity $E_T \sim 2\pi p_a(0)\eta$ as the *Thouless energy* of the system. Indeed, consider again the limit in which $\eta \ll 1$, and let $E \propto \eta^\delta$. For $E \ll E_T$ (i.e., $\delta > 1$), the eigenvalues organize in multiplets (or mini-bands [23]) and they repel each other as in the GOE ensemble – as a result, the level compressibility is zero at leading order (see Eq. (128)). For $E \gg E_T$ ($\delta < 1$), on the other hand, the various multiplets no longer interact, and we recover the Poisson statistics – see Eq. (130). Finally, the case $\delta = 1$ marks a crossover in which the level compressibility $\chi(E/E_T)$ assumes the universal scaling form given in Eq. (125). Indeed, the asymptotics of the function $\chi_T(y)$ in Eq. (125) can be checked to give

$$\chi_T(y) \simeq \begin{cases} \dfrac{y}{\pi}, & y \ll 1, \\ 1 - \dfrac{2(1 + \ln y)}{\pi y}, & y \gg 1, \end{cases} \tag{131}$$

showing that $\chi_T(y)$ interpolates between Wigner-Dyson statistics at low energy, and Poisson statistics at higher energy.
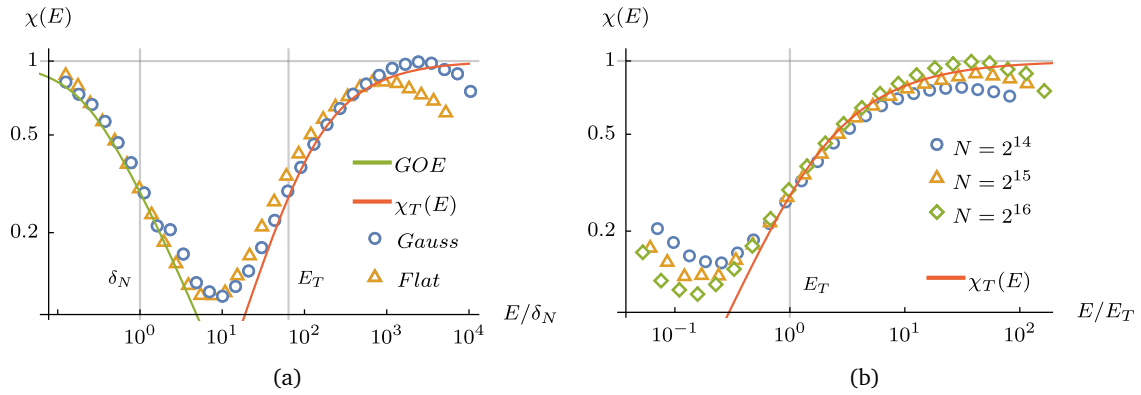
Figure 7: Behavior of the level compressibility $\chi(E)$ at low energies. The symbols correspond to numerical results, and we indicated with vertical lines the mean level spacing δ_N (see Eq. (4)) and the Thouless energy $E_T \sim N^{1-\gamma}$. In panel (a), $p_a(a)$ is chosen Gaussian with unit variance or uniform, and $N = 2^{16}$. The region $E \lesssim \delta_N$ is described by the GOE prediction in Eq. (E.25), while the crossover region $E \sim E_T$ is described by the universal scaling form in Eq. (125). In panel (b), $p_a(a)$ is Gaussian, and we show the approach to the universal curve $\chi_T(E)$ for increasing values of the matrix size N. We used $\gamma = 1.5$ throughout, and the simulations with $N = 2^{14}, 2^{15}$, or 2^{16} are averaged over $N_{\text{tot}} = 256, 64$, or 32 samples, respectively.

3.6 Behavior at low energy

In this final Section we use extensive numerical exact diagonalization of large random matrices in order to inspect the low-energy behavior of the level compressibility $\chi(E)$. Indeed, our prediction of Section 3.5 is expected to break down for energies of the order of the mean level spacing δ_N of the finite-sized matrix $\mathcal{H}$, which is given by $\delta_N \simeq [Np_a(0)]^{-1}$, Eq. (4). Equivalently, we expect that the replica calculation provides the correct result in the $N \to \infty$ limit, while for a matrix of size N and for sufficiently small E we should eventually recover the exact GOE result $\chi_{\text{GOE}}(E)$ [73–80],

whose derivation is reported in Appendix E.3 – see Eq. (E.25). The overall picture is thus the one we present in Fig. 7, and which we support by numerical results. The region $E \lesssim \delta_N$ is described by the GOE prediction in Eq. (E.25). For $1 < \gamma < 2$ the Thouless energy $E_T \sim N^{1-\gamma}$ is such that $\delta_N \ll E_T \ll 1$, so that the crossover region with $E \sim E_T$ is described by the universal function $\chi_T(E)$ given in Eq. (125). For larger $E \gtrsim \mathcal{O}(1)$, the level compressibility becomes model dependent and it is described by Eq. (99) (see also Fig. 5).

The datapoints[3] presented in panel (a) of Fig. 7 correspond to the choices of $p_a(a)$ Gaussian or uniform, which supports our claim of universality of $\chi_T(E)$ in the region $E \sim E_T$. Note, in fact, that the data follow the predicted curves (up to finite-size effects) with no adjustable parameters (i.e., no fitting was needed). The datapoints are eventually observed to deviate from the scaling prediction, as they reach a maximum and start decaying to zero as in Fig. 5. By increasing N, however, this maximum is observed to shift to larger values of E, and the plateau around $\chi(E) \sim 1$ broadens accordingly.

4 Conclusions

In this work we have used the replica method to study the average spectral density (Section 2) and the local level statistics (Section 3) of a deformed GOE random matrix ensemble known as the generalized RP model. We have focused in particular on its non-ergodic extended phase [23], which is conveniently characterized in terms of the level compressibility $\chi(E)$ (see Eq. (9)). In particular, we have shown that $\chi(E)$ assumes a universal form independent of the character $p_a(a)$ of the deformation matrix A (see Eq. (1)), provided that the system is probed over energy scales of the order of the Thouless energy E_T (see Section 3.5).

It is natural to conjecture that this universal regime should persist in structurally similar random matrix ensembles. For instance, one could numerically inspect the case in which the GOE matrix is replaced by a Wigner matrix (i.e., any real symmetric matrix with i.i.d. random entries taken from a probability distribution with finite variance), whose limiting average spectral density is still given by the semi-circle law [81]. Similarly, it would be interesting to study the effect of the diagonal deformation matrix A on a Wishart matrix (see, e.g., [82]), or else on a sparse (rather than dense) matrix B, such as those describing the Erdös-Rényi random graph [83–85], which can still be treated analytically (at least to some extent); if the average connectivity is chosen finite, the spectral density of Erdös-Rényi is no longer a semi-circle, but the local statistics is still of the Wigner-Dyson type. Finally, it is worth mentioning that the spectral properties of the intermediate phase of the generalized RP model are particularly simple: differently from realistic interacting quantum systems, the mini-bands are compact and the eigenvectors are fractal but not *multi*fractal (meaning that all the moments of the wave-functions' amplitudes are described by the same fractal exponent $D_\gamma = 2 - \gamma$). In order to overcome some of these issues, several extensions of the RP model have been proposed in the last few years, featuring either log-normal [46, 47] or power-law [49] distributed off-diagonal matrix elements. In these models the mini-band structure turns out to be non-compact but rather multifractal. It would be therefore illuminating to study the behavior of

---

[3]Although the scaling function in Eq. (125) has been derived under the assumption that the spectral density is symmetric, the numerical results are obtained by averaging the cumulants of the number of eigenvalues within many energy windows across the whole bandwidth, for which the symmetry with respect to the center of the window is lost. However, on the scale of the Thouless energy $E \simeq E_T \propto \eta$, the corrections due to the fact that $p_a(-E) \neq p_a(E)$ are of $\mathcal{O}(\eta)$, and therefore yield a contribution which is of the same order as the finite-size corrections, and which can be neglected for sufficiently large N.

the level compressibility at small energy and check whether or not the universal form discussed here is robust with respect to these modifications.

Acknowledgements

A   Number of i.i.d. variables in an interval

In this Appendix we revise the standard textbook result for the statistics of the number of eigenvalues in a finite interval, when such eigenvalues behave as independent and identically distributed random variables. This will help clarifying the behavior of the level compressibility $\chi(E)$ in the case of Poisson level statistics.

