Delocalization and Diffusion Profile for Random Band Matrices

László Erdős[1][*]   Antti Knowles[2][†]   Horng-Tzer Yau[3][‡]   Jun Yin[4][§]

July 19, 2013

We consider Hermitian and symmetric random band matrices $H = (h_{xy})$ in $d \geqslant 1$ dimensions. The matrix entries h_{xy}, indexed by $x, y \in (\mathbb{Z}/L\mathbb{Z})^d$, are independent, centred random variables with variances $s_{xy} = \mathbb{E}|h_{xy}|^2$. We assume that s_{xy} is negligible if $|x - y|$ exceeds the band width W. In one dimension we prove that the eigenvectors of H are delocalized if $W \gg L^{4/5}$. We also show that the magnitude of the matrix entries $|G_{xy}|^2$ of the resolvent $G = G(z) = (H - z)^{-1}$ is self-averaging and we compute $\mathbb{E}|G_{xy}|^2$. We show that, as $L \to \infty$ and $W \gg L^{4/5}$, the behaviour of $\mathbb{E}|G_{xy}|^2$ is governed by a diffusion operator whose diffusion constant we compute. Similar results are obtained in higher dimensions.

1. Introduction

Random band matrices $H = (h_{xy})_{x,y \in \Gamma}$ represent quantum systems on a large finite graph Γ with random quantum transition amplitudes effective up to distances of order W. The matrix entries are independent, centred random variables. The variance $s_{xy} := \mathbb{E}|h_{xy}|^2$ depends on the distance between the two sites x and y, and it typically decays with the distance on a characteristic length scale W, called the *band width* of H. This terminology comes from the simplest one-dimensional model where the graph $\Gamma = \{1, 2, \ldots, N\}$ is a path on N vertices, and the matrix entries h_{xy} are negligible if $|x - y| \geqslant W$. In particular, if $W = N$ and all variances are equal, we recover the well-known Wigner matrix, which corresponds to a mean-field model. Higher-dimensional models are obtained if Γ taken to be the box of linear size L in $\mathbb{Z}^d$. In this case the dimension of the matrix is $N = L^d$.

Typically, W is a mesoscopic scale, larger than the lattice spacing but smaller than the diameter L of the system: $1 \ll W \ll L$. These models are natural interpolations between random Schrödinger operators with short range quantum transitions such as the Anderson model [3] and mean-field random matrices such as Wigner matrices [40]. In particular, random band matrices may be used to model the *Anderson metal-insulator phase transition*, which we briefly outline.

The key physical parameter of all these models is the *localization length* ℓ, which describes the typical length scale of the eigenvectors of H. The system is said to be *delocalized* if the localization length is

comparable with the system size, $\ell \sim L$, and it is *localized* otherwise. Delocalized systems are electric conductors, while localized systems are insulators.

Nonrigorous supersymmetric calculations [28] show that for random band matrices the localization length is of order $\ell \sim W^2$ in $d = 1$ dimension. In $d = 2$, the localization length is expected to be exponentially growing in W, and in $d \geqslant 3$, it is macroscopic, $\ell \sim L$, i.e. the system is delocalized. We refer to the overview papers of Spencer [36, 38] and to the paper of Schenker [33] for more details on these conjectures.

These predictions are in accordance with those for the *Anderson model*, where the random matrix is of the form $-\Delta + \lambda V$; here Δ is the lattice Laplacian, V a random potential (i.e. a diagonal matrix with i.i.d. entries), and λ a small coupling constant. The localization length is $\ell \sim \lambda^{-2}$ in the regime of strong localization, which corresponds to the whole spectrum for $d = 1$ and a neighbourhood of the spectral edges for $d > 1$. This result follows from the rigorous multiscale analysis of Fröhlich and Spencer [27] as well as from the fractional moment method of Aizenman and Molchanov [2]. The two-dimensional Anderson model is conjectured to be in the weak localization regime with $\ell \sim \exp(\lambda^{-2})$ throughout the spectrum [1], but this has so far not been proved. In dimensions $d \geqslant 3$, the prediction is that there is a threshold energy, called the mobility edge, E_0, that separates the localized regime near the band edges from the delocalized regime in the bulk spectrum. The localization length is expected to diverge as the energy E approaches the mobility edge from the localization side. The increase of the localization length as an inverse power of $E - E_0$ has been rigorously established up to a certain scale in [5, 37], but this analysis does not allow E to actually reach the conjectured value of E_0. A key open question for the Anderson model is to establish the metal-insulator transition, i.e. to show that the mobility edge indeed exists.

For random band matrices, the metal-insulator transition can be investigated even in $d = 1$ by varying the band width W. The prediction that the localization length ℓ is of order W^2 can be recast in the form that the eigenvectors are delocalized if $W \geqslant L^{1/2}$. Currently only lower and upper bounds have been established for ℓ. On the side of localization, Schenker [33] proved that $\ell \leqslant W^8$, uniformly in the system size, by extending the methods of the proofs of the Anderson localization for random Schrödinger operators. As a lower bound, $\ell \geqslant W$ was proved in [24] by using a self-consistent equation for the diagonal matrix entries G_{xx} of the Green function $G = G(z) = (H - z)^{-1}$. In particular, Wigner matrices ($W = L$) are completely delocalized; in fact this has been proven earlier in [18, 19, 21] using a simpler self-consistent equation for the trace of the Green function, $\operatorname{Tr} G$. This lower bound was improved to $\ell \geqslant W^{7/6}$ in [6, 7] by using diagrammatic perturbation theory. In fact, not only was the lower bound on localization length established, but it was also shown that the unitary time evolution, e^{itH}, behaves diffusively on the spatial scale W, i.e. the typical propagation distance is $\sqrt{t}W$. Thus, the mechanism responsible for the delocalization of random band matrices is a random walk (in fact, a superposition of random walks) with step size of order W. Showing that the localization length is greater than the naive size W therefore requires a control of the random walk for large times. For technical reasons, in [6, 7] the time evolution could only be controlled up to time $t \leqslant W^{1/3}$, which corresponds to delocalization on the scale $W^{1+1/6}$. The work [6, 7] was partly motivated by a similar result for the long-time evolution for the Anderson model [15–17] combined with an algebraic renormalization using Chebyshev polynomials [26, 34].

