Technische Universität München
Zentrum Mathematik
Analysis und ihre Anwendungen

Hierarchical Random Matrices and Operators

Per von Soosten

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzende: Prof. Dr. Nina Gantert

Prüfer der Dissertation:

1. Prof. Dr. Simone Warzel
2. Prof. Dr. Herbert Spohn
3. Prof. Dr. László Erdős (nur schriftliche Beurteilung)

Die Dissertation wurde am 18.01.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 23.03.2018 angenommen.

This thesis is dedicated to Robi.

Acknowledgements

I could never have completed this project without the invaluable mentoring of Simone Warzel, who I thank for exercising endless patience, lending me her wisdom and experience, showing me how to think about mathematics, and reminding me that it's not rocket science at all the right moments. I offer my sincere gratitude to both the chair and the examining committee for organizing my defense taking the time to read this thesis. In the grander scheme of things, I would like to thank my family for supporting me unconditionally, my friends for great entertainment, and my previous teachers for getting me here. Finally, this work would not have happened without the generous financial support of the DFG.

Summary

This thesis is a conglomerate of our previous articles [89–92] studying three Dyson-hierarchical Hamiltonians with on-site disorder. The first of these, the hierarchical Anderson model, perturbs Dyson's hierarchical Laplacian by a random potential. We apply Feshbach-Krein-Schur renormalization techniques to establish a criterion on the single-site distribution which ensures exponential dynamical localization as well as positive inverse participation ratios and Poisson statistics of eigenvalues. Our criterion applies to all cases of exponentially decaying hierarchical hopping strengths and holds even for spectral dimension $d > 2$, which corresponds to the regime of transience of the underlying hierarchical random walk. This challenges recent numerical findings that the spectral dimension is significant as far as the Anderson transition is concerned.

Next, we study the ultrametric ensemble comprising a hierarchical analogue of power-law random band matrices. This symmetric ensemble consists of random matrices with independent entries whose variances decay exponentially in the metric induced by the tree topology on $\mathbb{N}$. We map out the entirety of the localization regime by proving the localization of eigenfunctions and Poisson statistics of the suitably scaled eigenvalues. Our results complement existing works on complete delocalization and random matrix universality, thereby proving the existence of a phase transition in this model. Along the way, we establish optimal stability results for the resolvent under Dyson Brownian motion up to times of order N^{-1} when the complex energy parameter is of order N^{-1}. These results go beyond norm-based continuity arguments for Dyson Brownian motion and complement existing proofs of equilibration of the local statistics for times greater than N^{-1}.

Finally, we consider the Rosenzweig-Porter model $H = V + \sqrt{T}\,\Phi$, where V is a $N \times N$ diagonal matrix, Φ is drawn from the $N \times N$ Gaussian Orthogonal Ensemble, and $N^{-1} \ll T \ll 1$. We prove that the eigenfunctions of H are typically supported in a set of approximately NT sites, thereby confirming the existence of a previously conjectured non-ergodic delocalized phase. Our proof is based on martingale estimates along the characteristic curves of the stochastic advection equation satisfied by the local resolvent of the Brownian motion representation of H.

Contents

1 Introduction

1.1 Localization Transitions

In this introductory section, we first provide some general context to localization problems in terms of the famous Anderson model [10]. We then give a heuristic picture of a possible mechanism behind the localization transition and briefly discuss two prominent toy models as motivation for the hierarchical approximations featured in this thesis. The discussion in this section is strongly influenced by the book [6], which should be kept in mind as a default reference when no others are given.

The Anderson model was designed to portray the dynamics of a single particle in a medium with impurities, which are described microscopically by a random potential energy at each location in space. This leads to a potential term in the relevant Hamiltonian which acts as multiplication by a function $V(x)$ constructed from classical random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The situation may be somewhat simplified by performing the tight-binding approximation, which amounts to reducing the physical space to a discrete lattice, insisting on a single state per site, and ignoring the interaction of the particle with the environment. The resulting Hamiltonian is defined on $\ell^2(\mathbb{Z}^d)$ as

$$H = -\varepsilon\Delta + V \tag{1.1.1}$$

where $\varepsilon > 0$ is a parameter governing the relative strengths of the kinetic and potential terms. The kinetic term is then the local averaging operator

$$(-\Delta\psi)(x) = \frac{1}{2d}\sum_{y\sim x}\psi(y),$$

where $y \sim x$ if y is a neighbor of x on the lattice $\mathbb{Z}^d$. Hence, Δ agrees with the finite difference Laplacian up to an affine transformation of the spectrum. Regarding the potential term, it is simplest to assume that its values at each site $x \in \mathbb{Z}^d$ are independent random variables $\{V_x\}$, resulting in a random multiplication operator defined by

$$(V\psi)(x) = V_x\psi(x).$$

A fundamental issue is how restricted the motion of a particle is under the dynamics generated by typical realizations of H. In particular, the question asked by Anderson is whether it is true with high classical probability that a particle started at some site $x \in \mathbb{Z}^d$ remains near x for all times with high quantum probability. Since the answer to this question might depend on the energy, one may ask the same question of particles subject to the filtered dynamics $P_I(H)e^{-itH}$, where $P_I(H)$ denotes the spectral projection onto the interval $I \subset \mathbb{R}$. In mathematical terms, the problem is whether

$$\sup_{t\in\mathbb{R}} |\langle \delta_y, P_I(H)e^{-itH}\delta_x\rangle|^2 \tag{1.1.2}$$

decays in $|x-y|$ for typical realizations of V, where δ_x denotes the canonical site-basis element

$$\delta_x(u) = \begin{cases} 1 & u = x \\ 0 & u \neq x \end{cases}.$$

The question can be answered by explicit diagonalization in the extreme cases $\varepsilon = 0$ with $H = V$ or "$\varepsilon = \infty$" with $H = \Delta$. In the first case, V is diagonalized by the basis $\{\delta_x\}$ which, up to phases, remains invariant under the evolution e^{itV} and therefore remains perfectly localized. In the second case, H exhibits localization in momentum space by the same token, and therefore exhibits delocalization in physical space. Hence, the question becomes how H interpolates these two extreme behaviors as ε varies in $(0, \infty)$. The emerging conjecture currently held by physicists is the following:

1. H exhibits localization whenever ε is small or I is near the edge of the spectrum,

2. H exhibits localization throughout its entire spectrum for all $\varepsilon > 0$ in dimension $d \leq 2$, and

3. H exhibits delocalization when I is in the bulk of the spectrum, ε is sufficiently large, and $d \geq 3$.

Mathematically rigorous derivations of any of these statements have proven difficult and cannot be achieved by filling the physical arguments with ε-s and δ-s. A major mathematical effort in the final quarter of the twentieth century succeeded in proving both the first point and that H is localized when $d = 1$. As particularly noteworthy in this regard, let us mention the works of Goldsheid-Molchanov-Pastur [53], Carmona-Klein-Martinelli [24], Fröhlich-Spencer [47], Simon-Wolff [84], and Aizenman-Molchanov [1]. The remaining questions are wide open. The author of this thesis is unaware of even the smallest bit of progress concerning the other two points for the lattice Anderson model.

The dynamical localization question (1.1.2) is of course intimately linked to the study of the spectral measures of H. Indeed, if the spectral measures are of pure-point type and the corresponding eigenfunctions are bounded by a profile

$$|\psi_\lambda(x)|^2 \leq C_\lambda f(|x - x_\lambda|),$$

then every state may be arbitrarily well approximated by a bound state for which the quantum probability of ever moving outside some finite set decays like f. In infinite volumes the mere fact that the spectral measure μ_x of δ_x for H is of pure-point type implies that

$$\lim_{R\to\infty}\sup_{t\in\mathbb{R}}\sum_{|x-y|>R}|\langle\delta_y, e^{-itH}\delta_x\rangle|^2 = 0$$

via the RAGE theorem (see, for example, [6]). Even stronger quantitative localization statements are produced if the spectral measures μ_{xy} for δ_x and δ_y decay in total-variation norm since

$$\sup_{t\in\mathbb{R}}|\langle\delta_y, P_I(H)e^{-itH}\delta_x\rangle|^2 \leq \sup_{t\in\mathbb{R}}\left|\int_I e^{-itE}\,\mu_{xy}(dE)\right| \leq |\mu_{xy}|(I).$$

The eigenfunctions of H are in turn linked to the Green functions

$$G(x, y; z) = \langle\delta_y, (H - z)^{-1}\delta_x\rangle,$$

which, in finite volumes, exist also for almost every $z = E \in \mathbb{R}$. The function $g(y) = G(x, y; E)$ satisfies

$$\left(H - \frac{1}{G(x, x; E)}|\delta_x\rangle\langle\delta_x|\right) g = Eg,$$

which shows that, when V is random and E is chosen judiciously, g is an eigenfunction corresponding to a resampled potential. One may thus hope to deduce the eigenfunction profiles from a profile for G. The Green function satisfies a formal path-expansion of the form

$$G(x, y; E) = \sum_{\gamma:x\to y}(-1)^{|\gamma|}\prod_{k=0}^{|\gamma|}\frac{1}{V(\gamma(k)) - E}, \tag{1.1.3}$$

the sum ranging over all finite paths from x to y. This provides some intuition as to the dependence of the localization transition on the geometry of the lattice and, in particular, makes it plausible that the Green function may not decay in high dimensions where there are many more paths joining any two sites. The main point of (1.1.3) is that high dimensions allow for a much larger number of potential combinations $\{V(x) : x \in \gamma\}$ resonating at a given energy E despite the interference of the randomness in V.

A less direct way of probing for localization is to consider the microscopic eigenvalue statistics defined by the random point process

$$\mu_N = \sum_{\lambda\in\sigma(H_N)}\delta_{|Q_N|(\lambda-E)},$$

where

$$H_N = 1_{Q_N} H 1_{Q_N}$$

are the restrictions of H to some appropriately increasing finite volumes Q_N. Since the canonical spacing between adjacent eigenvalues of H_N is approximately $|Q_N|^{-1}$, the measure μ_N captures the statistics of individual eigenvalues near $E \in \mathbb{R}$. If the eigenfunctions corresponding to eigenvalues close to E are localized, then the change of basis

$$U H_N U^* = \mathrm{diag}(\lambda_1, ..., \lambda_N)$$

can act only locally in space for energies close to E. This means that the eigenvalues near E inherit some independence from the potential V and makes it quite natural to expect that μ_N might converge to the universal limiting structure of asymptotically independent point processes, the Poisson point process. Which local statistics should be expected in delocalized regimes is a-priori much less clear, except that one would expect delocalized states to induce strong correlations in the spectrum. Nevertheless, it is believed that a large class of sufficiently complex quantum systems share a universal limiting behavior of the spectrum, which depends only on macrosopic parameters such as the symmetry class of the Hamiltonian. It is not completely unreasonable to suppose that delocalized random Schrödinger operators might also fall into the domain of these so called Wigner-Dyson-Mehta (WDM) statistics. While it is known that sufficiently strong localization bounds imply Poisson level statistics [69], a proof of WDM statistics remains obscure. There has however been spectacular progress in proving the emergence of WDM statistics in a variety of Hamiltonians whose randomness is equally distributed among the off-diagonal entries (see [40] and references therein).

In recent years, the realization that the rigorous understanding of the Anderson transition seems to be far beyond the currently available mathematical machinery has inspired the task of constructing simplified renditions of the transition in analytically tractable toy models. A prominent example, inspired by the successful analysis of the one-dimensional Anderson model, is the random Schödinger operator whose kinetic term consists of nearest-neighbor hopping on the Bethe lattice (the loopless graph of constant degree). For this model, the localization transition was established rigorously in [3, 57]. Because of the infinite branching from any given vertex, this model is sometimes interpreted as the infinite-dimensional Anderson model. We note however, that this is in stark contrast to the intuition provided by (1.1.3), since the Anderson model on the Bethe lattice is tractable precisely because there is only one self-avoiding walk between any two sites. It may be more appropriate to interpret the entire graph as a single infinite-dimensional cell and to note that the delocalization in the cell is due to tunneling enabled by the exponential growth of the volume in terms of the diameter. A further warning that this model may not fully capture the Anderson transition is given by the fact that the level statistics converge to a Poisson point process [4]. The closely related random regular graph, in which loops are also scarce, has also been studied [12, 13].

An alternative approach has been to effectively increase the number of paths from x to y on the one-dimensional lattice by studying Hamiltonians H on $\ell^2(\mathbb{Z})$ for which $\langle \delta_y, H\delta_x \rangle \neq 0$ also when $|x-y| > 1$. The hope is that this long-range hopping can induce a localization transition if $|\langle \delta_y, H\delta_x \rangle|$ decays sufficiently slowly. The random band matrices [25, 49] provide a central model of this type by letting $H_N : \ell^2(\{1, ..., N\}) \to \ell^2(\{1, ..., N\})$ consist of centered random variables with variance

$$\mathbb{E}\,|\langle \delta_y, H_N\delta_x \rangle|^2 = \begin{cases} W_N^{-1} & \text{if } |x-y| \leq W_N \\ 0 & \text{else} \end{cases}.$$

The localization transition is then conjectured to be governed by the size of W_N, with H_N being localized for $W_N \ll N^{1/2}$ and delocalized otherwise. While these models do not suffer from the defect outlined in the previous paragraph, they lose the analytic simplicity of the Bethe lattice with respect to path expansions. As a consequence, the rigorous results [19, 32, 40, 41, 80, 85] notwithstanding, it has been difficult to derive the behavior of H_N near the critical bandwidth $W_N = N^{1/2}$.

The hierarchical models featured in this thesis are attempts at further compromises between the complexities allowing for a real localization transition and the simplicity allowing for mathematical proofs. They are obtained by a performing a "coarse-graining", which tries to recursively compensate the removal of boundary conditions by adding small mean-field hopping components as follows. The configuration space of the hierarchical systems will be the natural numbers $\mathbb{N}_0$, on which we define nested partitions $\{\mathcal{P}_r\}$ by

$$\mathbb{N}_0 = \{0, ..., 2^r - 1\} \cup \{2^r, ..., 2 \cdot 2^r - 1\} \cup ...$$

It is often useful to describe this structure in terms of the hierarchical metric

$$d(x, y) = \min\{r \geq 0 : x, y \text{ contained in a common member of } \mathcal{P}_r\},$$

which encodes $\mathcal{P}_r$ as the closed balls of radius r. In one dimension, the idea is to mimick the nearest-neighbor averaging

$$(-\Delta\psi)(x) = \frac{1}{2}\left(\psi(x-1) + \psi(x+1)\right)$$

in the following fashion. First, we average ψ locally over the smallest non-trivial partition $\mathcal{P}_1$ with

$$(\Delta_1\psi)(x) = p_1 E_1\psi(x) = \frac{p_1}{2} \sum_{d(x,y)\leq 1} \psi(y)$$

thus removing the boundary conditions along $\mathcal{P}_1$. Next, we attempt to correct the error by adding the average of ψ on a coarser scale

$$(\Delta_2\psi)(x) = ((p_1E_1 + p_2E_2)\psi)(x) = (E_1\psi)(x) + \frac{p_2}{4}\sum_{d(x,y)\leq 2}\psi(y)$$

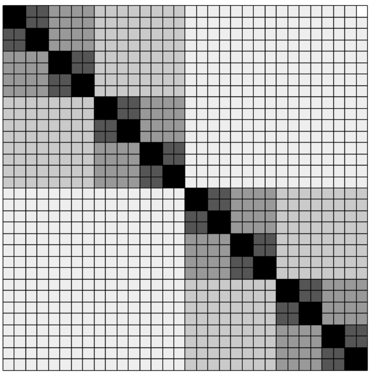
thereby (very roughly) compensating for the missing interactions between the sites on the boundary of member of $\mathcal{P}_1$ but on the interior of members of $\mathcal{P}_2$. Continuing in this way, we obtain the hierarchical Laplacian

$$\Delta_H = \sum_{r=1}^{\infty} p_r E_r \tag{1.1.4}$$

where

$$(E_r\psi)(x) = 2^{-r}\sum_{d(x,y)\leq r}\psi(y)$$

is the local averaging operator on scale 2^r and the coefficients $\{p_r\} \in \ell^1$ ensure the convergence of the sum. Hence, Δ_H is a sum of "layers" indexed by $r \geq 1$, each of which acts only locally on the length scale 2^r and is not subject to any constraints along the boundaries of the members of $\mathcal{P}_r$. Of course, it is also possible to repeat this procedure in higher dimensions, which results in partitions of $\mathbb{Z}^d$ into nested hypercubes of cardinality 2^{dr}. However, the hierarchical approximation can then only account for the dimension in terms of the faster growth of the partitions $\mathcal{P}_r$, all other geometric information being lost. Thus, we do not really obtain a new family of hierarchical Laplacians in higher dimensions as they can be obtained from (1.1.4) by setting $p_r = 0$ when r is not a multiple of d.



The Hierarchical Laplacian

Since they were introduced to the Ising model by Dyson [38,39], hierarchical approximations have enjoyed a long history during which they have consistently reproduced qualitative features when inserted into the central models of statistical physics. This has led to rigorous analysis of, among others, the hierarchical versions of Φ^4 perturbations [51, 52], self-avoiding random walks [23], and directed polymers [31]. The primary reason for the successful analysis is that real-space renormalization transformations, which usually somehow collapse the partition $\mathcal{P}_1$ into single effective spins, tend to transform the parameters into extreme phases while still yielding exact formulas recovering the behavior of the original model. Another important feature of the hierarchical Laplacian is that it allows for a tunable effective dimension defined by

$$d_s = \lim_{\lambda \downarrow 0} \frac{\ln \langle \delta_x, 1_{[\lambda_\infty - \lambda, \lambda_\infty]}(\Delta_H)\delta_x \rangle}{\ln \sqrt{\lambda}}$$

with $\lambda_\infty = \sup \sigma(\Delta_H)$. This quantity is equal to the true spatial dimension if Δ_H is replaced by the Laplacian on the d-dimensional lattice $\mathbb{Z}^d$ and often behaves quite analogously. For example, the random walk generated by the hierarchical Laplacian is reccurent if $d_s \leq 2$ and transitive otherwise [61]. With this history in mind, it is natural to attempt a hierarchical approximation of the localization transition. In this thesis, we consider three specific models of this type, the hierarchical Anderson model, the ultrametric ensemble, and the Rosenzweig-Porter model. The next three sections introduce these models and summarize the main results of our work.

1.2 The Hierarchical Anderson Model

The hierarchical Anderson model

$$H = \Delta + V,$$

which consists of adding a random potential to the hierarchical Laplacian as in (1.1.1), was introduced by Bovier in [22]. Here, we have dropped the subscript from the hierarchical Laplacian Δ_H and will continue to do so as we will have no further use for the Euclidean Laplacian. We will assume that the values $\{V_x : x \in \mathbb{N}_0\}$ of the potential are drawn independently from some bounded density $\varrho \in L^\infty$ which also satisfies

$$\varrho(v) \leq \frac{C}{1 + v^2} \tag{1.2.1}$$

for some $C < \infty$. Moreover, rather than letting the sequence $\{p_r\}$ in the definition of Δ be a general summable sequence, we will make the (largely artificial) restriction to sequences of the form

$$p_r = \varepsilon 2^{-cr}$$

with $\varepsilon, c \in (0, \infty)$. Notice that ε now governs the relative strength of the kinetic and potential terms as in (1.1.1). Furthermore, with this particular choice of $\{p_r\}$ it turns out [68] that the spectral dimension is

$$d_s = \frac{2}{c}.$$

The nature of the localization transition in the hierarchical Anderson model has hitherto been disputed. The original conjecture of Bovier was that the model is localized for $d_s < 4$ but may permit a delocalized phase in higher spectral dimensions. It was hence a surprise when Molchanov [71, 72] proved that H has pure-point spectrum almost surely for all spectral dimensions d_s in the special case that ϱ is the Cauchy distribution. Nevertheless, the conjecture remained consistent with the subsequent works of Kritchevski [60, 61], which proved that H has only pure-point spectrum for arbitrary ϱ when $d_s < 4$. The following theorem, which is an extension of Kritchevski's method, shows that the Cauchy distribution is not a pathological case and provides a clear answer regarding this debate in the infinite volume. In fact, it even shows that the associated eigenfunctions decay exponentially in the hierarchical metric.

Theorem 1.2.1. *The spectrum of H is almost surely of pure-point type with normalized eigenfunctions satisfying*

$$\sum_{x \in \mathbb{N}_0} 2^{\frac{c}{4} d(0,x)} |\psi_\lambda(x)|^2 < \infty \tag{1.2.2}$$

for any $\lambda \in \sigma(H)$.

The work of Molchanov and Kritchevski notwithstanding, numerical analysis led Metz et. al. [67, 68] to conjecture delocalization at a special energy in the regime $d_s > 2$. These authors considered the finite volume cutoff

$$H_n = 1_{B_n} H 1_{B_n},$$

where $B_n = B_n(0)$ is the ball (with respect to the hierarchical metric d) of radius n around 0 and 1_{B_n} is the canonical projection onto $\ell^2(B_n)$. In finite volumes, it is not completely absurd to look for some remnant of delocalization in spite of Theorem 1.2.1 because of the lack of control over the implied amplitudes of the wave functions in (1.2.2). In particular, the claim was that for weak Gaussian potentials

$$\lim_{\eta \to 0} \lim_{n \to \infty} \frac{1}{|B_n|} \mathbb{E}\left[\sum_{\lambda \in \sigma(H_n)} \|\psi_\lambda\|_4 \delta^\eta_{E-\lambda}\right] = 0$$

where $E = \sum p_r$ and δ^η denotes some suitable regularization of the Dirac delta whose width is $\mathcal{O}(\eta)$. Since there are then typically $\mathcal{O}(\eta|B_n|)$ eigenvalues contributing to the sum, the conclusion was that typically

$$\|\psi_\lambda\|_4 \ll 1,$$

which would indeed prohibit a localized regime where $\|\psi_\lambda\|_4 \approx 1$. The numerical analysis was based on comparing the model in a finite volume B_n to the model in B_{n-1} with the renormalized parameters

$$\mathcal{R}(\{p_r\}_{r\geq 1}, \varrho) = ((p_{r+1})_{r\geq 1}, T_{p_1}\varrho)\,, \tag{1.2.3}$$

where $T_p\varrho$ is the probability density of

$$\left(\frac{1}{2V} + \frac{1}{2V'}\right)^{-1} + p,$$

and V and V' are drawn independently from ϱ.

Despite this conjecture, our own numerical results seem to indicate that for any interval $I \subset \mathbb{R}$ there exists $\delta > 0$ such that

$$\sup_{E\in I} \|T_{p_r}...T_{p_1}\varrho(\cdot + E)\|_\infty = \mathcal{O}(2^{(c-\delta)r}) \tag{1.2.4}$$

quite generally when I lies in the spectrum of H and we were able to prove this in the special cases

- $d_s < 2$,
- V is Gaussian and $d_s < 4$, or
- V has a Cauchy component and $d_s < \infty$.

The point of (1.2.4) is that it implies that the effective disorder strength decays slower than the effective strength of the hierarchical hopping

$$(\mathcal{R}p)_r = p_{r+1} = 2^{-c}p_r,$$

which means that $\mathcal{R}$ drives the Hamiltonian into a high-disorder regime where one can prove localization. This idea can indeed be used to prove the decay of the spectral measures in total variation, that is, the decay of the eigenfunction correlator

$$Q_n(x, y; I) = \sup |\langle\delta_y, f(H_n)\delta_x\rangle|, \tag{1.2.5}$$

where the supremum ranges over those $f \in C_0$ with $\operatorname{supp} f \subset I$ and $\|f\|_\infty \leq 1$.

Theorem 1.2.2. *If* (1.2.4) *is satisfied in a bounded interval $I \subset \mathbb{R}$, then there exist $C, \mu \in (0,\infty)$ such that*

$$\sup_{n\in\mathbb{N}_0} \sup_{x\in\mathbb{N}_0} \sum_{y\in\mathbb{N}_0} 2^{\mu d(x,y)} \mathbb{E}\left[Q_n(x,y;I)\right] \le C|I|.$$

Repeated applications of Fatou's lemma show that this theorem implies the strong dynamical localization statement

$$\sum_{y:d(x,y)\ge R} \mathbb{E}\left|\langle \delta_y, 1_I(H)e^{itH}\delta_x\rangle\right|^2 \le C\, 2^{-\mu R}$$

in the spirit of Section 1.1. Moreover, it is easy to see that for non-degenerate spectra

$$Q_n(x,y;I) = \sum_{\lambda\in\sigma(H_n)\cap I} |\psi_\lambda(x)||\psi_\lambda(y)|,$$

which can be used to derive lower bounds on $\|\psi_\lambda\|_4$ with high probability, disproving the conjecture of Metz et. al. Indeed, our bounds are strong enough to draw conclusions about the inverse participation ratios (IPRs)

$$P_q(\psi) := \frac{\sum_x |\psi(x)|^{2q}}{\left[\sum_x |\psi(x)|^2\right]^q} = \frac{\|\psi\|_{2q}^{2q}}{\|\psi\|_2^{2q}}.$$

The IPRs are comparable for different values of $q \ge \frac{1}{2}$:

- for any $q \ge \frac{1}{2}$:
$$1 \le P_q(\psi)\,[P_{\frac{q}{2q-1}}(\psi)]^{2q-1}\,, \tag{1.2.6}$$

- for any $q \ge 1$ we have $r(\psi)^{2q} \le P_q(\psi) \le r(\psi)^{2(q-1)}$ where $r(\psi) = \|\psi\|_\infty/\|\psi\|_2$.

It therefore remains only to state a result concerning the most prominent case $q = 2$.

Corollary 1.2.3 (IPRs)**.** *If the assumption* (1.2.4) *is satisfied in a bounded open interval $I \subset \mathbb{R}$, then there exists some $C < \infty$ such that for any $E \in I$ and $W, \varepsilon > 0$*

$$\mathbb{P}\begin{pmatrix} \text{There is } \psi \in \ell^2(B_n) \text{ with } H_n\psi = \lambda\psi \text{ and} \\ |\lambda - E| \le 2^{-n-1}W \text{ such that } P_2(\psi) \le \varepsilon^4 \end{pmatrix} \le C\,W\varepsilon \tag{1.2.7}$$

for all $n \in \mathbb{N}_0$.

In order to appreciate this result, we stress that the smallness of the probability in (1.2.7) is not due to the fact that the interval $I_n = E + 2^{-n-1}[-W, W]$ is typically void of eigenvalues. In fact, as is proven in Theorem 1.2.4 below,

$$\lim_{n\to\infty} \mathbb{P}\left(\text{No eigenvalue of } H_n \text{ in } I_n\right) = \exp\left(-\nu(E)W\right)$$

at all Lebesgue points $E \in I$ of the (infinite-volume) density of states ν. The authors of [68] studied the averaged IPR

$$\Pi_n(I) = \frac{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I} \|\psi_\lambda\|_4^4}{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I} 1}$$

in the limit of vanishingly small intervals $I \subset \mathbb{R}$. From (2.5.4) in Section 2.5 one concludes

$$\Pi_n(I) \geq C^{-4}\left(\frac{\nu_n(I)}{|I|}\right)^4 \qquad \text{with } \nu_n(I) = \mathbb{E}\left\langle \delta_0, 1_I(H_n)\delta_0 \right\rangle \tag{1.2.8}$$

for any bounded $I \subset \mathbb{R}$ in which (1.2.4) is valid. Since the finite-volume density of states is bounded away from zero for all large enough n provided the interval I is strictly contained in the infinite-volume spectrum $\sigma(H)$, the right side can be shown to be strictly positive in the limit $n \to \infty$ (cf. [54]). In particular, if ϱ is a Gaussian distribution, $\sigma(H) = \mathbb{R}$, and this applies to all energies, which contradicts the conclusions in [68].

The condition (1.2.4) also guarantees that the level statistics converge to a Poisson point process, which was previously proven by Kritchevski [62] only for $d_s < 1$. In the following theorem,

$$\mu_n = \sum_{\lambda\in\sigma(H_n)} \delta_{2^n(\lambda - E)} \tag{1.2.9}$$

is the blown-up eigenvalue point process and ν is the density of states

$$\nu(f) = \mathbb{E}\left\langle \delta_0, f(H)\delta_0 \right\rangle, \tag{1.2.10}$$

which has a bounded density by the Wegner estimate (cf. Proposition 1.6.2).

Theorem 1.2.4. *Suppose* (1.2.4) *is satisfied in an open set $I \subset \mathbb{R}$ and $E \in I$ is a Lebesgue point of ν. Then, μ_n converges in distribution to a Poisson point process with intensity $\nu(E)$ as $n \to \infty$.*

The principal conclusion of our work is that the main appeal of the hierarchical Anderson model, the local separation of scales, is also its flaw since it completely prohibits the occurence of a bona-fide delocalized phase. A possible reason for this may be that the hierarchical structure forces transport to be aided by tunneling, but that the relevant tunneling amplitudes $p_r 2^{-r}$ are prohibitively small if one requires the existence of an infinite-volume limit. A proof of (1.2.4) for more general densities ϱ remains elusive, but given the proof for a dense subset of L^1, we find it hard to imagine that the hierarchical approximation in finite spectral dimension is fine enough to capture the Anderson transition on the lattice.