Given N random variables a_i distributed according to $p_a(a_i)$ (e.g., the eigenvalues of the matrix A in Eq. (1)), the number of variables contained in the interval $[\alpha, \beta]$ can be written as

$$I_N[\alpha, \beta] = \sum_{i=1}^{N} \theta_i \,, \tag{A.1}$$

where we introduced the indicator function

$$\theta_i \equiv \mathbb{1}_{[\alpha,\beta]}(a_i) \,, \qquad \mathbb{1}_{[\alpha,\beta]}(x) = \begin{cases} 1 & x \in [\alpha, \beta] \,, \\ 0 & x \notin [\alpha, \beta] \,. \end{cases} \tag{A.2}$$

Its cumulant generating function can be constructed by noting that

$$e^{-s I_N[\alpha,\beta]} = \prod_{i=1}^{N} e^{-s\theta_i} = \prod_{i=1}^{N} \left[1 + (e^{-s} - 1)\theta_i \right] \,, \tag{A.3}$$

and then

$$\ln \left\langle e^{-s I_N[\alpha,\beta]} \right\rangle = \sum_{i=1}^{N} \ln \left[1 + (e^{-s} - 1) \int_{\alpha}^{\beta} da \, p_a(a) \right] = N \ln \left[1 + (e^{-s} - 1) \frac{\langle I_N[\alpha,\beta] \rangle}{N} \right] \,. \tag{A.4}$$

Note that this coincides with the limiting case in Eq. (104) of our general result.

By expanding in powers of s the two sides of Eq. (A.4) and comparing with Eqs. (67) and (71), we can in particular extract the first two cumulants

$$\kappa_1 = \langle I_N[\alpha,\beta] \rangle \,, \qquad \kappa_2 = \langle I_N[\alpha,\beta] \rangle \left(1 - \frac{\langle I_N[\alpha,\beta] \rangle}{N} \right) \,, \tag{A.5}$$

and from Eq. (9) we obtain the level compressibility

$$\chi(E) = 1 - \frac{\langle I_N[-E,E]\rangle}{N}. \tag{A.6}$$

We thus generically expect $\chi(E) \sim 1$ for small E, and $\chi(E) \to 0$ for large E.

B   Details of the replica calculation for the spectral density

In this Appendix we fill in the missing steps which lead from Eq. (18) to Eq. (19) in Section 2. A replica-based calculation for the pure GOE ensemble can be found in Ref. [43], from which we partially adopt the notation. We start by expressing the average of the replicated partition function as

$$\langle \mathcal{Z}^n(\lambda) \rangle \propto \left\langle \int_{\mathbb{R}^{Nn}} \left(\prod_{\alpha=1}^{n} d\mathbf{r}_\alpha \right) \exp\left[-\frac{i}{2} \sum_{i,j=1}^{N} \sum_{\alpha=1}^{n} r_{i\alpha}(\lambda_\varepsilon \delta_{ij} - h_{ij}) r_{j\alpha} \right] \right\rangle_{A,B}, \tag{B.1}$$

where we indicated by $h_{ij} \equiv a_i \delta_{ij} + J b_{ij}$ the elements of the random matrix $\mathcal{H}$ given in Eq. (1), and $J = J(N) \equiv \nu N^{-\gamma/2}$. The average symbol means

$$\langle \bullet \rangle_{A,B} \equiv \int_{\mathbb{R}^N} \left(\prod_{i \leq j}^{N} db_{ij} \right) p_B(\{b_{ij}\}) \left(\prod_{i=1}^{N} \int da_i \, p_a(a_i) \right) (\bullet), \tag{B.2}$$

where the probability distribution of the elements b_{ij} of the GOE matrix B reads

$$p_B(\{b_{ij}\}) = \prod_{i=1}^{N} \frac{e^{-b_{ii}^2/2}}{\sqrt{2\pi}} \prod_{i \leq j} \frac{e^{-b_{ij}^2}}{\sqrt{\pi}}. \tag{B.3}$$

Here and henceforth, Latin indices run up to N in real space, while Greek indices run up to n in replica space. Computing the Gaussian integrals over b_{ij} gives (up to a numerical constant)

$$\langle \mathcal{Z}^n(\lambda) \rangle \propto \int_{\mathbb{R}^{Nn}} \left(\prod_{\alpha=1}^{n} d\mathbf{r}_\alpha \right) \left\langle \exp\left[-\frac{i}{2} \sum_{i=1}^{N} \sum_{\alpha=1}^{n} (\lambda_\varepsilon - a_i) r_{i\alpha}^2 \right] \right\rangle_A$$

$$\times \exp\left\{ -\frac{J^2}{4} \left[\frac{1}{2} \sum_{i=1}^{N} \left(\sum_{\alpha=1}^{n} r_{i\alpha}^2 \right)^2 + \sum_{i \leq j} \left(\sum_{\alpha=1}^{n} r_{i\alpha} r_{j\alpha} \right)^2 \right] \right\}, \tag{B.4}$$

where $\langle \bullet \rangle_A$ indicates the reduced averaged over the entries of A – see Eq. (B.2). The interacting term in the second line can be usually decoupled by means of the Hubbard-Stratonovich transformation [40], which is however ineffective in our case, for a generic choice of $p_a(a)$. We introduce instead the normalized density

$$\mu(\vec{y}) \equiv \frac{1}{N} \sum_{i=1}^{N} \prod_{\alpha=1}^{n} \delta(y_\alpha - r_{i\alpha}), \tag{B.5}$$

where $\vec{y} \in \mathbb{R}^n$ has components $y_a \in \mathbb{R}$, and we insert into Eq. (B.4) the functional integral representation of the identity

$$1 = N^{\dim(\mu)} \int \mathcal{D}\mu \, \mathcal{D}\hat{\mu} \, \exp\left\{ -i \int d\vec{y} \, \hat{\mu}(\vec{y}) \left[N\mu(\vec{y}) - \sum_{i=1}^{N} \prod_{\alpha=1}^{n} \delta(y_\alpha - r_{i\alpha}) \right] \right\}. \tag{B.6}$$

Here $\dim(\mu)$ is the dimension of the field μ, which renders the prefactor on the right hand side formally infinite – this will be of no consequence in the following calculation, since this prefactor is λ-independent. Equation (B.6) is useful because it allows us to rewrite

$$\frac{1}{2}\sum_{i=1}^{N}\left(\sum_{\alpha=1}^{n}r_{i\alpha}^2\right)^2 + \sum_{i\leq j}\left(\sum_{\alpha=1}^{n}r_{i\alpha}r_{j\alpha}\right)^2 = \frac{1}{2}\sum_{i,j=1}^{N}\left(\sum_{\alpha=1}^{n}r_{i\alpha}r_{j\alpha}\right)^2 \tag{B.7}$$

$$= \frac{N^2}{2}\int d\vec{y}\, d\vec{w}\, \mu(\vec{y})\mu(\vec{w})\left(\sum_{\alpha=1}^{n}y_\alpha w_\alpha\right)^2 = \frac{N^2}{2}\int d\vec{y}\, d\vec{w}\, \mu(\vec{y})\mu(\vec{w})(\vec{y}\cdot\vec{w})^2\,,$$

so that inserting the identity in Eq. (B.6) into Eq. (B.4) leads to

$$\langle \mathcal{Z}^n(\lambda)\rangle \propto \int \mathcal{D}\mu\,\mathcal{D}\hat{\mu}\,\exp\left\{-iN\int d\vec{y}\,\hat{\mu}(\vec{y})\mu(\vec{y}) - \frac{(JN)^2}{8}\int d\vec{y}\, d\vec{w}\, \mu(\vec{y})\mu(\vec{w})(\vec{y}\cdot\vec{w})^2\right\}$$

$$\times \int_{\mathbb{R}^{Nn}}\left(\prod_{\alpha=1}^{n}d\mathbf{r}_\alpha\right)\left\langle \exp\left[-\frac{i}{2}\sum_{i=1}^{N}\sum_{\alpha=1}^{n}(\lambda_\varepsilon - a_i)r_{i\alpha}^2 + i\sum_{i=1}^{N}\int d\vec{w}\,\hat{\mu}(\vec{w})\prod_{\alpha=1}^{n}\delta(w_\alpha - r_{i\alpha})\right]\right\rangle_A. \tag{B.8}$$

Staring at Eq. (B.8) for long enough, one realizes that the second line contains N copies of the same integral,

$$\int_{\mathbb{R}^{Nn}}\left(\prod_{\alpha=1}^{n}d\mathbf{r}_\alpha\right)\left\langle \exp\left[-\frac{i}{2}\sum_{i=1}^{N}\sum_{\alpha=1}^{n}(\lambda_\varepsilon - a_i)r_{i\alpha}^2 + i\sum_{i=1}^{N}\int d\vec{w}\,\hat{\mu}(\vec{w})\prod_{\alpha=1}^{n}\delta(w_\alpha - r_{i\alpha})\right]\right\rangle_A$$

$$= \left\{\int_{\mathbb{R}^n} d\vec{y}\int da\, p_a(a)\exp\left[-\frac{i}{2}\sum_{\alpha=1}^{n}(\lambda_\varepsilon - a)y_\alpha^2 + i\int d\vec{w}\,\hat{\mu}(\vec{w})\prod_{\alpha=1}^{n}\delta(w_\alpha - y_\alpha)\right]\right\}^N$$

$$= \left\{\int_{\mathbb{R}^n} d\vec{y}\int da\, p_a(a)\exp\left[-\frac{i}{2}(\lambda_\varepsilon - a)|\vec{y}|^2 + i\hat{\mu}(\vec{y})\right]\right\}^N. \tag{B.9}$$

Note that it was crucial to assume independent entries a_i, so that their distribution in Eq. (B.2) is factorized. Plugging this expression back into Eq. (B.8) allows us to rewrite $\langle \mathcal{Z}^n(\lambda)\rangle$ as reported in Eq. (19) of the main text.