In the current paper we develop a new self-consistent equation (see (4.2)) which keeps track of all matrix entries of the Green function G, and not only the diagonal ones as in [24]. We show that $|G_{xy}|^2$ is self-averaging and $\mathbb{E}|G_{xy}|^2$ behaves as the resolvent of a diffusion operator associated with a superposition of random walks with step size of order W. This result can then be translated into a lower bound on the localization length.

More precisely, for $d = 1$ we obtain full control on $|G_{xy}(z)|^2$ for relatively broad bands, $W \gg N^{4/5}$, and for $\eta = \operatorname{Im} z \geqslant (W/N)^2$. The condition $W \gg N^{4/5}$ is technical. The condition on η comes from the facts

that $t = \eta^{-1}$ corresponds to the time scale of the random walk, and a random walk with step size W in a box of size N reaches equilibrium in a time of order $(N/W)^2$. As a corollary, we prove that most eigenvectors are delocalized if $W \gg N^{4/5}$. This improves the exponent in [6,7], where delocalization for $W \gg N^{6/7}$ was proved. However, unlike in [6,7], here we do not obtain a lower bound on the localization length ℓ uniformly in N. We also prove analogous results in higher dimensions. In addition, we investigate the case where the variances s_{xy} of the matrix entries decay slowly according to the power law $|x - y|^{-(1+\beta)}$ for $0 < \beta < 2$. In this regime the system exhibits superdiffusive behaviour. In particular, we may allow decay of the form $|x - y|^{-2}$, which is critical in the sense of [30].

One key open question for random band matrices is to control the resolvent $G(z)$ for $\eta = \operatorname{Im} z \ll W^{-1}$. None of the results mentioned above yield a nontrivial control below W^{-1}. In the regime $\eta \geqslant W^{-1}$ robust pointwise bounds on G_{xy} have been obtained with high probability [24]. For $\frac{1}{N}\mathbb{E}\operatorname{Tr} G(z)$ and for $\eta \geqslant W^{-0.99}$, a more precise error estimate was derived for a special class of Bernoulli entries in [35]. However, controlling the quantity $\frac{1}{N}\mathbb{E}\operatorname{Tr} G(z)$ does not yield information on the localization length. In the current paper we obtain much more precise bounds on $|G_{xy}|^2$ in the regime $\eta \geqslant (W/N)^2 \geqslant W^{-1/2}$, which in particular imply delocalization bounds.

Supersymmetric (SUSY) methods offer a very attractive approach to studying the delocalization transition in band matrices, but rigorous control of the ensuing functional integrals away from the saddle points is difficult. This task has been performed for the density of states of a special three-dimensional Gaussian model [4]; this is the only result where a nontrivial control for $\eta \leqslant W^{-1}$ (in fact, uniform in η) was obtained. The SUSY method has so far only been applied to the expectation of single Green function, $\mathbb{E}G$, and not to its square, $\mathbb{E}|G|^2$.

The analysis of the trace of the single Green function yields the limiting spectral density of H which is the Wigner semicircle law provided the band width W diverges as $L \to \infty$. For band matrices the semicircle law on large scales, corresponding to spectral parameter $\eta > 0$ independent of N, was given in [31]. More recently, a *semicircle law on small scales*, in which $\eta \ll 1$, was derived in [24] and generalized in [10]. The results of [10, 24] are summarized in Lemma 3.4 below. As an application of our method, we prove a further improvement of the semiricle law in Theorem 2.2 below.

The main new ingredient in this paper is the self-consistent equation for the matrix T, whose entries

$$T_{xy} := \sum_i s_{xi}|G_{iy}|^2$$

are local averages of $|G_{xy}|^2$. We show in Theorem 4.1 below that T satisfies a self-consistent equation of the form

$$T = |m|^2 ST + |m|^2 S + \mathcal{E}, \tag{1.1}$$

where S is the matrix of variances (s_{xy}), $m \equiv m(z)$ is an explicit function of the spectral parameter $z = E + \mathrm{i}\eta$ (see (2.14) below), and $\mathcal{E}$ is an error term. Neglecting the error term $\mathcal{E}$, we obtain

$$T \approx \frac{|m|^2 S}{1 - |m|^2 S}.$$

In this paper we implement the band structure of H using a symmetric probability density f on $\mathbb{R}^d$, by requiring that $s_{xy} \approx W^{-d} f((i-j)/W)$ (see Section 2.1 below for the precise statement). Using translation invariance of S and the Taylor expansion of its Fourier transform $\widehat{S}(p)$ in the low momentum regime, we obtain for $|p| \ll W^{-1}$ that

$$\widehat{S}(p) \approx 1 - W^2 (p \cdot Dp) + \cdots, \tag{1.2}$$

3

where D is the matrix of second moments of f (see (8.1) below). In order to give the leading-order behaviour of T, we use $|m|^2 = 1 - \alpha\eta + O(\eta^2)$ (see (3.5) below), where

$$\alpha \equiv \alpha(E) := \frac{2}{\sqrt{4 - E^2}} \qquad (E = \operatorname{Re} z). \tag{1.3}$$

Therefore the Fourier transform of T is approximately given by

$$\frac{\alpha^{-1}}{\eta + W^2(p \cdot D_{\text{eff}}\, p)} \qquad \text{where} \qquad D_{\text{eff}} := \frac{D}{\alpha}, \tag{1.4}$$

in the regime $|p| \ll W^{-1}$ and $\eta \ll 1$. This corresponds to the diffusion approximation on scales larger than W with an effective diffusion constant D_{eff}. In the language of diagrammatic perturbation theory, the change from D to D_{eff} has the interpretation of a self-energy renormalization. This result coincides with Equation (1.5.5) of [36], which was obtained by computing the sum of ladder diagrams in a high-moment expansion.