1.3 The Ultrametric Ensemble

The failure of the hierarchical Anderson model in capturing the localization transition begs the question whether the transition is intrinsically forbidden by hierarchical structures. Hence, in addition to the hierarchical Anderson model, we will also consider the ultrametric ensemble of Fyodorov, Ossipov, and Rodriguez [50], in which the off-diagonal entries in the Hamiltonian are also random variables with variance decaying in the hierarchical metric. The realizations of this ensemble may therefore be interpreted as fully generic hierarchical disordered systems. The corresponding Hamiltonian, which we consider only in finite volume to begin with, is defined on $\ell^2(B_n)$ as

$$H_n = \frac{1}{Z_{n,c}} \sum_{r=0}^{n} 2^{-\frac{1+c}{2} r} \Phi_{n,r}. \tag{1.3.1}$$

The entries of the layers $\Phi_{n,r}$ are independent centered Gaussian random variables with variance profile

$$\mathbb{E}\, |\langle \delta_y, \Phi_{n,r} \delta_x \rangle|^2 = 2^{-r} \begin{cases} 2 & \text{if } d(x,y) = 0 \\ 1 & \text{if } 1 \le d(x,y) \le r\ . \\ 0 & \text{otherwise} \end{cases} \tag{1.3.2}$$

Thus $\Phi_{n,r}$ is a direct sum of 2^{n-r} random matrices drawn independently from the Gaussian Orthogonal Ensemble (GOE) of size 2^r. We choose the normalizing constant $Z_{n,c}$ such that

$$\sum_{y \in B_n} \mathbb{E}\, |\langle \delta_y, H_n \delta_x \rangle|^2 = 1, \tag{1.3.3}$$

which means that $Z_{n,c}$ grows exponentially in n in case $c < -1$ and $Z_{n,c}$ is asymptotically constant in case $c > -1$. The original definition in [50] contains an additional parameter governing the relative strengths of the diagonal and off-diagonal disorder, but this parameter does not significantly alter our analysis and so we omit it altogether. Moreover, the authors of [50] constructed the block matrices $\Phi_{n,r}$ from the Gaussian Unitary Ensemble (GUE), and our results apply to both GOE and GUE blocks with only slight changes.

Notice that the variance matrix of H_n has the structure of a hierarchical Laplacian and that the layer $r = 0$ in (1.3.1) plays the role of a random potential with Gaussian distribution, although our analysis does not change significantly upon inserting some other regular distribution into the layer $r = 0$. The typical size of the entries in H_n is

$$\mathbb{E}\, |\langle \delta_y, H_n \delta_x \rangle|^2 \approx 2^{-(2+c)d(x,y)},$$

so this model can also be thought of as a hierarchical analogue of the power-law random band matrices [70, 78] for which the variances of the entries decay like $|x - y|^{-(2+c)}$. The exponential decay in the ultrametric ensemble is the correct analogue of the algebraic decay in the power-law random band matrices because our definition of the hierarchical metric grows only logarithmically in the volume, while the one-dimensional Euclidean metric grows linearly in the volume. In particular, as c varies in $\mathbb{R}$, H_n interpolates between a perfectly localized random potential at "$c = \infty$" and a perfectly delocalized Wigner random matrix at "$c = -\infty$".

The article [50] conjectured a localization transition at $c = 0$ based on arguments with a physics level of rigor and numerical simulation. Here, we wish to adopt a point of view based on a dynamical representation, which is more susceptible to a mathematically acceptable proof. The self-similar structure of H_n shows that

$$H_n = \sum_{r=0}^{n-1} 2^{-\frac{(1+c)}{2}r} \Phi_{n,r} + 2^{-\frac{(1+c)}{2}n} \Phi_{n,n} = H_{n-1} \oplus H'_{n-1} + 2^{-\frac{(1+c)}{2}n} \Phi_{n,n}$$

so that it is possible to construct H_n recursively from Gaussian perturbations. The idea of Dyson [37] was to study $N \times N$ Gaussian perturbations by representing them as matrix-valued stochastic processes whose entries

$$\langle \delta_u, \Phi_t \delta_v \rangle = \sqrt{\frac{1 + \delta_{uv}}{N}} B_{uv}(t)$$

are rescaled Brownian motions independent up to the symmetry constraint. Using this idea with $N = 2^n$ and $T = 2^{-(1+c)n}$ we obtain that

$$H_n = H_{n-1} \oplus H'_{n-1} + \Phi_T$$

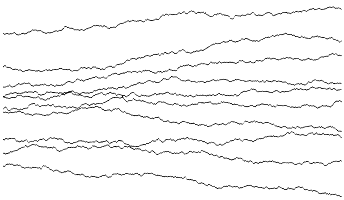
in distribution. Dyson showed that the evolution of the eigenvalues of Φ_t is given by

$$d\lambda_j(t) = \sqrt{\frac{2}{N}} dB_j(t) + \frac{1}{N} \sum_{i \neq j} \frac{dt}{\lambda_j(t) - \lambda_i(t)}, \tag{1.3.4}$$

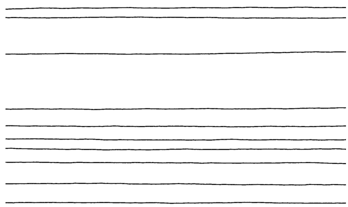
which is now called Dyson Brownian motion (DBM). To construct the spectrum of H_n one may therefore initialize $\sigma(H_0)$ to a Gaussian random variable and iterate the following two steps.

1. Sample an independent copy $\sigma(H'_{k-1})$ of $\sigma(H_{k-1})$
2. Let $\sigma(H_k)$ be the evolution of $\sigma(H_{k-1}) \cup \sigma(H'_{k-1})$ under DBM with duration $T = 2^{-(1+c)k}$.

These two operations are in inherent competition with each other, the new sample in the first step increasing the amount of independence in the spectrum, and the the long-range Coulomb interaction of the DBM in the second step increasing the correlations. If one assumes that the typical distance between adjacent eigenvalues is of order N^{-1}, then the heuristic effect of the DBM on the spectrum becomes negligible with respect to the spacing precisely when $T \ll N^{-1}$.



Trajectories until $T = 1$ of 10-particle DBM



10 Trajectories until $T = 100^{-1}$ of 100-particle DBM

Hence, when $T \ll N^{-1}$, one might hope that the DBM in the second step above is not running for a long enough time to compensate the fluctuations introduced in the first step. It is indeed possible to make this intuition rigorous, yielding a proof of Poisson statistics for H_n in the regime $c > 0$. For the following theorem, the level statistics μ_n and the density of states ν are defined as in (1.2.9) and (1.2.10).

Theorem 1.3.1. *Suppose $c > 0$ and $E \in \mathbb{R}$ is a Lebesgue point of ν. Then, μ_n converges in distribution to a Poisson point process with intensity $\nu(E)$ as $n \to \infty$.*

It is possible to adapt this argument to the stability of the associated eigenfunctions, which implies the following localization bounds for the eigenfunction correlator (defined as in (1.2.5)).

Theorem 1.3.2. *Suppose $c > 0$ and let $E \in \mathbb{R}$. Then, there exist $w, \mu, \kappa > 0$, $C < \infty$, and a sequence m_n with $n - m_n \to \infty$ such that for every $x \in B_n$ the*

ℓ^2-normalized eigenfunctions satisfy

$$\mathbb{P}\left(\sum_{y \in B_n \setminus B_{m_n}(x)} Q_n(x, y; W) > 2^{-\mu n} \right) \leq C\, 2^{-\kappa n}$$

with

$$W = \left[E_0 - 2^{-(1-w)n}, E_0 + 2^{-(1-w)n} \right].$$

This theorem is much weaker than Theorem 1.2.2 because the width of the spectral window W can only be mesoscopic. Nevertheless, this theorem is an indication of localization. For if the eigenfunctions in W were completely extended, we would obtain

$$\sum_{y \in B} Q_n(x, y; W) = \sum_{y \in B} \sum_{\lambda \in \sigma(H_n) \cap W} |\psi_\lambda(x)||\psi_\lambda(y)| \approx |B| 2^n |W| 2^{-n} = |B||W|$$

since there are typically $2^n|W|$ eigenvalues in W. Since $|B| \approx 2^n$ and $|W| = 2^{-(1-w)n}$ in the previous theorem this quantity would then be large rather than small with high probability.

The regime $c < 0$ corresponds to perturbations Φ_T with $T \gg N^{-1}$ in the iterative construction of H_n. For such times, the DBM does have a significant effect on the local statistics of the spectrum. Landon, Sosoe, and Yau [64, 65] showed that if V has a sufficiently regular spectrum, then the local statistics of $V + \Phi_T$ agree asymptotically with the WDM statistics of the GOE Φ_1 as $N \to \infty$. While this behavior is a good indicator of a phase transition at $c = 0$, it has proven difficult on a technical level to verify the required regularity for the unperturbed matrices $H_{n-1} \oplus H'_{n-1}$. However, when $c < -1$, the ultrametric ensemble has an essential mean field character and techniques originally developed for Wigner matrices show that the energy levels agree asymptotically with those of the GOE and that the eigenfunctions are completely delocalized. We will now roughly sketch how to apply these results in the present situation and state the corresponding theorems. The key observation is that the normalizing factor $Z_{n,c}$, which scales the spectrum to $\mathcal{O}(1)$, is given by

$$Z_{n,c}^2 = \sum_{y \in B_n} \mathbb{E} \left| \langle \delta_y, \left(\sum_{r=0}^{n} 2^{-\frac{(1+c)}{2} r} \Phi_{n,r} \right) \delta_x \rangle \right|^2 = \left(1 - 2^{-(1+c)(n+1)} \right) \frac{1 + \mathcal{O}(1)}{1 - 2^{-(1+c)}},$$

so that the spread

$$M_n := \left(\max_{x, y \in B_n} \mathbb{E} \left| \langle \delta_y, H_n \delta_x \rangle \right|^2 \right)^{-1} = \begin{cases} Z_{n,c}^2\, 2^{-o(n)} & \text{if } c \geq -2, \\ 2^{(1+o(1))n} & \text{if } c < -2, \end{cases}$$

grows like a positive power of the system size 2^n when $c < -1$. The results of [42] then show that the semicircle law (i.e. $\nu(E) = \sqrt{(4-E^2)_+}/(2\pi)$) is valid on scales of order M_n^{-1} even for the matrices

$$\widetilde{H}_n = \frac{1}{Z_{n,c}} \sum_{r=0}^{n-1} 2^{-\frac{(1+c)}{2}r} \Phi_{n,r} + \frac{1-\sqrt{T_n}}{Z_{n,c}} 2^{-\frac{(1+c)}{2}n} \Phi_{n,n}$$

with a small part of the final $\mathcal{O}(1)$ Gaussian component removed. We set $T_n = M_n^{-1+\delta}$ with $\delta \in (0,1)$. The validity of the local semicircle law already implies the complete delocalization of the eigenfunctions in mesoscopic windows in the bulk of the spectrum (see [40, Thm. 2.21]).

Theorem 1.3.3 (cf. [40, 42])**.** *Let $c < -1$. For any compact interval $I \subset (-2,2)$ there exist $\kappa, \varepsilon > 0$ such that for all $E \in I$ the ℓ^2-normalized eigenfunctions of H_n in $[E - M_n^{-1}, E + M_n^{-1}]$ satisfy*

$$\|\psi_\lambda\|_\infty = \mathcal{O}(M_n^{-1/2+\varepsilon})$$

with probability $1 - \mathcal{O}(N^{-\kappa})$.

Random matrix universality of the local statistics may be expressed by saying that the k-point correlation functions

$$\varrho_{H_n}^{(k)}(\lambda_1, ..\lambda_k) = \int_{\mathbb{R}^{2^n-k}} \varrho_{H_n}(\lambda_1, ..., \lambda_{2^n})\, d\lambda_{k+1} ... \, d\lambda_{2^n},$$

the k-th marginals of the symmetrized eigenvalue density ϱ_{H_n}, locally agree with the corresponding objects for the GOE asymptotically. For this, we employ the previously mentioned work [64, Thm. 2.2] concerning the universality of Gaussian perturbations for

$$H_n = \widetilde{H}_n + \frac{\sqrt{T_n}}{Z_{n,c}} 2^{-\frac{(1+c)}{2}n} \Phi_{n,n}.$$

For the statement of the theorem, let

$$\begin{aligned} \Psi_{n,E}^{(k)}(\alpha_1, ..., \alpha_k) = {} & \varrho_{H_n}^{(k)} \left(E + 2^{-n} \frac{\alpha_1}{\varrho_{sc}(E)}, ..., E + 2^{-n} \frac{\alpha_k}{\varrho_{sc}(E)} \right) \\ & - \varrho_{GOE}^{(k)} \left(E + 2^{-n} \frac{\alpha_1}{\varrho_{sc}(E)}, ..., E + 2^{-n} \frac{\alpha_k}{\varrho_{sc}(E)} \right), \end{aligned}$$

where $\varrho_{GOE}^{(k)}$ is the k-point correlation function of the $2^n \times 2^n$ GOE and ϱ_{sc} is the density of the semicircle law.

Theorem 1.3.4 (cf. [42, 64]). *Suppose $c < -1$, $E \in (-2, 2)$ and $k \geq 1$. Then,*

$$\lim_{n\to\infty} \int_{\mathbb{R}^k} O(\alpha)\Psi^{(k)}_{n,E}(\alpha)\, d\alpha = 0$$

for every $O \in C_c^\infty(\mathbb{R}^k)$.

Summing up, these results rigorously prove the existence of a metal-insulator transition in the ensemble of ultrametric random matrices. In particular, our results allow an approach all the way to the critical point from the localized side $c > 0$, which improves upon the best known corresponding result for random band matrices [80]. However, the above arguments do not cover the regime $c \in [-1, 0)$, in which the local eigenvalue statistics are still expected to be of Wigner-Dyson-Mehta type as in the case $c < -1$ [50]. The regime $c \in (-1, 0)$ would be particularly interesting because it would contain a delocalized Hamiltonian which does not have a mean-field character and whose infinite-volume limit exists almost surely.

1.4 The Rosenzweig-Porter Model

The last section detailed how the study of the ultrametric ensemble amounts to the detailed study of Gaussian perturbations, making it natural to also consider the ultimate hierarchical simplification, the Gaussian perturbation of a random potential. The *Rosenzweig-Porter model* is thus defined as the $N \times N$ matrix

$$H_T = V + \Phi_T, \tag{1.4.1}$$

where Φ_t is the stochastic process of the previous section and V is a diagonal matrix whose entries are drawn independently from $\varrho \in L^\infty$. Hence, the Rosenzweig-Porter model is the simplest possible random matrix with a non-trivial spatial structure and provides a standard interpolation between localization and delocalization. Moreover, the Rosenzweig-Porter model and its relatives, the Anderson models on the Bethe lattice and on the random regular graph, have recently received a renewed surge of interest related to the many-body localization transition [8, 46, 59]. In that context, they provide basic examples of phases in which eigenfunctions delocalize over a large number of sites, but not uniformly over the entire volume.

Clearly, one expects the spectral behavior of H_T to interpolate between V and Φ_T as T increases, with $T = N^{-1}$ being a critical point at least as far as the local statistics are concerned. Since any difficulties arising from the need to propagate regularity estimates of the spectrum through the middle layers of a true hierarchical model disappear for the Rosenzweig-Porter model, it is possible to verify the regularity conditions required by [64] as input to the proof of equilibration of the DBM.

Theorem 1.4.1 (cf. [64]). *Suppose $E \in \mathbb{R}$ lies in the interior of* $\operatorname{supp} \varrho$ *and* $T = N^{-1+\varepsilon}$ *for some* $\varepsilon > 0$. *Then,*

$$\Psi^{(k)}_{N,E}(\alpha_1, ..., \alpha_k) = \varrho^{(k)}_{H_T}\left(E + N^{-1}\frac{\alpha_1}{\varrho_{sc}(E)}, ..., E + N^{-1}\frac{\alpha_k}{\varrho_{sc}(E)}\right)$$
$$- \varrho^{(k)}_{GOE}\left(E + N^{-1}\frac{\alpha_1}{\varrho_{sc}(E)}, ..., E + N^{-1}\frac{\alpha_k}{\varrho_{sc}(E)}\right),$$

satisfies

$$\lim_{n\to\infty}\int_{\mathbb{R}^k} O(\alpha)\Psi^{(k)}_{n,E}(\alpha)\, d\alpha = 0$$

for every $k \geq 1$ *and* $O \in C^\infty_c(\mathbb{R}^k)$.

On the other hand, if $T \ll N^{-1}$ and the analysis leading to Theorems 1.3.1 and 1.3.2 may also be used to prove the following theorem.

Theorem 1.4.2. *If* $T \leq N^{-(1+\varepsilon)}$ *with* $\varepsilon > 0$ *and* $E_0 \in \mathbb{R}$, *then:*

1. *As* $N \to \infty$, *the random measure defined by*

$$\mu_N(f) = \sum_{\lambda\in\sigma(H_T)} f(N(\lambda - E_0))$$

   converges in distribution to a Poisson point process with intensity $\varrho(E_0)$ *provided* E_0 *is a Lebesgue point of* ϱ.

2. *There exist* $w, \mu, \kappa > 0$ *and* $C < \infty$ *such that for every* $x \in \{1, ..., N\}$ *the* ℓ^2*-normalized eigenfunctions satisfy*

$$\mathbb{P}\left(\sum_{\lambda\in\sigma(H_T)\cap W}\sum_{y\neq x}|\psi_\lambda(x)\psi_\lambda(y)| > N^{-\mu}\right) \leq CN^{-\kappa}$$

   with

$$W = \left[E_0 - N^{-(1-w)}, E_0 + N^{-(1-w)}\right].$$

The proof of this theorem is contained in Section 3.5. Theorem 1.4.1 and the first point of Theorem 1.4.2 combine to prove a sharp transition in the local statistics at $T = N^{-1}$. The second point of Theorem 1.4.2 asserts that if a state near E_0 carries some mass at $x \in \{1, ..., N\}$, then it doesn't carry any mass in $\{1, .., N\} \setminus \{x\}$ with high probability. This result should be compared with the scenario in which the eigenfunctions ψ_λ are completely extended, where one would expect that typically

$$\sum_{\lambda\in\sigma(H_t)\cap W}\sum_{y\neq x}|\psi_\lambda(x)\psi_\lambda(y)| \approx \sum_{\lambda\in\sigma(H_t)\cap W} 1 \approx N|W| = N^w \qquad (1.4.2)$$

becomes very large for mesoscopic spectral windows. The theorem thus proves localization in the sense that this quantity actually vanishes asymptotically for small enough mesoscopic intervals W. The proof of Theorem 1.4.2 yields an explicit relation between w, μ, κ and ε, which shows that w, μ, κ may increase if ε increases as well.

If $T \geq 1$, the complete delocalization of the eigenfunctions was proved by Lee and Schnelli [66] as a corollary to a local law, but there were hitherto no previous rigorous results concerning the behavior of the eigenfunctions in the intermediate regime $N^{-1} \ll T \ll 1$. Moreover, the nature of the transition in the eigenfunctions in the related Anderson models on the Bethe lattice and random regular graph has been widely disputed even in the physics literature [7, 16, 28, 86]. Here, we confirm the picture conjectured by Kravtsov et. al. in [59], by proving that in the intermediate regime a normalized eigenfunction ψ_λ corresponding to $\lambda \in \sigma(H_T)$ delocalizes across approximately those $NT \gg 1$ sites for which V_x is closest to λ. This means that the mass of each eigenfunction spreads to a large number of sites. These sites nevertheless form a vanishing fraction of the entire volume $\{1, ..., N\}$, indicating the existence of a non-ergodic delocalized phase.

Our method is quite different from the ideas contained in [59] and is inspired by the clarifying analysis of Facoetti, Vivo, and Biroli [46]. This article also explains how the abrupt transition in the local statistics does not contradict the gradual transition in the degree of eigenfunction localization, by arguing that the statistics retain a Poissonian character at mesoscopic scales greater than T. For rigorous results of a similar character we refer to [36, 55].

Another method for studying the eigenfunctions of H_t was devised by Bourgade and Yau [21] and developed further by Bourgade, Huang, and Yau [20], whose Theorem 2.1 may also be used to derive the second point of Theorem 1.4.3 below. The method was adapted to the present problem by Benigni [14]. Here, it yields the local eigenvector statistics even for mesoscopic Wigner perturbations, covering Theorem 1.4.3, albeit with lower probability.

Theorem 1.4.3. *Suppose $T = N^{-1+\delta}$ with $\delta > 0$ and let W be contained in the interior of* $\operatorname{supp} \varrho$. *Let $\kappa > \delta > \theta$ and set*

$$X_\lambda = \{x \in \{1, ..., N\} : |\lambda - V_x| > N^{-1+\kappa}\}.$$

Then, there exists $\gamma > 0$ such that for any $p > 0$ and all sufficiently large N

1. *The normalized eigenfunctions in W carry only negligible mass inside X_λ:*

$$\mathbb{P}\left(\sup_{\lambda \in \sigma(H_T) \cap W} \sum_{x \in X_\lambda} |\psi_\lambda(x)|^2 > N^{-\gamma}\right) \leq N^{-p}.$$

2. The normalized eigenfunctions in W are maximally extended outside X_λ:

$$\mathbb{P}\left(\sup_{\lambda\in\sigma(H_T)\cap W}\|\psi_\lambda\|_\infty > N^{-\theta/2}\right)\leq N^{-p}.$$

Theorem 1.4.3 becomes meaningful when κ and θ are chosen close to δ. Indeed, the number of sites outside X_λ is then approximately

$$|\{V_x : |\lambda - V_x| \leq N^{-1+\kappa}\}| \approx NN^{-1+\kappa} = N^\kappa \approx NT.$$

Moreover, the fact that

$$|\psi_\lambda(x)|^2 \leq N^{-\theta} \approx (NT)^{-1}$$

shows that the eigenfunctions are maximally extended inside the subvolume $\{1, ..., N\} \setminus X_\lambda$.

1.5 Ideas from the Proofs

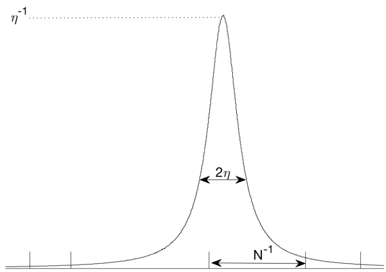
Having outlined the main contributions of this thesis, we now sketch some of the principal ideas behind the proofs. We have already mentioned that our localization results for the ultrametric ensemble may be understood as a consequence of the stability of repeated Gaussian perturbations. The high degree of rotational symmetry in Gaussian ensembles facilitates the analysis by enabling the very explicit DBM (1.3.4) for the evolution of the eigenvalues. Nevertheless, the Coulomb interaction in (1.3.4) is quite singular and forces the direct study of this process to be technically involved. Therefore, we prefer to study the empirical measure

$$\nu_t = \frac{1}{N}\sum_{\lambda\in\sigma(H_t)}\delta_\lambda$$

in terms of its Stieltjes transform

$$S_t(z) = \int \frac{1}{\lambda - z}\,\nu_t(d\lambda) = \frac{1}{N}\mathrm{Tr}\,(H_t - z)^{-1}.$$

Since $\operatorname{Im} S_t(z)$ is then the Poisson integral of ν_t evaluated at z, the function S_t encodes detailed local information on the locations of the eigenvalues when $\operatorname{Im} z \to 0$. For typical realizations of H_t, the distance between adjacent eigenvalues is approximately N^{-1} so tracking the profile of $\operatorname{Im} S_t$ with $\operatorname{Im} z \ll N^{-1}$ is the same as tracking the trajectories of individual eigenvalues.



The Poisson kernel at submicroscopic scale $\eta \ll N^{-1}$

The following theorem shows that $S_t(z)$ remains stable at such submicroscopic scales provided that also $t \ll N^{-1}$. The theorem is valid for operators of the form

$$H_t = \tilde{H} + V + \Phi_t, \tag{1.5.1}$$

where Φ_t is defined as before. We will assume that

$$V = \sum_x V(x)|\delta_x\rangle\langle\delta_x|$$

is an independent random potential such that the conditional distributions are uniformly Lipschitz continuous, i.e.,

$$\mathbb{P}\left(V(x) \in I \,|\, \{V(y)\}_{y \neq x}\right) \leq C_V |I|, \tag{1.5.2}$$

for all Borel sets $I \subset \mathbb{R}$ and $x \in \{1, ..., N\}$ with a constant $C_V < \infty$ independent of N. Finally, $\tilde{H}$ is some real symmetric $N \times N$ matrix, which may also be random provided $\tilde{H}$, V, and Φ_t remain independent.

Theorem 1.5.1. *For every $\varepsilon > 0$, there exists $C < \infty$, depending only on ε and C_V, such that*

$$\mathbb{E}\left|S_T\left(E + i\eta\right) - S_0\left(E + i\eta\right)\right| \leq CN^{-\varepsilon/2}\left(1 + \frac{1}{N\eta} + \frac{1}{(N\eta)^3}\right)$$

for all $T \leq N^{-(1+\varepsilon)}$ and $E \in \mathbb{R}$.

In essence, Theorem 1.5.1 asserts that, on scales much larger than T, the empirical eigenvalue measure is unaffected by the flow (1.5.1). This result goes beyond norm-based continuity arguments for the DBM. The example $H_0 = 0$ shows that at least some regularity of the initial condition is needed for Theorem 1.5.1 to remain true. However, Theorem 1.5.1 can be proved for slightly more general H_0 under the weaker assumption that H_t satisfies the Wegner and Minami estimates (cf. Propositions 1.6.2 and 1.6.3)

$$\mathbb{E}\,\nu_t(I) \leq C|I|, \qquad \mathbb{E}\,\nu_t(I)(\nu_t(I) - N^{-1}) \leq C|I|^2$$

with a constant $C < \infty$ uniform in N and t. This is easily seen from the proof below. It is also possible to present Theorem 1.5.1 (and the following Theorem 1.5.2) as explicit bounds for arbitrary $T > 0$, but we artificially restrict to $T \leq N^{-(1+\varepsilon)}$ in order to keep the right hand side simple.

The properties of the eigenfunctions of H_t are encoded in the spectral measures

$$\mu_{xy} = \sum_{\lambda\in\sigma(H_t)} \psi_\lambda(x)\overline{\psi}_\lambda(y)\delta_\lambda$$

where $\{\psi_\lambda\}$ is an orthonormal basis of eigenfunctions of H_t and we have eased the notational burden by keeping the dependence of ψ_λ and μ_{xy} on t implicit. Hence, the Green functions

$$G_t(x, y; z) = \langle \delta_y, R_t(z)\delta_x\rangle = \int \frac{1}{\lambda - z}\,\mu_{xy}(d\lambda)$$

at scales $\operatorname{Im} z \approx N^{-1}$ describe the eigenfunctions of H_t locally near $\operatorname{Re} z$. The stability result analogous to Theorem 1.5.1 for $G_t(x, y; z)$ is contained in the following theorem.

Theorem 1.5.2. *For every $\varepsilon > 0$, there exists $C < \infty$, depending only on ε and C_V, such that*

$$\frac{1}{N}\sum_y \mathbb{E}\,|G_T\,(x, y; E + i\eta) - G_0\,(x, y; E + i\eta)| \leq CN^{-\varepsilon/2}\left(1 + \frac{1}{N\eta} + \frac{1}{(N\eta)^3}\right)$$

for all $T \leq N^{-(1+\varepsilon)}$, $E \in \mathbb{R}$, and $x \in \{1, ..., N\}$.

The proofs of Theorems 1.5.1 and 1.5.2 are based on the small algebraic miracle that the Green functions obey the stochastic advection equation

$$\begin{aligned} dG_t(x, y; z) = &\left(S_t(z)\frac{\partial}{\partial z}G_t(x, y; z) + \frac{1}{2N}\frac{\partial^2}{\partial z^2}G_t(x, y; z)\right)\,dt \\ &+ dM_t(x, y; z), \end{aligned} \tag{1.5.3}$$

where $M_t(x, y; z)$ is a martingale that can be given explicitly in terms of $(H_t - z)^{-1}$. This equation perfectly encodes the cancellations of the Coulomb repulsion and lets one proceed by smoothing $G_t(x, y; z)$ and its quadratic variation using the randomness of V.