C   Connection with the Zee formula

In this Appendix we show why Eq. (44) is hiddenly the Zee formula. In Ref. [62], the recipe for computing the spectrum $\rho_{1+2}(\lambda)$ of the sum of two random matrices $M_1 + M_2$ is given as follows:

(i) Compute the resolvents (or Green's functions) associated to $\rho_1(\lambda)$ and $\rho_2(\lambda)$, i.e., $\mathcal{G}_1(z)$ and $\mathcal{G}_2(z)$.

(ii) Compute their functional inverses $B_1(z)$ and $B_2(z)$, or Blue's functions, defined by

$$B(\mathcal{G}(z)) = z\,. \tag{C.1}$$

(iii) Apply the sum rule

$$B_{1+2}(z) = B_1(z) + B_2(z) - 1/z\,. \tag{C.2}$$

(iv) Invert the result back (see Eq. (42)) to find

$$B_{1+2}(z) \to \mathcal{G}_{1+2}(z) \to \rho_{1+2}(\lambda). \tag{C.3}$$

Another interesting object is however the *R*-function, which is simply defined as

$$R(z) \equiv B(z) - 1/z, \tag{C.4}$$

and which is easily seen to satisfy the free-sum rule [43, 61, 86]

$$R_{1+2}(z) = R_1(z) + R_2(z). \tag{C.5}$$

It follows that

$$B_1(x) = B_{1+2}(x) - R_2(x), \tag{C.6}$$

which we can choose to apply in particular on $x = \mathcal{G}_{1+2}(z)$, yielding by construction

$$B_1(\mathcal{G}_{1+2}(z)) = z - R_2(\mathcal{G}_{1+2}(z)). \tag{C.7}$$

Applying $\mathcal{G}_1$ on both sides finally yields

$$\mathcal{G}_{1+2}(z) = \mathcal{G}_1(z - R_2(\mathcal{G}_{1+2}(z))). \tag{C.8}$$

The analogy with Eq. (44) is readily established once we recall that, if M_2 is a GOE matrix, then its *R*-function is simply $R_2(z) = z$ [87].

D   Details of the replica calculation of the level compressibility

In this Appendix we provide the technical steps for the derivation of the cumulant generating function and the level compressibility presented in Section 3. A similar calculation for the pure GOE/GUE ensemble can be found in Ref. [70], while the derivation in the case of the Erdös-Rényi graph and the Anderson model on a random regular graph was reported in Ref. [69].

D.1   Functional representation

The target of this Section is to express $Q_{[\alpha,\beta]}(n_\pm)$ given in Eq. (69) within the replica formalism, as we did in Appendix B. The first step is to choose a suitable representation for the partition function $\mathcal{Z}(z)$ which appears in Eq. (69): indeed, the one we introduced in Eq. (17) is only appropriate if $\text{Im}\{z\} < 0$, being the integral not convergent otherwise. If on the contrary $\text{Im}\{z\} > 0$, then one should choose instead

$$\mathcal{Z}_+(z) \equiv \left(\frac{i}{2\pi}\right)^{N/2} \int_{\mathbb{R}^N} d\mathbf{r}\, e^{\frac{i}{2}\mathbf{r}^T(z\mathbb{1}-\mathcal{H})\mathbf{r}}. \tag{D.1}$$

Since the various prefactors in front of the integral in Eq. (17) will cancel out in Eq. (69) after we take the analytic continuation to $n_\pm \to \pm is/\pi$, we will not need to keep track of them in the following.

In analogy with the representation in Eq. (B.1), we can still write

$$Q_{[\alpha,\beta]}(n_\pm) \propto \left\langle \int_{\mathbb{R}^{Nn}} \left(\prod_{\sigma=1}^{n} d\mathbf{r}_\sigma\right) \exp\left[-\frac{i}{2}\sum_{i,j=1}^{N}\sum_{\sigma=1}^{n} r_{i\sigma}(\Lambda_{\sigma\sigma}\delta_{ij} - L_{\sigma\sigma}h_{ij})r_{j\sigma}\right]\right\rangle_{A,B}, \tag{D.2}$$

but now we interpret

$$n = 2(n_+ + n_-),$$ (D.3)

because each of the four partition functions in Eq. (69) requires its own set of replicas (here labelled by the Greek index σ, to avoid confusion with the left boundary α of the interval). We have also replaced the eigenvalue λ_ε by the block matrix $\hat{\Lambda}$, which is defined in Eq. (74) together with the block matrix $\hat{L}$. Notice that the elements $\bar{\alpha}_\varepsilon = -\alpha_\varepsilon^*$ of the matrix $\hat{\Lambda}$ follow from the representation in Eq. (D.1).

The very same steps which in Appendix B led us to Eq. (19) of the main text now give

$$Q_{[\alpha,\beta]}(n_\pm) \propto \int \mathcal{D}\mu \, \mathcal{D}\hat{\mu} \, \exp\{N\mathcal{S}_{n_\pm}[\mu,\hat{\mu};\hat{\Lambda}]\},$$ (D.4)

with the action

$$\mathcal{S}_{n_\pm}[\mu,\hat{\mu};\hat{\Lambda}] \equiv -i \int \mathrm{d}\vec{\tau}\, \mu(\vec{\tau})\hat{\mu}(\vec{\tau}) - \frac{\eta}{2} \int \mathrm{d}\vec{\tau} \, \mathrm{d}\vec{\tau}' \, \mu(\vec{\tau})\mu(\vec{\tau}') \left(\vec{\tau}\,\hat{L}\,\vec{\tau}'\right)^2$$ (D.5)

$$+ \ln \int \mathrm{d}\vec{\tau}\, \exp\left[-\frac{i}{2}\vec{\tau}\,\hat{\Lambda}\,\vec{\tau} + i\hat{\mu}(\vec{\tau})\right] \int \mathrm{d}a \, p_a(a) \exp\left(\frac{i}{2}a\vec{\tau}\,\hat{L}\,\vec{\tau}\right).$$

This generalizes the action in Eq. (20), and the vector $\vec{\tau} \in \mathbb{R}^n$ plays the same role as the vector $\vec{y}$ but in an extended replica space, with n given in Eq. (D.3). By noting that the action in Eq. (D.5) is quadratic in μ, we can evaluate the Gaussian functional integral in $\mathcal{D}\mu$ to obtain

$$Q_{[\alpha,\beta]}(n_\pm) \propto \int \mathcal{D}\hat{\mu} \, \exp\{N\mathcal{S}_{n_\pm}[\hat{\mu};\hat{\Lambda}]\},$$ (D.6)

$$\mathcal{S}_{n_\pm}[\hat{\mu};\hat{\Lambda}] \equiv -\frac{1}{2\eta} \int \mathrm{d}\vec{\tau} \, \mathrm{d}\vec{\tau}' \, \hat{\mu}(\vec{\tau})M^{-1}(\vec{\tau},\vec{\tau}')\hat{\mu}(\vec{\tau}')$$ (D.7)

$$+ \ln \int \mathrm{d}\vec{\tau}\, \exp\left[-\frac{i}{2}\vec{\tau}\,\hat{\Lambda}\,\vec{\tau} + i\hat{\mu}(\vec{\tau})\right]\psi_a\left(-\frac{1}{2}\vec{\tau}\,\hat{L}\,\vec{\tau}\right),$$

where $\psi_a(z)$ was given in Eq. (30), and we introduced the function $M(\vec{\tau},\vec{\tau}')$ as in Eq. (75). We omitted from Eq. (D.6) a $\hat{\Lambda}$−independent prefactor coming from the Gaussian integration, which will in general depend on n. However, one can check that all the prefactors cancel out smoothly by including the Jacobian of the variable transformations (see later), and after taking the functional integral over the Gaussian fluctuations in Appendix D.2. We will thus avoid reporting these prefactors, so as to lighten the notation.