The main result of this paper is a justification of this heuristic argument in a certain range of parameters. The error term $\mathcal{E}$ contains fluctuations of local averages. Roughly speaking, we need to control the size of $\sum_x \left[|G_{xy}|^2 - P_x |G_{xy}|^2 \right]$, where P_x denotes partial expectation with respect to the matrix entries in the x-th row (see $\widetilde{T}_{xy}$ in (4.5) below). Unfortunately, $|G_{xy}|^2$ and $|G_{x'y}|^2$ for $x \neq x'$ are not independent; in fact they are strongly correlated for small η, and they do not behave like independent random variables. Estimating high moments of these averages requires an unwrapping of the hierarchical correlation structure among several resolvent matrix entries. The necessary estimates are quite involved. They are a special case of the more general Fluctuation Averaging Theorem that is published separately [8], and was originally developed for application in the current paper. There have been several previous results in this direction; see [23, Lemma 5.2], [25, Lemma 4.1], [11, Theorem 5.6], and [32, Theorem 3.2]. The Fluctuation Averaging Theorem generalizes these ideas to arbitrary monomials of G and exploits an additional cancellation mechanism in averages of $|G_{xy}|^2$ that is not present in averages of G_{xx}. For more details, see [8].

2. Formulation of the results

2.1. Setup. Fix $d \in \mathbb{N}$ and let f be a smooth and symmetric (i.e. $f(x) = f(-x)$) probability density on $\mathbb{R}^d$. Let L and W be integers satisfying

$$L^\delta \leqslant W \leqslant L \tag{2.1}$$

for some fixed $\delta > 0$. The parameter L is the fundamental large quantity of our model. Define the d-dimensional discrete torus

$$\mathbb{T}_L^d := [-L/2, L/2)^d \cap \mathbb{Z}^d.$$

Thus, $\mathbb{T}_L^d$ has $N := L^d$ lattice points. For the following we fix an (arbitrary) ordering of $\mathbb{T}_L^d$, which allows us to identify it with $\{1, \ldots, N\}$. We define the canonical representative of $i \in \mathbb{Z}^d$ through

$$[i]_L := (i + L\mathbb{Z}^d) \cap \mathbb{T}_L^d,$$

and introduce the periodic distance

$$|i|_L := \big| [i]_L \big|,$$

where $|\cdot|$ denotes Euclidean distance in $\mathbb{R}^d$.

4

Define the $N \times N$ matrix $S(L, W) \equiv S = (s_{ij} : i, j \in \mathbb{T}_L^d)$ through

$$s_{ij} := \frac{1}{Z_{L,W}} f\left(\frac{[i-j]_L}{W}\right), \tag{2.2}$$

where $Z_{L,W}$ is a normalization constant chosen so that S is a stochastic matrix:

$$\sum_j s_{ij} = 1 \tag{2.3}$$

for all $i \in \mathbb{T}_L^d$. Unless specified otherwise, summations are always over the set $\mathbb{T}_L^d$. By symmetry of f we find that S is symmetric: $s_{ij} = s_{ji}$. As a stochastic matrix, the spectrum of S lies in $[-1, 1]$. In fact it is proved in Lemma A.1 of [24] that there exists a positive constant δ, depending only on f, such that

$$-1 + \delta \leqslant S \leqslant 1. \tag{2.4}$$

We let $(\zeta_{ij} : i \leqslant j)$, where $i, j \in \mathbb{T}_L^d$, be a family of independent, complex-valued, centred random variables $\zeta_{ij} \equiv \zeta_{ij}^{(N)}$ satisfying

$$\mathbb{E}\zeta_{ij} = 0, \qquad \mathbb{E}|\zeta_{ij}|^2 = 1, \qquad \zeta_{ii} \in \mathbb{R}. \tag{2.5}$$

For $i > j$ we define

$$\zeta_{ij} := \bar{\zeta}_{ji}.$$

We define the band matrix $H = (h_{ij})_{i,j \in \mathbb{T}_L^d}$ through

$$h_{ij} := (s_{ij})^{1/2} \zeta_{ij}. \tag{2.6}$$

Thus we have $H = H^*$ and

$$\mathbb{E}|h_{ij}|^2 = s_{ij}. \tag{2.7}$$

In particular, we may consider the two classical symmetry classes of random matrices: real symmetric and complex Hermitian. For *real symmetric band matrices* we assume

$$\zeta_{ij} \in \mathbb{R} \quad \text{for all} \quad i \leqslant j. \tag{2.8}$$

For *complex Hermitian band matrices* we assume

$$\mathbb{E}\zeta_{ij}^2 = 0 \quad \text{for all} \quad i < j. \tag{2.9}$$

in addition to (2.5). A common way to satisfy (2.9) is to choose the real and imaginary parts of ζ_{ij} to be independent with identical variance. As in [8], our results also hold without this assumption, but we omit the details of this generalization to avoid needless complications.

We introduce the parameter

$$M \equiv M_N := \frac{1}{\max_{i,j} s_{ij}}. \tag{2.10}$$

From the definition of S it is easy to see that $Z_{N,W} = W^d + O(W^{d-1})$. In particular,

$$M = \left(W^d + O(W^{d-1})\right)/\|f\|_\infty.$$

We assume that the random variables ζ_{ij} have finite moments, uniformly in N, i, and j, in the sense that for all $p \in \mathbb{N}$ there is a constant μ_p such that

$$\mathbb{E}|\zeta_{ij}|^p \ \leqslant \ \mu_p \tag{2.11}$$

for all N, i, and j.

The following definition introduces a notion of a high-probability bound that is suited for our purposes.

DEFINITION 2.1 (STOCHASTIC DOMINATION). *Let* $X = \big(X^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\big)$ *be a family of random variables, where* $U^{(N)}$ *is a possibly N-dependent parameter set. Let* $\Psi = \big(\Psi^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\big)$ *be a deterministic family satisfying* $\Psi^{(N)}(u) \geqslant 0$. *We say that X is* stochastically dominated by Ψ, *uniformly in* u, *if for all $\varepsilon > 0$ and $D > 0$ we have*

$$\sup_{u \in U^{(N)}} \mathbb{P}\Big[\big|X^{(N)}(u)\big| > N^\varepsilon \Psi^{(N)}(u)\Big] \ \leqslant \ N^{-D}$$

for large enough $N \geqslant N_0(\varepsilon, D)$. Unless stated otherwise, throughout this paper the stochastic domination will always be uniform in all parameters apart from the parameter δ in (2.1) and the sequence of constants μ_p in (2.11); thus, $N_0(\varepsilon, D)$ also depends on δ and μ_p. If X is stochastically dominated by Ψ, uniformly in u, we use the equivalent notations

$$X \ \prec \ \Psi \qquad and \qquad X \ = \ O_\prec(\Psi)\,.$$

For example, using Chebyshev's inequality and (2.11) one easily finds that

$$h_{ij} \ \prec \ (s_{ij})^{1/2} \ \leqslant \ M^{-1/2}\,, \tag{2.12}$$

so that we may also write $h_{ij} = O_\prec((s_{ij})^{1/2})$. The relation $\prec$ satisfies the familiar algebraic rules of order relations. The general statements are formulated later in Lemma 3.3.