The proof of Theorem 1.4.3 also uses (1.5.3) to control the local resolvent

$$G_t(x,z) := G_t(x,x;z).$$

If we retain only the leading term on the right hand side of (1.5.3), we obtain an equation transporting $G_0(x,z)$ along the characteristic curve defined by

$$\dot{z}_t = -S_t(z_t). \tag{1.5.4}$$

We will prove that the remaining terms are negligible in the regime $\operatorname{Im} z \gg N^{-1}$, which, combined with the previous observation, yields

$$G_t(x,z_t) \approx G_0(x,z). \tag{1.5.5}$$

Our bounds are strong enough to conclude that for every $z \in \mathbb{C}_+$ with $\operatorname{Im} z \gg N^{-1}$ there exists $w \in \mathbb{C}_+$ with $|w-z| = \mathcal{O}(t)$ such that

$$G_t(x,z) \approx G_0(x,w). \tag{1.5.6}$$

This means that the effect on the eigenfunctions of perturbing V by Φ_t locally consists of a shift in the energy followed by a smearing of the scales below t. In essence, the change in the local resolvent on the given time scale is through an energy renormalization.

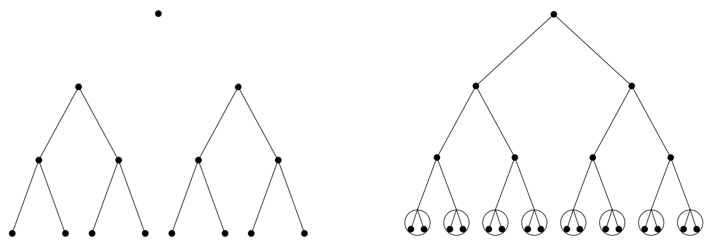
The relations (1.5.4)–(1.5.6) amount to strong finite volume versions of the famous semi-circular flow of Pastur [74] localized to a single site and energy. They also bear some similarity to a preliminary Schur complement relation in Erdős, Schlein, and Yau's proof of the local semi-circle law [43], although our results are clearly only valid for Gaussian ensembles. The fact that (1.5.6) merely changes the spectral parameter at which the local resolvent is evaluated can also be seen as a particular instance of the subordination relations in free probability [15, 87].

Finally, let us note that the analogue of (1.5.1) for perturbations drawn from the GUE,

$$\langle \delta_y, \tilde{\Phi}_t \delta_x \rangle = \sqrt{\frac{1}{N}} \begin{cases} \frac{1}{\sqrt{2}}(B_{xy}(t) + i\tilde{B}_{xy}(t)) & \text{if } x < y \\ B_{xx}(t) & \text{if } x = y \end{cases}$$

with $\tilde{B}_{xy}$ independent of B_{xy}, has also been widely studied. The analysis of this model is usually simpler because the additional symmetry enables explicit integration formulas (see [40] and references therein for a summary) and all the results and methods mentioned here require only minor modifications to treat also the GUE flow.

The procedure reducing the ultrametric ensemble to repeated Gaussian perturbations may be visualized by thinking of the layers in H_n as being indexed by a binary tree whose $(N-r)$-th generation is the partition $\mathcal{P}_r$ with the root removed. For the hierarchical Anderson model, the analogous procedure is much less powerful because it results in a decomposition of H_n into a succession of rank-one perturbations that are a-priori no more useful than any other such decomposition. The point is to realize that the ranges of the rank-one perturbations are nested so that Δ_H retains a simple hierarchical structure in the eigenbasis of the bottom-most layer E_1. Hence, the analysis proceeds by collapsing the pairs of sites in $\mathcal{P}_1$ into effective single sites, much in the spirit of real-space renormalization arguments in classical statistical mechanics.



Removing the root　　Collapsing the leaves

To be more specific, we consider the basis $\{e_0, e_1, ..., f_0, f_1, ...\}$ of $\ell^2(\mathbb{N}_0)$ whose members are

$$e_k = \frac{1}{\sqrt{2}}\left(\delta_{2k} + \delta_{2k+1}\right), \quad f_k = \frac{1}{\sqrt{2}}\left(\delta_{2k} - \delta_{2k+1}\right)$$

and let U be the unitary change to the canonical basis. Then

$$U\Delta_H U^* = 2^{-c}\begin{pmatrix}\Delta_H & 0\\ 0 & 0\end{pmatrix}$$

so that the strength of the kinetic term has decreased in the new basis. At the same time, (1.2.4) shows that the strength of the disorder decays at a much slower rate, so that we reach a well-understood high-disorder regime after finitely many steps. Performing this change of basis on the full Hamiltonian H, a computation using the Feshbach-Krein-Schur map yields a formal relation of the form

$$G(0, 2y; 0) = 2\frac{V_1}{V_0 + V_1}\frac{V_{2y+1}}{V_{2y} + V_{2y+1}}\mathcal{R}G(0, y; 0)$$

where $\mathcal{R}G$ is the Green function of the Hamiltonian with renormalized parameters as in (1.2.3). If we ignore the somewhat singular prefactor, this formula tells us that the localization length of any profile for the Green function increases by at most a factor of two after a single renormalization step. Hence, the localization length of the original Green function is a finite multiple of the localization length of an exponentially decaying renormalized Green function in the high disorder regime. This result then implies the dynamical localization bound of Theorem 1.2.2 in a standard way.

1.6 Spectral Averaging

In this brief section we mention a simple but powerful mechanism for smoothing the spectral measures of H in the presence of a random potential, which features prominently in our localization proofs for the hierarchical Anderson model and the ultrametric ensemble. We will consider a general Anderson-type random matrix of the form

$$H = \tilde{H} + V$$

acting on $\ell^2(\{1, ..., N\})$, where $\tilde{H}$ is some Hermitian $N \times N$ matrix and V is a random potential with entries drawn independently from a bounded density $\varrho \in L^\infty$.

From the heuristic principle that the eigenvalues depend most on the diagonal entries of a matrix, we expect that the addition of $O(1)$-randomness on the diagonal should have a generic smoothing effect on the typical realization of the spectral measures $\mu_x := \mu_{\delta_x}$. This intuition is confirmed by the spectral averaging principle of Kotani [58]. In the following formulation, which is taken from [6], $\mathbb{E}[\cdot \mid \cdot]$ denotes the conditional expectation and $|\cdot|$ denotes the Lebesgue measure on $\mathbb{R}$.

Proposition 1.6.1. *Let $B \subset \mathbb{R}$ be a Borel set. Then,*

$$\mathbb{E}\left[\mu_x(B) \,|\, \{V_y\}_{y \neq x}\right] \leq \|\varrho\|_\infty |B|.$$

The proof consists of showing that the left hand side defines a bounded linear functional on $L^1(\mathbb{R})$ by testing with the dense set of Poisson kernels. Indeed, let H_0 denote the operator obtained from H by setting V_x to zero, so that

$$H = H_0 + V_x |\delta_x\rangle\langle\delta_x|.$$

Then, writing $\mu(B) = \mathbb{E}\left[\mu_x(B) \,|\, \{V_y\}_{y \neq x}\right]$ and $\Gamma = -\left(\langle\delta_x, (H_0 - z)^{-1}\delta_x\rangle\right)^{-1}$, standard rank-one perturbation formulas show that

$$\begin{aligned}
\int \frac{\operatorname{Im} z}{(t - \operatorname{Re} z)^2 + (\operatorname{Im} z)^2}\, \mu(dt) &= \int \operatorname{Im} \langle\delta_x, (H - z)^{-1}\delta_x\rangle \varrho(V_x)\, dV_x \\
&= \int \operatorname{Im} \frac{1}{V_x - \Gamma} \varrho(V_x)\, dV_x \\
&\leq \|\varrho\|_\infty \pi = \|\varrho\|_\infty \int \frac{\operatorname{Im} z}{(t - \operatorname{Re} z)^2 + (\operatorname{Im} z)^2}\, dt.
\end{aligned}$$

Since the empirical eigenvalue measure ν is just the average of μ_x over $x \in \{1, ..., N\}$, we immediately obtain the following corollary due to Wegner [93].

Corollary 1.6.2. *For every $f \in L^1$ we have*

$$\mathbb{E}\left|\frac{1}{N}\sum_{\lambda\in\sigma(H)} f(\lambda)\right| = \mathbb{E}\left|\frac{1}{N}\operatorname{Tr} f(H)\right| \leq \|\varrho\|_\infty \|f\|_1.$$

Applying Markov's inequality to the Wegner estimate bounds the probability of a Borel set $B \subset \mathbb{R}$ containing one or more eigenvalues by

$$\mathbb{P}(|\sigma(H) \cap B| \geq 1) \leq \|\varrho\|_\infty N|B|.$$

This bound was extended to the case of two or more eigenvalues, and k or more eigenvalues by Minami [69] and Combes, Germinet and Klein [26], respectively.

Proposition 1.6.3. *Let $B \subset \mathbb{R}$ be a Borel set. Then,*

$$\mathbb{P}(|\sigma(H) \cap B| \geq k) \leq \frac{C_\varrho}{k!}(N|B|)^k.$$

The proof of this proposition is based on the Chebyshev bound

$$\mathbb{P}(|\sigma(H) \cap B| \geq k) \leq \frac{1}{k!}\mathbb{E}\prod_{j=0}^{k}(\operatorname{Tr} 1_B(H) - j),$$

the right hand side of which can be controlled by peeling off factors via conditional expectations. We illustrate the gist of this method later in the proof of Lemma 3.1.1.

2 Renormalization Group Analysis of the Hierarchical Anderson Model

In this chapter we prove the results stated in Section 1.2 using the renormalization group ideas sketched in Section 1.5. We recall that the hierarchical Anderson model depends on essentially two parameters: the sequence $\mathbf{p} = (p_r)$ of coefficients in the hierarchical Laplacian and the density ϱ of the diagonal disorder.

2.1 The Renormalization Group

The content of this section is the investigation of a relationship between the resolvents of H and the resolvents of an operator $\mathcal{R}H$ whose parameters $(\mathbf{p}, \varrho)$ have effectively been renormalized to $\mathcal{R}(\mathbf{p}, \varrho)$ (cf. (1.2.3)). This is achieved by considering the new Hamiltonian $\mathcal{R}H = \mathcal{R}\Delta + \mathcal{R}V$ with components

$$\mathcal{R}\Delta = \sum_{r=1}^{\infty} p_{r+1} E_r \tag{2.1.1}$$

and

$$(\mathcal{R}V)_y = \left(\frac{1}{2V_{2y}} + \frac{1}{2V_{2y+1}} \right)^{-1} + p_1 . \tag{2.1.2}$$

The definition (2.1.2) guarantees that the renormalized potential $\mathcal{R}V$ consists of independent random variables whose common density is $T_{p_1}\varrho$.

It will be useful in our proof of Theorem 1.2.2 to also consider a slightly more general situation, in which the disorder remains independent, but is allowed to have different distributions at different sites. We will thus suppose that $V_y \sim \varrho_y$, where $\{\varrho_y \,|\, y \in \mathbb{N}_0\}$ is a collection of probability densities with $\sup_y \|\varrho_y\|_\infty < \infty$. The renormalization transformations (2.1.1) and (2.1.2) extend to this setting directly, the only difference being that the renormalized potential values $(\mathcal{R}V)_y$ are now drawn from the densities $T_{p_1}(\varrho_{2y}, \varrho_{2y+1})$ of the random variables defined in (2.1.2).

To obtain a relation between H and $\mathcal{R}H$, we set $L = 2^{n-1} - 1$ and consider the orthonormal basis $\{e_0, ..., e_L, f_0, ..., f_L\}$ of $\ell^2(B_n)$ whose members are

$$e_y = \frac{1}{\sqrt{2}} \left(\delta_{2y} + \delta_{2y+1} \right), \quad f_y = \frac{1}{\sqrt{2}} \left(\delta_{2y} - \delta_{2y+1} \right).$$

Thus $\ell^2(B_n) = E \oplus F$, where E and F are the linear spans of $\{e_0, ..., e_L\}$ and $\{f_0, ..., f_L\}$, respectively. Let $U_e : \ell^2(B_{n-1}) \to E$ and $U_f : \ell^2(B_{n-1}) \to F$ be the isomorphisms defined by

$$U_e \delta_y = e_y, \quad U_f \delta_y = f_y$$

and let $U = U_e \oplus U_f$. A direct computation shows that a matrix representation of the form

$$U^* H_n U = \begin{pmatrix} (\mathcal{R}\Delta)_{n-1} + p_1 + V_{ee} & V_{fe} \\ V_{ef} & V_{ff} \end{pmatrix} \tag{2.1.3}$$

is valid in the site basis of $\ell^2(B_{n-1}) \oplus \ell^2(B_{n-1})$. The entries occurring on the right of (2.1.3) are the operators defined by

$$V_{ee}\delta_y = V_{ff}\delta_y = \frac{1}{2}(V_{2y} + V_{2y+1})\delta_y \tag{2.1.4}$$

and

$$V_{fe}\delta_y = V_{ef}\delta_y = \frac{1}{2}(V_{2y} - V_{2y+1})\delta_y \tag{2.1.5}$$

acting between the appropriate factors of $\ell^2(B_{n-1}) \oplus \ell^2(B_{n-1})$. Let us emphasize that (2.1.4) cannot be taken completely literally because V_{ee} maps only the first factor of $\ell^2(B_{n-1}) \oplus \ell^2(B_{n-1})$ into itself, whereas V_{ff} maps only the second factor into itself. Similar considerations apply to (2.1.5).

The Schur complement of $(\mathcal{R}\Delta)_{n-1} + p_1 + V_{ee}$ in (2.1.3) is the operator

$$(\mathcal{R}\Delta)_{n-1} + p_1 + V_{ee} - V_{fe}V_{ff}^{-1}V_{ef} = (\mathcal{R}H)_{n-1}$$

and thus the Schur complement formula for the inverse yields the following proposition.

Proposition 2.1.1. *The formula*

$$U^* H_n^{-1} U = \begin{pmatrix} 1 & 0 \\ -V_{ff}^{-1}V_{ef} & 1 \end{pmatrix} \begin{pmatrix} (\mathcal{R}H)_{n-1}^{-1} & 0 \\ 0 & V_{ff}^{-1} \end{pmatrix} \begin{pmatrix} 1 & -V_{fe}V_{ff}^{-1} \\ 0 & 1 \end{pmatrix}$$

is valid whenever $H_n, (\mathcal{R}H)_{n-1}$ and V_{ff} are invertible.

We will show in Section 2.4 that $T_p(\varrho, \tilde{\varrho}) \in L^\infty$ whenever $\varrho, \tilde{\varrho} \in L^\infty$ so the Wegner estimate applies to both H and $\mathcal{R}H$. Therefore, H_n and $(\mathcal{R}H)_{n-1}$ are almost surely invertible and V_{ff} is almost surely invertible because it is a multiplication operator whose entries are independent continuously distributed random variables. In terms of the operator

$$S = U_e^* - V_{fe}V_{ff}^{-1}U_f^*,$$

where we have identified U_e and U_f with $U_e \oplus 0$ and $0 \oplus U_f$, respectively, this proves the following important formula.

Corollary 2.1.2. *Let $\varphi, \psi \in \ell^2(B_n)$. Then*

$$\langle \varphi, H_n^{-1}\psi \rangle = \langle S\varphi, (\mathcal{R}H)_{n-1}^{-1} S\psi \rangle + \langle U_f^*\varphi, V_{ff}^{-1}U_f^*\psi \rangle$$

almost surely.

We will now use Corollary 2.1.2 to bound the fractional moments of the Green function

$$G_n(0,y;E) = \langle \delta_y, (H_n - E)^{-1} \delta_0 \rangle$$

by the fractional moments of its renormalized counterpart

$$\mathcal{R}G_n(0,y;E) = \langle \delta_y, ((\mathcal{R}H)_n - E)^{-1} \delta_0 \rangle \,.$$

To simplify the analysis, we will restrict ourselves to the case that some of the potential values V_y have a Cauchy distribution

$$\varrho_y(v) = P_{\mu+i\sigma}(v) := \frac{1}{\pi} \frac{\sigma}{(v-\mu)^2 + \sigma^2} \,. \tag{2.1.6}$$

In this case, a decoupling inequality becomes available (see [6, Thm. 8.7]), which states that for every $s \in (0,1)$ and $z \in \mathbb{C}_+$ there exists a constant $D_s(z) \in (0,\infty)$ with the property that

$$\frac{1}{D_s(z)} \int \frac{1}{|v-\gamma|^s} \, P_z(v) \, dv \le \int \frac{|v|^s}{|v-\gamma|^s} \, P_z(v) \, dv \tag{2.1.7}$$

$$\le D_s(z) \int \frac{1}{|v-\gamma|^s} \, P_z(v) \, dv \tag{2.1.8}$$

uniformly in $\gamma \in \mathbb{C}$. The proofs of these inequalities are based on the simple observation that the integrals occuring there define continuous functions of γ with the same limiting behavior as $|\gamma| \to \infty$. The restriction to the Cauchy case is possible thanks to a partial comparison trick which we will devise in the proof of Theorem 1.2.2.

Theorem 2.1.3. *Let $s \in (0,1)$ and $y \in B_{n-1} \setminus \{0\}$. If*

$$\varrho_0 = \varrho_1 = \varrho_{2y} = \varrho_{2y+1} = P_z,$$

then both

$$\mathbb{E}\,|G_n(0,2y;0)|^s \le D_s(z)^4 \, \mathbb{E}\,|\mathcal{R}G_{n-1}(0,y;0)|^s$$

and

$$\mathbb{E}\,|G_n(0,2y+1;0)|^s \le D_s(z)^4 \, \mathbb{E}\,|\mathcal{R}G_{n-1}(0,y;0)|^s.$$

Proof. We will prove only the estimate for $\mathbb{E}\,|G_n(0,2y;0)|^s$ since the analysis of $\mathbb{E}\,|G_n(0,2y+1;0)|^s$ then reduces to swapping some indices and changing some signs. The formulas

$$S\delta_{2\ell} = \frac{1}{\sqrt{2}} \left(1 - \frac{V_{2\ell} - V_{2\ell+1}}{V_{2\ell} + V_{2\ell+1}} \right) \delta_\ell$$

and

$$U_f^* \delta_{2\ell} = \delta_\ell$$

are valid for both $\ell = 0$ and $\ell = y$. Since $y \neq 0$, we necessarily have that $\langle \delta_0, V_{ff}^{-1} \delta_y \rangle = 0$ and hence Corollary 2.1.2 asserts that

$$G_n(0, 2y; 0) = 2 \frac{V_1}{V_0 + V_1} \frac{V_{2y+1}}{V_{2y} + V_{2y+1}} \mathcal{R}G_{n-1}(0, y; 0). \tag{2.1.9}$$

Consider a term of the form

$$X(\ell) = \mathbb{E}_{\,\ell} \left| \left(\frac{V_{2\ell+1}}{V_{2\ell} + V_{2\ell+1}} \right) \mathcal{R}G_{n-1}(0, y; 0) \right|^s,$$

where $\ell \in \{0, y\}$ and $\mathbb{E}_{\,\ell}$ denotes the conditional expectation with respect to $\{V_i \,|\, i \neq 2\ell, 2\ell + 1\}$. The Green function is of the form

$$|\mathcal{R}G_{n-1}(0, y; 0)|^s = \left| \frac{\alpha}{(\mathcal{R}V)_\ell - \beta} \right|^s$$

for some $\alpha, \beta \in \mathbb{C}$ which are independent of $(\mathcal{R}V)_\ell$ (cf. [6, Sec. 5.5.]). Writing

$$u = \left(\frac{1}{2v} + \frac{1}{2w} \right)^{-1},$$

it follows that

$$\begin{aligned} X(\ell) &= \iint \left| \frac{v}{v+w} \frac{\alpha}{u - \beta} \right|^s P_z(v) P_z(w) \, dv \, dw \\ &= \iint \left| \frac{\alpha v}{2vw - \beta(v+w)} \right|^s P_z(v) P_z(w) \, dv \, dw \\ &= \int \left| \frac{\alpha v}{2v - \beta} \right|^s \int \left| \frac{1}{w - \beta v (2v - \beta)^{-1}} \right|^s P_z(w) \, dw \, P_z(v) \, dv, \end{aligned}$$

where we have absorbed the shift of the renormalized potential by p_1 into the constant β. Applying the decoupling inequality (2.1.7) to the inner integral and reversing the previous calculations shows that

$$\begin{aligned} X(\ell) &\leq D_s(z) \int \left| \frac{\alpha v}{2v - \beta} \right|^s \int \left| \frac{w}{w - \beta v (2v - \beta)^{-1}} \right|^s P_z(w) \, dw \, P_z(v) \, dv \\ &= D_s(z) \iint \left| \frac{vw}{v+w} \frac{\alpha}{u - \beta} \right|^s P_z(v) P_z(w) \, dv \, dw \\ &= 2^{-s} D_s(z) \iint \left| \frac{\alpha u}{u - \beta} \right|^s P_z(v) P_z(w) \, dv \, dw \\ &= 2^{-s} D_s(z) \int \left| \frac{\alpha u}{u - \beta} \right|^s (T_0 P_z)(u) \, du. \end{aligned}$$

It is easy to see that $T_0 P_z = P_z$ (see also Section 2.4), so applying the decoupling inequality (2.1.8) yields

$$
\begin{aligned}
X(\ell) &\le 2^{-s} D_s(z)^2 \int \left| \frac{\alpha}{u-\beta} \right|^s (T_0 P_z)(u)\, du \\
&= 2^{-s} D_s(z)^2 \iint \left| \frac{\alpha}{u-\beta} \right|^s P_z(v) P_z(w)\, dv\, dw \\
&= 2^{-s} D_s(z)^2\, \mathbb{E}_{\,\ell} |\mathcal{R}G_{n-1}(0,y;0)|^s .
\end{aligned}
$$

Since V_0 and V_1 are independent of V_{2y} and V_{2y+1}, combining the bound for $X(\ell)$ with (2.1.9) implies

$$
\begin{aligned}
\mathbb{E}\, |G_n(0,2y,0)|^s &= 2^s\, \mathbb{E} \left(\left| \frac{V_1}{V_0+V_1} \right|^s X(y) \right) \\
&\le D_s(z)^2\, \mathbb{E} \left(\left| \frac{V_1}{V_0+V_1} \right|^s \mathbb{E}_{\,y} \, |\mathcal{R}G_{n-1}(0,y;0)|^s \right) \\
&= D_s(z)^2\, \mathbb{E}\, X(0) \\
&\le D_s(z)^4\, \mathbb{E}\, |\mathcal{R}G_{n-1}(0,y;0)|^s .
\end{aligned}
$$

□

Along with the restricted operators H_n, there is another sequence of truncations

$$
H_{n,m} = 1_{B_n} \left(\sum_{r=1}^{m} p_r E_r + V \right) 1_{B_n}, \tag{2.1.10}
$$

which will be useful in our proof of Theorem 1.2.4. Notice that

$$
H_n = H_{n,n} + \alpha |\varphi_n\rangle\langle\varphi_n|
$$

with

$$
\varphi_n = |B_n|^{-1/2} 1_{B_n}, \qquad \alpha = \sum_{r>n} 2^{n-r} p_r,
$$

so reasoning analogous to Corollary 2.1.2 shows that also

$$
\langle \varphi, H_{n,n}^{-1} \psi \rangle = \langle S\varphi, (\mathcal{R}H)_{n-1,n-1}^{-1} S\psi \rangle + \langle U_f^* \varphi, V_{ff}^{-1} U_f^* \psi \rangle \tag{2.1.11}
$$

almost surely. The formula (2.1.11) lets us determine the distribution of the quantity

$$
\frac{1}{\Phi_n(E)} = \langle \varphi_n, (H_{n,n} - E)^{-1} \varphi_n \rangle
$$

explicitly in terms of the operators T_p when the the disorder has the same distribution at each site, that is, $\varrho_y = \varrho$ for every $y \in \mathbb{N}_0$.

Corollary 2.1.4. *The density of $\Phi_n(E)$ is given by $T_{p_n}...T_{p_1}\varrho_E$.*

Proof. Notice that $S\varphi_n = \varphi_{n-1}$ and $U_f^*\varphi_n = 0$ so (2.1.11) shows that

$$\frac{1}{\Phi_n(0)} = \langle \varphi_{n-1}, (\mathcal{R}H)^{-1}_{n-1,n-1}\varphi_{n-1}\rangle = \frac{1}{\mathcal{R}\Phi_{n-1}(0)}.$$

We can continue renormalizing in this fashion until we reach a Hamiltonian consisting of a 1×1 random matrix whose element is distributed as $T_{p_n}...T_{p_1}\varrho$. This proves the result for $E = 0$. The general case follows by shifting the density of the original potential by $-E$. □

2.2 Proof of Localization

We begin our proof of Theorem 1.2.2 by considering Hamiltonians with single-site densities $\{\varrho_y \,|\, y \in \mathbb{N}_0\}$ which may vary from site to site, and proving a uniform high-disorder bound for the Green function in terms of the relative strengths of the hopping $|p_r| \leq \varepsilon 2^{-cr}$ and the disorder $\sup_{y\in\mathbb{N}_0} \|\varrho_y\|_\infty$.

Proposition 2.2.1. *If $s \in (0,1)$ and $\mu > 0$ satisfy*

$$1 + \mu < s(1 + c),$$

then there exist $\varepsilon_0 > 0$ and $C \in (0,\infty)$ such that

$$\sup_{y\in\mathbb{N}_0} 2^{(1+\mu)d(0,y)} \left(\sup_{n\geq 1} \mathbb{E}\,|G_n(0,y;0)|^s\right) \leq C$$

for any collection of single-site densities satisfying $\varepsilon\left(\sup_{i\in\mathbb{N}_0} \|\varrho_i\|_\infty\right) < \varepsilon_0$.

Proof. Since our method of proof is completely standard, and every detail of the argument can be found in a general setting in [6, Ch. 10], we provide only a sketch of the proof. Let G_Λ denote the Green function of the restriction of H to a finite volume $\Lambda \subset \mathbb{N}_0$. If $y \neq 0$, deleting the matrix elements $\{\Delta(y,x), \Delta(x,y) \,|\, x \neq y\}$ from Δ and applying the resolvent identity yields the formula

$$G_\Lambda(0,y;E) = -\sum_{x\neq y} G_\Lambda(y,y;E)\Delta(y,x)G_{\Lambda\setminus\{y\}}(0,x;E).$$

Let $M = \sup_{i\in\mathbb{N}_0} \|\varrho_i\|_\infty$. Factoring the expectation through the conditional expectation $\mathbb{E}_y$ with respect to V_y, we obtain

$$\mathbb{E}\,|G_\Lambda(0,y;E)|^s \leq \frac{2M^s}{1-s}\sum_{x\neq y} |\Delta(y,x)|^s \mathbb{E}\,|G_{\Lambda\setminus\{y\}}(0,x;E)|^s, \tag{2.2.1}$$

because $G_{\Lambda\setminus\{y\}}(0,x,E)$ does not depend on V_y and

$$\mathbb{E}_y|G_\Lambda(y,y;E)|^s \le \frac{2\|\varrho_y\|_\infty^s}{1-s}.$$

Setting

$$f(y) = \sup_{|\Lambda|<\infty} \mathbb{E}\,|G_\Lambda(0,y;0)|^s < \infty$$

and taking the supremum over all finite $\Lambda \subset \mathbb{N}_0$ in (2.2.1) yields

$$f(y) \le AM^s\left(\delta_{0,y} + \sum_{x\neq y}|\Delta(y,x)|^s f(x)\right) \tag{2.2.2}$$

with $A = 2/(1-s)$. The definition of Δ and the inequality $1+\mu < s(1+c)$ show that

$$\begin{aligned} AM^s \sup_{y\in\mathbb{N}_0}\sum_{x\in\mathbb{N}_0}\frac{2^{\mu d(0,x)}}{2^{\mu d(0,y)}}|\Delta(y,x)|^s &\le AM^s\sum_{x\in\mathbb{N}_0}2^{\mu d(0,x)}|\Delta(0,x)|^s \\ &\le A'\varepsilon^s M^s\sum_{x\in\mathbb{N}_0}2^{\mu d(0,x)}2^{-s(1+c)d(0,x)} \\ &= A'\varepsilon^s M^s\sum_{x\in\mathbb{N}_0}2^{(\mu-s(1+c))d(0,x)} \\ &\le A'\varepsilon^s M^s\sum_{r\ge 0}2^{(1+\mu-s(1+c))r} < 1 \end{aligned}$$

provided $\varepsilon M < \varepsilon_0$ is small enough. Hence, by iterating (2.2.2),

$$C = \sum_{y\in\mathbb{N}_0}2^{\mu d(0,y)}f(y) < \infty$$

which implies

$$\sup_n \sum_{y\in\mathbb{N}_0}2^{\mu d(0,y)}\mathbb{E}\,|G_n(0,y;0)|^s \le C.$$

The theorem now follows by observing that $|B_r \setminus B_{r-1}| = 2^{r-1}$ for all $r \ge 1$ and that $\mathbb{E}\,|G_n(0,y;0)|^s$ depends on y only in terms of $d(0,y)$. $\square$

We will now return to the setting of Theorem 1.2.2 in which the potential was identically distributed with a common density ϱ. Our strategy is to extend the conclusion of Theorem 2.2.1 to the entire parameter range by renormalizing into the high-disorder regime. This is based on the observation that, when $|p_r| \le \varepsilon 2^{-cr}$, the renormalized hopping $\mathcal{R}\mathbf{p}$ satisfies

$$|(\mathcal{R}\mathbf{p})_r| \le 2^{-c}\varepsilon 2^{-cr} \approx 2^{-c}|p_r| \tag{2.2.3}$$

so that the renormalization has effectively decreased ε by a factor 2^{-c}.