The saddle-point equation follows simply from Eq. (D.7) as

$$\hat{\mu}(\vec{\tau}) = i\eta \frac{\int \mathrm{d}\vec{\tau}'\, M(\vec{\tau},\vec{\tau}')\exp\left[-\frac{i}{2}\vec{\tau}'\,\hat{\Lambda}\,\vec{\tau}' + i\hat{\mu}(\vec{\tau}')\right]\psi_a\left(-\frac{1}{2}\vec{\tau}'\,\hat{L}\,\vec{\tau}'\right)}{\int \mathrm{d}\vec{\tau}'\, \exp\left[-\frac{i}{2}\vec{\tau}'\,\hat{\Lambda}\,\vec{\tau}' + i\hat{\mu}(\vec{\tau}')\right]\psi_a\left(-\frac{1}{2}\vec{\tau}'\,\hat{L}\,\vec{\tau}'\right)},$$ (D.8)

which is analogous to Eq. (26). In the following, we will look for an explicit rotationally-invariant solution: to this end, it is useful to introduce the new variable $\varphi(\vec{\tau})$ defined via

$$\hat{\mu}(\vec{\tau}) = i \int \mathrm{d}\vec{\tau}'\, M(\vec{\tau},\vec{\tau}')\varphi(\vec{\tau}').$$ (D.9)

Changing variables from $\hat{\mu}$ to φ in Eq. (D.6) leads to the expression reported in Eq. (73).

D.2 Gaussian fluctuations around the saddle-point

In order to go beyond the saddle-point approximation, we introduce the fluctuation $\phi(\vec{\tau})$ around the saddle-point $\varphi_0(\vec{\tau})$ in the form

$$\varphi(\vec{\tau}) = \varphi_0(\vec{\tau}) + \phi(\vec{\tau}). \tag{D.10}$$

Calling for brevity $\mathcal{S}_{n_\pm}[\varphi; \hat{\Lambda}] \equiv \mathcal{S}[\varphi]$, we then have up to $\mathcal{O}(N^{-2})$

$$Q_{[\alpha,\beta]}(n_\pm) \propto e^{N\mathcal{S}[\varphi_0]} \int \mathcal{D}(i\phi) \exp\left\{ \frac{N}{2} \int d\vec{\tau}_1 d\vec{\tau}_2 \, \phi(\vec{\tau}_1) \left. \frac{\delta^2 \mathcal{S}[\varphi]}{\delta\varphi(\vec{\tau}_1)\delta\varphi(\vec{\tau}_2)} \right|_{\varphi=\varphi_0} \phi(\vec{\tau}_2) \right\}, \tag{D.11}$$

and one can check that we can express

$$\left. \frac{\delta^2 \mathcal{S}[\varphi]}{\delta\varphi(\vec{\tau}_1)\delta\varphi(\vec{\tau}_2)} \right|_{\varphi=\varphi_0} = \frac{1}{\eta} M(\vec{\tau}_1, \vec{\tau}_2)[\mathbb{1}(\vec{\tau}_1, \vec{\tau}_2) + T(\vec{\tau}_1, \vec{\tau}_2)] \tag{D.12}$$

in terms of the functions M and T given in Eqs. (75) and (79), respectively. Computing the Gaussian integral in Eq. (D.11) we thus find

$$Q_{[\alpha,\beta]}(n_\pm) = \exp\left\{ N\mathcal{S}[\varphi_0] - \frac{1}{2} \ln\det(\mathbb{1} + T) \right\} + \mathcal{O}(1/N^2). \tag{D.13}$$

Expanding the logarithm in series as

$$\ln(1+x) = -\sum_{k=1}^\infty \frac{(-x)^k}{k} \tag{D.14}$$

we finally get Eq. (78), where the trace and matrix operations are intended over the replica vectors as

$$\operatorname{Tr} T = \int d\vec{\tau} \, T(\vec{\tau}, -\vec{\tau}), \qquad T^2(\vec{\tau}_1, \vec{\tau}_2) = \int d\vec{\tau} \, T(\vec{\tau}_1, \vec{\tau}) T(\vec{\tau}, \vec{\tau}_2). \tag{D.15}$$

We may try to specialize Eq. (78) to the rotationally invariant Ansatz in Eq. (82). The fluctuation matrix in Eq. (D.12) becomes

$$T(\vec{\tau}_1, \vec{\tau}_2) = \varphi_0(\vec{\tau}_1) \left[\left(\vec{\tau}_1 \hat{L} \, \vec{\tau}_2 \right)^2 - \left(\vec{\tau}_2 \hat{K} \, \vec{\tau}_2 \right) \right], \tag{D.16}$$

and by using Wick's theorem together with Eq. (84) we obtain, for instance,

$$\operatorname{Tr} T = \int d\vec{\tau} \, T(\vec{\tau}, \vec{\tau}) = \eta \left[\sum_i \left(2K_{ii}^{(2)} - K_{ii}^2 \right) + \sum_{ij} L_{ii} L_{jj} K_{ij}^{(2)} \right], \tag{D.17}$$

where we introduced the matrix

$$K_{ij}^{(2)} \equiv \int da \, p_a(a) \left(\frac{\hat{C}}{1 - ia\hat{L}\hat{C}} \right)_{ii} \left(\frac{\hat{C}}{1 - ia\hat{L}\hat{C}} \right)_{jj}. \tag{D.18}$$

We recognize in the last expression a generalization of the resolvent (see Eq. (41)) which encodes higher order correlations. The next terms $\operatorname{Tr} T^k$ with $k > 1$ in the series of Eq. (78) will involve

some matrices $K_{ij}^{(k+1)}$ with increasingly higher order correlations, which are nontrivial to compute in general. However, it is straightforward to show that $\text{Tr}\, T^k = \mathcal{O}(\eta^k)$, so that when η is small the series in Eq. (78) is dominated by its first few terms. To the best of our efforts, it has not been possible to resum the whole series in Eq. (78), as it happens instead in the pure GOE case – see Appendix E and Ref. [70].

Specializing to the Cauchy distribution in Eq. (45) with $\mu = 0$, one finds for instance that K_{ii} is given in Eq. (107) in terms of the matrix elements of $\hat{C}$, while using complex integration in Eq. (D.18) yields

$$K_{ij}^{(2)} = \begin{cases} \dfrac{C_{ii}C_{jj}}{(\omega C_{ii}-1)(\omega C_{jj}-1)}, & L_{ii} \cdot L_{jj} > 0, \\[4mm] \dfrac{C_{ii}C_{jj}}{(\omega C_{ii}-1)(\omega C_{jj}-1)}\left[1 - \dfrac{2\omega C_{ii}C_{jj}}{C_{ii}+C_{jj}}\right], & L_{ii} \cdot L_{jj} < 0. \end{cases} \tag{D.19}$$

E   Level compressibility in the pure GOE case

In this Appendix we recover the results of Ref. [70] concerning the level compressibility for a GOE matrix. This allows us to inspect the similarities and the differences with respect to the GRP case analyzed in this manuscript. In the final section E.3, we will repeat the derivation of $\chi_{\text{GOE}}(E)$ using more standard techniques in order to address the low-energy region $E \ll \delta_N$ (see Eq. (4) and Section 3.6).

The pure GOE case can be formally obtained from Eq. (1) by letting the distribution $p_a(a)$ of the diagonal elements of the matrix A tend to a delta function, so that $\psi_a(z) \to 1$. The calculation then becomes analogous to that reported in Ref. [70], whose main steps we detail here for completeness. By replacing the replico-symmetric Ansatz of Eq. (80) into the saddle point equation (76) one first obtains $\mathcal{N} = \eta/Z_\varphi$, where Z_φ is given in Eq. (83). By using Gaussian integration one can then show that

$$\int d\vec{\tau}' \, M(\vec{\tau}, \vec{\tau}')\varphi_0(\vec{\tau}') = \eta \vec{\tau} \, \hat{C} \, \vec{\tau}, \tag{E.1}$$

and thus the remaining free parameters in Eq. (80) can be determined by solving the set of four self-consistency equations which follow from Eq. (76) as

$$\hat{C}^{-1} = 2\eta \, \hat{C} + i \, \hat{\Lambda}. \tag{E.2}$$

Note that this can be recovered from Eqs. (85) and (86) by using the fact that the resolvent corresponding to $p_a(a) = \delta(a)$ is $\mathcal{G}_a(z) = 1/z$.