We remark that Definition 2.1 is tailored to the assumption that (2.11) holds for any p. If (2.11) only holds for some large but fixed p then all of our results still hold, but in a somewhat weaker sense. Indeed, the control of the exceptional events in our theorems is expressed via the relation $\prec$. If only finitely many moments are assumed to be finite in (2.11), then the exponents ε and D in the definition of $\prec$ cannot be chosen to be arbitrary, and will in fact depend on p. Repeating our arguments under this weaker assumption would require us to follow all of these exponents through the entire proof. Our assumption that (2.11) holds for any p streamlines our statements and proofs, by avoiding the need to keep track of the precise values of these parameters.

Throughout the following we make use of a spectral parameter

$$z \ = \ E + \mathrm{i}\eta\,, \qquad E \in \mathbb{R}\,, \qquad \eta > 0\,.$$

We choose and fix two arbitrary (small) global constants $\gamma > 0$ and $\kappa > 0$. All of our estimates will depend on κ and γ, and we shall often omit the explicit mention of this dependence. Set

$$\mathbf{S} \ \equiv \ \mathbf{S}^{(N)}(\kappa, \gamma) \ := \ \big\{ E + \mathrm{i}\eta : -2 + \kappa \leqslant E \leqslant 2 - \kappa\,, \ M^{-1+\gamma} \leqslant \eta \leqslant 10 \big\}\,. \tag{2.13}$$

We shall always assume that the spectral parameter z lies in $\mathbf{S}(\kappa, \gamma)$. In this paper we always consider families $X^{(N)}(u) = X_i^{(N)}(z)$ indexed by $u = (z, i)$, where $z \in \mathbf{S}(\kappa, \gamma)$ and i takes on values in some finite (possibly N-dependent or empty) index set.

We introduce the Stieltjes transform of Wigner's semicircle law, defined by

$$m(z) := \frac{1}{2\pi} \int_{-2}^{2} \frac{\sqrt{4 - \xi^2}}{\xi - z} \, \mathrm{d}\xi \, . \tag{2.14}$$

It is well known that the Stieltjes transform m is characterized by the unique solution of

$$m(z) + \frac{1}{m(z)} + z = 0 \tag{2.15}$$

with $\operatorname{Im} m(z) > 0$ for $\operatorname{Im} z > 0$. Thus we have

$$m(z) = \frac{-z + \sqrt{z^2 - 4}}{2} \, . \tag{2.16}$$

To avoid confusion, we remark that the Stieltjes transform m was denoted by m_{sc} in the papers [9, 11–14, 18–25], in which m had a different meaning from (2.14).

We define the *resolvent* of H through

$$G \equiv G(z) := (H - z)^{-1} \, ,$$

and denote its entries by $G_{ij}(z)$. In the following sections we list our main results on the resolvent matrix entries.

We conclude this section by introducing some notation that will be used throughout the paper. We use C to denote a generic large positive constant, which may depend on some fixed parameters and whose value may change from one expression to the next. Similarly, we use c to denote a generic small positive constant. For two positive quantities A_N and B_N we sometimes use the notation $A_N \asymp B_N$ to mean $cA_N \leqslant B_N \leqslant CA_N$. Moreover, we use $A_N \ll B_N$ to mean that there exists a constant $c > 0$ such that $A_N \leqslant N^{-c}B_N$; we also use $A_N \gg B_N$ to denote $B_N \ll A_N$. (Note that these latter conventions are nonstandard.) Finally, we introduce the Japanese bracket $\langle x \rangle := \sqrt{1 + |x|^2}$. Most quantities in this paper depend on the spectral parameter z, which we however mostly omit from the notation.

For simplicity, here we state our main results assuming that $d = 1$ and that f satisfies the decay condition

$$|f(x)| \leqslant C_n \langle x \rangle^{-n} \qquad \text{for all } n \in \mathbb{N} \, . \tag{2.17}$$

Since $d = 1$, we have $N = L$ and we shall consistently use N instead of L. Similarly, $M \asymp W$, and we shall consistently use W in estimates. We also abbreviate $\mathbb{T}_N^1 \equiv \mathbb{T}$.

The generalization of our results to $d > 1$ and slowly decaying f is straightforward, and will be given in Section 8. We emphasize that the core of our argument, given in Sections 3–5, is valid in general, independent of the dimension.

2.2. Improved local semicircle law for resolvent entries and delocalization. Throughout this section we assume $d = 1$ and (2.17). The Wigner semicircle law states that the normalized trace, $\frac{1}{N} \operatorname{Tr} G(z)$, is asymptotically given by $m(z)$. In fact, this asymptotics holds even for individual matrix entries. Our first theorem controls the (z-dependent) random variable

$$\Lambda(z) := \max_{x,y} \left| G_{xy}(z) - \delta_{xy} m(z) \right| .$$

For the following we introduce the deterministic control parameter $\Phi \equiv \Phi^{(N)}(z)$ through

$$\Phi^2 := \max \left\{ \frac{1}{N\eta}, \frac{1}{W\sqrt{\eta}} \right\} . \tag{2.18}$$

THEOREM 2.2 (IMPROVED LOCAL SEMICIRCLE LAW). *Assume $d = 1$ and* (2.17). *Suppose moreover that*

$$N \ll W^{5/4}, \qquad \eta \gg N^2/W^3. \tag{2.19}$$

Then we have

$$\Lambda^2 \prec \Phi^2 \tag{2.20}$$

for $z \in \mathbf{S}$.