Theorem 2.2.2. *Suppose the assumption* (1.2.4) *is true in a bounded interval* $I \subset \mathbb{R}$*. If* $s \in (0,1)$ *and* $\mu > 0$ *satisfy*

$$1 + \mu < s(1 + c),$$

then there exists $C < \infty$ *such that*

$$\sup_{n \geq 1} \mathbb{E} \left| G_n(0, y; E) \right|^s \leq C\, 2^{-(1+\mu) d(0,y)}$$

for all $y \in \mathbb{N}_0$ *and* $E \in I$*.*

Proof. Since I is bounded, the requirement (1.2.1) means that there exist $z \in \mathbb{C}_+$ and $C_I < \infty$ such that

$$\varrho_E(v) \leq C_I P_z(v) \tag{2.2.4}$$

for all $E \in I$ and $v \in \mathbb{R}$, where P_z is the Poisson kernel defined in (2.1.6). The following bound for $G_n(0, y; 0)$ will depend only on z, C_I, $\|\varrho\|_\infty$, and the constants occuring in the assumption (1.2.4), which implies that we can restrict ourselves to the situation where $E = 0 \in I$ without any loss of generality.

Suppose $n \geq N \geq 1$ and let $y \in B_n$. We will first consider the Hamiltonian $H' = \Delta + V'$ which is obtained from H by replacing the potential values in $B_N(0) \cup B_N(y)$ by random variables with the Cauchy distribution P_z. Thus V_i' has the density

$$\varrho_i = \begin{cases} P_z & \text{if } i \in B_N(0) \cup B_N(y) \\ \varrho & \text{else} \end{cases}.$$

Since $T_p P_z = P_{z+p}$, the renormalized potential $\mathcal{R}^N V'$ has densities

$$\varrho_i = \begin{cases} P_{z+p_1+\ldots+p_N} & \text{if } i \in \left\{0, \lfloor 2^{-N} y \rfloor\right\} \\ T_{p_N} \ldots T_{p_1} \varrho & \text{else} \end{cases},$$

and by iterating the observation (2.2.3), $\mathcal{R}^N \Delta$ has a hopping strength

$$\left| (\mathcal{R}^N \mathbf{p})_r \right| = |p_{r+N}| \leq \varepsilon_N 2^{-cr}, \quad \varepsilon_N = 2^{-cN} \varepsilon.$$

Because ϱ satisfies the assumption (1.2.4) in I and $\|P_{z+p}\|_\infty \leq (\operatorname{Im} z)^{-1}$ for all $p \in \mathbb{R}$, this implies that the hypothesis

$$\varepsilon \left(\sup_{i \in \mathbb{N}_0} \|\varrho_i\|_\infty \right) < \varepsilon_0$$

of Proposition 2.2.1 is eventually satisfied by $\mathcal{R}^N H'$ for some sufficiently large N which depends on z and the constants in the assumption (1.2.4). Hence, when $1+\mu < s(1+c)$, there is some $C_0 < \infty$ such that

$$\sup_{n\geq 1} \mathbb{E}\left|\mathcal{R}^N G'_n\left(0, \lfloor 2^{-N} y\rfloor; 0\right)\right|^s \leq C_0\, 2^{-(1+\mu)d\left(0,\lfloor 2^{-N}y\rfloor\right)},$$

where G'_n denotes the Green function of H'_n. If $y \in B_n \setminus B_N$, then N successive applications of Theorem 2.1.3 show that

$$\mathbb{E}\,|G'_n(0,y;0)|^s \leq D\, C_0\, 2^{-(1+\mu)d\left(0,\lfloor 2^{-N}y\rfloor\right)} \leq D\, C_0\, 2^{-(1+\mu)(d(0,y)-N)}, \quad (2.2.5)$$

with

$$D = \left[D_s(z)D_s(z+p_1)...D_z(z+p_1+...+p_N)\right]^4.$$

Since H' is obtained from H by replacing $\{V_i \mid i \in B_N(0) \cup B_N(y)\}$ with random variables distributed according to P_z, (2.2.4) and (2.2.5) show that

$$\mathbb{E}\,|G_n(0,y;0)|^s \leq C_I^{2|B_N|}\mathbb{E}\,|G'_n(0,y;0)|^s \leq C_1\, 2^{-(1+\mu)d(0,y)} \quad (2.2.6)$$

for some $C_1 < \infty$ which depends on z, C_I, and the constants occurring in the assumption (1.2.4). If $y \in B_N$, then the a priori bound

$$\mathbb{E}\,|G_n(0,y;0)|^s \leq \frac{4\|\varrho\|_\infty^s}{1-s}$$

is valid so (2.2.6) implies that

$$\mathbb{E}\,|G_n(0,y;0)|^s \leq C 2^{-(1+\mu)d(0,y)}$$

with a constant $C < \infty$ depending only on z, C_I, $\|\varrho\|_\infty$, and the constants occurring in the assumption (1.2.4). $\square$

Theorem 1.2.2 is a consequence of the relationship between eigenfunction correlators and Green functions. Indeed, Theorem 1.2.2 follows from the standard result (see Chapter 7 in [6])

$$\mathbb{E}\, Q_n(x,y;I) \leq C_s\, \mathbb{E}\int_I |G_n(x,y;E)|^s\, dE$$

and the fact that $\mathbb{E}\,|G_n(x,y;E)|^s$ depends on x and y only in terms of $d(x,y)$.

2.3 Proof of Poisson Statistics

This section is devoted to the proof of Theorem 1.2.4 concerning the convergence of the random point measure

$$\mu_n(f) = \sum_{\lambda \in \sigma(H_n)} f(2^n(\lambda - E))$$

to a Poisson point process with intensity $\nu(E)$ when E is a Lebesgue point of the density of states. In this setting, H was a Hamiltonian with fixed hopping $|p_r| \leq \varepsilon 2^{-cr}$ and a single-site density $\varrho \in L^\infty$ such that the assumption (1.2.4) is valid in a neighborhood of E for some $\delta > 0$. Our argument is based on the following fundamental fact [6, Prop. 17.5], which essentially characterizes Poisson point processes as simple point processes consisting of infinitely many independent components.

Proposition 2.3.1. *Consider a sequence of point processes of the form* $\mu_n = \sum_j \mu_{n,j}$*, where* $\{\mu_{n,j} \,|\, j = 1, ..., N_n\}$ *is a triangular array of point processes with the following properties:*

i. The point processes $\{\mu_{n,1}, ..., \mu_{n,N_n}\}$ *are independent for all* $n \geq 1$*.*

ii. If $B \subset \mathbb{R}$ *is a bounded Borel set, then*

$$\lim_{n\to\infty} \sup_{j \leq N_n} \mathbb{P}(\mu_{n,j}(B) \geq 1) = 0.$$

iii. There exists some $c \geq 0$ *such that if* $B \subset \mathbb{R}$ *is a bounded Borel set with* $|\partial B| = 0$*, then*

$$\lim_{n\to\infty} \sum_{j=1}^{N_n} \mathbb{P}(\mu_{n,j}(B) \geq 1) = c|B|$$

and

$$\lim_{n\to\infty} \sum_{j=1}^{N_n} \mathbb{P}(\mu_{n,j}(B) \geq 2) = 0.$$

Then μ_n *converges in distribution to a Poisson point process with intensity* c*.*

Among the several equivalent options available [27, 56], we choose the definition that a sequence of point processes μ_n converges in distribution to μ provided

$$\lim_{n\to\infty} \mathbb{E}\, e^{-\mu_n(P_z)} = \mathbb{E}\, e^{-\mu(P_z)}$$

for all $z \in \mathbb{C}_+$, where P_z is the Poisson kernel (2.1.6). Hence, Theorem 1.2.4 can be established by finding a sequence $\tilde{\mu}_n$ such that Proposition 2.3.1 applies to $\tilde{\mu}_n$ and

$$\lim_{n\to\infty} \mathbb{E}\, e^{-\tilde{\mu}_n(P_z)} = \lim_{n\to\infty} \mathbb{E}\, e^{-\mu_n(P_z)}$$

for all $z \in \mathbb{C}_+$. The truncated operators $H_{n,m}$ (cf. (2.1.10)) provide a valuable tool in this endeavor because, for any $m \le k \le n$,

$$H_{n,m} = \bigoplus_{j=1}^{2^{n-k}} H_{k,m}^{(j)}, \tag{2.3.1}$$

and each $H_{k,m}^{(j)}$ is an independent copy of $H_{k,m}$. The relationship between $H_n, H_{n,n}$, and $H_{n,n-1}$ is essentially controlled by the quantity featured in the next lemma.

Lemma 2.3.2. *Let*

$$F_n(z) := \langle \varphi_n, (H_{n,n-1} - z)^{-1} \varphi_n \rangle$$

with $\varphi_n = 2^{-n/2} 1_{B_n}$ *and* $z \in \mathbb{C}_+$. *Then:*

i. φ_n *is almost surely cyclic for* $H_{n,n-1}$.

ii. If the assumption (1.2.4) *holds for* $I \subset \mathbb{R}$ *then there exists* $C < \infty$ *such that*

$$\mathbb{P}(|F_n(t)| \ge |\alpha|^{-1}) \le C\, 2^{(c-\delta)n}\, |\alpha|$$

for all $t \in I$ *and* $\alpha \neq 0$.

Proof. The vector φ_n is cyclic for $H_{n,n-1}$ if and only if

$$\operatorname{span}\{ f(H_{n,n-1})\varphi \mid f \in C_0 \} = \ell^2(B_n),$$

which is clearly true almost surely when $n = 1$. Now suppose the result is true for $H_{n,n-1}$. Since

$$H_{n,n} = H_{n,n-1} + p_n |\varphi_n\rangle\langle\varphi_n|,$$

φ_n is cyclic for $H_{n,n}$ whenever it is cyclic for $H_{n,n-1}$ [82]. It follows that $\varphi_{n+1} = \frac{1}{\sqrt{2}}(\varphi_n \oplus \varphi_n)$ is cyclic for $H_{n+1,n} = H_{n,n}^{(1)} \oplus H_{n,n}^{(2)}$ when the spectrum is simple, as is almost surely the case by the Minami estimate [69] (see also [6]).

For the second part, recall Lemma 2.1.4, which asserts that

$$\Phi_{n-1}(t) = \left(\langle \varphi_{n-1}, (H_{n-1,n-1} - t)^{-1} \varphi_{n-1} \rangle \right)^{-1}$$

is a random variable with density $T_{p_{n-1}} ... T_{p_1} \varrho_t$. Since $F_n(t)$ is an average of two independent copies of $(\Phi_{n-1}(t))^{-1}$, we have:

$$\begin{aligned} \mathbb{P}(|F_n(t)| \ge |\alpha|^{-1}) &\le 2\, \mathbb{P}(|\Phi_{n-1}(t)| \le |\alpha|) \\ &\le 4|\alpha|\, \|T_{p_{n-1}} ... T_{p_1} \varrho_t\|_\infty \\ &= |\alpha| \cdot \mathcal{O}\left(2^{(c-\delta)n} \right) \end{aligned}$$

where the last estimate holds uniformly in $t \in I$ thanks to the assumption (1.2.4). □

Our next goal is to understand how the finite-volume density of states

$$\nu_n(f) = 2^{-n}\mathrm{Tr}\, f(H_n)$$

is approximated by its analogue

$$\nu_{n,m}(f) = 2^{-n}\mathrm{Tr}\, f(H_{n,m}),$$

which is the content of Theorem 2.3.3 below. For its statement, we introduce the notation

$$z_\ell := E + 2^{-\ell} z$$

for all $z \in \mathbb{C}_+$ and $\ell \geq 0$. The connection between Theorem 1.2.4 and ν_n is through the formula

$$\mu_n(P_z) = \nu_n(P_{z_n}). \tag{2.3.2}$$

Theorem 2.3.3. *Suppose the assumption* (1.2.4) *is satisfied in an open set* $I \subset \mathbb{R}$ *and* $E \in I$*. Let* $z \in \mathbb{C}_+$ *and set* $o_\ell(z) := \int_{I^c} P_{z_\ell}(t)dt$*. Then:*

i. There is some $\varepsilon > 0$ *such that* $o_\ell(z) \leq \frac{2\mathrm{Im}\, z}{\varepsilon\, \pi}\, 2^{-\ell}$*. In particular,* $o_\ell(z)$ *is a null sequence for any* $z \in \mathbb{C}_+$ *as* $\ell \to \infty$*.*

ii. There is some $C < \infty$*, which does not depend on* n, m, ℓ *or* z*, such that for all* $m \leq n$*:*

$$\mathbb{E}\, |\nu_{n,n}(P_{z_\ell}) - \nu_{n,m}(P_{z_\ell})| \leq C\, (\mathrm{Im}\, z)^{-1}\, 2^{\ell-m} \left(2^{-\delta m} + o_\ell(z)\right), \tag{2.3.3}$$

$$\mathbb{E}\, |\nu_n(P_{z_\ell}) - \nu_{n,n}(P_{z_\ell})| \leq C\, (\mathrm{Im}\, z)^{-1}\, 2^{\ell-n} \left(2^{-\delta n} + o_\ell(z)\right). \tag{2.3.4}$$

Proof. The first assertion follows from the fact that there is a ε-neighborhood of $E \in I$ which is fully contained in I together with a simple explicit computation.

For a proof of the second assertion we set $\alpha = p_n$ so that

$$H_{n,n} = H_{n,n-1} + \alpha|\varphi_n\rangle\langle\varphi_n|.$$

Since φ_n is almost surely cyclic for $H_{n,n-1}$, the theory of rank-one perturbations [82] shows that the following statements are valid:

- The eigenvalues of $H_{n,n-1}$ coincide with the set of poles of F_n.
- The eigenvalues of $H_{n,n}$ coincide with the set $\{E \in \mathbb{R} \,|\, F_n(E) = -\alpha^{-1}\}$.
- The function F_n is monotone increasing between its poles.

For the sake of clarity, let us spell out the proof only in case $\alpha > 0$ (the case $\alpha < 0$ being similar). Setting $W = \{t \in \mathbb{R} \,|\, F_n(t) \leq -\alpha^{-1}\}$, the fundamental theorem of calculus implies that

$$\nu_{n,n}(P_{z_\ell}) - \nu_{n,n-1}(P_{z_\ell}) = 2^{-n} \int 1_W(t) P'_{z_\ell}(t)\, dt.$$

Since $|P'_z(t)| \leq (\operatorname{Im} z)^{-1} P_z(t)$, taking the expected value yields

$$\begin{aligned}\mathbb{E}\, |\nu_{n,n}(P_{z_\ell}) - \nu_{n,n-1}(P_{z_\ell})| &\leq 2^{\ell-n} (\operatorname{Im} z)^{-1}\, \mathbb{E} \int 1_W(t) P_{z_\ell}(t)\, dt \\ &= 2^{\ell-n} (\operatorname{Im} z)^{-1} \int P_{z_\ell}(t)\, \mathbb{P}(F_n(t) \leq -\alpha^{-1})\, dt.\end{aligned}$$

Because $\alpha = \mathcal{O}(2^{-cn})$, Lemma 2.3.2 asserts that $\mathbb{P}(F_n(t) \leq -\alpha^{-1}) \leq C\, 2^{-\delta n}$ for all $t \in I$ so that

$$\int P_{z_\ell}(t)\, \mathbb{P}(F_n(t) \leq -\alpha^{-1})\, dt \leq C\, 2^{-\delta n} + o_l(z). \tag{2.3.5}$$

This proves (2.3.3) when $m = n - 1$. Moreover, setting $\alpha = \sum_{r>n} 2^{n-r} p_r = \mathcal{O}(2^{-cn})$ and repeating the argument above with ν_n in place of $\nu_{n,n}$ and $\nu_{n,n}$ in place of $\nu_{n,n-1}$ proves (2.3.4).

For a proof of (2.3.3), we expand in a telescopic sum, i.e., for any $f \in C_0$

$$\nu_n(f) - \nu_{n,m}(f) = \nu_n(f) - \nu_{n,n-1}(f) + \sum_{k=m+1}^{n-1} (\nu_{n,k}(f) - \nu_{n,k-1}(f))$$

and

$$\begin{aligned}\nu_{n,k}(f) - \nu_{n,k-1}(f) &= 2^{-n} \left(\operatorname{Tr} f(H_{n,k}) - \operatorname{Tr} f(H_{n,k-1})\right) \\ &= 2^{-(n-k)} \sum_{j=1}^{2^{n-k}} 2^{-k} \left(\operatorname{Tr} f(H^{(j)}_{k,k}) - \operatorname{Tr} f(H^{(j)}_{k,k-1})\right) \\ &= 2^{-(n-k)} \sum_{j=1}^{2^{n-k}} \left(\nu^{(j)}_{k,k}(f) - \nu^{(j)}_{k,k-1}(f)\right) \qquad (2.3.6)\end{aligned}$$

because of the decomposition (2.3.1). Taking moments and noticing that each term in (2.3.6) has the same distribution yields

$$\begin{aligned}\mathbb{E}\, |\nu_n(P_{z_\ell}) - \nu_{n,m}(P_{z_\ell})| &\leq C\, (\operatorname{Im} z)^{-1}\, 2^{\ell-n} \left(2^{-\delta n} + o_\ell(z)\right) \\ &\quad + \sum_{k=m+1}^{n-1} C\, (\operatorname{Im} z)^{-1}\, 2^{\ell-(k-1)} \left(2^{-\delta(k-1)} + o_\ell(z)\right) \\ &\leq C\, (\operatorname{Im} z)^{-1}\, 2^{\ell-m} \left(2^{-\delta m} + o_\ell(z)\right).\end{aligned}$$

□

By the Wegner estimate, the measures $\mathbb{E}\,\nu_n$ and $\mathbb{E}\,\nu_{n,m}$ are absolutely continuous with densities that are uniformly bounded independently of n and m, and (2.3.2) shows that the same is true of $\mathbb{E}\,\mu_n$. Moreover, by ergodicity [6, 62, 63],

$$\lim_{n\to\infty} \mathbb{E}\,\nu_n(f) = \nu(f) \tag{2.3.7}$$

for all $f \in C_0$. We will now show that this limit also exists with ν_n replaced by μ_n.

Corollary 2.3.4. *If E is a Lebesgue point of ν, then*

$$\lim_{n\to\infty} \mathbb{E}\,\mu_n(B) = \nu(E)|B|$$

for all bounded Borel sets $B \subset \mathbb{R}$.

Proof. That E is a Lebesgue point of ν means that

$$\lim_{n\to\infty} 2^n \nu(2^{-n}B + E) = \nu(E)|B|$$

so it suffices to prove the relation

$$\lim_{n\to\infty} \big(\mathbb{E}\,\mu_n(B) - 2^n\nu(2^{-n}B + E)\big) = 0. \tag{2.3.8}$$

Since B is bounded, 1_B can be approximated arbitrarily well in L^1 by finite linear combinations from the set $\{P_z \,|\, z \in \mathbb{C}_+\}$. Moreover, since the measures occurring in (2.3.8) are absolutely continuous with densities bounded uniformly in n, we conclude that it is enough to show (2.3.8) with B replaced by P_z. By (2.3.2), this is equivalent to

$$\lim_{n\to\infty} (\mathbb{E}\,\nu_n(P_{z_n}) - \nu(P_{z_n})) = 0.$$

Applying (2.3.7) and the fact that $\mathbb{E}\,\nu_{p,n} = \mathbb{E}\,\nu_{n,n}$ for any $p \geq n$ (cf. (2.3.1)) we conclude from Theorem 2.3.3 that

$$\begin{aligned}\lim_{n\to\infty} |\mathbb{E}\,\nu_n(P_{z_n}) - \nu(P_{z_n})| &= \lim_{n\to\infty}\lim_{p\to\infty} |\mathbb{E}\,[\nu_n(P_{z_n}) - \nu_p(P_{z_n})]| \\ &= \lim_{n\to\infty}\lim_{p\to\infty} |\mathbb{E}\,[\nu_{p,n}(P_{z_n}) - \nu_p(P_{z_n})]| = 0.\end{aligned}$$

□

The next corollary defines the approximating processes $\tilde{\mu}_n$ alluded to earlier.

Corollary 2.3.5. *There exists a sequence m_n with $m_n \to \infty$ and $0 < n-m_n \to \infty$ such that the measure defined by*

$$\tilde{\mu}_n(P_z) = \nu_{n,m_n}(P_{z_n})$$

satisfies

$$\lim_{n\to\infty} \mathbb{E}\, |\mu_n(P_z) - \tilde{\mu}_n(P_z)| = 0$$

for all $z \in \mathbb{C}_+$.

Proof. Using (2.3.2) and Theorem 2.3.3 we see that

$$\begin{aligned}\mathbb{E}\, |\mu_n(P_z) - \tilde{\mu}_n(P_z)| &= \mathbb{E}\, |\nu_n(P_{z_n}) - \nu_{n,m_n}(P_{z_n})| \\ &\leq C(\operatorname{Im} z)^{-1}\left[2^{-\delta n} + 2^{n-(1+\delta)m_n} + \frac{2\operatorname{Im} z}{\varepsilon\,\pi}\left(2^{-n} + 2^{-m_n}\right)\right].\end{aligned}$$

Since $\delta > 0$, we can choose m_n such that $m_n \to \infty$, $n - m_n \to \infty$ and $n - (1+\delta)m_n \to -\infty$ which proves the result. $\square$

Combining the fact that $|e^{-t_1} - e^{-t_2}| \leq |t_1 - t_2|$ when $t_1, t_2 \geq 0$ with Corollary 2.3.5 implies that $\tilde{\mu}_n$ satisfies

$$\lim_{n\to\infty} \mathbb{E}\, e^{-\tilde{\mu}_n(P_z)} = \lim_{n\to\infty} \mathbb{E}\, e^{-\mu_n(P_z)}.$$

It thus remains to show that $\tilde{\mu}_n$ satisfies the hypothesis of Proposition 2.3.1. In the interest of readability, we will suppress the dependence on n and write simply m in place of m_n for the remainder of this section. By (2.3.1), $\tilde{\mu}_n$ is a sum of independent point processes

$$\tilde{\mu}_n = \sum_{j=1}^{2^{n-m}} \tilde{\mu}_{n,j}$$

with

$$\tilde{\mu}_{n,j}(B) = \operatorname{Tr} 1_{2^{-n}B+E}(H^{(j)}_{m,m})$$

for all Borel sets $B \subset \mathbb{R}$. By a theorem of Combes-Germinet-Klein [26] (cf. [6, 69]),

$$\mathbb{P}(\operatorname{Tr} 1_B(H_{m,m}) \geq \ell) \leq \frac{(C\, 2^m |B|)^\ell}{\ell!}$$

which implies

$$\mathbb{P}(\tilde{\mu}_{n,j}(B) \geq \ell) \leq \frac{(C|B|\, 2^{m-n})^\ell}{\ell!}. \tag{2.3.9}$$

Since $n-m \to \infty$, this shows immediately that the first requirement of Proposition 2.3.1 is satisfied. For the other requirements, let us abbreviate

$$X(n,\ell) = \sum_{j=1}^{2^{n-m}} \mathbb{P}(\tilde{\mu}_{n,j}(B) \geq \ell)$$

so that (2.3.9) implies

$$X(n,\ell) \leq 2^{n-m} \frac{(C|B|\, 2^{m-n})^\ell}{\ell!} \to 0$$

when $\ell \geq 2$. In particular, $X(n,2) \to 0$ and the last assumption of Proposition 2.3.1 is satisfied. Since $\tilde{\mu}_{n,j}(B)$ takes values in the non-negative integers

$$\lim_{n\to\infty} X(n,1) = \lim_{n\to\infty} \sum_{j=1}^{2^{n-m}} \mathbb{E}\, \tilde{\mu}_{n,j}(B) - \lim_{n\to\infty} \sum_{\ell\geq 2} X(n,\ell).$$

By (2.3.9) and the dominated convergence theorem

$$\lim_{n\to\infty} X(n,1) = \lim_{n\to\infty} \sum_{j=1}^{2^{n-m}} \mathbb{E}\, \tilde{\mu}_{n,j}(B) = \lim_{n\to\infty} \mathbb{E}\, \tilde{\mu}_n(B),$$

so to finish the proof it suffices to derive the identity

$$\lim_{n\to\infty} \mathbb{E}\, \tilde{\mu}_n(B) = \nu(E)|B| \tag{2.3.10}$$

for any bounded Borel set B. Since $\tilde{\mu}_n(B)$ and μ_n are easily seen to have uniformly bounded densities, we can approximate 1_B by linear combinations from $\{P_z \mid z \in \mathbb{C}_+\}$ and use Corollary 2.3.5 to see that

$$\lim_{n\to\infty} \mathbb{E}\, |\tilde{\mu}_n(B) - \mu_n(B)| = 0$$

for all bounded Borel sets B. Thus we can replace $\mathbb{E}\, \tilde{\mu}_n(B)$ with $\mathbb{E}\, \mu_n(B)$ in (2.3.10) and Corollary 2.3.4 concludes the proof of Theorem 1.2.4.

2.4 The Renormalized Density

This section consists of the proofs of several previous claims regarding the renormalized densities $T_{p_r} ... T_{p_1} \varrho_E$. Let us start by proving the claim in Section 2.1, that $T_p(\varrho, \tilde{\varrho})$ is bounded if ϱ and $\tilde{\varrho}$ are.

Lemma 2.4.1. *Suppose $\varrho, \tilde{\varrho} \in L^\infty$ are probability densities. Then*

$$\|T_p(\varrho, \tilde{\varrho})\|_\infty \leq \|\varrho\|_\infty + \|\tilde{\varrho}\|_\infty$$

for any $p \in \mathbb{R}$.

Proof. Notice that

$$\mathbb{E}_{T_0 \varrho} f := \int f(v) T_0(\varrho, \tilde{\varrho})(v)\, dv = \int f\left(\frac{2vw}{v+w}\right) \varrho(v)\tilde{\varrho}(w)\, dv\, dw$$

for any sufficiently regular $f \in L^1$ and

$$\left(\frac{\partial}{\partial v} + \frac{\partial}{\partial w}\right)\left(\frac{2vw}{v+w}\right) = 2\frac{v^2+w^2}{(v+w)^2} \geq 1.$$

Thus $\mathbb{E}_{T_0 \varrho} f$ is bounded by

$$\begin{aligned} &\int f\left(\frac{2vw}{v+w}\right) \varrho(v)\tilde{\varrho}(w) \frac{\partial}{\partial v}\left(\frac{2vw}{v+w}\right) dv\, dw \\ &\quad + \int f\left(\frac{2vw}{v+w}\right) \varrho(v)\tilde{\varrho}(w) \frac{\partial}{\partial w}\left(\frac{2vw}{v+w}\right) dv\, dw \\ &= \int f(x)\varrho(v(x))\tilde{\varrho}(w) dx\, dw + \int f(x)\varrho(v)\tilde{\varrho}(w(x)) dx\, dv \\ &\leq (\|\varrho\|_\infty + \|\tilde{\varrho}\|_\infty)\, \|f\|_1. \end{aligned}$$

Hence $\|T_0(\varrho, \tilde{\varrho})\|_\infty \leq \|\varrho\|_\infty + \|\tilde{\varrho}\|_\infty$ and the lemma follows from the translation invariance of the norm. □

We will now consider the validity of the assumption (1.2.4) for the special cases

- $c > 1$,
- V has a Gaussian distribution and $c > 1/2$,
- V has a Cauchy component and $c > 0$,

as mentioned in the introduction. The case $c > 1$ is an easy consequence of Lemma 2.4.1 since

$$\|T_{p_r} \ldots T_{p_1} \varrho_E\|_\infty \leq 2^r \|\varrho_E\|_\infty = 2^r \|\varrho\|_\infty$$

so the assumption (1.2.4) is true with $I = \mathbb{R}$ and $\delta = c - 1 > 0$.

Our analysis of the Gaussian distribution $\mathcal{N}(\mu, \sigma)$ is based on the following observations:

- If $V \in \mathbb{R}^L$ is a random vector with independent $\mathcal{N}(0, \sigma)$ entries and $O : \mathbb{R}^L \to \mathbb{R}^L$ is an orthogonal matrix, then OV also consists of independent $\mathcal{N}(0, \sigma)$ entries.

- If $F : \mathbb{C}_+ \to \mathbb{C}_+$ is a singular Herglotz function and $A \subset \mathbb{R}$ is a Borel set, then
$$\int 1_A(F(t+i0))P_z(t)\,dt = \int 1_A(t)P_{F(z)}(t)\,dt$$
where P_z is the Poisson kernel corresponding to $z \in \mathbb{C}_+$ (cf. [5]).