E.1   Action and fluctuations around the saddle-point

Both the action and its Gaussian fluctuation matrix T can now be computed in correspondence of the saddle-point solution in Eq. (80). By using the definition of the action given in Eq. (73) together with the saddle-point equation (76) and the property in Eq. (E.1), one can easily deduce

$$\mathcal{S}_{n_\pm}[\varphi_0; \hat{\Lambda}] = \frac{\eta}{2}\left[n_+\left(\Delta_\alpha^2 + \bar{\Delta}_\beta^2\right) + n_-\left(\Delta_\beta^2 + \bar{\Delta}_\alpha^2\right)\right] + \frac{1}{2}(n_+ + n_-)\ln(2\pi)$$
$$+ \frac{1}{2}n_+ \ln\left(\Delta_\alpha \bar{\Delta}_\beta\right) + \frac{1}{2}n_- \ln\left(\Delta_\beta \bar{\Delta}_\alpha\right). \tag{E.3}$$

The computation of $\operatorname{Tr} T^k$ in Eq. (78) requires more work. First, we rewrite in correspondence of the Ansatz in Eq. (80)

$$T(\vec{\tau}_1, \vec{\tau}_2) = \varphi_0(\vec{\tau}_1) \left[\left(\vec{\tau}_1 \,\hat{L}\, \vec{\tau}_2 \right)^2 - \left(\vec{\tau}_2 \,\hat{C}\, \vec{\tau}_2 \right) \right], \tag{E.4}$$

where we have used the definition of the function $M(\vec{\tau}_1, \vec{\tau}_2)$ in Eq. (75) and the property in Eq. (E.1). The first few powers of T can then be computed by applying Wick's theorem: by introducing the notation

$$\langle \bullet \rangle_1 \equiv \frac{1}{\eta} \int \mathrm{d}\vec{\tau}_1 \, (\bullet) \varphi_0(\vec{\tau}_1), \tag{E.5}$$

it is sufficient to note that

$$\left\langle \tau_{1i} \tau_{1j} \right\rangle_1 = C_{ij}, \tag{E.6}$$

so that more complicated averages can be handled as

$$\left\langle \tau_{1i} \tau_{1j} \tau_{1k} \tau_{1l} \right\rangle_1 = C_{ij} C_{kl} + C_{ik} C_{jl} + C_{il} C_{jk}. \tag{E.7}$$

Upon noting that $\hat{L}^2 = \mathbb{1}_n$, one can then prove by induction the relation

$$T^{k+1}(\vec{\tau}_1, \vec{\tau}_2) = (2\eta)^k \varphi_0(\vec{\tau}_1) \left[\left(\vec{\tau}_1 \,\hat{L}^{k+1} \,\hat{C}^k\, \vec{\tau}_2 \right)^2 - \left(\vec{\tau}_2 \,\hat{C}^{2k+1}\, \vec{\tau}_2 \right) \right]. \tag{E.8}$$

Using Eq. (D.15) now yields

$$\operatorname{Tr} T^k = 2^{k-1} \eta^k \left[\left(\sum_i L_{ii}^k C_{ii}^k \right)^2 + \sum_i C_{ii}^{2k} \right], \tag{E.9}$$

and inserting the definition of the matrices $\hat{L}$ and $\hat{C}$ given in Eqs. (74) and (81) gives

$$\operatorname{Tr} T^k = 2^{k-1} \eta^k \Big\{ \left[n_+ \left(\Delta_\alpha^k + (-\bar{\Delta}_\beta)^k \right) + n_- \left(\Delta_\beta^k + (-\bar{\Delta}_\alpha)^k \right) \right]^2$$
$$+ n_+ \left(\Delta_\alpha^{2k} + \bar{\Delta}_\beta^{2k} \right) + n_- \left(\Delta_\beta^{2k} + \bar{\Delta}_\alpha^{2k} \right) \Big\}. \tag{E.10}$$

Taking the limit $n_\pm \to \pm is/\pi$ in Eqs. (E.3) and (E.10) then results in

$$\mathcal{S}_{\pm \frac{is}{\pi}}[\varphi_0; \hat{\Lambda}] = \frac{is}{2\pi} \left[\eta \left(\Delta_\alpha^2 + \bar{\Delta}_\beta^2 - \Delta_\beta^2 - \bar{\Delta}_\alpha^2 \right) + \ln\left(\frac{\Delta_\alpha \bar{\Delta}_\beta}{\Delta_\beta \bar{\Delta}_\alpha} \right) \right], \tag{E.11}$$

$$\operatorname{Tr} T^k \Big|_{\varphi = \varphi_0} = 2^{k-1} \eta^k \left\{ \frac{is}{\pi} \left(\Delta_\alpha^{2k} + \bar{\Delta}_\beta^{2k} - \Delta_\beta^{2k} - \bar{\Delta}_\alpha^{2k} \right) - \frac{s^2}{\pi^2} \left[\Delta_\alpha^k + (-\bar{\Delta}_\beta)^k - \Delta_\beta^k - (-\bar{\Delta}_\alpha)^k \right]^2 \right\}. \tag{E.12}$$

Comparing with the definitions of the cumulant generating function and the cumulants in Eqs. (67) and (71), respectively, we can finally identify

$$\frac{\kappa_1}{N} = -\frac{i}{2\pi} \left[\eta \left(\Delta_\alpha^2 + \bar{\Delta}_\beta^2 - \Delta_\beta^2 - \bar{\Delta}_\alpha^2 \right) + \ln\left(\frac{\Delta_\alpha \bar{\Delta}_\beta}{\Delta_\beta \bar{\Delta}_\alpha} \right) \right]$$
$$- \frac{i}{4\pi N} \sum_{k=1}^{\infty} \frac{(-2\eta)^k}{k} \left(\Delta_\alpha^{2k} + \bar{\Delta}_\beta^{2k} - \Delta_\beta^{2k} - \bar{\Delta}_\alpha^{2k} \right), \tag{E.13}$$

$$\frac{\kappa_2}{N} = -\frac{i}{2\pi^2 N} \sum_{k=1}^{\infty} \frac{(-2\eta)^k}{k} \left[\Delta_\alpha^k + (-\bar{\Delta}_\beta)^k - \Delta_\beta^k - (-\bar{\Delta}_\alpha)^k \right]^2. \tag{E.14}$$

In the next Section we will specialize these results to the case in which the interval $[\alpha, \beta]$ is chosen symmetric.

E.2 Case of a symmetric interval

We consider here the case in which $\alpha = -E$ and $\beta = E$. With this choice, solving Eq. (E.2) gives[4]

$$\Delta_\alpha \bigg|_{a=-E} = \frac{1}{4\eta} \left[-(\varepsilon - iE) \pm \sqrt{8\eta + (\varepsilon - iE)^2} \right] \equiv re^{i\theta}, \tag{E.15}$$

where we choose the positive branch of the square root so that $\mathrm{Re}\,\Delta_\alpha \geq 0$ for any positive η (recall that Δ_α represents the variance of a Gaussian distribution, see Eq. (80)). Similarly, from Eq. (E.2) one finds for the entries of $\hat{C}$ the same symmetries as in Eq. (90). The first two cumulants in Eqs. (E.13) and (E.14) are then found to yield

$$\frac{\kappa_1}{N} = \frac{x}{\pi} \sin 2\theta + \frac{2\theta}{\pi} + \frac{i}{2\pi N} \ln\left(\frac{1 + xe^{2i\theta}}{1 + xe^{-2i\theta}} \right), \tag{E.16}$$

$$\frac{\kappa_2}{N} = \frac{1}{\pi^2 N} \ln\left[1 + \left(\frac{2x \sin 2\theta}{1 - x^2} \right)^2 \right], \tag{E.17}$$

where we called $x \equiv 2\eta r^2$. As a first check, one can easily verify that both $\kappa_1, \kappa_2 \to 0$ in the limit of a vanishing interval $E \to 0$.

By choosing $2\eta = 1$, we obtain in the $N \to \infty$ limit an eigenvalue spectrum distributed within the interval $[-2, 2]$. Sending $\varepsilon \to 0^+$ as prescribed by Eq. (67), one can check that $r \to 1$ (hence $x \to 1$), while

$$\theta \xrightarrow[\varepsilon \to 0^+]{} \theta_0 \equiv \arctan\left(\frac{E}{\sqrt{4 - E^2}} \right) \in \left[-\frac{\pi}{2}, \frac{\pi}{2} \right]. \tag{E.18}$$

This concludes the calculation of κ_1 (see Eq. (E.16)), which includes both the leading order term and its $\mathcal{O}(N^0)$ correction (note that the latter is actually real-valued). However, the second cumulant κ_2 is seen to diverge in the limit $\varepsilon \to 0^+$; the problem is addressed in Ref. [70] by introducing a N-dependent regularization of the infinite sum which appears in Eq. (E.14). Nonetheless, we have shown that such infinite sum (and hence κ_2 itself in Eq. (E.17)) does *not* diverge for any finite value of ε. This hints at the well-known fact that the limit $\varepsilon \to 0^+$ and that for $N \to \infty$ in Eq. (67) are not interchangeable. It is then useful to expand for small ε

$$x = r^2 \simeq 1 - \frac{2\varepsilon}{\sqrt{4 - E^2}}, \tag{E.19}$$

so as to rewrite

$$\kappa_2 \simeq \frac{2}{\pi^2} \ln\left[\frac{\sqrt{4 - E^2} \sin 2\theta}{2\varepsilon} \right]. \tag{E.20}$$

In order to recover the leading order result $\kappa_2 \sim \ln N$ found in previous literature [88–92], one has to assume some type of functional relation $\varepsilon = \varepsilon(N)$, so that the limit $\varepsilon \to 0^+$ is taken by controlling the product εN [89]. This goes however beyond the scope of the present paper.