Clearly, the assumption $\eta \gg N^2/W^3$ can be replaced with the stronger assumption $\eta \gg W^{-1/2}$. The assumption $N \ll W^{5/4}$ is technical; to see why it is needed, see (6.3) in the proof of Theorem 2.2 below. In the regime (2.19), Theorem 2.2 improves the earlier result

$$\Lambda^2 \prec \frac{1}{M\eta} \tag{2.21}$$

proved in [24] (see Lemma 3.4 below). In fact, the estimate (2.20) is optimal, as may be seen from (2.38) and the first estimate of (2.30) below. By spectral decomposition of G one easily finds that

$$\frac{1}{N^2} \sum_{x,y} |G_{xy}|^2 = \frac{1}{N^2} \operatorname{Tr} G^* G = \frac{1}{N\eta} \operatorname{Im} \frac{\operatorname{Tr} G}{N} = \frac{\operatorname{Im} m}{N\eta} + O_\prec\left(\frac{\Lambda}{N\eta}\right).$$

Thus, in the regime where Λ is bounded, the average of $|G_{xy}|^2$ is of order $(N\eta)^{-1}$. Here we introduced the notation $G^*(z) := (G(z))^* = (H - \bar{z})^{-1}$, which we shall use throughout the following.

REMARK 2.3. The bound (2.20) implies an estimate on the Stieltjes transform of the empirical spectral density, $m_N(z) := N^{-1} \operatorname{Tr} G(z)$. Under the assumptions of Theorem 2.2 and the conditions (2.19), we have

$$m_N(z) - m(z) \prec \Phi^2 \tag{2.22}$$

for $z \in \mathbf{S}$. Once $\Lambda \prec \Phi$ is established, (2.22) easily follows from

$$\frac{1}{N} \sum_k Q_k G_{kk} \prec \Phi^2 ; \tag{2.23}$$

we leave the details to the reader. We remark that (2.23) is the simplest form of the fluctuation averaging mechanism (see Section 3.1). A concise proof of (2.23) can be found in [10, Theorem 4.6].

For $\eta \leqslant (W/N)^2$ we have $\Phi^2 = (N\eta)^{-1}$, and the bound (2.20) therefore shows that all off-diagonal entries of G have a magnitude comparable with the average of their magnitudes. We say that the resolvent is *completely delocalized*. Complete delocalization of the resolvent implies that the eigenvectors are completely delocalized in a weak sense. The precise formulation is given in Proposition 7.1 below. By choosing η such that $W^{-1/2} \leqslant \eta \leqslant (W/N)^2$ and invoking Proposition 7.1 we obtain the following corollary.

COROLLARY 2.4 (EIGENVECTOR DELOCALIZATION). *Assume $d = 1$ and* (2.17). *If $N \ll W^{5/4}$ then the eigenvectors of H are completely delocalized in the sense of Proposition 7.1 below.*

This corollary improves the result in [6,7], where complete eigenvector delocalization (in a slightly weaker sense; see Remark 2.7 below) was proved under the condition $N \ll W^{7/6}$. It was observed in Section 11 of [7] that the graphical perturbative renormalization scheme of [6,7] faces a fundamental barrier at $N = W^{6/5}$. The reason for this barrier is that a large family of graphs whose contribution was subleading for $N \ll W^{6/5}$ in fact yield a leading-order contribution for $N \geqslant W^{6/5}$ if estimated individually. The cancellation mechanism among these subleading graphs has so far not been identified. As evidenced by Corollary 2.4, our present approach goes beyond this barrier.

2.3. Diffusion profile. In this section we assume $d = 1$ and (2.17). In the previous section we saw that for $\eta \leqslant (W/N)^2$ the profile of $|G_{xy}|^2$ is essentially flat. In the complementary regime, $\eta \geqslant (W/N)^2$, the averaged resolvent $\mathbb{E}|G_{xy}|^2$ is determined by a non-constant deterministic profile given by the diffusion approximation

$$\Theta_{xy} := \left(\frac{|m|^2 S}{1 - |m|^2 S} \right)_{xy} \qquad (x, y \in \mathbb{T}). \tag{2.24}$$

Note that the matrix $\Theta = (\Theta_{xy})$ solves the equation

$$\Theta = |m|^2 S\Theta + |m|^2 S,$$

which is obtained from (1.1) by dropping the error term $\mathcal{E}$. Clearly, Θ_{xy} is translation invariant, i.e. $\Theta_{xy} = \Theta_{u0}$ with $u = [x - y]_N$. Moreover, $\Theta_{xy} > 0$ for all x, y. Indeed, this follows immediately from the geometric series representation

$$\Theta_{xy} = \sum_{n \geqslant 1} |m|^{2n} (S^n)_{xy}, \tag{2.25}$$

which converges by $|m| < 1$ (see (3.6) below) and the trivial bound $0 \leqslant (S^n)_{xy} \leqslant 1$, as follows from (2.3).

The representation (2.25) in fact provides the following interpretation of Θ_{xy} in terms of random walks. From (3.5) below we find that $|m|^2 \approx e^{-\alpha\eta}$ (recall the definition of α from (1.3)). Thus the right-hand side of (2.25) may be approximately written as $\sum_{n \geqslant 1} e^{-n\alpha\eta} (S^n)_{xy}$. By definition, S is a doubly stochastic matrix – the transition matrix of a random walk on $\mathbb{T}$ whose steps are of size W and whose transition probabilities are given by $p(x \to y) = s_{xy}$. The normalized variance of each step is given by the *unrenormalized diffusion constant*

$$D \equiv D_W := \frac{1}{2} \sum_{u \in \mathbb{T}} \left(\frac{u}{W} \right)^2 s_{u0}. \tag{2.26}$$

(We normalize by W^{-2} to account for the fact that the distribution s_{u0} has variance $O(W^2)$.) It is easy to see that

$$D = D_\infty + O(W^{-1}) \qquad \text{where} \qquad D_\infty := \frac{1}{2} \int x^2 f(x) \, \mathrm{d}x. \tag{2.27}$$

We conclude that Θ_{xy} is a superposition of random walks up to times of order $(\alpha\eta)^{-1}$. In this superposition the random walk with n steps carries a weight $|m|^{2n} \approx e^{-n\alpha\eta}$, so that walks with times larger than $(\alpha\eta)^{-1}$ are strongly suppressed. The total weight of Θ_{u0} is

$$\sum_u \Theta_{u0} = \sum_{n \geqslant 1} |m|^{2n} \approx (\alpha\eta)^{-1}; \tag{2.28}$$

a precise computation is given in (5.2) below.