Let $\varphi_r = 2^{-r/2}(1,...,1) \in \mathbb{R}^{|B_r|}$ be the unit vector with constant entries. By rotation invariance, there exists a random vector $Z \in \varphi_r^\perp$ and an independent scalar Gaussian $g \sim \mathcal{N}(0,\sigma)$ such that

$$V = g\varphi_r + \mu_r\varphi_r + Z$$

where $\mu_r = 2^{r/2}\mu$. Since there exist some $z \in \mathbb{C}_+$ and $C < \infty$ such that the $\mathcal{N}(0,\sigma)$ density is dominated pointwise by $C\,P_z$, this implies that for any bounded Borel set $A \subset \mathbb{R}$

$$\mathbb{E}\,1_A(\mathcal{R}^r V) \leq C \int 1_A(\mathcal{R}^r(t\varphi_r + \mu_r\varphi_r + Z))\,P_z(t)\,dt\,\xi(dZ),$$

where ξ is some probability distribution on $\varphi_r^\perp$. Notice that $\mathcal{R}^r(V)$ is a singular Herglotz function of each of the variables $V_0, ..., V_{2^r-1}$ with the property

$$\operatorname{Im}\mathcal{R}^r(V) \geq \min\{\operatorname{Im}V_k \mid 0 \leq k \leq 2^r - 1\}$$

which follows from the definition of $\mathcal{R}(V)$ and the fact that

$$\operatorname{Im}\left(\frac{1}{2z} + \frac{1}{2w}\right)^{-1} \geq \min\{\operatorname{Im}z, \operatorname{Im}w\}.$$

Thus $F(t) = \mathcal{R}^r(t\varphi_r + \mu_r\varphi_r + Z)$ is a singular Herglotz function of t when μ and Z are fixed. Hence

$$\begin{aligned}\mathbb{E}\,1_A(\mathcal{R}^r V) &\leq C \iint 1_A(t)P_{F(z)}(t)\,dt\,\xi(dZ)\\ &\leq \frac{C}{\operatorname{Im}F(z)}|A|\\ &\leq C2^{r/2}|A|,\end{aligned}$$

which proves that $\|T_{p_r}...T_{p_0}\varrho_E\|_\infty \leq C\,2^{r/2}$ uniformly in $E \in \mathbb{R}$ because the previous estimates did not depend on μ. Thus the assumption (1.2.4) is true with $I = \mathbb{R}$ and $\delta = c - 1/2$.

Finally, we consider the case where ϱ is a mixture of Poisson kernels, i.e.,

$$\varrho = \int_{\mathbb{C}_+} P_z\,\mu(dz) \tag{2.4.1}$$

for some probability measure $\mu \in M(\mathbb{C}_+)$. A simple calculation, which is described in some detail in [60], shows that

$$T_p\left(\int_{\mathbb{C}_+} P_z\,\mu(dz)\right) = \int_{\mathbb{C}_+} P_z\,T_p\mu(dz)$$

and that $\operatorname{supp} T_p\mu \subset \{z \in \mathbb{C}_+ \,|\, \operatorname{Im} z > \varepsilon\}$ if $\operatorname{supp}\mu \subset \{z \in \mathbb{C}_+ \,|\, \operatorname{Im} z > \varepsilon\}$. In particular, if ϱ is of the form (2.4.1) with $\operatorname{supp}\mu \subset \{z \in \mathbb{C}_+ \,|\, \operatorname{Im} z > \varepsilon\}$, then

$$\|T_{p_r}...T_{p_1}\varrho_E\|_\infty \leq \varepsilon^{-1},$$

which proves the assumption (1.2.4) with $I = \mathbb{R}$ and $\delta = c$. By definition, V has a Cauchy component if $\varrho = \mu * P_z$ for some $z \in \mathbb{C}_+$ and some probability measure $\mu \in M(\mathbb{R})$, which is a special case of (2.4.1).

2.5 Eigenfunction Correlators and IPRs

The purpose of this section is to prove two statements made in the introduction regarding the behavior of the IPR in a regime of eigenfunction correlator localization. The arguments here do not rely on the specifics of the hierarchical model. First, let us present the proof of Corollary 1.2.3, which bounds the probability of the event

$$A = \left\{ \begin{array}{ll} \text{There is } \psi \in \ell^2(B_n) \text{ with } H_n\psi = \lambda\psi \text{ and} & \\ |\lambda - E| \leq 2^{-n-1}W \text{ such that } \ P_2(\psi) \leq \varepsilon^4 & \end{array} \right\}.$$

Proof of Corollary 1.2.3. Let $J_n = \{|\lambda - E| \leq 2^{-n-1}W\}$ and let ψ_λ denote the eigenfunction associated to an eigenvalue $\lambda \in \sigma(H_n)$. Then it follows from

$$1 = \|\psi_\lambda\|_2^2 \leq \|\psi_\lambda\|_4 \|\psi_\lambda\|_1 \tag{2.5.1}$$

that

$$\begin{aligned} \mathbb{P}(A) &\leq \mathbb{P}\left(\sum_{\lambda\in\sigma(H_n)\cap J_n} \frac{1}{P_2(\psi_\lambda)^{1/4}} > \frac{1}{\varepsilon}\right) \qquad (2.5.2)\\ &\leq \varepsilon\,\mathbb{E} \sum_{\lambda\in\sigma(H_n)\cap J_n} \frac{1}{\|\psi_\lambda\|_4} \leq \varepsilon\,\mathbb{E} \sum_{\lambda\in\sigma(H_n)\cap J_n} \|\psi_\lambda\|_1. \end{aligned}$$

Since $\sum_{y\in B_n}|\psi_\lambda(y)|^2=1$, we have

$$\begin{aligned}\mathbb{E}\,2^{-n}\sum_{\lambda\in\sigma(H_n)\cap J_n}\|\psi_\lambda\|_1 &= 2^{-n}\sum_{y\in B_n}\mathbb{E}\sum_{x\in B_n}\sum_{\lambda\in\sigma(H_n)\cap J_n}|\psi_\lambda(y)|^2|\psi_\lambda(x)| \\ &\le 2^{-n}\sum_{y\in B_n}\sum_{x\in B_n}Q_n(y,x;J_n) \\ &\le 2^{-n}\sum_{y\in B_n}C|J_n| = C|J_n|\end{aligned} \tag{2.5.3}$$

in regimes of eigenfunction correlator localization. Plugging (2.5.3) into (2.5.2), we obtain

$$\mathbb{P}(A)\le\varepsilon\,\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap J_n}\|\psi_\lambda\|_1\le\varepsilon\,2^nC|J_n| = C\,W\varepsilon.$$

□

Our other goal is to prove that eigenfunction correlator localization implies the lower bound (1.2.8) for the averaged IPR

$$\Pi_n(I)=\frac{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}\|\psi_\lambda\|_4^4}{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}1}.$$

Using (2.5.1) term by term yields

$$\Pi_n(I)\ge\frac{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}\|\psi_\lambda\|_1^{-4}}{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}1}.$$

We now apply Jensen's inequality with the probability measure defined by

$$\mu(f)=\frac{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}f(\lambda)}{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}1}$$

and the convex function $\Phi(x)=x^{-4}$ to see that

$$\begin{aligned}\Pi_n(I) &\ge \left(\frac{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}\|\psi_\lambda\|_1}{\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}1}\right)^{-4} \\ &= \left(\frac{2^{-n}\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}1}{2^{-n}\mathbb{E}\sum_{\lambda\in\sigma(H_n)\cap I}\|\psi_\lambda\|_1}\right)^{4}.\end{aligned}$$

The numerator of this expression is equal to

$$
\begin{aligned}
2^{-n}\mathbb{E}\,\mathrm{Tr}\,1_I(H_n) &= 2^{-n}\sum_{y\in B_n}\mathbb{E}\,\langle\delta_y, 1_I(H_n)\delta_y\rangle \\
&= \mathbb{E}\,\langle\delta_0, 1_I(H_n)\delta_0\rangle = \nu_n(I)
\end{aligned}
$$

so repeating the calculation (2.5.3) with I in place of J_n shows that

$$
\Pi_n(I) \geq C^{-4}\left(\frac{\nu_n(I)}{|I|}\right)^4 \tag{2.5.4}
$$

as desired.

2.6 Spectral Localization

This section contains the completion of an argument by E. Kritchevski [60], which proves that the spectrum of H is almost surely of pure-point type with eigenfunctions satisfying

$$
\sum_{y\in\mathbb{N}_0} 2^{\frac{c}{4}d(0,y)}|\psi(y)|^2 < \infty
$$

for any parameters $(\mathbf{p},\varrho)$ without relying on the assumption (1.2.4). The following should be regarded as an accompanying note to [60] and thus we will not present the entire argument in detail, but simply cite the theorems of [60] as necessary. The argument makes use of the truncations $H_{n,m}$ in the $n\to\infty$ limit:

$$
H_{\infty,n} = \sum_{r=1}^{n} p_r E_r + V.
$$

Notice that, for fixed realizations of V,

$$
\|H - H_{\infty,n}\| \leq \sum_{r=n+1}^{\infty}|p_r|
$$

and hence

$$
\lim_{n\to\infty}(H_{\infty,n}-z)^{-1}\delta_x = (H-z)^{-1}\delta_x
$$

for any $x\in\mathbb{N}_0$ and $z\in\mathbb{C}_+$. We will be particularly interested in the quantities

$$
\begin{aligned}
G_n(x,y;z) &= \langle\delta_y, (H_{\infty,n}-z)^{-1}\delta_x\rangle \\
g_n(x;z) &= 2^{-n}\sum_{y\in B_n(x)}\langle\delta_y, (H_{\infty,n}-z)^{-1}\delta_x\rangle \\
Q_n(x,z) &= \langle\varphi_n(x), (H_{\infty,n}-z)^{-1}\varphi_n(x)\rangle
\end{aligned}
$$

with $y \in B_n(x)$ and $\varphi_n(x) = 2^{-n/2} 1_{B_n(x)}$. Proposition 2.2 of [60] contains the formula

$$G_n(x, y; z) = G_0(x, y; z) - \sum_{r=d(x,y)}^{n} 2^{r-1} p_r g_{r-1}(x; z) g_r(y; z).$$

Letting $w(y) = 2^{\mu d(x,y)}$ and using the triangle inequality for the w-weighted ℓ^2-norm, this implies that

$$S(x, n, \mu) := \left(\sum_{y \in \mathbb{N}_0} w(y) |G_n(x, y; E)|^2 \right)^{1/2}$$

is bounded by

$$\begin{aligned} & |G_0(x, x; E)| + \sum_{r=1}^{n} 2^{r-1} |p_r| |g_{r-1}(x; E)| \left(\sum_{d(x,y) \le r} w(y) |g_r(y; E)|^2 \right)^{1/2} \\ \le & |G_0(x, x; E)| + \sum_{r=1}^{n} |p_r| 2^{\mu r} 2^{r-1} |g_{r-1}(x; E)| \left(\sum_{d(x,y) \le r} |g_r(y; E)|^2 \right)^{1/2}. \end{aligned}$$

Provided that

$$\sum_{r=1}^{\infty} |p_r Q_r(x, E)| < \infty \tag{2.6.1}$$

for $\mathbb{P} \otimes m$-almost every $(\omega, E) \in \Omega \times \mathbb{R}$, Proposition 2.3 and the proof of Proposition 2.4 in [60] show that for almost every (ω, E) there exist constants $C(\omega, E), C'(\omega, E) < \infty$ such that

$$\left(\sum_{d(x,y) \le r} |g_r(y; E)|^2 \right)^{1/2} \le C(\omega, E)\, 2^{\frac{c}{4} r}$$

and

$$2^{r-1} |g_{r-1}(x; E)| \le C'(\omega, E)$$

for all $r \ge 1$. It follows that

$$S(x, n, \mu) \le |G_0(x, x; E)| + C(\omega, E) C'(\omega, E) \sum_{r=1}^{n} 2^{-cr} 2^{\mu r} 2^{\frac{c}{4} r}. \tag{2.6.2}$$

and since $G_0(x,x;E)$ exists and is finite for almost every (ω, E), choosing $\mu = 0$ we obtain

$$\sup_{n\geq 1} \|(H_{\infty,n} - E)^{-1}\delta_x\| = \sup_{n\geq 1} S(x,n,0) < \infty$$

for almost every (ω, E). Applying the monotone convergence theorem to the spectral measures of δ_x for H and $H_{\infty,n}$ shows that

$$\begin{aligned}
\lim_{\varepsilon\to 0} \|(H - E - i\varepsilon)^{-1}\delta_x\| &= \sup_{\varepsilon>0} \|(H - E - i\varepsilon)^{-1}\delta_x\| \\
&\leq \sup_{\varepsilon>0} \sup_{n\geq 1} \|H_{\infty,n} - E - i\varepsilon)^{-1}\delta_x\| \\
&= \sup_{n\geq 1} \sup_{\varepsilon>0} \|(H_{\infty,n} - E - i\varepsilon)^{-1}\delta_x\| \\
&= \sup_{n\geq 1} \|(H_{\infty,n} - E)^{-1}\delta_x \ < \infty,
\end{aligned}$$

and thus the Simon-Wolff Criterion [84] asserts that the spectrum of H is almost surely of pure-point type. If G denotes the Green function of the full operator H, then

$$\begin{aligned}
|G(x,y;z) - G_n(x,y;z)| &= \left|\langle \delta_x, (H_{\infty,n} - z)^{-1}(H - H_{\infty,n})(H - z)^{-1}\delta_y\rangle\right| \\
&\leq \|H - H_{\infty,n}\| \|(H_{\infty,n} - z)^{-1}\delta_x\| \|(H - z)^{-1}\delta_y\|
\end{aligned}$$

and the preceding argument proves that for almost all (ω, E) we can take first $\operatorname{Im} z \to 0$ and then $n \to \infty$ so that $G_n(x,y;E) \to G(x,y;E)$ for almost all (ω, E). Applying Fatou's lemma to (2.6.2) with $\mu = \frac{c}{4}$ we see that

$$\begin{aligned}
\left(\sum_{y\neq 0} 2^{\frac{c}{4}d(x,y)}|G(x,y;E)|^2\right)^{1/2} &\leq C(\omega,E)C'(\omega,E) \sup_{n\geq 0} \sum_{r=1}^{n} 2^{-\frac{c}{2}r} \\
&< \infty,
\end{aligned}$$

and a trivial modification of Theorem 9 from [84] now shows that the eigenfunctions of H satisfy

$$\sum_{y\in\mathbb{N}_0} 2^{\frac{c}{4}d(0,y)}|\psi(y)|^2 < \infty$$

as well.

Thus it remains to prove (2.6.1). Since $Q_r(x;z)$ is the Borel transform of a singular probability measure, Boole's inequality [6, Prop. 8.2] shows that

$$(\mathbb{P}\otimes m)(\{|Q_n(x;E)| > 2^{\frac{c}{2}r}\}) = 2\cdot 2^{-\frac{c}{2}r}.$$

It follows from the Borel-Cantelli lemma that

$$|Q_n(x;E)| \leq C''(\omega,E)2^{\frac{c}{2}r}$$

with $C''(\omega,E) < \infty$ on a set of full $\mathbb{P}\otimes m$ measure, and this implies (2.6.1).

3 The Ultrametric Ensemble and Local Stability of Dyson Brownian Motion

In this chapter we prove results on Gaussian perturbations stated in Section 1.5 and then use them to derive the localization theorems for the ultrametric ensemble and Rosenzweig-Porter model.

3.1 Smoothing Effects of Random Potentials

Throughout this section we will let H be a general $N \times N$ random matrix of the form

$$H = \tilde{H} + V,$$

where V is a potential satisfying the assumption (1.5.2) and $\tilde{H}$ is some symmetric random matrix independent of V, which should be thought of as $\tilde{H} + \Phi_T$ from (1.5.1). Our goal is to use the smoothing effects of V on the spectral measures μ_{xy} of δ_x and δ_y for H and the empirical eigenvalue measure

$$\nu(f) = \frac{1}{N} \sum_{\lambda \in \sigma(H)} f(\lambda)$$

to control the resolvent flow (1.5.3). The following lemma is a simple extension of the proof of Proposition 1.6.3 by Combes, Germinet, and Klein. We write $|\mu|$ for the total variation measure of μ.

Lemma 3.1.1. *There exists $C < \infty$, depending only on C_V, such that*

1. $\mathbb{E}\,|\mu_{xy}|(I) \leq C|I|$ *and*
2. $\mathbb{E}\,[\nu(I)|\mu_{xy}|(J)] \leq C\left(|I| + \frac{2}{N}\right)|J|$

for all Borel sets $I, J \subset \mathbb{R}$ and $x, y \in \{1, ..., N\}$.

Proof. Notice that

$$|\mu_{xy}|(I) = \sum_{\lambda \in \sigma(H) \cap I} |\psi_\lambda(x)\psi_\lambda(y)|\,,$$

so the Cauchy-Schwarz inequality implies

$$|\mu_{xy}|(I) \leq \sqrt{\mu_x(I)\mu_y(I)}.$$

Applying the Cauchy-Schwarz inequality to the expectation $\mathbb{E}_{xy}$ conditioned on $\{V(k) : k \neq x, y\}$ and using Proposition 1.6.1 then yield

$$\mathbb{E}_{xy}\,|\mu_{xy}|(I) \leq \mathbb{E}_{xy}\sqrt{\mu_x(I)\mu_y(I)} \leq \sqrt{\mathbb{E}_{xy}\,\mu_x(I)\mathbb{E}_{xy}\,\mu_y(I)} \leq C|I|, \qquad (3.1.1)$$

which implies the first assertion of the Lemma.

For the second claim, notice that for fixed values $\{V(k) : k \neq x, y\}$ of the potential away from x and y, the number of eigenvalues in I can change by at most two as $V(x)$ and $V(y)$ vary in $\mathbb{R}$. Hence

$$\begin{aligned}\mathbb{E}\left[\nu(I)|\mu_{xy}|(J)\right] &\leq \mathbb{E}\left[\left(\nu(I)+\frac{2}{N}\right)\mathbb{E}_{xy}\,|\mu_{xy}|(J)\right] \\ &\leq C|J|\mathbb{E}\left[\nu(I)+\frac{2}{N}\right] \leq C\left(|I|+\frac{2}{N}\right)|J|,\end{aligned}$$

by Proposition 1.6.2 and (3.1.1). □

Intuitively, Lemma 3.1.1 asserts that the joint measure $\mathbb{E}\left[\nu \times |\mu_{xy}|\right]$ is continuous down to scales of order N^{-1}, which clearly has consequences for the integrals of test functions in terms of their variations on scales of order N^{-1}. The next results are a quantitative manifestation of this idea for the Stieltjes transforms

$$G(x,y;z) = \int \frac{1}{\lambda - z}\,\mu_{xy}(d\lambda)$$

and

$$S(z) = \int \frac{1}{\lambda - z}\,\nu(d\lambda),$$

which occur naturally in our study of the resolvent flow. In particular, the following theorem gives bounds for the drift.

Theorem 3.1.2. *There exists $C < \infty$, depending only on C_V, such that*

$$\mathbb{E}\left|\frac{1}{2N}\frac{\partial^2}{\partial z^2}G(x,y;z)\right| \leq \frac{C}{N(\operatorname{Im} z)^2}$$

and

$$\mathbb{E}\left|S(z)\frac{\partial}{\partial z}G(x,y;z)\right| \leq CN\left(\log N + \frac{1}{N\operatorname{Im} z}\right)\left(1+\frac{1}{(N\operatorname{Im} z)^2}\right) + \frac{C}{\operatorname{Im} z}$$

for all $x, y \in \{1, ..., N\}$ and $z \in \mathbb{C}_+$.

Proof. The first point of Lemma 3.1.1 implies that

$$\begin{aligned}\mathbb{E}\left|\frac{1}{2N}\frac{\partial^2}{\partial z^2}G(x,y;z)\right| &\leq \frac{1}{N\operatorname{Im} z}\mathbb{E}\int\frac{1}{|\lambda - z|^2}\,|\mu_{x,y}|(d\lambda) \\ &\leq \frac{C}{N\operatorname{Im} z}\int\frac{1}{|\lambda - z|^2}\,d\lambda \\ &\leq \frac{C}{N(\operatorname{Im} z)^2},\end{aligned} \tag{3.1.2}$$

which is the first assertion of the theorem.

Next, let us introduce

$$f(\lambda) = \frac{1_{|\lambda - \operatorname{Re} z| \le 1}}{|\lambda - z|}, \quad \tilde{f}(\lambda) = \frac{1_{|\lambda - \operatorname{Re} z| > 1}}{|\lambda - z|}, \quad g(\lambda) = \frac{1}{|\lambda - z|^2}$$

so that

$$\mathbb{E} \left| S(z) \frac{\partial}{\partial z} G(x, y; z) \right| \le \mathbb{E} \iint \left(f(\lambda_1) + \tilde{f}(\lambda_1) \right) g(\lambda_2) \, \nu(d\lambda_1) |\mu_{xy}|(d\lambda_2).$$

Setting $I_\alpha = \operatorname{Re} z + [\alpha/N, (\alpha+1)/N)$,

$$\begin{aligned}
&\mathbb{E} \iint f(\lambda_1) g(\lambda_2) \, \nu(d\lambda_1) |\mu_{xy}|(d\lambda_2) \\
&\le \sum_{\alpha, \beta \in \mathbb{Z}} \left(\sup_{\lambda \in I_\alpha} f(\lambda) \right) \left(\sup_{\lambda \in I_\beta} g(\lambda) \right) \mathbb{E} \left[\nu(I_\alpha) |\mu_{xy}|(I_\beta) \right] \\
&\le \frac{C}{N^2} \sum_{\alpha, \beta \in \mathbb{Z}} \left(\sup_{\lambda \in I_\alpha} f(\lambda) \right) \left(\sup_{\lambda \in I_\beta} g(\lambda) \right),
\end{aligned}$$

where we used the second part of Lemma 3.1.1 to bound the expectations. Since f and g are symmetric about $\operatorname{Re} z$ and monotone decreasing in $|\lambda - \operatorname{Re} z|$, the previous chain of inequalities continues

$$\begin{aligned}
&\le \frac{4C}{N^2} \sum_{\alpha, \beta \in \mathbb{N}_0} f\left(\operatorname{Re} z + \frac{\alpha}{N} \right) g\left(\operatorname{Re} z + \frac{\beta}{N} \right) \\
&= CN \sum_{\alpha=0}^{N} \frac{1}{\sqrt{\alpha^2 + (N \operatorname{Im} z)^2}} \sum_{\beta \in \mathbb{N}_0} \frac{1}{\beta^2 + (N \operatorname{Im} z)^2} \\
&\le CN \left(\log N + \frac{1}{N \operatorname{Im} z} \right) \left(1 + \frac{1}{(N \operatorname{Im} z)^2} \right).
\end{aligned}$$

Finally, because $|\tilde{f}| \le 1$, the remaining summands satisfy

$$\mathbb{E} \iint \tilde{f}(\lambda_1) g(\lambda_2) \, \nu(d\lambda_1) |\mu_{xy}|(d\lambda_2) \le \mathbb{E} \int \frac{1}{|\lambda - z|^2} \, |\mu_{xy}|(d\lambda) \le \frac{C}{\operatorname{Im} z},$$

arguing as in (3.1.2). □

Evaluating the trace defining $S(z)$ in the site basis,

$$S(z) = \frac{1}{N} \sum_y G(y, y; z),$$

we may average the bounds furnished by Theorem 3.1.2 to obtain the following corollary, which gives the corresponding bounds for the drift of the trace of the resolvent.

Corollary 3.1.3. *There exists $C < \infty$, depending only on C_V, such that*

$$\mathbb{E}\left|\frac{1}{2N}\frac{\partial^2}{\partial z^2}S(z)\right| \leq \frac{C}{N(\operatorname{Im} z)^2}$$

and

$$\mathbb{E}\left|S(z)\frac{\partial}{\partial z}S(z)\right| \leq CN\left(\log N + \frac{1}{N\operatorname{Im} z}\right)\left(1 + \frac{1}{(N\operatorname{Im} z)^2}\right) + \frac{C}{\operatorname{Im} z}$$

for all $z \in \mathbb{C}_+$.

We conclude this section with a bound in the same spirit as the previous results for a term which does not explicitly occur in the resolvent flow, but which will nevertheless prove useful in controlling the diffusion of (1.5.3).

Theorem 3.1.4. *There exists $C < \infty$, depending only on C_V, such that*

$$\mathbb{E}\left[\operatorname{Im} G(x,x;z)\operatorname{Im} S(z)\right] \leq C\left(N\operatorname{Im} z + \frac{1}{N\operatorname{Im} z}\right)^2$$

for all $x \in \{1, ..., N\}$ and $z \in \mathbb{C}_+$.

Proof. The proof follows along the same lines as that of Theorem 3.1.2. Setting $I_\alpha = \operatorname{Re} z + [\alpha/N, (\alpha+1)/N)$, letting

$$P_z(\lambda) = \operatorname{Im}\frac{1}{\lambda - z} = \frac{\operatorname{Im} z}{(\lambda - \operatorname{Re} z)^2 + (\operatorname{Im} z)^2} \tag{3.1.3}$$

denote the rescaled Poisson kernel, and using Lemma 3.1.1, we see that

$$\begin{aligned}
\mathbb{E}\left[\operatorname{Im} G(x,x;z)\operatorname{Im} S(z)\right] &= \mathbb{E}\iint P_z(\lambda_1)P_z(\lambda_2)\,\nu(d\lambda_1)\mu_x(d\lambda_2) \\
&\leq \sum_{\alpha,\beta\in\mathbb{Z}}\left(\sup_{\lambda\in I_\alpha} P_z(\lambda)\right)\left(\sup_{\lambda\in I_\beta} P_z(\lambda)\right)\mathbb{E}\left[\nu(I_\alpha)\mu_x(I_\beta)\right] \\
&\leq \frac{C}{N^2}\sum_{\alpha,\beta\in\mathbb{Z}}\left(\sup_{\lambda\in I_\alpha} P_z(\lambda)\right)\left(\sup_{\lambda\in I_\beta} P_z(\lambda)\right).
\end{aligned}$$

Since P_z is symmetric about $\operatorname{Re} z$ and monotone decreasing in $|\lambda - \operatorname{Re} z|$, the last term is in turn bounded by

$$\begin{aligned}
&\leq \frac{4C}{N^2} \sum_{\alpha,\beta \in \mathbb{N}_0} P_z\left(\operatorname{Re} z + \frac{\alpha}{N}\right) P_z\left(\operatorname{Re} z + \frac{\beta}{N}\right) \\
&= C \sum_{\alpha \in \mathbb{N}_0} \frac{N \operatorname{Im} z}{\alpha^2 + (N \operatorname{Im} z)^2} \sum_{\beta \in \mathbb{N}_0} \frac{N \operatorname{Im} z}{\beta^2 + (N \operatorname{Im} z)^2} \\
&\leq C \left(N \operatorname{Im} z + \frac{1}{N \operatorname{Im} z}\right)^2 .
\end{aligned}$$

□

3.2 Local Stability of Dyson Brownian Motion

In this section, we turn to the proofs of our main results for Gaussian perturbations, Theorems 1.5.1 and 1.5.2. We start by deriving the stochastic differential equations (1.5.3) for the resolvent $R_t(z)$ in terms of the Green functions and the normalized trace. For this, we define the martingales

$$dM_t(x, y; z) = -\frac{1}{\sqrt{N}} \sum_{u \leq v} \langle \delta_y, R_t(z) P_{uv} R_t(z) \delta_x \rangle \, dB_{uv}(t),$$

where

$$P_{uv} = \frac{1}{\sqrt{1 + \delta_{uv}}} \left(|\delta_u\rangle\langle\delta_v| + |\delta_v\rangle\langle\delta_u| \right) = \sqrt{N} \frac{\partial}{\partial B_{uv}} H_t$$

denotes the symmetric matrix element corresponding to $\{\delta_u, \delta_v\}$.

Theorem 3.2.1. *The Green function satisfies*

$$dG_t(x, y; z) = \left(S_t(z) \frac{\partial}{\partial z} G_t(x, y; z) + \frac{1}{2N} \frac{\partial^2}{\partial z^2} G_t(x, y; z) \right) dt + dM_t(x, y; z)$$

for all $x, y \in \{1, ..., N\}$ and $z \in \mathbb{C}_+$.