E.3 Derivation of the scaling function $\chi_{\mathrm{GOE}}(y)$

The prediction for $\chi(E)$ which we have pursued in the previous Section using the replica method is in any case expected to fail if E lies below the mean level spacing, i.e., $E \ll \delta_N$ (as discussed in Section 3.6). In this Appendix, we thus derive in this limit an explicit expression for the level

---

[4]The angle θ should be compared with θ_L in Ref. [70].

compressibility for a pure $N \times N$ GOE matrix (such as the matrix B in Eq. (1)) using an alternative and more standard technique. We denote by $\rho_B(\lambda)$ (defined as in Eq. (7)) the average density of eigenvalues normalized to unity. In the large N limit, $\rho_B(\lambda)$ has a finite support over $[-\sqrt{2N}, \sqrt{2N}]$ and it is given by the standard Wigner semi-circle. We focus here on the number of eigenvalues $I_N[-E, E]$ in a symmetric interval $[-E, E]$ in the bulk of the spectrum, i.e., we choose E of the order of the mean inter-particle spacing $[N\rho_B(0)]^{-1} = O(N^{-1/2})$. The mean number $\langle I_N[-E, E]\rangle$ is easily obtained as

$$\langle I_N[-E, E]\rangle = N \int_{-E}^{E} \rho_B(\lambda)\, d\lambda \approx 2N\rho_B(0)E = 2\tilde{E}, \qquad \tilde{E} = N\rho_B(0)E. \tag{E.21}$$

The variance of $I_N[-E, E]$ has been studied since the pioneering works of Dyson and Mehta [74]. However, an explicit expression for it, valid for any value of E in the bulk, seems hard to find in the literature. This observable was recently revisited in the context of full counting statistics of interacting fermions in Ref. [93], which provides a useful starting point, namely [see Eqs. (31)-(37) therein]

$$\langle (I_N[-E, E])^2 \rangle - \langle I_N[-E, E]\rangle^2 = -N\rho_B(0)^2 \int_{-E}^{E} dx \left(\int_{-\infty}^{-E} dy + \int_{E}^{+\infty} dy \right) C(x, y),$$

$$C(x, y) = -N\rho_B(0)^2 Y_{21}(N\rho_B(0)|x - y|), \tag{E.22}$$

where the "cluster" function $Y_{21}(r)$ is given by [94]

$$Y_{21}(r) = \left(\frac{\sin(\pi r)}{\pi r} \right)^2 - \left(\frac{\mathrm{Si}(\pi r)}{\pi} - \frac{1}{2} \right) \left(\frac{\pi r \cos(\pi r) - \sin(\pi r)}{\pi r^2} \right), \tag{E.23}$$

with $\mathrm{Si}(z) = \int_0^z \sin(t)/t\, dt$ denoting the sine-integral function. By inserting this explicit expression (E.23) in Eq. (E.22) and performing explicitly the integrals over x and y, one obtains for the level compressibility

$$\chi(E) = \frac{\langle (I_N[-E, E])^2 \rangle - \langle I_N[-E, E]\rangle^2}{\langle I_N[-E, E]\rangle} \sim \chi_{\mathrm{GOE}}(y = N\rho_B(0)E), \tag{E.24}$$

where the function $\chi_{\mathrm{GOE}}(y)$ is given by

$$\chi_{\mathrm{GOE}}(y) = \frac{1}{2\pi^2 y} \Big\{ [\mathrm{Si}(2\pi y)]^2 - 2\,\mathrm{Ci}(4\pi y) - \pi\,\mathrm{Si}(2\pi y) \tag{E.25}$$

$$+ 2\Big[-4\pi y\,\mathrm{Si}(4\pi y) + 2\pi^2 y + \log(4\pi y) - \cos(4\pi y) + \gamma_E + 1 \Big] \Big\},$$

where $\mathrm{Ci}(z) = -\int_z^{\infty} \cos(t)/t\, dt$ is the cosine integral function and $\gamma_E = 0.577216\ldots$ is the Euler-gamma constant. Its asymptotic behaviors are given by

$$\chi_{\mathrm{GOE}}(y) = \begin{cases} 1 - 2y + O(y^2), & y \to 0 \\ \dfrac{\ln y}{\pi^2 y} + O(1/y), & y \to \infty. \end{cases} \tag{E.26}$$

References

[1] P. W. Anderson, *Absence of diffusion in certain random lattices*, Phys. Rev. **109**, 1492 (1958), doi:10.1103/PhysRev.109.1492.

[2] F. Wegner, *Inverse participation ratio in 2+ ε dimensions*, Z. Phys. B Cond. Mat. **36**(3), 209 (1980), doi:10.1007/BF01325284.

[3] A. Rodriguez, L. J. Vasquez, K. Slevin and R. A. Römer, *Multifractal finite-size scaling and universality at the Anderson transition*, Phys. Rev. B **84**, 134209 (2011), doi:10.1103/PhysRevB.84.134209.

[4] M. Srednicki, *Chaos and quantum thermalization*, Phys. Rev. E **50**, 888 (1994), doi:10.1103/PhysRevE.50.888.

[5] M. Rigol, V. Dunjko and M. Olshanii, *Thermalization and its mechanism for generic isolated quantum systems*, Nature **452**, 854 (2008), doi:10.1038/nature06838.

[6] N. Macé, F. Alet and N. Laflorencie, *Multifractal scalings across the many-body localization transition*, Phys. Rev. Lett. **123**, 180601 (2019), doi:10.1103/PhysRevLett.123.180601.

[7] G. De Tomasi, I. M. Khaymovich, F. Pollmann and S. Warzel, *Rare thermal bubbles at the many-body localization transition from the fock space point of view*, Phys. Rev. B **104**, 024202 (2021), doi:10.1103/PhysRevB.104.024202.

[8] I. Gornyi, A. Mirlin, D. Polyakov and A. Burin, *Spectral diffusion and scaling of many-body delocalization transitions*, Ann. Phys. **529**, 1600360 (2017), doi:10.1002/andp.201600360.

[9] M. Tarzia, *Many-body localization transition in hilbert space*, Phys. Rev. B **102**, 014208 (2020), doi:10.1103/PhysRevB.102.014208.

[10] D. M. Basko, I. L. Aleiner and B. L. Altshuler, *Metal–insulator transition in a weakly interacting many-electron system with localized single-particle states*, Ann. Phys. **321**, 1126 (2006), doi:10.1016/j.aop.2005.11.014.

[11] B. L. Altshuler, Y. Gefen, A. Kamenev and L. S. Levitov, *Quasiparticle lifetime in a finite system: A nonperturbative approach*, Phys. Rev. Lett. **78**, 2803 (1997), doi:10.1103/PhysRevLett.78.2803.

[12] L. Faoro, M. V. Feigel'man and L. Ioffe, *Non-ergodic extended phase of the quantum random energy model*, Ann. Phys. **409**, 167916 (2019), doi:10.1016/j.aop.2019.167916.

[13] C. L. Baldwin and C. R. Laumann, *Quantum algorithm for energy matching in hard optimization problems*, Phys. Rev. B **97**, 224201 (2018), doi:10.1103/PhysRevB.97.224201.

[14] G. Biroli, D. Facoetti, M. Schiró, M. Tarzia and P. Vivo, *Out-of-equilibrium phase diagram of the quantum random energy model*, Phys. Rev. B **103**, 014204 (2021), doi:10.1103/PhysRevB.103.014204.

[15] T. Parolini and G. Mossi, *Multifractal dynamics of the QREM*, doi:10.48550/arxiv.2007.00315 (2020).

[16] K. Kechedzhi, V. Smelyanskiy, J. R. McClean, V. S. Denchev, M. Mohseni, S. Isakov, S. Boixo, B. Altshuler and H. Neven, *Efficient population transfer via non-ergodic extended states in quantum spin glass*, doi:10.48550/arxiv.1807.04792 (2018).