The following theorem shows that an averaged version of $|G_{xy}|^2$ is asymptotically given by Θ_{xy} with high probability. The averaging can be done in two ways. First, we can take the expectation $\mathbb{E}|G_{xy}|^2$. In fact, taking *partial expectation* $P_x|G_{xy}|^2$ is enough; here P_x denotes partial expectation in the randomness of the x-th row of H (see Definition 3.2 below). Second, we can average in the index x (or y or both) on a scale of W; for simplicity we consider the weighted average

$$T_{xy} := \sum_i s_{xi} |G_{iy}|^2. \tag{2.29}$$

Note that T is not symmetric, but our results also hold for T_{xy} replaced with the quantities $\sum_j s_{yj} |G_{xj}|^2$ or $\sum_{i,j} s_{xi} s_{yj} |G_{ij}|^2$.

9

THEOREM 2.5 (DIFFUSION PROFILE). *Assume $d = 1$ and (2.17). Suppose that $N \ll W^{5/4}$ and $(W/N)^2 \leqslant \eta \leqslant 1$. Then*

$$|T_{xy} - \Theta_{xy}| \; \prec \; \frac{1}{N\eta}, \qquad \left| P_x |G_{xy}|^2 - \delta_{xy}|m|^2 - |m|^2 \Theta_{xy} \right| \; \prec \; \frac{1}{N\eta} + \frac{\delta_{xy}}{\sqrt{W}} . \tag{2.30}$$

In addition, we have the upper bounds

$$T_{xy} \; \prec \; \Upsilon_{xy} \tag{2.31}$$

and

$$\left| G_{xy} - \delta_{xy} m \right|^2 \; \prec \; \Upsilon_{xy} , \tag{2.32}$$

where we defined

$$\Upsilon_{xy} \; \equiv \; \Upsilon_{xy}^{(K)} \; := \; \frac{1}{N\eta} + \frac{1}{W\sqrt{\eta}} \exp\left[-\frac{\sqrt{\alpha\eta}}{W\sqrt{D}} |x - y|_N \right] + \frac{1}{W}\left\langle \frac{\sqrt{\eta}|x - y|_N}{W} \right\rangle^{-K} . \tag{2.33}$$

Here K is an arbitrary, fixed, positive integer. All estimates are uniform in $z \in \mathbf{S}$ and $x, y \in \mathbb{T}$.

Note that the total mass of the distribution $|G_{x0}|^2$ may be computed explicitly by spectral decomposition of G: assuming $\Lambda \prec \Psi$ we have

$$\sum_x T_{x0} \; = \; \sum_x |G_{x0}|^2 \; = \; \frac{\operatorname{Im} G_{00}}{\eta} \; = \; \frac{\operatorname{Im} m}{\eta}(1 + O_\prec(\Psi)) , \tag{2.34}$$

in agreement with the corresponding statement (2.28) for the deterministic limiting profile.

REMARK 2.6. We expect that (2.30) should in fact hold under the weaker conditions $\eta \gg \frac{1}{N}$ and $N \ll W^2$. The improved local semicircle law (2.20) should also hold under these weaker conditions. In particular, this would imply complete delocalization of the eigenvectors for all $N \ll W^2$. One obstacle is that a non-trivial control on Λ in the regime $\eta \leqslant \frac{1}{W}$ is difficult to obtain.

REMARK 2.7. In [6, 7] a diffusion approximation was proved for $\mathbb{E}\left| (e^{-itH})_{xy} \right|^2$ up to times $t \ll W^{1/3}$; this result was established only in a *weak sense*, i.e. by integrating against a test function in $x - y$, living on the diffusive scale $Wt^{1/2}$. The formula

$$\frac{1}{H - E - i\eta} \; = \; i \int_0^\infty e^{-it(H - E - i\eta)} \, dt \tag{2.35}$$

relates the resolvent with the unitary time evolution. Notice that the time integration can be truncated at $t \leqslant t_0$ with t_0 slightly larger than η^{-1}. Hence, controlling the resolvent whose spectral parameter has imaginary part greater than η is basically equivalent to controlling the unitary time evolution up to time $t = \eta^{-1}$. Although it was not explicitly worked out in [6, 7], the control on e^{-itH} up to $t \ll W^{1/3}$ allows one to control the resolvent for $\eta \gg W^{-1/3}$. Theorem 2.5 (combined with Theorem 2.2) is thus stronger than the results of [6, 7] in the following three senses.

(i) The resolvent is controlled for $\eta \geqslant W^{-1/2}$ (instead of $\eta \gg W^{-1/3}$).

(ii) The control on the profile is pointwise in x and y (instead of in a weak sense on the scale $W\eta^{-1/2}$).

(iii) The estimates hold with high probability (instead of in expectation).

However, the result in the current paper is not uniform in N, unlike that of [6,7].

We conclude this section with an asymptotic result on the deterministic profile Θ_{x0}. Since we are interested in large values of x, we need to consider the small-momentum behaviour of the Fourier transform of Θ_{x0}. Using the small-p expansion (1.2) and (1.4), we therefore find that $\Theta_{x0} \approx \theta_x$, where we defined the N-periodic function

$$\theta_x \; := \; \frac{|m|^2}{N} \sum_{p \in \frac{2\pi}{N}\mathbb{Z}} \mathrm{e}^{\mathrm{i}px} \frac{1}{\alpha\eta + W^2 D p^2} \; = \; \frac{|m|^2}{2W\sqrt{D\alpha\eta}} \sum_{k \in \mathbb{Z}} \exp\left[-\frac{\sqrt{\alpha\eta}}{W\sqrt{D}} \left|x + kN\right|\right]; \qquad (2.36)$$

here the second equality follows by Poisson summation and the Fourier transform $\int \mathrm{e}^{\mathrm{i}px}(1+p^2)^{-1}\,\mathrm{d}p = \pi\mathrm{e}^{-|x|}$. The following proposition, proved in Appendix A, gives the precise statement.

PROPOSITION 2.8 (DETERMINISTIC DIFFUSION PROFILE). *Assume $d = 1$ and (2.17). For each $K \in \mathbb{N}$ we have*

$$\Theta_{xy} \; = \; \theta_{x-y} + O\left(\frac{1}{W^2}\right) + O_K\left(\frac{1}{W}\left\langle\frac{\sqrt{\eta}\,|x-y|_N}{W}\right\rangle^{-K}\right) \qquad (2.37)$$

uniformly for x, y, and $z \in \mathbf{S}$ with $\eta \ll 1$.
   In particular,

$$\max_{x,y} \Theta_{xy} \; \asymp \; \Phi^2. \qquad (2.38)$$

Moreover, if $(W/N)^2 \leqslant \eta \leqslant 1$ and $N \leqslant W^2$, we have the sharp upper bound $\Theta_{xy} \leqslant C\Upsilon_{xy}$.