Proof. By the resolvent equation,

$$\frac{\partial}{\partial B_{uv}} R_t(z) = -\frac{1}{\sqrt{N}} R_t(z) P_{uv} R_t(z),$$

so using Itô's Lemma shows that

$$\begin{aligned}
dG_t(x,y;z) &= \frac{1}{N}\sum_{u\le v}\langle \delta_y, R_t(z)P_{uv}R_t(z)P_{uv}R_t(z)\delta_x\rangle\, dt \\
&\quad - \frac{1}{\sqrt{N}}\sum_{u\le v}\langle \delta_y, R_t(z)P_{uv}R_t(z)\delta_x\rangle\, dB_{uv}(t) \\
&= \frac{1}{N}\sum_{u\le v}\langle \delta_y, R_t(z)P_{uv}R_t(z)P_{uv}R_t(z)\delta_x\rangle\, dt + dM_t(x,y;z).
\end{aligned}$$

We expand the drift term as

$$\begin{aligned}
&\frac{1}{N}\sum_{u<v}\langle \delta_y, R_t(z)\delta_v\rangle\langle \delta_u, R_t(z)\delta_u\rangle\langle \delta_v, R_t(z)\delta_x\rangle + \langle \delta_y, R_t(z)\delta_u\rangle\langle \delta_v, R_t(z)\delta_v\rangle\langle \delta_u, R_t(z)\delta_x\rangle \\
&+ \frac{1}{N}\sum_{u<v}\langle \delta_y, R_t(z)\delta_v\rangle\langle \delta_u, R_t(z)\delta_v\rangle\langle \delta_u, R_t(z)\delta_x\rangle + \langle \delta_y, R_t(z)\delta_u\rangle\langle \delta_v, R_t(z)\delta_u\rangle\langle \delta_v, R_t(z)\delta_x\rangle \\
&+ \frac{2}{N}\sum_{u}\langle \delta_y, R_t(z)\delta_u\rangle\langle \delta_u, R_t(z)\delta_u\rangle\langle \delta_u, R_t(z)\delta_x\rangle
\end{aligned}$$

and exploit that the second term in each sum is the same as the first term with u and v interchanged to rewrite these sums as

$$\begin{aligned}
&= \frac{1}{N}\sum_{u,v}\langle \delta_y, R_t(z)\delta_v\rangle\langle \delta_u, R_t(z)\delta_u\rangle\langle \delta_v, R_t(z)\delta_x\rangle \\
&+ \frac{1}{N}\sum_{u,v}\langle \delta_y, R_t(z)\delta_v\rangle\langle \delta_u, R_t(z)\delta_v\rangle\langle \delta_u, R_t(z)\delta_x\rangle.
\end{aligned}$$

In the second sum, we use that the spectral measures μ_{vu} are real to replace $\langle \delta_u, R_t(z)\delta_v\rangle$ with $\langle \delta_v, R_t(z)\delta_u\rangle$, which yields

$$\begin{aligned}
&= \langle \delta_y, R_t(z)^2\delta_x\rangle \frac{1}{N}\operatorname{Tr} R_t(z) + \frac{1}{N}\langle \delta_y, R_t(z)^3\delta_x\rangle \\
&= S_t(z)\frac{\partial}{\partial z}G_t(x,y;z) + \frac{1}{2N}\frac{\partial^2}{\partial z^2}G_t(x,y;z).
\end{aligned}$$

□

We remark that the applicability of these arguments to GUE perturbations in place of GOE perturbations is not affected by the last part of the proof, which made use of the fact that the spectral measures are real in the GOE case. This is because the additional unitary symmetry ensures that the third order term involving $\langle \delta_y, R_t(z)^3\delta_x\rangle$ vanishes completely for the GUE flow.

By averaging the evolution of $G_t(x,x;z)$ over $x \in \{1,...,N\}$, we obtain an equation with a diffusion given by

$$M_t(z) = \frac{1}{N}\sum_x M_t(x,x;z),$$

which is the familiar complex Burgers equation for $S_t(z)$ [9].

Corollary 3.2.2. *The normalized trace satisfies*

$$dS_t(z) = \left(S_t(z)\frac{\partial}{\partial z}S_t(z) + \frac{1}{2N}\frac{\partial^2}{\partial z^2}S_t(z)\right) dt + dM_t(z)$$

for all $z \in \mathbb{C}_+$.

We will now employ the results of Section 3.1 to smooth the resolvent flow of Theorem 3.2.1. Theorem 3.1.2 and Corollary 3.1.3 already accomplish this for the drift, but some further analysis based on spatial averaging is required to control the diffusion and this is the content of the next two theorems.

Theorem 3.2.3. *There exists a constant $C < \infty$, depending only on C_V, such that*

$$\frac{1}{N}\sum_y \mathbb{E}\,|M_T(x,y;z)| \le C\sqrt{\frac{T}{N(\operatorname{Im} z)^2}}\left(N\operatorname{Im} z + \frac{1}{N\operatorname{Im} z}\right)$$

for all $x \in \{1,...,N\}$, $z \in \mathbb{C}_+$ and $T \ge 0$.

Proof. The quadratic variation of $M_t(x,y;z)$ satisfies

$$\begin{aligned}
[M_T(x,y;z)] &= \frac{1}{N}\int_0^T \sum_{u\le v} |\langle \delta_y, R_s(z)P_{uv}R_s(z)\delta_x\rangle|^2\, ds \\
&\le \frac{2}{N}\int_0^T \sum_{u,v} |\langle \delta_y, R_s(z)\delta_u\rangle\langle \delta_v, R_s(z)\delta_x\rangle|^2\, ds \\
&= \frac{2}{N}\int_0^T \left(\sum_u |\langle \delta_y, R_s(z)\delta_u\rangle|^2\right)\left(\sum_v |\langle \delta_v, R_s(z)\delta_x\rangle|^2\right) ds \\
&= \frac{2}{N(\operatorname{Im} z)^2}\int_0^T \operatorname{Im} G_s(x,x;z)\operatorname{Im} G_s(y,y;z)\, ds,
\end{aligned}$$

where we combined the symmetrization argument of Theorem 3.2.1 with the inequality $(a+b)^2 \le 2(a^2+b^2)$. Hence

$$\begin{aligned}
\frac{1}{N}\sum_y \mathbb{E}\,|\,[M_t(x,y;z)]\,| &\le \frac{2}{N(\operatorname{Im} z)^2}\int_0^T \mathbb{E}\,[\operatorname{Im} G_s(x,x;z)\operatorname{Im} S_s(z)]\, ds \\
&\le \frac{CT}{N(\operatorname{Im} z)^2}\left(N\operatorname{Im} z + \frac{1}{N\operatorname{Im} z}\right)^2
\end{aligned}$$

by Theorem 3.1.4. Combining the Burkholder-Davis-Gundy inequality with Jensen's inequality for $\frac{1}{N}\sum_y \mathbb{E}$ shows that

$$\frac{1}{N}\sum_y \mathbb{E}\,|M_T(x,y;z)| \leq C\left(\frac{1}{N}\sum_y \mathbb{E}\,[M_T(x,y;z)]\right)^{1/2}$$
$$\leq C\sqrt{\frac{T}{N(\operatorname{Im} z)^2}}\left(N\operatorname{Im} z + \frac{1}{N\operatorname{Im} z}\right).$$

□

Next, we state the corresponding result for the averaged martingale

$$M_t(z) = \frac{1}{N}\sum_x M_t(x,x;z)$$

occuring in Corollary 3.2.2.

Theorem 3.2.4. *There exists a constant $C < \infty$, depending only on C_V, such that*

$$\mathbb{E}\,|M_T(z)| \leq \sqrt{\frac{CT}{N^2(\operatorname{Im} z)^3}}$$

for all $z \in \mathbb{C}_+$ and $T \geq 0$.

Proof. By symmetrization,

$$\begin{aligned}
M_T(z) &= \frac{1}{N}\sum_x M_T(x,x;z) \\
&= -\frac{1}{N^{3/2}}\sum_{u,v}\frac{1}{\sqrt{1+\delta_{uv}}}\int_0^T \sum_x \langle \delta_v, R_s(z)\delta_x\rangle\langle\delta_x, R_s(z)\delta_u\rangle\, dB_{uv}(s) \\
&= -\frac{1}{N^{3/2}}\sum_{u,v}\frac{1}{\sqrt{1+\delta_{uv}}}\int_0^T \frac{\partial}{\partial z}\langle \delta_v, R_s(z)\delta_u\rangle\, dB_{uv}(s),
\end{aligned}$$

so the quadratic variation may be expressed as

$$\begin{aligned}
[M_T(z)] &= \frac{1}{N^3}\int_0^T \sum_{u,v}\frac{1}{1+\delta_{uv}}\left|\frac{\partial}{\partial z}\langle\delta_v, R_s(z)\delta_u\rangle\right|^2 ds \\
&\leq \frac{1}{N^3(\operatorname{Im} z)^2}\int_0^T \sum_{u,v}|\langle\delta_v, R_s(z)\delta_u\rangle|^2\, ds \\
&= \frac{1}{N^2(\operatorname{Im} z)^3}\int_0^T \operatorname{Im} S_s(z)\, ds.
\end{aligned}$$

Using, in order, the Burkholder-Davis-Gundy inequality, Jensen's inequality, and Proposition 1.6.2 yields

$$
\begin{aligned}
\mathbb{E}\,|M_T(z)| &\le C\left(\mathbb{E}\,[M_T(z)]\right)^{1/2} \\
&\le C\left(\frac{1}{N^2(\operatorname{Im} z)^3}\int_0^T \mathbb{E}\operatorname{Im} S_s(z)\,ds\right)^{1/2} \\
&\le \sqrt{\frac{CT}{N^2(\operatorname{Im} z)^3}}.
\end{aligned}
$$

□

The proofs of Theorems 1.5.1 and 1.5.2 now reduce to plugging the various previous estimates into the integrated forms of Theorem 3.2.1 and Corollary 3.2.2. For the sake of completeness, we illustrate this with the proof of Theorem 1.5.2, but omit the very similar proof of Theorem 1.5.1.

Proof of Theorem 1.5.2. By Theorem 3.2.1,

$$
\begin{aligned}
&\frac{1}{N}\sum_y \mathbb{E}\,\left|G_T\left(x,y;E+i\eta\right)-G_0\left(x,y;E+i\eta\right)\right| \\
&\le \frac{1}{N}\sum_y\int_0^T \mathbb{E}\left|S_s(z)\frac{\partial}{\partial z}G_s(x,y;z)+\frac{1}{2N}\frac{\partial^2}{\partial z^2}G_s(x,y;z)\right| ds \\
&+\frac{1}{N}\sum_y \mathbb{E}\,|M_T(x,y;z)|,
\end{aligned}
$$

which by Theorems 3.1.2 and 3.2.3 is bounded by

$$
\begin{aligned}
&\le CTN\left(\log N+\frac{1}{N\eta}\right)\left(1+\frac{1}{(N\eta)^2}\right)+\frac{CT}{\eta}+\frac{CT}{N\eta^2} \\
&+C\sqrt{\frac{T}{N\eta^2}}\left(N\eta+\frac{1}{N\eta}\right).
\end{aligned}
$$

After taking a factor $N^{-\varepsilon/2}$ from $T \le N^{-(1+\varepsilon)}$ to control the $\log N$ term, each term is dominated by either $1+(N\eta)^{-1}$ or $1+(N\eta)^{-3}$, which proves the theorem. □

3.3 Proof of Poisson Statistics in the Ultrametric Ensemble

In the remainder of this chapter, we will show how to apply Theorems 1.5.1 and 1.5.2 to the ultrametric ensemble H_n defined in (1.3.1), thereby obtaining Theorems 1.3.1 and 1.3.2. When $c > 0$, the limit $\lim_{n\to\infty} Z_{n,c} \in (0,\infty)$ exists, and thus we may drop the normalizing constant $Z_{n,c}$ from the definition of H_n without any loss of generality. We will prove Theorem 1.3.1 by approximating $H_n \equiv \sum_{r=0}^{n} 2^{-\frac{1+c}{2}r}\Phi_{n,r}$ with the truncated Hamiltonian

$$H_{n,m} = \sum_{r=0}^{m} 2^{-\frac{1+c}{2}r}\Phi_{n,r}, \tag{3.3.1}$$

which has the property that, for any $m \leq k \leq n$,

$$H_{n,m} = \bigoplus_{j=1}^{2^{n-k}} H_{k,m}^{(j)}, \tag{3.3.2}$$

where each $H_{k,m}^{(j)}$ is an independent copy of H_k. Therefore

$$\mu_{n,m}(f) = \sum_{\lambda\in\sigma(H_{n,m})} f(2^n(\lambda - E))$$

consists of 2^{n-m} independent components, a fact whose relevance to Theorem 1.3.1 is contained in the characterization of Poisson point processes given by Proposition 2.3.1.

Hence, as with the hierarchical Anderson model, Theorem 1.3.1 follows by furnishing a sequence m_n such that Proposition 2.3.1 applies to μ_{n,m_n} and

$$\lim_{n\to\infty} \mathbb{E}\, e^{-\mu_{n,m_n}(P_z)} = \lim_{n\to\infty} \mathbb{E}\, e^{-\mu_n(P_z)} \tag{3.3.3}$$

for all $z \in \mathbb{C}_+$. The difference $H_n - H_{n,n-1} = \sqrt{T}\,\Phi_{n,n}$ is a Gaussian perturbation with time parameter $T = 2^{-(1+c)n}$. Therefore, Theorem 1.5.1 shows that there exists $C_z < \infty$ such that for all $\ell \geq n$ we have

$$\begin{aligned}\frac{1}{2^n}\,\mathbb{E}\,\left|\mathrm{Tr}\,(H_n - z_\ell)^{-1} - \mathrm{Tr}\,(H_{n,n-1} - z_\ell)^{-1}\right| &\leq C_z\, 2^{-\frac{c}{2}n-1}\left(1 + 2^{3(\ell-n)}\right)\\ &\leq C_z\, 2^{-\frac{c}{2}n}\, 2^{3(\ell-n)}\end{aligned} \tag{3.3.4}$$

with $z_\ell = E + 2^{-\ell}z$. Our strategy in achieving (3.3.3) thus consists of applying (3.3.4) to the finite-volume density of states measures

$$\nu_n(f) = 2^{-n}\mathrm{Tr}\, f\,(H_n)\,, \quad \nu_{n,m}(f) = 2^{-n}\mathrm{Tr}\, f\,(H_{n,m})$$

in an iterative fashion.

Theorem 3.3.1. *There exist $C_z < \infty$ and $\delta > 0$ such that*

$$\mathbb{E}\,\left|\nu_n\left(P_{z_\ell}\right) - \nu_{n,m}\left(P_{z_\ell}\right)\right| \leq C_z\, 2^{3(\ell-(1+\delta)m)}$$

for all $\ell \geq n$.

Proof. The estimate (3.3.4) proves that

$$\mathbb{E}\,\left|\nu_k\left(P_{z_\ell}\right)-\nu_{k,k-1}\left(P_{z_\ell}\right)\right|\le C_z\,2^{3(\ell-(1+\delta)k)} \tag{3.3.5}$$

with $\delta=c/6$ when $\ell\ge k$. Since $\nu_n-\nu_{n,m}$ is given by a telescopic sum,

$$\nu_n(P_{z_\ell})-\nu_{n,m}(P_{z_\ell})=\sum_{k=m+1}^{n}\left(\nu_{n,k}(P_{z_\ell})-\nu_{n,k-1}(P_{z_\ell})\right),$$

the decomposition (3.3.2) implies that

$$\nu_{n,k}(P_{z_\ell})-\nu_{n,k-1}(P_{z_\ell})=2^{-(n-k)}\sum_{j=1}^{2^{n-k}}\left(\nu_k(P_{z_\ell})-\nu_{k,k-1}(P_{z_\ell})\right). \tag{3.3.6}$$

Applying (3.3.5) to each term in (3.3.6) yields

$$\mathbb{E}\,\left|\nu_n\left(P_{z_\ell}\right)-\nu_{n,m}\left(P_{z_\ell}\right)\right|\le\sum_{k=m+1}^{n}C_z\,2^{3(\ell-(1+\delta)k)}\le C_z\,2^{3(\ell-(1+\delta)m)}.$$

□

Theorem 3.3.1 enables us to find a suitable sequence μ_{n,m_n} satisfying (3.3.3).

Corollary 3.3.2. *There exists a sequence m_n with $m_n\to\infty$ and $n-m_n\to\infty$ such that*

$$\lim_{n\to\infty}\mathbb{E}\,\left|\mu_n(P_z)-\mu_{n,m_n}(P_z)\right|=0$$

for all $z\in\mathbb{C}_+$.

Proof. Since $\delta>0$, there exists a sequence m_n with $m_n\to\infty$, $n-m_n\to\infty$ and $n-(1+\delta)m_n\to-\infty$. By applying Theorem 3.3.1 with $\ell=n$, we obtain

$$\mathbb{E}\,\left|\mu_n(P_z)-\mu_{n,m_n}(P_z)\right|\le C_z2^{3(n-(1+\delta)m_n)}\to 0.$$

□

To finish the proof of Theorem 1.3.1, we need to show that μ_{n,m_n} satisfies the hypothesis of Proposition 2.3.1, which works exactly the same way as in the case of the hierarchical Anderson model above. We omit the details, which may be copied verbatim from the previous analysis.

3.4 Proof of Localization in the Ultrametric Ensemble

In this section, we prove Theorem 1.3.2 by comparing the eigenfunctions of H_n with the obviously localized eigenfunctions of $H_{n,m}$. Nevertheless, we again start by considering a more general $N \times N$ random matrix $H = \tilde{H} + V$ with a potential satisfying (1.5.2) and proving an implication of local resolvent bounds for the eigenfunction correlator

$$Q(x, y; W) = \sum_{\lambda \in \sigma(H) \cap W} |\psi_\lambda(x) \psi_\lambda(y)|$$

in some mesoscopic spectral window

$$W = \left[E_0 - N^{-(1-w)}, E_0 + N^{-(1-w)}\right]$$

with $w > 0$.

Theorem 3.4.1. *Let $\eta = N^{-(1+\ell)}$ with $\ell > w > 0$ and let $Y \subset \{1, ..., N\}$. Then, there exists a constant $C < \infty$, depending only on C_V, such that*

$$\mathbb{P}\left(\sum_{y \in Y} Q(x, y; W) > \frac{2}{\pi} \sum_{y \in Y} \int_W |\operatorname{Im} G(x, y; E + i\eta| \, dE + \frac{\log N}{N^w} \right) \leq C N^{w - \ell}$$

for all $x \in \{1, ..., N\}$.

The proof of Theorem 3.4.1 is based on the following two lemmas, the first of which is formulated in terms of the the Poisson kernel P_z defined in (3.1.3).

Lemma 3.4.2. *There exists a constant $C < \infty$, depending only on C_V, such that*

$$\mathbb{E} \sum_y |\mu_{xy}| \left(1_{W^c}(1_W * P_{i\eta})\right) \leq C N \eta \left(1 + \log \sqrt{1 + \eta^{-2}|W|^2}\right)$$

for all intervals $W \subset \mathbb{R}$ and $\eta > 0$.

Proof. By spectral averaging (Lemma 3.1.1),

$$\begin{aligned}
\sum_y \mathbb{E}\, |\mu_{xy}|(1_{W^c}(1_W * P_{i\eta})) &\leq CN \int_{W^c} (1_W * P_{i\eta})(\lambda)\, d\lambda \\
&= CN \int_{W^c} \int_W \frac{\eta}{(u - v)^2 + \eta^2}\, du\, dv \\
&= CN\eta \int_{\eta^{-1} W^c} \int_{\eta^{-1} W} \frac{1}{1 + (u - v)^2}\, du\, dv.
\end{aligned}$$

Without loss of generality, we may assume that $\eta^{-1}W = [-a, a]$, so

$$\begin{aligned}\int_{\eta^{-1}W^c}\int_{\eta^{-1}W}\frac{1}{1+(u-v)^2}\,du\,dv &= \int_{\eta^{-1}W^c}\arctan(a-v)-\arctan(v+a)\,dv\\ &= 2\int_a^\infty \arctan(v+a)-\arctan(v-a)\,dv\end{aligned}$$

since $\arctan v$ is an odd function of v. After the appropriate translations, this last integral is

$$\begin{aligned}&= 2\lim_{R\to\infty}\int_{R-a}^{R+a}\arctan v\,dv - 2\int_0^{2a}\arctan v\,dv\\ &= 2\left(\frac{2\pi a}{2} - \int_0^{2a}\arctan v\,dv\right)\\ &= 2a\left(\frac{\pi}{2}-\arctan(2a)\right) + \log\sqrt{1+4a^2}.\end{aligned}$$

The proof is completed by noting $|\arctan(x) - \pi/2| \leq 1/x$ and inserting $a = \eta^{-1}|W|/2$. $\square$

The second lemma needed for the proof of Theorem 3.4.1 controls the generic spacing between the eigenvalues of H in the interval W.

Lemma 3.4.3. *Let $W \subset \mathbb{R}$ be an interval and $|W| \geq S > 0$. Then, there exists a constant $C < \infty$, depending only on C_V, such that the event*

$$\mathcal{E} = \left\{\min_{\lambda\in\sigma(H)\cap W} d\left(\lambda, \partial W \cup \sigma(H)\setminus\{\lambda\}\right) > 2S\right\}$$

satisfies

$$\mathbb{P}(\mathcal{E}^c) \leq CSN(1+|W|N).$$

Proof. We split W into a disjoint union of adjacent intervals

$$W = I_1 \cup \ldots \cup I_p$$

with $|I_k| = 2S$ for $1 \leq k \leq p-1$ and $|I_p| \leq 2S$, and let $\tilde{I}_k$ denote the fattened interval $I_k + [-2S, 2S]$. Then $\mathcal{E}^c$ can only occur if

1. $\tilde{I}_k$ contains at least two eigenvalues of H for some $1 \leq k \leq p$, or

2. $\partial W + [-2S, 2S]$ contains an eigenvalue of H.

Therefore, the Wegner and Minami estimates show that

$$\begin{aligned}\mathbb{P}(\mathcal{E}^c) &\le \mathbb{P}\left(|(\partial W + [-2S, 2S]) \cap \sigma(H)| \ge 1\right) + \sum_{k=1}^{p} \mathbb{P}\left(\left|\tilde{I}_k \cap \sigma(H)\right| \ge 2\right)\\ &\le CSN + Cp\,(SN)^2\,,\end{aligned}$$

and since $p \le 2|W|/S$, this proves the lemma. □

Proof of Theorem 3.4.1. Let $S = \frac{8}{\pi}\eta$ so that the event $\mathcal{E}$ defined in Lemma 3.4.3 satisfies

$$\mathbb{P}(\mathcal{E}^c) \le CN^{w-\ell}.$$

Since the spectral measures μ_{xy} are real, we can construct the function

$$f(E) = \sum_{\lambda \in \sigma(H) \cap W} \operatorname{sgn}\,[\psi_\lambda(x)\psi_\lambda(y)]\, I_\lambda(E),$$

where I_λ denotes the indicator function of the interval $[\lambda - S, \lambda + S]$. We will prove that on the event $\mathcal{E}$ we have $\|f\|_\infty \le 1$ and

$$\sum_{y \in Y} |\mu_{xy}|(W) \le \frac{2}{\pi} \sum_{y \in Y} \mu_{xy}(f * P_{i\eta}) + \sum_{y} |\mu_{xy}|\left(1_{W^c}(1_W * P_{i\eta})\right), \tag{3.4.1}$$

so, since

$$\begin{aligned}\mu_{xy}(f * P_{i\eta}) &= \iint f(E) P_{\lambda + i\eta}(E)\, dE\, \mu_{xy}(d\lambda)\\ &= \int f(E) \int P_{E+i\eta}(\lambda)\, \mu_{xy}(d\lambda)\, dE\\ &\le \|f\|_\infty \int_W |\operatorname{Im} G(x, y; E + i\eta)|\, dE,\end{aligned}$$

the theorem follows from Lemma 3.4.2 and Markov's inequality.

On $\mathcal{E}$, the intervals I_λ are disjoint and contained in W, so $|f| \le 1_W$ and, in particular, $\|f\|_\infty \le 1$. To verify (3.4.1), we note that

$$\begin{aligned}\mu_{xy}(f * P_{i\eta}) &= \mu_{xy}\left(1_W(f * P_{i\eta})\right) + \mu_{xy}\left(1_{W^c}(f * P_{i\eta})\right)\\ &\ge \sum_{\lambda \in \sigma(H) \cap W} \psi_\lambda(x)\psi_\lambda(y)(f * P_{i\eta})(\lambda) - |\mu_{xy}|\left(1_{W^c}(1_W * P_{i\eta})\right)\end{aligned}$$

on $\mathcal{E}$ and hence it remains only to prove that

$$\operatorname{sgn}\,[\psi_\lambda(x)\psi_\lambda(y)]\,(f * P_{i\eta})(\lambda) \ge \frac{\pi}{2}$$

for all $\lambda \in \sigma(H) \cap W$. This is based on the fact that

$$\int (1 - I_\lambda(E)) P_{\lambda + i\eta}(E)\, dE \leq \frac{2\eta}{S} = \frac{\pi}{4}$$

and hence

$$\int I_\lambda(E) P_{\lambda+i\eta}(E)\, dE \geq \pi - \int (1 - I_\lambda(E)) P_{\lambda+i\eta}(E)\, dE \geq \frac{3\pi}{4}.$$

If $\lambda \in \sigma(H) \cap W$ with sgn $[\psi_\lambda(x)\psi_\lambda(y)] = 1$, it follows that

$$\begin{aligned} (f * P_{i\eta})(\lambda) &= \int f(E) P_{\lambda+i\eta}(E)\, dE \\ &\geq \int I_\lambda(E) P_{\lambda+i\eta}(E)\, dE - \int (1 - I_\lambda(E)) P_{\lambda+i\eta}(E)\, dE \\ &\geq \frac{\pi}{2}, \end{aligned}$$

and similarly

$$(f * P_{i\eta})(\lambda) \leq -\frac{\pi}{2}$$

if sgn $[\psi_\lambda(x)\psi_\lambda(y)] = -1$. □

The proof of the last theorem made use of the fact that the spectral measures μ_{xy} are always real for the GOE flow. It is possible to extend this result to models with complex off-diagonal spectral measures, such as the GUE flow, by using the fact that

$$\langle \delta_y, \operatorname{Im}(H - z)^{-1} \delta_x \rangle + \langle \delta_x, \operatorname{Im}(H - z)^{-1} \delta_y \rangle = \operatorname{Im} G(x, y; z) + \operatorname{Im} G(y, x; z),$$

but we omit these complications here.

With Theorem 3.4.1 in hand, we now turn to the proof of Theorem 1.3.2. As in Section 3.3, we drop the normalizing constant $Z_{n,c}$ from the definition of H_n. The core of this argument again consists of resolvent bounds for Gaussian perturbations, and thus we consider the Green functions

$$G_n(x, y; z) = \langle \delta_y, (H_n - z)^{-1} \delta_x \rangle, \quad G_{n,m}(x, y; z) = \langle \delta_y, (H_{n,m} - z)^{-1} \delta_x \rangle.$$

If $\eta = 2^{-(1+\ell)n}$ for some $\ell > 0$, Theorem 1.5.2 proves that there exists $C < \infty$ such that

$$\begin{aligned} &2^{-k} \sum_{y \in B_k(x)} \mathbb{E}\, |G_k(x, y; E + i\eta) - G_{k,k-1}(x, y; E + i\eta)| \\ &\leq C\, 2^{-\frac{c}{2}k} \left(1 + 2^{3((1+\ell)n - k)}\right) \\ &= C\, 2^{3(1+\ell)n - 3(1+\delta)k} \end{aligned}$$

with $\delta = c/6$ whenever $k \leq n$. Iterating this result, we see that

$$
\begin{aligned}
&2^{-n} \sum_{y \in B_n} \mathbb{E} \left| G_n \left(x, y; E + i\eta\right) - G_{n,m} \left(x, y; E + i\eta\right) \right| \\
&\leq 2^{-n} \sum_{k=m+1}^{n} \sum_{y \in B_n} \mathbb{E} \left| G_{n,k} \left(x, y; E + i\eta\right) - G_{n,k-1} \left(x, y; E + i\eta\right) \right| \\
&= 2^{-n} \sum_{k=m+1}^{n} \sum_{y \in B_k(x)} \mathbb{E} \left| G_k \left(x, y; E + i\eta\right) - G_{k,k-1} \left(x, y; E + i\eta\right) \right| \\
&\leq 2^{-n} \sum_{k=m+1}^{n} 2^k C\, 2^{3(1+\ell)n - 3(1+\delta)k} \leq C\, 2^{(3(1+\ell)-1)n}\, 2^{-(3(1+\delta)-1)m}.
\end{aligned}
$$

Since $\delta > 0$, we can choose $\ell > 0$, $\varepsilon \in (0,1)$, and $w \in (0, \ell)$ such that

$$2\mu := (1-\varepsilon)(3(1+\delta)-1) - (3(1+\ell)-1) - w > 0.$$

Thus, setting $m_n = (1-\varepsilon)n$ and

$$W = \left[E - 2^{-(1-w)n}, E + 2^{-(1-w)n}\right],$$

and using that $G_{n,m}(x,y;z) = 0$ if $y \notin B_m(x)$ show that

$$\sum_{y \in B_n \setminus B_{m_n}(x)} \mathbb{E} \int_W \left| \operatorname{Im} G_n \left(x, y; E + i\eta\right) \right| \, dE \leq C\, 2^{-2\mu n}.$$

Applying Markov's inequality, we arrive at

$$\mathbb{P} \left(\sum_{y \in B_n \setminus B_{m_n}(x)} \int_W \left| \operatorname{Im} G_n \left(x, y; E + i\eta\right) \right| \, dE > 2^{-\mu n} \right) \leq C\, 2^{-\mu n},$$

so Theorem 1.3.2 follows from Theorem 3.4.1, which says that

$$\sum_{y \in B_n \setminus B_{m_n}(x)} Q_n(x, y; W) \leq \sum_{y \in B_n \setminus B_{m_n}(x)} \int_W \left| \operatorname{Im} G_n(x, y; E + i\eta \right| \, dE + \frac{\log 2^n}{2^{wn}}$$

with probability $1 - \mathcal{O}\left(2^{(w-\ell)n}\right)$.