[17] M. Pino, V. E. Kravtsov, B. L. Altshuler and L. B. Ioffe, *Multifractal metal in a disordered josephson junctions array*, Phys. Rev. B **96**, 214205 (2017), doi:10.1103/PhysRevB.96.214205.

[18] M. Pino, L. B. Ioffe and B. L. Altshuler, *Nonergodic metallic and insulating phases of Josephson junction chains*, Proc. Natl. Acad. Sci. USA **113**, 536 (2016), doi:10.1073/pnas.1520033113.

[19] V. N. Smelyanskiy, K. Kechedzhi, S. Boixo, S. V. Isakov, H. Neven and B. Altshuler, *Nonergodic delocalized states for efficient population transfer within a narrow band of the energy landscape*, Phys. Rev. X **10**, 011017 (2020), doi:10.1103/PhysRevX.10.011017.

[20] T. Micklitz, F. Monteiro and A. Altland, *Nonergodic extended states in the Sachdev-Ye-Kitaev model*, Phys. Rev. Lett. **123**, 125701 (2019), doi:10.1103/PhysRevLett.123.125701.

[21] F. Monteiro, T. Micklitz, M. Tezuka and A. Altland, *Minimal model of many-body localization*, Phys. Rev. Research **3**, 013023 (2021), doi:10.1103/PhysRevResearch.3.013023.

[22] N. Rosenzweig and C. E. Porter, *Repulsion of energy levels in complex atomic spectra*, Phys. Rev. **120**, 1698 (1960), doi:10.1103/PhysRev.120.1698.

[23] V. E. Kravtsov, I. M. Khaymovich, E. Cuevas and M. Amini, *A random matrix model with localization and ergodic transitions*, New J. Phys. **17**, 122002 (2015), doi:10.1088/1367-2630/17/12/122002.

[24] P. von Soosten and S. Warzel, *Non-ergodic delocalization in the Rosenzweig–Porter model*, Lett. Math. Phys. **109**, 905 (2019), doi:10.1007/s11005-018-1131-7.

[25] D. Facoetti, P. Vivo and G. Biroli, *From non-ergodic eigenvectors to local resolvent statistics and back: A random matrix perspective*, Europhys. Lett. **115**, 47003 (2016), doi:10.1209/0295-5075/115/47003.

[26] K. Truong and A. Ossipov, *Eigenvectors under a generic perturbation: Non-perturbative results from the random matrix approach*, Europhys. Lett. **116**, 37002 (2016), doi:10.1209/0295-5075/116/37002.

[27] E. Bogomolny and M. Sieber, *Eigenfunction distribution for the Rosenzweig-Porter model*, Phys. Rev. E **98**, 032139 (2018), doi:10.1103/PhysRevE.98.032139.

[28] G. D. Tomasi, M. Amini, S. Bera, I. M. Khaymovich and V. E. Kravtsov, *Survival probability in Generalized Rosenzweig-Porter random matrix ensemble*, SciPost Phys. **6**, 14 (2019), doi:10.21468/SciPostPhys.6.1.014.

[29] M. Amini, *Spread of wave packets in disordered hierarchical lattices*, Europhy. Lett. **117**, 30003 (2017), doi:10.1209/0295-5075/117/30003.

[30] M. Pino, J. Tabanera and P. Serna, *From ergodic to non-ergodic chaos in Rosenzweig-Porter model*, J. Phys. A - Mat. Theor. **52**, 475101 (2019), doi:10.1088/1751-8121/ab4b76.

[31] R. Berkovits, *Super-Poissonian behavior of the Rosenzweig-Porter model in the nonergodic extended regime*, Phys. Rev. B **102**, 165140 (2020), doi:10.1103/PhysRevB.102.165140.

[32] A. Pandey, *Brownian-motion model of discrete spectra*, Chaos Soliton. Frac. **5**, 1275 (1995), doi:10.1016/0960-0779(94)E0065-W.

[33] T. Guhr, *Transition from Poisson regularity to chaos in a time-reversal noninvariant system*, Phys. Rev. Lett. **76**, 2258 (1996), doi:10.1103/PhysRevLett.76.2258.

[34] T. Guhr and A. Müller-Groeling, *Spectral correlations in the crossover between GUE and Poisson regularity: on the identification of scales*, J. Math. Phys. **38**, 1870 (1997), doi:10.1063/1.531918.

[35] E. Brézin and S. Hikami, *Correlations of nearby levels induced by a random potential*, Nucl. Phys. B **479**, 697 (1996), doi:10.1016/0550-3213(96)00394-X.

[36] H. Kunz and B. Shapiro, *Transition from Poisson to Gaussian unitary statistics: the two-point correlation function*, Phys. Rev. E **58**, 400 (1998), doi:10.1103/PhysRevE.58.400.

[37] A. Altland, M. Janssen and B. Shapiro, *Perturbation theory for the Rosenzweig-Porter matrix model*, Phys. Rev. E **56**, 1471 (1997), doi:10.1103/PhysRevE.56.1471.

[38] A. Krajenbrink, P. Le Doussal and N. O'Connell, *Tilted elastic lines with columnar and point disorder, non-Hermitian quantum mechanics, and spiked random matrices: pinning and localization*, Phys. Rev. E **103**, 042120 (2021), doi:10.1103/PhysRevE.103.042120.

[39] P. Mergny and S. N. Majumdar, *Stability of large complex systems with heterogeneous relaxation dynamics*, J. Stat. Mech.: Theor. Exp. **2021**, 123301 (2021), doi:10.1088/1742-5468/ac3b47.

[40] M. Mézard, G. Parisi and M. Virasoro, *Spin Glass Theory and Beyond*, Lecture Notes in Physics Series. World Scientific, ISBN 9789971501150 (1987).

[41] H. Ikeda, *Bose-Einstein-Like condensation of deformed random matrix: a replica approach*, doi:10.48550/arxiv.2208.01848 (2022).

[42] T. Claeys, A. B. J. Kuijlaars, K. Liechty and D. Wang, *Propagation of singular behavior for Gaussian perturbations of random matrices*, Commun. Math. Phys. **362**, 1 (2018), doi:10.1007/s00220-018-3195-8.

[43] G. Livan, M. Novaes and P. Vivo, *Introduction to Random Matrices*, Springer International Publishing, doi:10.1007/978-3-319-70885-0 (2018).

[44] B. Altshuler and B. Shklovskii, *Repulsion of energy levels and conductivity of small metal samples*, Sov. Phys. JETP **64**, 127 (1986), http://www.jetp.ras.ru/cgi-bin/dn/e_064_01_0127.pdf.

[45] E. Cuevas and V. E. Kravtsov, *Two-eigenfunction correlation in a multifractal metal and insulator*, Phys. Rev. B **76**, 235119 (2007), doi:10.1103/PhysRevB.76.235119.

[46] V. E. Kravtsov, I. M. Khaymovich, B. L. Altshuler and L. B. Ioffe, *Localization transition on the Random Regular Graph as an unstable tricritical point in a log-normal Rosenzweig-Porter random matrix ensemble*, doi:10.48550/arxiv.2002.02979 (2020).

[47] I. M. Khaymovich, V. E. Kravtsov, B. L. Altshuler and L. B. Ioffe, *Fragile extended phases in the log-normal Rosenzweig-Porter model*, Phys. Rev. Res. **2**, 043346 (2020), doi:10.1103/PhysRevResearch.2.043346.

[48] C. Monthus, *Multifractality of eigenstates in the delocalized non-ergodic phase of some random matrix models: Wigner–Weisskopf approach*, J. Phys. A - Mat. Theor. **50**, 295101 (2017), doi:10.1088/1751-8121/aa77e1.

[49] G. Biroli and M. Tarzia, *Lévy-Rosenzweig-Porter random matrix ensemble*, Phys. Rev. B **103**, 104205 (2021), doi:10.1103/PhysRevB.103.104205.

[50] W. Buijsman and Y. Bar Lev, *Circular rosenzweig-porter random matrix ensemble*, SciPost Phys. **12**, 082 (2022), doi:10.21468/SciPostPhys.12.3.082.

[51] I. M. Khaymovich and V. E. Kravtsov, *Dynamical phases in a "multifractal" Rosenzweig-Porter model*, SciPost Phys. **11**, 045 (2021), doi:10.21468/SciPostPhys.11.2.045.

[52] M. Sarkar, R. Ghosh, A. Sen and K. Sengupta, *Mobility edge and multifractality in a periodically driven Aubry-André model*, Phys. Rev. B **103**, 184309 (2021), doi:10.1103/PhysRevB.103.184309.

[53] S. Roy, I. M. Khaymovich, A. Das and R. Moessner, *Multifractality without fine-tuning in a Floquet quasiperiodic chain*, SciPost Phys. **4**, 025 (2018), doi:10.21468/SciPostPhys.4.5.025.