REMARK 2.9. The leading-order behaviour of (2.37) is given by (2.36). If $\eta \ll (W/N)^2$ then θ is essentially a constant, i.e. the profile is flat. Conversely, if $\eta \gg (W/N)^2$ then the leading term on the right-hand side of (2.36) is given by the term $k = 0$ (by periodicity of θ we assume that $x \in \mathbb{T}$). This is an exponentially decaying profile on the scale $|x| \sim W\eta^{-1/2}$. The shape of the profile is therefore nontrivial if and only if $\eta \gg (W/N)^2$. Note that in both of the above regimes the error terms in (2.37) are negligible compared with the main term.

The total mass of the profile $\sum_{x \in \mathbb{T}} \theta_x$ is given by N times the term $p = 0$ in the first sum of (2.36):

$$\sum_{x \in \mathbb{T}} \theta_x \; = \; \frac{|m|^2}{\alpha\eta} \; = \; \frac{\mathrm{Im}\,m}{\eta}\big(1 + O(\eta)\big), \qquad (2.39)$$

where in the last step we used the elementary identities (3.3) and (3.5) below. In fact, the calculation (2.39) is a mere consistency check (to leading order) since $\sum_x \Theta_{x0} = \frac{\mathrm{Im}\,m}{\eta}$; see (5.2) below. We conclude that the average height of the profile is of order $(N\eta)^{-1}$. The peak of the exponential profile has height of order $(W\sqrt{\eta})^{-1}$, which dominates over the average height if and only if $\eta \gg (W/N)^2$. The regime $\eta \gg (W/N)^2$ corresponds to the regime where η is sufficiently large that the complete delocalization has not taken place, and the profile is mostly concentrated in the region $|x - y| \leqslant W\eta^{-1/2} \ll N$.

These scenarios are best understood in a dynamical picture in which η is decreased down from 1. The ensuing dynamics of θ corresponds to the diffusion approximation, where the quantum problem is replaced with a random walk of step-size of order W. On a configuration space consisting of N sites, such a random walk will reach an equilibrium beyond time scales $(N/W)^2$. As observed in Remark 2.7, η^{-1} plays the role of time t, so that in this dynamical picture equilibrium is reached for $t \sim \eta^{-1} \gg (N/W)^2$. Figure 1 illustrates this diffusive spreading of the profile for different values of η.
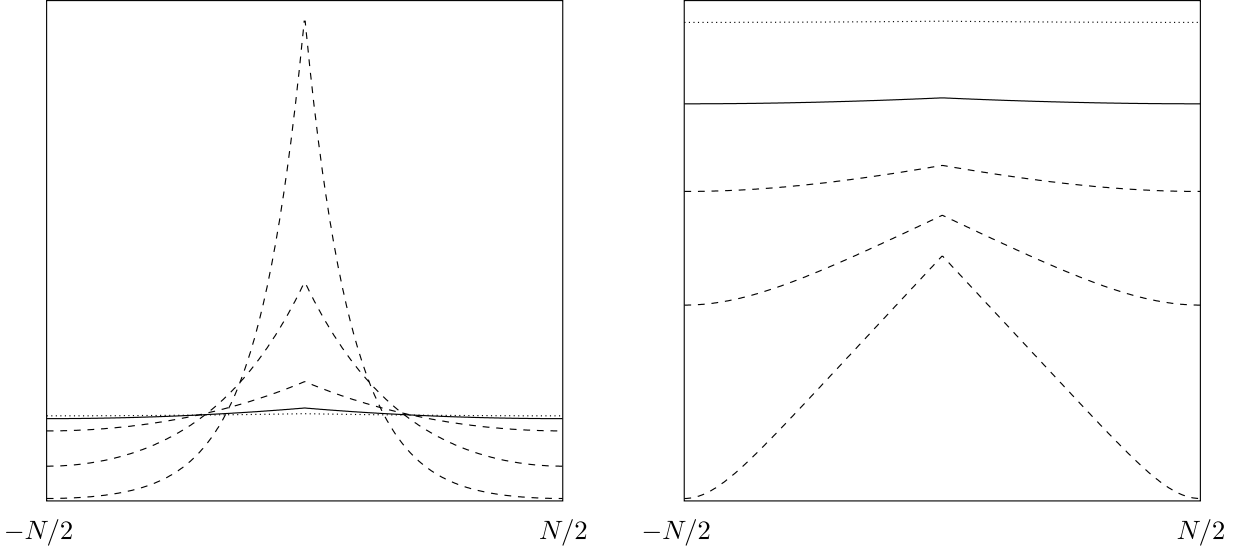
FIGURE 1. A plot of the diffusion profile function at five different values of η, where the argument x ranges over the torus $\mathbb{T}$. Left: the graph $x \mapsto \eta\theta_x$ (see (2.39) for the choice of normalization). Right: the graph $x \mapsto \log\theta_x$. Here we chose $N = 25W$ and $\eta = 5^{-k}$ for $k = 1, 2, 3, 4, 5$. The cases $k = 1, 2, 3$ (where $\eta > (W/N)^2$) are drawn using dashed lines, the case $k = 4$ (where $\eta = (W/N)^2$) using solid lines, and the case $k = 5$ (where $\eta < (W/N)^2$) using dotted lines.

REMARK 2.10. The W-dependent quantity D in the definition (2.36) may be replaced with the constant D_∞ on the right-hand side of (2.36), at the expense of a multiplicative error $(1 + O(W^{-1} \log W))$ and an additive error $O(W^{-2})$. Indeed, the replacement $D \mapsto D_\infty$ in the prefactor is trivial by (2.27). In order to estimate the error arising from the replacement $D \mapsto D_\infty$ in the exponent, we use the estimate $\mathrm{e}^{-\xi(1+O(1/W))} = \mathrm{e}^{-\xi}(1 + O(x/W))$ with $\xi \sim \frac{\sqrt{\eta}}{W}|x - y + kN|$. If $\xi \leqslant C_0 \log W$ then $x/W \leqslant C_0 W^{-1} \log W$, which is small enough. On the other hand, if $\xi \geqslant C_0 \log W$ then $\mathrm{e}^{-\xi} \leqslant W^{-C_0}$, so that the resulting error is an additive error $O(W^{-2})$.