3.5 Localization Regime of the Rosenzweig-Porter model

Proof of Theorem 1.4.2. As $N \to \infty$, the random measure defined by

$$\mu_{N,0}(f) = \sum_{\lambda \in \sigma(H_0)} f(N(\lambda - E_0))$$

converges in distribution to a Poisson point process with intensity $\varrho(E_0)$. Setting $z_N = E_0 + z/N$, a simple calculation yields

$$\mu_N(P_z) = \operatorname{Im} S_T(z_N).$$

Thus,

$$\left| \mathbb{E}\, e^{-\mu_N(P_z)} - \mathbb{E}\, e^{-\mu_{N,0}(P_z)} \right| \leq \mathbb{E}\, |S_T(z_N) - S_0(z_N)| \leq C N^{-\varepsilon/2},$$

which shows that the characteristic functionals of μ_N and $\mu_{N,0}$ asymptotically agree on the set $\{P_z : z \in \mathbb{C}_+\}$ whose linear span is dense in C_0. This proves the first point.

For the second assertion, choose $\ell > w > 0$ and $\mu_0 > 0$ such that

$$3\ell + w + 2\mu_0 \leq \varepsilon/2.$$

Since $G_0(x, y; z) = 0$ for $x \neq y$, Theorem 1.5.2 shows that with $\eta = N^{-(1+\ell)}$ we have

$$\begin{aligned} \mathbb{E} \sum_{y \neq x} \int_W |\operatorname{Im} G_T(x, y; E + i\eta)|\, dE &\leq C|W| N N^{-\varepsilon/2} (\eta N)^{-3} \\ &\leq C N^{w + 3\ell - \varepsilon/2} \leq N^{-2\mu_0}. \end{aligned}$$

By Markov's inequality,

$$\mathbb{P}\left(\sum_{y \neq x} \int_W |\operatorname{Im} G_T(x, y; E + i\eta)|\, dE \geq N^{-\mu_0} \right) \leq C N^{-\mu_0}$$

so choosing $0 < \mu < \min\{w, \mu_0\}$ and $\kappa = \min\{w - \ell, \mu_0\}$, Theorem 3.4.1 shows that

$$\mathbb{P}\left(\sum_{y \neq x} Q_N(x, y; W) > N^{-\mu} \right) \leq C N^{-\kappa}.$$

□

4 Non-Ergodic Delocalization in the Rosenzweig-Porter Model

In this chapter, we will prove Theorem 1.4.3. Throughout, we will fix the time $T = N^{-1+\delta}$ from the statement of Theorem 1.4.3. and a spectral domain of the form

$$D = W + i[\eta, 1]$$

where $W \subset \mathbb{R}$ is a bounded interval and $\eta = N^{-1+\alpha}$ is a spectral scale whose parameter $\alpha > 0$ is fixed but may be arbitrarily small. To simplify the exposition, we will assume that V is a *deterministic* potential, which possesses some regularity that will be expressed in terms of the resolvent-like functionals

$$F_I(z) = \frac{1}{N} \sum_{V_x \notin I} \frac{1}{V_x - z},$$

where $I \subset \mathbb{R}$ is a possibly empty interval. Notice that $F_\emptyset(z) = N^{-1} \sum_x (V_x - z)^{-1}$ coincides with the Stieltjes transform of the empirical eigenvalue measure of V.

Assumption 4.0.1. *There exist $\varepsilon > 0$ and constants $K_m, K_\ell, K_i \in (0, \infty)$ such that*

1. $|F_\emptyset(z)| \le K_m \log N$ *uniformly in* $\operatorname{Im} z > \eta$,
2. $\operatorname{Im} F_\emptyset(z) \ge K_l$ *uniformly in* $z \in \mathbb{C}_+$ *with* $\operatorname{dist}(z, D) \le \varepsilon$, *and,*
3. *if* $z \in D$ *with* $\operatorname{Im} z > K_l T/2$ *and* $I \subset W$ *is an interval with* $\operatorname{Re} z \in I$ *and* $\operatorname{dist}(\operatorname{Re} z, \partial I) > N^{-1+\kappa}$ *for some* $\kappa > \delta$, *then*

$$\operatorname{Im} F_I(z) \le K_i \frac{\operatorname{Im} z}{N^{-1+\kappa}} + N^{-\delta/4}.$$

If the entries V_x are drawn independently from a compactly supported density $\varrho \in L^\infty$, we will show in Section 4.3 that Assumption 4.0.1 is satisfied with overwhelming probability for any interval W on which ϱ is bounded below. The restriction of the conclusion of Theorem 1.4.3 to the states in W is then only a mild condition since ϱ coincides with the asymptotic density of states of H_T (Theorem 1.5.1) and hence one expects that the majority of $\sigma(H_T)$ typically lies in $\operatorname{supp} \varrho$.

4.1 Characteristic Curves

In this section, we study the properties of the characteristic curve

$$\dot{z}_t = -S_t(z_t), \qquad z_0 = z \tag{4.1.1}$$

of the transport equation (1.5.3). However, it is technically more convenient to consider instead the process

$$\xi_t(z) = z_{t \wedge \tau_z}$$

which is stopped at

$$\tau_z = \{\inf t > 0 : \operatorname{Im} z_t \leq \eta/2\}.$$

Regarding $R_t(\xi_t(z))$ as a function of the processes $\{B_{uv}(t)\}$ and $\xi_t(z)$, Itô's lemma shows that

$$\begin{aligned} dG_t(x, \xi_t(z)) = &\frac{1}{N} \sum_{u \leq v} \langle \delta_x, R_t(\xi_t(z)) P_{uv} R_t(\xi_t(z)) P_{uv} R_t(\xi_t(z)) \delta_x \rangle \, dt \\ &- \frac{1}{\sqrt{N}} \sum_{u \leq v} \langle \delta_x, R_t(\xi_t(z)) P_{uv} R_t(\xi_t(z)) \delta_x \rangle \, dB_{uv}(t) + \dot{\xi}_t(z) \frac{\partial}{\partial \xi} G_t(x, \xi_t(z)) \, dt \end{aligned}$$

with

$$P_{uv} = \frac{1}{\sqrt{1 + \delta_{uv}}} \left(|\delta_u\rangle\langle\delta_v| + |\delta_v\rangle\langle\delta_u| \right).$$

The piecewise C^1 process $\xi_t(z)$ has vanishing covariation with all the $B_{uv}(t)$. The calculations in the proof of Theorem 3.2.1 then show that

$$\begin{aligned} dG_t(x, \xi_t(z)) = &\left(S_t(\xi_t(z)) \frac{\partial}{\partial \xi} G_t(x, \xi_t(z)) + \frac{1}{2N} \frac{\partial^2}{\partial \xi^2} G_t(x, \xi_t(z)) \right) dt \\ &+ \dot{\xi}_t(z) \frac{\partial}{\partial \xi} G_t(x, \xi_t(z)) \, dt + dM_t(x, z) \end{aligned}$$

with

$$dM_t(x, z) = -\frac{1}{\sqrt{N}} \sum_{u \leq v} \langle \delta_x, R_t(\xi_t(z)) P_{uv} R_t(\xi_t(z)) \delta_x \rangle \, dB_{uv}(t). \tag{4.1.2}$$

If τ is any stopping time such that $\tau \leq \tau_z$ almost surely, (4.1.1) yields

$$\begin{aligned} G_\tau(x, \xi_\tau(z)) - G_0(x, z) = &\int_0^\tau \frac{1}{2N} \frac{\partial^2}{\partial \xi^2} G_t(x, \xi_t(z)) \, dt \\ &- \frac{1}{\sqrt{N}} \sum_{u \leq v} \int_0^\tau \langle \delta_x, R_t(\xi_t(z)) P_{uv} R_t(\xi_t(z)) \delta_x \rangle \, dB_{uv}(t) \end{aligned} \tag{4.1.3}$$

for the change in the local resolvent along the characteristic curve.

Our next goal is to show that with high probability the change in S_t along the curve $\xi_t(z)$ is small for a sufficiently dense set of initial points z. Let

$$\tilde{D} \subset \{z \in \mathbb{C}_+ : \operatorname{Im} z > \eta\}$$

be some finite set. The next theorem bounds the probability of the event

$$\mathcal{A}_S = \left\{ \sup_{z\in\tilde{D}} \sup_{t\le\tau_z} |S_t(\xi_t(z)) - S_0(z)| > \frac{4}{\sqrt{N\eta}} \right\},$$

showing that with high probability $S_t(\xi_t(z))$ is approximately constant if $|\tilde{D}|$ grows only polynomially in N. In the statement of the theorem, and throughout, $C, c \in (0,\infty)$ denote deterministic constants that are independent of N but whose value may change from instance to instance.

Theorem 4.1.1. *For every $z \in \mathbb{C}_+$ with $\operatorname{Im} z > \eta$ we have*

$$\mathbb{P}\left(\sup_{t\le\tau_z} |S_t(\xi_t(z)) - S_0(z)| > \frac{4}{\sqrt{N\eta}} \right) \le 2e^{-\frac{1}{2}N\eta} \tag{4.1.4}$$

and therefore $\mathbb{P}(\mathcal{A}_S) \le C|\tilde{D}|e^{-\frac{1}{2}N\eta}$.

Proof. By averaging (4.1.3) over all x, we see that the process

$$\tilde{S}_t = S_{t\wedge\tau_z}(\xi_t(z))$$

satisfies

$$\begin{aligned}\tilde{S}_t - \tilde{S}_0 &= \int_0^{t\wedge\tau_z} \frac{1}{2N}\frac{\partial^2}{\partial\xi^2} S_s(\xi_s(z))\,ds \\ &\quad - \frac{1}{\sqrt{N^3}}\sum_x\sum_{u\le v}\int_0^{t\wedge\tau_z} \langle \delta_x, R_s(\xi_s(z))P_{uv}R_s(\xi_s(z))\delta_x\rangle\,dB_{uv}(s).\end{aligned}$$

The drift component of $\tilde{S}$ is bounded by

$$\begin{aligned}\int_0^{t\wedge\tau_z} \frac{1}{2N}\left|\frac{\partial^2}{\partial\xi^2}S_s(\xi_s(z))\right| ds &\le \frac{1}{N}\int_0^{t\wedge\tau_z} \frac{\operatorname{Im} S_s(\xi_s(z))}{(\operatorname{Im}\xi_s(z))^2}\,ds \\ &= \frac{1}{N}\int_0^{t\wedge\tau_z} \frac{-d(\operatorname{Im}\xi_s(z))}{(\operatorname{Im}\xi_s(z))^2} \\ &= \frac{1}{N\operatorname{Im}\xi_t(z)} - \frac{1}{N\operatorname{Im} z} \le \frac{2}{N\eta}.\end{aligned}$$

The martingale part of $\tilde{S}$ is given by

$$\begin{aligned}M_t &= -\frac{1}{\sqrt{N^3}}\sum_x\sum_{u\le v}\int_0^{t\wedge\tau_z}\langle\delta_x, R_s(\xi_s(z))P_{uv}R_s(\xi_s(z))\delta_x\rangle\,dB_{uv}(s) \\ &= -\frac{1}{\sqrt{N^3}}\sum_{u,v}\sqrt{1+\delta_{uv}}\int_0^{t\wedge\tau_z}\langle\delta_v, R_s(\xi_s(z))^2\delta_u\rangle\,dB_{uv}(s)\,.\end{aligned}$$

Its quadratic variation may be expressed as

$$
\begin{aligned}
[M]_t &\le \frac{2}{N^3}\int_0^{t\wedge\tau_z}\sum_{u,v}\left|\langle\delta_v, R_s(\xi_s(z))^2\delta_u\rangle\right|^2\,ds\\
&\le \frac{2}{N^2}\int_0^{t\wedge\tau_z}\frac{\operatorname{Im}S_s(\xi_s(z))}{(\operatorname{Im}\xi_s(z))^3}\,ds.\\
&= \frac{1}{N^2}\int_0^{t\wedge\tau_z}\frac{-2}{(\operatorname{Im}\xi_s(z))^3}\,d(\operatorname{Im}\xi_s(z))\\
&= \frac{1}{(N\operatorname{Im}\xi_t(z))^2}-\frac{1}{(N\operatorname{Im}z)^2}\le\frac{4}{(N\eta)^2}.
\end{aligned}
$$

It follows that there exists a Brownian motion $\tilde{B}$ such that

$$
\sup_t|\tilde{S}_t-\tilde{S}_0|\le\sup_t\left(\frac{2}{N\eta}+\left|\tilde{B}_{[M]_t}\right|\right)\le\frac{2}{\sqrt{N\eta}}+\sup_{t\le 4/(N\eta)^2}|\tilde{B}_t|.
$$

Applying the reflection principle to $\tilde{B}$ we obtain (4.1.4) and the second assertion follows from the union bound. □

Once we know that $S_t(\xi_t(z))$ is approximately constant, this term can be inserted into integrals involving $\xi_t(z)$ more or less at will, and the substitution trick from Theorem 4.1.1 gives bounds improving on the trivial bound by a factor of η. We illustrate this in the following corollary, which will prove useful in extending our method to the local resolvents.

Corollary 4.1.2. *If $\mathcal{A}_S$ does not occur, then*

$$
\int_0^{t\wedge\tau_z}\frac{1}{(\operatorname{Im}\xi_s(z))^2}\,ds\le\frac{4}{K_l\eta}
$$

for all $t>0$ and $z\in D\cap\tilde{D}$.

Proof. If $\mathcal{A}_S$ does not occur and $z\in D\cap\tilde{D}$, then for sufficiently large N

$$
\inf_{s\le t\wedge\tau_z}\operatorname{Im}S_s(\xi_s(z))\ge\operatorname{Im}S_0(z)-\frac{4}{\sqrt{N\eta}}\ge\frac{K_l}{2}
$$

where K_l is the lower bound from Assumption 4.0.1. Hence,

$$
\begin{aligned}
\int_0^{t\wedge\tau_z}\frac{1}{(\operatorname{Im}\xi_s(z))^2}\,ds &\le \frac{2}{K_l}\int_0^{t\wedge\tau_z}\frac{\operatorname{Im}S_s(\xi_s(z))}{(\operatorname{Im}\xi_s(z))^2}\,ds\\
&= \frac{2}{K_l}\int_0^{t\wedge\tau_z}\frac{d(\operatorname{Im}\xi_s(z))}{(\operatorname{Im}\xi_s(z))^2}\le\frac{4}{K_l\eta}.
\end{aligned}
$$

□

The Picard-Lindelöf theorem and the Herglotz property of S_s imply that, almost surely, for every $z \in D$ there exists a $w \in \mathbb{C}_+$ with $\xi_T(w) = z$ satisfying the a-priori deterministic bound

$$|w - z| \leq \int_0^T |S_s(\xi_s(w))|\, ds \leq \frac{T}{\eta}.$$

In order for Theorem 4.1.1 to be useful in the study of the function S_T, we need to guarantee that a sufficiently dense subset of D is of the form $\xi_T(w)$ with $w \in \tilde{D}$. To this end, we define the distance

$$r = \min\left\{KT\eta^2, N^{-2\theta}\eta^3, N^{-(1+2\gamma)}\eta^3\right\}, \tag{4.1.5}$$

where

$$K = \sup_{\operatorname{Im} z > \eta} |S_0(z)| + \frac{4}{\sqrt{N\eta}} \leq C \log N, \tag{4.1.6}$$

and $\gamma, \theta > 0$ are the parameters from the statement of Theorem 1.4.3. We now require $\tilde{D}$ to be such that

$$\operatorname{dist}(z, \tilde{D}) \leq r$$

for all $z \in \mathbb{C}_+$ with $\operatorname{Im} z > \eta$ and $\operatorname{dist}(z, D) \leq T/\eta$. The grid $\tilde{D}$ can hence be chosen such that its cardinality is bounded by

$$|\tilde{D}| \leq C(\eta r)^{-2}.$$

The following theorem provides a Lipschitz constant for the characteristic flow which grows only polynomially in η. The resulting bound is a significant improvement on the exponential bound provided by the direct application of Grönwall's inequality and enables us to keep the cardinality of $\tilde{D}$ polynomial in N.

Theorem 4.1.3. *Suppose $\mathcal{A}_S$ does not occur and N is sufficiently large. Then for every $z \in D$ there exists $w \in \tilde{D}$ such that:*

1. $|\xi_T(w) - z| \leq C\eta^{-2} r$,
2. $|w - z| \leq CKT$ *with K as in* (4.1.6), *and*
3. $\operatorname{Im} w \geq \frac{1}{2} K_l T$ *with K_l defined in Assumption 4.0.1.*

Proof. By the construction of $\tilde{D}$, for any $z \in D$ there exist $w_0 \in \mathbb{C}_+$ with $\xi_T(w_0) = z$ and $w \in \tilde{D}$ with $|w - w_0| \leq r$. If $t \leq \tau_{w_0} \wedge \tau_w$, the evolution (4.1.1) yields

$$\begin{aligned}
|\xi_t(w_0) - \xi_t(w)| &\leq |w_0 - w| + \int_0^t |S_s(\xi_s(w_0)) - S_s(\xi_s(w))|\, ds \\
&\leq |w_0 - w| + \frac{1}{N} \int_0^t \sum_i \frac{|\xi_s(w_0) - \xi_s(w)|}{|\lambda_i(s) - \xi_s(w_0)||\lambda_i(s) - \xi_s(w)|}\, ds.
\end{aligned}$$

Using the inequality $ab \le \frac{1}{2}(a^2+b^2)$, the integral in the last term is bounded by

$$\frac{1}{2}\int_0^t |\xi_s(w_0)-\xi_s(w)|\frac{1}{N}\sum_i\left(\frac{1}{|\lambda_i(s)-\xi_s(w_0)|^2}+\frac{1}{|\lambda_i(s)-\xi_s(w)|^2}\right)ds$$
$$\le \frac{1}{2}\int_0^t |\xi_s(w_0)-\xi_s(w)|\left(\frac{\operatorname{Im} S_s(\xi_s(w_0))}{\operatorname{Im}\xi_s(w_0)}+\frac{\operatorname{Im} S_s(\xi_s(w))}{\operatorname{Im}\xi_s(w)}\right)ds$$

so Grönwall's inequality shows that

$$\begin{aligned}\log\frac{|\xi_t(w_0)-\xi_t(w)|}{|w_0-w|} &\le \frac{1}{2}\int_0^t \frac{\operatorname{Im} S_s(\xi_s(w_0))}{\operatorname{Im}\xi_s(w_0)}+\frac{\operatorname{Im} S_s(\xi_s(w))}{\operatorname{Im}\xi_s(w)}\,ds\\ &= -\frac{1}{2}\int_0^t\frac{d(\operatorname{Im}\xi_s(w_0))}{\operatorname{Im}\xi_s(w_0)}-\frac{1}{2}\int_0^t\frac{d(\operatorname{Im}\xi_s(w))}{\operatorname{Im}\xi_s(w)}\\ &= \log\sqrt{\frac{\operatorname{Im} w_0}{\operatorname{Im}\xi_t(w_0)}\frac{\operatorname{Im} w}{\operatorname{Im}\xi_t(w)}}.\end{aligned}$$

Thus, using $\operatorname{Im} w \le \operatorname{Im} w_0 + r$, $\operatorname{Im} w_0 \le 1+T\eta^{-1}\le c\eta^{-1}$, and the stopping rules, we obtain

$$|\xi_t(w_0)-\xi_t(w)| \le C\eta^{-2}|w_0-w| \le C\eta^{-2}r \tag{4.1.7}$$

for all $t\le\tau_{w_0}\wedge\tau_w$. Since $\operatorname{Im}\xi_T(w_0)=z$ and $\operatorname{Im}\xi_t(w_0)$ is decreasing, $\tau_{w_0}>T$, so (4.1.7) and the definition of r shows that for sufficiently large N we have $|\xi_t(w_0)-\xi_t(w)|\le\eta/4$ for all $t\le T\wedge\tau_w$. If it were true that $\tau_w<T$, we would obtain the contradiction

$$\frac{\eta}{2}=\operatorname{Im}\xi_{\tau_w}(w)\ge\operatorname{Im}\xi_{\tau_w}(w_0)-\frac{\eta}{4}\ge\eta-\frac{\eta}{4}.$$

Hence (4.1.7) is valid for $t=T$, establishing the first claim of the theorem. If $\mathcal{A}_S$ does not occur, then

$$|\xi_T(w)-w|\le\int_0^T|S_s(\xi_s(w))|\,ds\le\int_0^T|S_0(w)|+\frac{4}{\sqrt{N\eta}}\,ds\le KT$$

since $w\in\tilde{D}$. Hence the definition of r and (4.1.7) yield

$$|w-z|\le|w-\xi_T(w)|+|\xi_T(w)-z|\le KT+CKT=CKT.$$

proving the second claim of the theorem. The second claim also implies that $\operatorname{dist}(w,D)\le\varepsilon$ for sufficiently large N so that Assumption 4.0.1 guarantees $\operatorname{Im} S_0(w)\ge K_l$. On the complement of $\mathcal{A}_S$ this yields

$$\operatorname{Im} w=\operatorname{Im}\xi_T(w)+\int_0^T\operatorname{Im} S_s(\xi_s(w))\,ds\ge T\left(\operatorname{Im} S_0(w)-\frac{4}{\sqrt{N\eta}}\right)\ge\frac{K_lT}{2}$$

for sufficiently large N. □

4.2 Local Resolvent Bounds

Since $S_t(z)$ is entirely featureless regarding a possible localization transition when $\operatorname{Im} z \gg N^{-1}$, we now turn our attention to controlling the local resolvents $G_t(x,z)$ along the characteristic $\xi_t(z)$. Unlike S_t, the function $G_t(x,\cdot)$ may be heavily concentrated around certain energies in non-ergodic regimes. Therefore, its derivative may be large in all directions and we cannot expect an exact analogue of Theorem 4.1.1 to hold true for all energies. However, one may hope that the change in $G_t(x,z)$ along the characteristic is small in those regions where $G_t(x,z)$ itself is small. We encode this phenomenon in the event

$$\mathcal{A}_G(\ell) = \left\{ \sup_x \sup_{z\in D\cap\tilde{D}} \sup_{s\le\tau_z} \frac{\operatorname{Im} G_s(x,\xi_s(z))}{\operatorname{Im} G_0(x,z)} > N^\ell \right\},$$

whose probability does in fact decay as $N\to\infty$. The proof is somewhat reminiscent of a Grönwall-type argument for martingales, which is greatly facilitated by the built-in control of the running maximum. Still, the basic mechanism behind the following argument is somewhat different from the stochastic Grönwall lemmas that previously appeared in [81, 88].

Theorem 4.2.1. *For every $\ell > 0$ and $p > 0$ we have*

$$\mathbb{P}\left(\mathcal{A}_G(\ell)\right) \le N^{-p}$$

for all sufficiently large N.

Proof. Fix $z \in D\cap\tilde{D}$ and consider the stopping time

$$\tau = \tau_z \wedge \inf\left\{ t\ge 0 : \int_0^{t\wedge\tau_z} \frac{1}{(\operatorname{Im}\xi_s(z))^2}\, ds \ge \frac{5}{K_l\eta} \right\},$$

where K_l is the lower bound from Assumption 4.0.1. As in Theorem 4.1.1, the stopped process $\tilde{G}_t = G_{t\wedge\tau}(x,\xi_{t\wedge\tau}(z))$ satisfies

$$\begin{aligned}\tilde{G}_t - \tilde{G}_0 &= \int_0^{t\wedge\tau} \frac{1}{2N}\frac{\partial^2}{\partial\xi^2} G_s(x,\xi_s(z))\, ds \\ &\quad - \frac{1}{\sqrt{N}}\sum_{u\le v}\int_0^{t\wedge\tau} \langle \delta_x, R_s(\xi_s(z))P_{uv}R_s(\xi_s(z))\delta_x\rangle\, dB_{uv}(s).\end{aligned}$$

The drift component of $\tilde{G}$ is bounded by

$$\begin{aligned}\int_0^{t\wedge\tau} \frac{1}{2N}\left|\frac{\partial^2}{\partial\xi^2}G_s(x,\xi_s(z))\right| ds &\le \left(\sup_{s\le T}\operatorname{Im}\tilde{G}_s\right)\int_0^{T\wedge\tau}\frac{1}{N(\operatorname{Im}\xi_s(z))^2}\, ds \\ &\le \frac{5}{K_l N\eta}\left(\sup_{s\le T}\operatorname{Im}\tilde{G}_s\right),\end{aligned}$$

and, letting M denote the martingale part of $\tilde{G}$, its quadratic variation is bounded as follows,

$$\begin{aligned}
[M]_T &\le \frac{2}{N}\int_0^{T\wedge\tau} \sum_{u,v} |\langle \delta_x, R_s(\xi_s(z))\delta_u\rangle\langle\delta_v, R_s(\xi_s(z))\delta_x\rangle|^2 \, ds \\
&= \frac{2}{N}\int_0^{T\wedge\tau} \left(\sum_u |\langle\delta_x, R_s(\xi_s(z))\delta_u\rangle|^2\right)\left(\sum_v |\langle\delta_v, R_s(\xi_s(z))\delta_x\rangle|^2\right) ds \\
&= \frac{2}{N}\int_0^{T\wedge\tau} \left(\frac{\operatorname{Im} G_s(x,\xi_s(z))}{\operatorname{Im}\xi_s(z)}\right)^2 ds \\
&\le \left(\sup_{s\le T\wedge\tau} \operatorname{Im} G_s(x,\xi_s(z))\right)^2 \int_0^{T\wedge\tau} \frac{2}{N(\operatorname{Im}\xi_s(z))^2}\, ds \\
&\le \frac{10}{K_l N\eta}\left(\sup_{s\le T}\operatorname{Im}\tilde{G}_s\right)^2 .
\end{aligned}$$

Hence,

$$\sup_{s\le T}\operatorname{Im}\tilde{G}_s \le \operatorname{Im}\tilde{G}_0 + \frac{5}{K_l N\eta}\sup_{s\le T}\operatorname{Im}\tilde{G}_s + \sup_{s\le T}|M_s|$$

so the Burkholder-Davis-Gundy inequality (with exponent $q > 0$ and constant C_q) yields

$$\begin{aligned}
\left(1 - \frac{5}{K_l N\eta}\right)\left(\mathbb{E}\left|\sup_{s\le T}\operatorname{Im}\tilde{G}_s\right|^q\right)^{1/q} &\le \operatorname{Im}\tilde{G}_0 + \left(\mathbb{E}\left|\sup_{s\le T}|M_s|\right|^q\right)^{1/q} \\
&\le \operatorname{Im}\tilde{G}_0 + C_q\left(\mathbb{E}\,[M]_T^{q/2}\right)^{1/q} \\
&\le \operatorname{Im}\tilde{G}_0 + C_q\sqrt{\frac{10}{K_l N\eta}}\left(\mathbb{E}\left|\sup_{s\le T}\operatorname{Im}\tilde{G}_s\right|^q\right)^{1/q} .
\end{aligned}$$

Since $N\eta\to\infty$, we can choose N large enough such that $(1+C_q)\sqrt{\frac{10}{K_l N\eta}} < 1/2$. Rearranging and applying Markov's inequality shows

$$\mathbb{P}\left(\sup_{s\le T\wedge\tau}\operatorname{Im} G_s(x,\xi_s(z)) > 4N^{\ell}\operatorname{Im} G_0(x,z)\right) \le N^{-\ell q}.$$

By Corollary 4.1.2, $\tau = \tau_z$ on the event $\mathcal{A}_S$ and we conclude that

$$\mathbb{P}\left(\sup_{s\le T\wedge\tau_z}\operatorname{Im} G_s(x,\xi_s(z)) > 4N^{\ell}\operatorname{Im} G_0(x,z)\right) \le N^{-\ell q} + \mathbb{P}(\mathcal{A}_S),$$

so, choosing q large enough, the theorem follows from the union bound. $\square$

To prove Theorem 1.4.3, it remains only to combine the previous results with the fact that $G_T(x,\cdot)$ is the Stieltjes transform of the spectral measure at x.