[54] J. Wang, X.-J. Liu, G. Xianlong and H. Hu, *Phase diagram of a non-Abelian Aubry-André-Harper model with p-wave superfluidity*, Phys. Rev. B **93**, 104504 (2016), doi:10.1103/PhysRevB.93.104504.

[55] P. A. Nosov, I. M. Khaymovich and V. E. Kravtsov, *Correlation-induced localization*, Phys. Rev. B **99**, 104203 (2019), doi:10.1103/PhysRevB.99.104203.

[56] A. Duthie, S. Roy and D. E. Logan, *Anomalous multifractality in quantum chains with strongly correlated disorder*, Phys. Rev. B **106**, L020201 (2022), doi:10.1103/PhysRevB.106.L020201.

[57] A. G. Kutlin and I. M. Khaymovich, *Emergent fractal phase in energy stratified random models*, SciPost Phys. **11**, 101 (2021), doi:10.21468/SciPostPhys.11.6.101.

[58] V. Motamarri, A. S. Gorsky and I. M. Khaymovich, *Localization and fractality in disordered Russian Doll model*, doi:10.48550/arxiv.2112.05066 (2021).

[59] W. Tang and I. M. Khaymovich, *Non-ergodic delocalized phase with Poisson level statistics*, Quantum **6**, 733 (2022), doi:10.22331/q-2022-06-09-733.

[60] M. Tarzia, *Fully localized and partially delocalized states in the tails of Erdös-Rényi graphs in the critical regime*, Phys. Rev. B **105**, 174201 (2022), doi:10.1103/PhysRevB.105.174201.

[61] D. V. Voiculescu, K. J. Dykema and A. Nica, *Free random variables*, American Mathematical Soc. (1992).

[62] A. Zee, *Law of addition in random matrix theory*, Nucl. Phys. B **474**, 726 (1996), doi:10.1016/0550-3213(96)00276-3.

[63] U. Fano, *Ionization yield of radiations. II. the fluctuations of the number of ions*, Phys. Rev. **72**, 26 (1947), doi:10.1103/PhysRev.72.26.

[64] S. N. Majumdar, P. von Bomhard and J. Krug, *Exactly solvable record model for rainfall*, Phys. Rev. Lett. **122**, 158702 (2019), doi:10.1103/PhysRevLett.122.158702.

[65] S. N. Majumdar, P. Mounaix and G. Schehr, *Universal record statistics for random walks and Lévy flights with a nonzero staying probability*, J. Phys. A - Math. Theor. **54**, 315002 (2021), doi:10.1088/1751-8121/ac0a2f.

[66] S. F. Edwards and R. C. Jones, *The eigenvalue spectrum of a large symmetric random matrix*, J. Phys. A - Math. Gen. **9**, 1595 (1976), doi:10.1088/0305-4470/9/10/011.

[67] J.-P. Bouchaud, *Random matrix theory and (big) data analysis*, In *Stochastic Processes and Random Matrices: Lecture Notes of the Les Houches Summer School: Volume 104, July 2015*. Oxford University Press, ISBN 9780198797319, doi:10.1093/oso/9780198797319.003.0006 (2017).

[68] G. Parisi, *Statistical Field Theory*, Addison-Wesley (1988).

[69] F. L. Metz and I. Pérez Castillo, *Large deviation function for the number of eigenvalues of sparse random graphs inside an interval*, Phys. Rev. Lett. **117**, 104101 (2016), doi:10.1103/PhysRevLett.117.104101.

[70] F. L. Metz, *Replica-symmetric approach to the typical eigenvalue fluctuations of Gaussian random matrices*, J. Phys. A - Math. Theor. **50**, 495002 (2017), doi:10.1088/1751-8121/aa94f8.

[71] D. H. (https://math.stackexchange.com/users/55051/david h), *Asymptotic expansion of integral involving an ArcTan*, Mathematics Stack Exchange, https://math.stackexchange.com/q/4523130.

[72] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Elsevier/Academic Press, Amsterdam, 7th edn., ISBN 978-0-12-373637-6 (2007).

[73] F. J. Dyson, *Statistical Theory of the Energy Levels of Complex Systems. I*, J. Math. Phys. **3**, 140 (1962), doi:10.1063/1.1703773.

[74] F. J. Dyson and M. L. Mehta, *Statistical Theory of the Energy Levels of Complex Systems. IV*, J. Math. Phys. **4**, 701 (1963), doi:10.1063/1.1704008.

[75] M. L. Mehta, *Random Matrices*, Academic Press, New York, 3rd edn. (2004).

[76] P. J. Forrester, *Log-Gases and Random Matrices (LMS-34)*, London Mathematical Society Monographs. Princeton University Press, ISBN 9781400835416 (2010).

[77] A. D. Mirlin, *Statistics of energy levels and eigenfunctions in disordered systems*, Phys. Rep. **326**, 259 (2000), doi:10.1016/S0370-1573(99)00091-5.

[78] R. Marino, S. N. Majumdar, G. Schehr and P. Vivo, *Phase transitions and edge scaling of number variance in Gaussian random matrices*, Phys. Rev. Lett. **112**, 254101 (2014), doi:10.1103/PhysRevLett.112.254101.

[79] R. Marino, S. N. Majumdar, G. Schehr and P. Vivo, *Number statistics for β-ensembles of random matrices: Applications to trapped fermions at zero temperature*, Phys. Rev. E **94**, 032115 (2016), doi:10.1103/PhysRevE.94.032115.

[80] K. S. Tikhonov and A. D. Mirlin, *Statistics of eigenstates near the localization transition on random regular graphs*, Phys. Rev. B **99**, 024202 (2019), doi:10.1103/PhysRevB.99.024202.

[81] G. B. Arous and A. Guionnet, *Wigner matrices*, The Oxford Handbook of Random Matrix Theory pp. 433–451 (2011).

[82] M. Potters and J.-P. Bouchaud, *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*, Cambridge University Press (2020).

[83] G. Semerjian and L. F. Cugliandolo, *Sparse random matrices: the eigenvalue spectrum revisited*, J. Phys. A - Math. Gen. **35**, 4837 (2002), doi:10.1088/0305-4470/35/23/303.

[84] T. Rogers, I. Pérez Castillo, R. Kühn and K. Takeda, *Cavity approach to the spectral density of sparse symmetric random matrices*, Phys. Rev. E **78**, 031116 (2008), doi:10.1103/PhysRevE.78.031116.

[85] F. L. Metz, G. Parisi and L. Leuzzi, *Finite-size corrections to the spectrum of regular random graphs: An analytical solution*, Phys. Rev. E **90**, 052109 (2014), doi:10.1103/PhysRevE.90.052109.

[86] D. Voiculescu, *Limit laws for random matrices and free products*, Invent. Math. **104**(1), 201 (1991), doi:10.1007/BF01245072.

[87] P. Biane, *On the free convolution with a semi-circular distribution*, Indiana U. Math. J. **46**, 705 (1997), https://hal.archives-ouvertes.fr/hal-00536164.

[88] F. J. Dyson, *Statistical theory of the energy levels of complex systems. III*, J. Math. Phys. **3**, 166 (1962), doi:10.1063/1.1703775.

[89] A. Cavagna, J. P. Garrahan and I. Giardina, *Index distribution of random matrices with an application to disordered systems*, Phys. Rev. B **61**, 3960 (2000), doi:10.1103/PhysRevB.61.3960.

[90] S. N. Majumdar, C. Nadal, A. Scardicchio and P. Vivo, *Index distribution of Gaussian random matrices*, Phys. Rev. Lett. **103**, 220603 (2009), doi:10.1103/PhysRevLett.103.220603.

[91] S. N. Majumdar, C. Nadal, A. Scardicchio and P. Vivo, *How many eigenvalues of a Gaussian random matrix are positive?*, Phys. Rev. E **83**, 041105 (2011), doi:10.1103/PhysRevE.83.041105.

[92] I. Pérez Castillo, *Spectral order statistics of Gaussian random matrices: Large deviations for trapped fermions and associated phase transitions*, Phys. Rev. E **90**, 040102 (2014), doi:10.1103/PhysRevE.90.040102.

[93] N. R. Smith, P. Le Doussal, S. N. Majumdar and G. Schehr, *Full counting statistics for interacting trapped fermions*, SciPost Phys. **11**, 110 (2021), doi:10.21468/SciPostPhys.11.6.110.

[94] A. Pandey, *Statistical properties of many-particle spectra: III. Ergodic behavior in random-matrix ensembles*, Ann. Phys. **119**, 170 (1979), doi:10.1016/0003-4916(79)90254-9.