Proof of Theorem 1.4.3. We now specify the parameters $\alpha, \gamma, \ell > 0$ occuring in the spectral scale $\eta = N^{-1+\alpha}$, the definition of r in (4.1.5), and the event $\mathcal{A}_G(\ell)$ of Theorem 4.2.1 by requiring that

$$\alpha + \ell + \delta < \kappa, \quad \alpha + \ell < \delta/4, \quad \gamma < \kappa - (\alpha + \ell + \delta).$$

Suppose that neither of the events $\mathcal{A}_S, \mathcal{A}_G(\ell)$ of Theorems 4.1.1 and 4.2.1 occur, which is the case with probability $1 - N^{-p}$ provided N is sufficiently large. For every $\lambda \in \sigma(H_T) \cap W$,

$$\sum_{x \in X_\lambda} |\psi_\lambda(x)|^2 \leq \sum_{x \in X_\lambda} \sum_{E \in \sigma(H_T)} \frac{\eta^2}{(E-\lambda)^2 + \eta^2} |\psi_E(x)|^2 = \eta \sum_{x \in X_\lambda} \operatorname{Im} G_T(x, z)$$

with $z = \lambda + i\eta$. By Theorem 4.1.3, there exists $w \in \tilde{D}$ is such that $|w - z| \leq CKT$, $\operatorname{Im} w > K_l T/2$, and $|\xi_T(w) - z| \leq C\eta N^{-(1+2\gamma)}$. Hence, for sufficiently large N,

$$\operatorname{Re} w \in I := 1\{|V_x - \lambda| > N^{-1+\kappa}\}$$

and $\operatorname{dist}(\operatorname{Re} w, \partial I) > \frac{1}{2} N^{-1+\kappa}$. Using Assumption 4.0.1 and the η^{-2}-Lipschitz continuity of $G_T(x,z)$, this yields

$$\begin{aligned}
\sum_{x \in X_\lambda} |\psi_\lambda(x)|^2 &\leq \eta \sum_{x \in X_\lambda} \operatorname{Im} G_T(x, \xi_T(w)) + CN^{-2\gamma} \\
&\leq \eta N^\ell \sum_{x \in X_\lambda} \operatorname{Im} G_0(x, w) + CN^{-2\gamma} \\
&= N^{\alpha+\ell} \operatorname{Im} F_I(w) + CN^{-2\gamma} \\
&\leq CN^{\alpha+\ell} \left(\frac{\operatorname{Im} w}{N^{-1+\kappa}} + N^{-\delta/4} \right) + CN^{-2\gamma}.
\end{aligned}$$

Since $\operatorname{Im} w \leq \eta + KT \leq CN^{-1+\delta} \log N$, the last term is bounded by $N^{-\gamma}$ if and N is large enough, proving the first claim of the theorem.

If, in addition, we require that $\alpha + \ell < \delta - \theta$, the second claim follows by the same token. Combining the Lipschitz continuity of $G_T(x,z)$ with $|\xi_T(w) - z| \leq C\eta N^{-2\theta}$, we obtain

$$\begin{aligned}
|\psi_\lambda(x)|^2 &\leq \eta \operatorname{Im} G_T(x, \lambda + i\eta) \\
&\leq \eta \operatorname{Im} G_T(x, \xi_T(w)) + CN^{-2\theta} \\
&\leq \eta N^\ell \operatorname{Im} G_0(x, w) + CN^{-2\theta} \\
&\leq CN^{\alpha+\ell-\delta} + CN^{-2\theta} \leq N^{-\theta}
\end{aligned}$$

since $\operatorname{Im} w > K_l T/2$. $\square$

4.3 Regularity Estimates for Random Potentials

This section is devoted to the verification of Assumption 4.0.1 in the case that the $\{V_x\}$ are drawn independently from a compactly supported density $\varrho \in L^\infty$. We will assume that ϱ is bounded below in a neighborhood of W, i.e. there exists $\varepsilon > 0$ such that

$$\inf_{v \in W(\varepsilon)} \varrho(v) > 0$$

with $W(\varepsilon) = W + [-\varepsilon, \varepsilon]$. We start by proving a concentration inequality in the spirit of Cramér's theorem for F_I, which is uniform in spectral domains of the form

$$D(J,\zeta) = \{z \in \mathbb{C}_+ : \operatorname{Re} z \in J, \quad \zeta \le \operatorname{Im} z \le 1\}.$$

Theorem 4.3.1. *Let $I \subset \mathbb{R}$ and let $J \subset \mathbb{R}$ be bounded. Then*

$$\mathbb{P}\left(\sup_{z \in D(J,\zeta)} |\operatorname{Im} F_I(z) - \mathbb{E} \operatorname{Im} F_I(z)| > \mu \right) \le C|J|\mu^{-2}\zeta^{-4} e^{-c\mu\sqrt{N\zeta}}$$

for all $\mu > 0$.

Proof. Let $z = \alpha + i\beta$. Performing the substitution $v = (\tilde{v} - \alpha)/\beta$ and denoting the indicator of $\mathbb{R} \setminus I$ by χ, we obtain

$$\begin{aligned}
\mathbb{E}\, e^{t \operatorname{Im} F_I(z)} &= \left(\beta \int \varrho(\alpha + \beta v) \exp\left(\frac{t}{N\beta} \chi(\alpha + \beta v) \frac{1}{1+v^2} \right) dv \right)^N \\
&\le \left(1 + \frac{t \mathbb{E} \operatorname{Im} F_I(z)}{N} + \frac{t^2 \|\varrho\|_\infty}{N^2 \beta} \int \left(\frac{1}{1+v^2} \right)^2 \exp\left(\frac{t}{N\beta} \frac{1}{1+v^2} \right) dv \right)^N
\end{aligned}$$

by Taylor's theorem. We choose $t = \sqrt{N\beta}$. Since $(1+v^2)^{-2} \in L^1$ and

$$\frac{t}{N\beta} \frac{1}{1+v^2} \le \sqrt{2},$$

there exists an absolute constant $C < \infty$ such that

$$\begin{aligned}
\mathbb{E}\, e^{t \operatorname{Im} F_I(z)} &\le \left(1 + \frac{t \mathbb{E} \operatorname{Im} F_I(z)}{N} + C \left(\frac{t}{N\beta} \right)^2 \beta \right)^N \\
&\le \exp\left(N \left(\frac{t \mathbb{E} \operatorname{Im} F_I(z)}{N} + C \left(\frac{t}{N\beta} \right)^2 \beta \right) \right) \\
&= \exp(t \mathbb{E} \operatorname{Im} F_I(z)) \exp\left(\frac{Ct^2}{N\beta} \right).
\end{aligned}$$

Using an exponential Chebyshev argument, we conclude that

$$\begin{aligned}\mathbb{P}\left(\operatorname{Im} F_I(z) \geq \mathbb{E} \operatorname{Im} F_I(z) + \mu\right) &\leq e^{-t(\mathbb{E} \operatorname{Im} F_I(z)+\mu)} \mathbb{E}\, e^{t \operatorname{Im} F_I(z)} \\ &\leq e^{-t\mu} \exp\left(\frac{Ct^2}{N\beta}\right) \\ &\leq C\, e^{-c\mu\sqrt{N\zeta}}.\end{aligned}$$

The proof of the lower bound works the same way. Replacing the previous Chebyshev bound with

$$\mathbb{P}\left(\operatorname{Im} F_I(z) \leq \mathbb{E} \operatorname{Im} F_I(z) - \mu\right) \leq e^{-t(\mathbb{E} \operatorname{Im} F_I(z)-\mu)} \mathbb{E}\, e^{-t \operatorname{Im} F_I(z)},$$

yields that for every fixed $z \in D(J, \zeta)$

$$\mathbb{P}\left(|\operatorname{Im} F_I(z) - \mathbb{E} \operatorname{Im} F_I(z)| > \mu\right) \leq C\, e^{-c\mu\sqrt{N\zeta}}.$$

Since $D(J, \zeta)$ is bounded, there exists a set of at most $C|J|\mu^{-2}\zeta^{-4}$ points $\{z_k\} \subset D(J, \zeta)$ such that for every $z \in D(J, \zeta)$ there exists k with $|z - z_k| \leq \frac{\mu\zeta^2}{12}$. By the union bound,

$$\mathbb{P}\left(\sup_k |\operatorname{Im} F_I(z_k) - \mathbb{E} \operatorname{Im} F_I(z_k)| > \frac{\mu}{3}\right) \leq C|J|\mu^{-2}\zeta^{-4} e^{-c\mu\sqrt{N\zeta}}.$$

But $\operatorname{Im} F_I$ and $\mathbb{E} \operatorname{Im} F_I$ are $(2/\zeta)^2$-Lipschitz continuous in $D(J, \zeta)$ and thus

$$\begin{aligned}|\operatorname{Im} F_I(z) - \mathbb{E} \operatorname{Im} F_I(z)| \leq |\operatorname{Im} F_I(z) - \operatorname{Im} F_I(z_k)| + |\operatorname{Im} F_I(z_k) - \mathbb{E} \operatorname{Im} F_I(z_k)| \\ + |\mathbb{E} \operatorname{Im} F_I(z_k) - \mathbb{E} \operatorname{Im} F_I(z)| \leq \mu,\end{aligned}$$

extending the bound to all $z \in D(J, \zeta)$. □

Since ϱ was assumed to be bounded below in $W(\varepsilon)$, the corresponding lower bound for $\operatorname{Im} F_\emptyset$ in the ε-fattening of the original spectral domain D follows immediately, proving the second point in Assumption 4.0.1.

Corollary 4.3.2. *There exists $K_l \in (0, \infty)$ such that*

$$\mathbb{P}\left(\inf_{\operatorname{dist}(z,D)\leq\varepsilon} \operatorname{Im} F_\emptyset(z) < K_l\right) \leq C\eta^{-4} e^{-c\sqrt{N\eta}}.$$

Next, we combine the previous estimates with a standard argument for the Hilbert transform to produce a logarithmic bound for $|F_\emptyset|$, proving the first point in Assumption 4.0.1.

Corollary 4.3.3. *There exists $K_m \in (0, \infty)$ such that*

$$\mathbb{P}\left(\sup_{\operatorname{Im} z>\eta} |F_\emptyset(z)| > K_m + K_m \log\left(1 + \eta^{-2}\right)\right) \leq C\eta^{-4} e^{-c\sqrt{N\eta}}.$$

Proof. Using that ϱ is compactly supported and the trivial estimate

$$\operatorname{Im} S_0(z) \leq \frac{1}{\operatorname{dist}(z, \operatorname{supp} \varrho)}$$

Theorem 4.3.1 with $J = \operatorname{supp} \varrho + [-1, 1]$ and $\zeta = \eta/2$ shows that there exist $C, c, K_m \in (0, \infty)$ such that

$$\mathbb{P}\left(\sup_{\operatorname{Im} z > \frac{\eta}{2}} \operatorname{Im} S_0(z) > K_m \right) \leq C\eta^{-4} e^{-c\sqrt{N\eta}}.$$

Letting

$$Q_z = \frac{1}{\pi} \frac{t - \operatorname{Re} z}{(t - \operatorname{Re} z)^2 + (\operatorname{Im} z)^2}$$

be the conjugate Poisson kernel and writing $z = \alpha + i(\beta/2)$, we see that on the complement of this event

$$\begin{aligned}
\operatorname{Re} S_0(\alpha + i\beta) &= \int \operatorname{Im} S_0\,(t - z)\, Q_{i\frac{\beta}{2}}(t)\, dt \\
&= \int_{[-1,1]} \operatorname{Im} S_0\,(t - z)\, Q_{i\frac{\beta}{2}}(t)\, dt + \int_{\mathbb{R}\setminus[-1,1]} \operatorname{Im} S_0\,(t - z)\, Q_{i\frac{\beta}{2}}(t)\, dt \\
&\leq K_m \frac{1}{\pi} \int_{-1}^{1} \frac{|t|}{t^2 + \beta^2}\, dt + \frac{1}{\pi} \int \operatorname{Im} S_0 \left(t + i\frac{\beta}{2} \right) dt \\
&\leq K_m \log\left(1 + \beta^{-2}\right) + 1.
\end{aligned}$$

□

Finally, we use an entropy argument to prove the third point of Assumption 4.0.1. Let $\{I_k\}$ be a collection of at most CN adjacent intervals covering $W(\varepsilon)$ with $|I_k| = \frac{1}{4} N^{-1+\kappa}$ and set

$$J_k = I_{k-1} \cup I_k \cup I_{k+1}.$$

We prove the desired inequality on the event

$$\bigcup_k \left\{ \sup_{z \in D(J,\zeta)} \operatorname{Im} F_{J_k}(z) - \mathbb{E} \operatorname{Im} F_{J_k}(z) \leq N^{-\delta/4} \right\}$$

with $J = W(\varepsilon)$ and $\zeta = K_l T/2$. By Theorem 4.3.1, the probability of this event is close to one since $N^{\delta/4} \ll N^{\delta/2} = \sqrt{NT}$. Now let $I \subset W$ and $z \in D$ be such that $\operatorname{Re} z \in I$ and $\operatorname{dist}(\operatorname{Re} z, \partial I) > \frac{1}{2} N^{-1+\kappa}$. Then there exists k such that $\operatorname{Re} z \in I_k$ and $J_k \subset I$, so that

$$\operatorname{Im} F_I(z) \leq \operatorname{Im} F_{J_k}(z) \leq \mathbb{E} \operatorname{Im} F_{J_k}(z) + N^{-\delta/4}.$$

But $\operatorname{dist}(\operatorname{Re} z, \partial J_k) \geq \frac{1}{4} N^{-1+\kappa}$ and hence

$$\mathbb{E}\, F_{J_k}(z) = \int_{\mathbb{R} \setminus J_k} \frac{\operatorname{Im} z}{(v - \operatorname{Re} z)^2 + (\operatorname{Im} z)^2}\, \varrho(v)\, dv \leq K_i \frac{\operatorname{Im} z}{N^{-1+\kappa}}.$$

References

[1] M. Aizenman and S. Molchanov. Localization at large disorder and at extreme energies: an elementary derivation. *Comm. Math. Phys.*, 157(2):245–278, 1993.

[2] M. Aizenman, M. Shamis, and S. Warzel. Resonances and partial delocalization on the complete graph. *Ann. Henri Poincaré*, 16(9):1969–2003, 2015.

[3] M. Aizenman, R. Sims, and S. Warzel. Stability of the absolutely continuous spectrum of random Schrödinger operators on tree graphs. *Probab. Theory Related Fields*, 136(3):363–394, 2006.

[4] M. Aizenman and S. Warzel. The canopy graph and level statistics for random operators on trees. *Math. Phys. Anal. Geom.*, 9(4):291–333 (2007), 2006.

[5] M. Aizenman and S. Warzel. On the ubiquity of the Cauchy distribution in spectral problems. *Probab. Theory Related Fields*, 163(1-2):61–87, 2015.

[6] M. Aizenman and S. Warzel. *Random operators: Disorder effects on quantum spectra and dynamics*, volume 168 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015.

[7] B. L. Altshuler, E. Cuevas, L. B. Ioffe, and V. E. Kravtsov. Nonergodic phases in strongly disordered random regular graphs. *Phys. Rev. Lett.*, 117:156601, Oct 2016.

[8] M. Amini. Spread of wave packets in disordered hierarchical lattices. *EPL (Europhysics Letters)*, 117(3):30003, 2017.

[9] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[10] P. W. Anderson. Absence of diffusion in certain random lattices. *Phys. Rev.*, 109:1492–1505, Mar 1958.

[11] Z. D. Bai and Y. Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.*, 16(4):1729–1741, 1988.

[12] R. Bauerschmidt, J. Huang, A. Knowles, and H.-T. Yau. Bulk eigenvalue statistics for random regular graphs. *Ann. Probab.*, 45(6A):3626–3663, 2017.

[13] R. Bauerschmidt, A. Knowles, and H.-T. Yau. Local semicircle law for random regular graphs. *Comm. Pure Appl. Math.*, 70(10):1898–1960, 2017.

[14] L. Benigni. Eigenvectors distribution and quantum unique ergodicity for deformed Wigner matrices. Preprint available at arXiv:1711.07103, 2017.

[15] P. Biane. Processes with free increments. *Math. Z.*, 227(1):143–174, 1998.

[16] G. Biroli, A. C. Ribeiro-Teixera, and M. Tarzia. Difference between level statistics, ergodicity and localization transitions on the Bethe lattice. Preprint available at arXiv:1211.7334.

[17] P. M. Bleher and J. G. Sinai. Investigation of the critical point in models of the type of Dyson's hierarchical models. *Comm. Math. Phys.*, 33(1):23–42, 1973.

[18] P. M. Bleher and Y. G. Sinai. Critical indices for Dyson's asymptotically-hierarchical models. *Comm. Math. Phys.*, 45(3):247–278, 1975.

[19] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin. Universality for a class of random band matrices. *Adv. Theor. Math. Phys.*, 21(3):739–800, 2017.

[20] P. Bourgade, J. Huang, and H.-T. Yau. Eigenvector statistics of sparse random matrices. *Electron. J. Probab.*, 22:38 pp., 2017.

[21] P. Bourgade and H.-T. Yau. The eigenvector moment flow and local quantum unique ergodicity. *Comm. Math. Phys.*, 350(1):231–278, 2017.

[22] A. Bovier. The density of states in the Anderson model at weak disorder: a renormalization group analysis of the hierarchical model. *J. Statist. Phys.*, 59(3-4):745–779, 1990.

[23] D. Brydges, S. N. Evans, and J. Z. Imbrie. Self-avoiding walk on a hierarchical lattice in four dimensions. *Ann. Probab.*, 20(1):82–124, 1992.

[24] R. Carmona, A. Klein, and F. Martinelli. Anderson localization for Bernoulli and other singular potentials. *Comm. Math. Phys.*, 108(1):41–66, 1987.

[25] G. Casati, L. Molinari, and F. Izrailev. Scaling properties of band random matrices. *Phys. Rev. Lett.*, 64:1851–1854, Apr 1990.

[26] J.-M. Combes, F. Germinet, and A. Klein. Generalized eigenvalue-counting estimates for the Anderson model. *J. Stat. Phys.*, 135(2):201–216, 2009.

[27] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. Springer Series in Statistics. Springer-Verlag, New York, 1988.

[28] A. De Luca, B. L. Altshuler, V. E. Kravtsov, and A. Scardicchio. Anderson localization on the Bethe lattice: Nonergodicity of extended states. *Phys. Rev. Lett.*, 113:046806, Jul 2014.

[29] R. Del Rio, N. Makarov, and B. Simon. Operators with singular continuous spectrum. II. Rank one operators. *Comm. Math. Phys.*, 165(1):59–67, 1994.

[30] F. Delyon, H. Kunz, and B. a. Souillard. One-dimensional wave equations in disordered media. *J. Phys. A*, 16(1):25–42, 1983.

[31] B. Derrida and R. B. Griffiths. Directed polymers on disordered hierarchical lattices. *EPL (Europhysics Letters)*, 8(2):111, 1989.

[32] M. Disertori, H. Pinson, and T. Spencer. Density of states for random band matrices. *Comm. Math. Phys.*, 232(1):83–124, 2002.

[33] J. Dombrowski. Tridiagonal matrix representations of cyclic selfadjoint operators. *Pacific J. Math.*, 114(2):325–334, 1984.

[34] J. Dombrowski. Tridiagonal matrix representations of cyclic selfadjoint operators. II. *Pacific J. Math.*, 120(1):47–53, 1985.

[35] J. Dombrowski. Cyclic operators, commutators, and absolutely continuous measures. *Proc. Amer. Math. Soc.*, 100(3):457–463, 1987.

[36] M. Duits and K. Johansson. On mesoscopic equilibrium for linear statistics in Dyson's Brownian motion. *To appear in Memoirs of the American Mathematical Society*, 2017.

[37] F. J. Dyson. A Brownian-motion model for the eigenvalues of a random matrix. *J. Mathematical Phys.*, 3:1191–1198, 1962.

[38] F. J. Dyson. Existence of a phase-transition in a one-dimensional Ising ferromagnet. *Comm. Math. Phys.*, 12(2):91–107, 1969.

[39] F. J. Dyson. An Ising ferromagnet with discontinuous long-range order. *Comm. Math. Phys.*, 21:269–283, 1971.

[40] L. Erdős. Universality of Wigner random matrices: a survey of recent results. *Russian Mathematical Surveys*, 66(3):507, 2011.

[41] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Delocalization and diffusion profile for random band matrices. *Comm. Math. Phys.*, 323(1):367–416, 2013.

[42] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. The local semicircle law for a general class of random matrices. *Electron. J. Probab.*, 18:no. 59, 58, 2013.

[43] L. Erdős, B. Schlein, and H.-T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287(2):641–655, 2009.

[44] L. Erdős, B. Schlein, and H.-T. Yau. Universality of random matrices and local relaxation flow. *Invent. Math.*, 185(1):75–119, 2011.

[45] F. Evers and A. D. Mirlin. Anderson transitions. *Rev. Mod. Phys.*, 80:1355–1417, Oct 2008.

[46] D. Facoetti, P. Vivo, and G. Biroli. From non-ergodic eigenvectors to local resolvent statistics and back: A random matrix perspective. *EPL (Europhysics Letters)*, 115(4):47003, 2016.

[47] J. Fröhlich and T. Spencer. Absence of diffusion in the Anderson tight binding model for large disorder or low energy. *Comm. Math. Phys.*, 88(2):151–184, 1983.

[48] Y. V. Fyodorov, A. Kupiainen, and C. Webb. Towards rigorous analysis of the Levitov-Mirlin-Evers recursion. *Nonlinearity*, 29(12):3871–3896, 2016.

[49] Y. V. Fyodorov and A. D. Mirlin. Scaling properties of localization in random band matrices: A σ-model approach. *Phys. Rev. Lett.*, 67:2405–2409, Oct 1991.

[50] Y. V. Fyodorov, A. Ossipov, and A. Rodriguez. The Anderson localization transition and eigenfunction multifractality in an ensemble of ultrametric random matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(12):L12001, 2009.

[51] K. Gawędzki and A. Kupiainen. Renormalization group study of a critical lattice model. I. Convergence to the line of fixed points. *Comm. Math. Phys.*, 82(3):407–433, 1981/82.

[52] K. Gawędzki and A. Kupiainen. Renormalization group study of a critical lattice model. II. The correlation functions. *Comm. Math. Phys.*, 83(4):469–492, 1982.

[53] I. J. Goldsheid, S. A. Molchanov, and L. A. Pastur. A random homogeneous Schrödinger operator has a pure point spectrum. *Funkcional. Anal. i Priložen.*, 11(1):1–10, 96, 1977.

[54] P. D. Hislop and P. Müller. A lower bound for the density of states of the lattice Anderson model. *Proc. Amer. Math. Soc.*, 136(8):2887–2893, 2008.

[55] J. Huang and B. Landon. Local law and mesoscopic fluctuations of Dyson Brownian motion for general β and potential. *arXiv:1612.06306*, 2017.

[56] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[57] A. Klein. Absolutely continuous spectrum in the Anderson model on the Bethe lattice. *Math. Res. Lett.*, 1(4):399–407, 1994.

[58] S. Kotani. Lyapunov exponents and spectra for one-dimensional random Schrödinger operators. In *Random matrices and their applications (Brunswick, Maine, 1984)*, volume 50 of *Contemp. Math.*, pages 277–286. Amer. Math. Soc., Providence, RI, 1986.

[59] V. E. Kravtsov, I. M. Khaymovich, E. Cuevas, and M. Amini. A random matrix model with localization and ergodic transitions. *New Journal of Physics*, 17(12):122002, 2015.

[60] E. Kritchevski. Hierarchical Anderson model. In *Probability and mathematical physics*, volume 42 of *CRM Proc. Lecture Notes*, pages 309–322. Amer. Math. Soc., Providence, RI, 2007.

[61] E. Kritchevski. Spectral localization in the hierarchical Anderson model. *Proc. Amer. Math. Soc.*, 135(5):1431–1440 (electronic), 2007.

[62] E. Kritchevski. Poisson statistics of eigenvalues in the hierarchical Anderson model. *Ann. Henri Poincaré*, 9(4):685–709, 2008.

[63] S. Kuttruf and P. Müller. Lifshits tails in the hierarchical Anderson model. *Ann. Henri Poincaré*, 13(3):525–541, 2012.

[64] B. Landon, P. Sosoe, and H.-T. Yau. Fixed energy universality for Dyson Brownian motion. Preprint available at arXiv:1609.09011.

[65] B. Landon and H.-T. Yau. Convergence of Local Statistics of Dyson Brownian Motion. *Comm. Math. Phys.*, 355(3):949–1000, 2017.

[66] J. O. Lee and K. Schnelli. Local deformed semicircle law and complete delocalization for Wigner matrices with random potential. *J. Math. Phys.*, 54(10):103504, 62, 2013.

[67] F. L. Metz, L. Leuzzi, and G. Parisi. Renormalization flow of the hierarchical Anderson model at weak disorder. *Phys. Rev. B*, 89:064201, Feb 2014.

[68] F. L. Metz, L. Leuzzi, G. Parisi, and V. Sacksteder. Transition between localized and extended states in the hierarchical Anderson model. *Phys. Rev. B*, 88:045103, Jul 2013.

[69] N. Minami. Local fluctuation of the spectrum of a multidimensional Anderson tight binding model. *Comm. Math. Phys.*, 177(3):709–725, 1996.

[70] A. D. Mirlin, Y. V. Fyodorov, F.-M. Dittes, J. Quezada, and T. H. Seligman. Transition from localized to extended eigenstates in the ensemble of power-law random banded matrices. *Phys. Rev. E*, 54:3221–3230, Oct 1996.

[71] S. Molchanov. Lectures on random media. In *Lectures on probability theory (Saint-Flour, 1992)*, volume 1581 of *Lecture Notes in Math.*, pages 242–411. Springer, Berlin, 1994.

[72] S. Molchanov. Hierarchical random matrices and operators. Application to Anderson model. In *Multidimensional statistical analysis and theory of random matrices (Bowling Green, OH, 1996)*, pages 179–194. VSP, Utrecht, 1996.

[73] C. Monthus and T. Garel. A critical Dyson hierarchical model for the Anderson localization transition. *J. Stat. Mech.*, P05005:1–27, 2011.

[74] L. A. Pastur. The spectrum of random matrices. *Teoret. Mat. Fiz.*, 10(1):102–112, 1972.

[75] M. Reed and B. Simon. *Methods of modern mathematical physics. III.* Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1979. Scattering theory.

[76] M. Reed and B. Simon. *Methods of modern mathematical physics. I.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, second edition, 1980. Functional analysis.

[77] N. Rosenzweig and C. E. Porter. "Repulsion of energy levels" in complex atomic spectra. *Phys. Rev.*, 120:1698–1714, Dec 1960.

[78] I. Rushkin, A. Ossipov, and Y. V. Fyodorov. Universal and non-universal features of the multifractality exponents of critical wavefunctions. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(03):L03001, 2011.

[79] C. Sadel. Anderson transition at two-dimensional growth rate on antitrees and spectral theory for operators with one propagating channel. *Ann. Henri Poincaré*, 17(7):1631–1675, 2016.

[80] J. Schenker. Eigenvector localization for random band matrices with power law band width. *Comm. Math. Phys.*, 290(3):1065–1097, 2009.

[81] M. Scheutzow. A stochastic Gronwall lemma. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 16(2):1350019, 4, 2013.

[82] B. Simon. *Trace ideals and their applications*, volume 120 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, second edition, 2005.

[83] B. Simon and T. Spencer. Trace class perturbations and the absence of absolutely continuous spectra. *Comm. Math. Phys.*, 125(1):113–125, 1989.

[84] B. Simon and T. Wolff. Singular continuous spectrum under rank one perturbations and localization for random Hamiltonians. *Comm. Pure Appl. Math.*, 39(1):75–90, 1986.

[85] S. Sodin. The spectral edge of some random band matrices. *Ann. of Math. (2)*, 172(3):2223–2251, 2010.

[86] K. S. Tikhonov, A. D. Mirlin, and M. A. Skvortsov. Anderson localization and ergodicity on random regular graphs. *Phys. Rev. B*, 94:220203, Dec 2016.

[87] D. Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory. I. *Comm. Math. Phys.*, 155(1):71–92, 1993.

[88] M.-K. von Renesse and M. Scheutzow. Existence and uniqueness of solutions of stochastic functional differential equations. *Random Oper. Stoch. Equ.*, 18(3):267–284, 2010.

[89] P. von Soosten and S. Warzel. Non-ergodic delocalization in the Rosenzweig-Porter model. Preprint available at arXiv:1709.10313.

[90] P. von Soosten and S. Warzel. The phase transition in the ultrametric ensemble and local stability of Dyson Brownian motion. Preprint available at arXiv:1705.00923.

[91] P. von Soosten and S. Warzel. Singular spectrum and recent results on hierarchical operators. To appear in Contemporary Mathematics.

[92] P. von Soosten and S. Warzel. Renormalization Group Analysis of the Hierarchical Anderson Model. *Ann. Henri Poincaré*, 18(6):1919–1947, 2017.

[93] F. Wegner. Bounds on the density of states in disordered systems. *Z. Phys. B*, 44(1-2):9–15, 1981.