2.4. Delocalization with a small mean-field component. In this section we continue to assume $d = 1$ and (2.17). We now consider a related model

$$H_\varepsilon = (1 - \varepsilon)^{1/2} H + \sqrt{\varepsilon} U, \tag{2.40}$$

where H is the band matrix from Section 2.1, $U = (u_{ij})$ is a standard Wigner matrix independent of H, and $\varepsilon \leqslant \frac{1}{2}$ is a small parameter. We assume that U has the same symmetry type as H, i.e. either (2.8) or (2.9). Its matrix entries are normalized such that $\mathbb{E}u_{ij} = 0$ and $\mathbb{E}|u_{ij}|^2 = \frac{1}{N}$. Moreover, in analogy to (2.11), we make the technical assumption that for each p there exists a constant μ_p such that $\mathbb{E}|N^{1/2}u_{ij}|^p \leqslant \mu_p$ for all N, i, and j.

Let $S_\varepsilon = (s_{ij}^{(\varepsilon)})$ denote the matrix of variances of the entries of $H_\varepsilon = (h_{ij}^{(\varepsilon)})$, i.e. $s_{ij}^{(\varepsilon)} := \mathbb{E}|h_{ij}^{(\varepsilon)}|^2$. We find

$$S_\varepsilon = (1 - \varepsilon)S + \varepsilon\mathbf{e}\mathbf{e}^*,$$

12

where we introduced the vector $\mathbf{e} := N^{-1/2}(1,1,\dots,1)^T$. (Hence $\mathbf{ee}^*$ is the matrix of the variances of U.) Clearly, $0 \leqslant S_\varepsilon \leqslant 1$ and S_ε is a symmetric stochastic matrix satisfying (2.3) and (2.10).

The effect of adding a small Wigner component of size ε is that the imaginary part of the spectral parameter effectively increases from η to $\eta + \varepsilon$ in the local semicircle law and in the diffusion approximation. In particular, we can eliminate the condition $N \leqslant W^{5/4}$ and still obtain delocalization for H_ε provided ε is not too small. These results are summarized in the following theorem. In order to state it, we introduce the control parameter

$$\Phi_\varepsilon^2 := \max\left\{\frac{1}{N\eta}, \frac{1}{W\sqrt{\varepsilon+\eta}}\right\}, \tag{2.41}$$

which is analogous to Φ defined in (2.18).

THEOREM 2.11 (DELOCALIZATION WITH SMALL MEAN-FIELD COMPONENT). *Assume $d = 1$ and (2.17). The following estimates hold uniformly for $z \in \mathbf{S}$.*

(i) *Suppose that $\eta(\eta + \varepsilon) \gg W^{-1}$. Moreover, suppose that $N \ll W^{5/4}$ or $\eta + \varepsilon \gg W^{-1/2}$. Then*

$$\Lambda^2 \prec \Phi_\varepsilon^2. \tag{2.42}$$

(ii) *Suppose that $\varepsilon + \eta \gg W^{-1/2}$ and*

$$\frac{1}{W(\varepsilon+\eta)} \ll \eta \leqslant \frac{W\sqrt{\varepsilon+\eta}}{N}. \tag{2.43}$$

Then the resolvent is completely delocalized:

$$\Lambda^2 \prec \frac{1}{N\eta}.$$

(iii) *If $\varepsilon \gg (N/W^2)^{2/3}$ then the eigenvectors of H_ε are completely delocalized in the sense of Proposition 7.1.*

This theorem formulates only the bounds concerning delocalization, i.e. the counterparts of Theorem 2.2 and Corollary 2.4. Similarly to Theorem 2.5, a non-trivial profile can be proved for the average of $|G_{xy}|^2$. The profile is visible in the regime $N\eta \geqslant W\sqrt{\varepsilon+\eta}$, and it is given by

$$\Theta_{xy}^{(\varepsilon)} := \left(\frac{|m|^2 S_\varepsilon}{1 - |m|^2 S_\varepsilon}\right)_{xy} \approx \frac{|m|^2(1-\varepsilon)}{W\sqrt{D\big((1-\varepsilon)\alpha\eta+\varepsilon\big)}} \exp\left[-\frac{\sqrt{(1-\varepsilon)\alpha\eta+\varepsilon}}{W\sqrt{D}}|x-y|\right], \tag{2.44}$$

where the approximation is valid in the regime $|x - y| \ll N$. The details of the precise formulation and the proof are left to the reader.

3. Preliminaries

In this subsection we introduce some further notations and collect some basic facts that will be used throughout the paper. Throughout this section we work in the general d-dimensional setting of Section 2.1.

DEFINITION 3.1 (MINORS). *For $T \subset \{1, \ldots, N\}$ we define $H^{(T)}$ by*

$$(H^{(T)})_{ij} := \mathbf{1}(i \notin T)\mathbf{1}(j \notin T)h_{ij}.$$

Moreover, we define the resolvent of $H^{(T)}$ through

$$G_{ij}^{(T)}(z) := (H^{(T)} - z)_{ij}^{-1}.$$

We also set

$$\sum_{i}^{(T)} := \sum_{i:i \notin T}.$$

When $T = \{a\}$, we abbreviate $(\{a\})$ by (a) in the above definitions; similarly, we write (ab) instead of $(\{a, b\})$. Unless specified otherwise, summations are over the set $\{1, 2, \ldots, N\}$. Recalling the identification of $\mathbb{T}_L^d$ with $\{1, 2, \ldots, N\}$, this convention is in agreement with the one given after (2.3).

DEFINITION 3.2 (PARTIAL EXPECTATION AND INDEPENDENCE). *Let $X \equiv X(H)$ be a random variable. For $i \in \{1, \ldots, N\}$ define the operations P_i and Q_i through*

$$P_i X := \mathbb{E}(X|H^{(i)}), \qquad Q_i X := X - P_i X.$$

We call P_i partial expectation in the index i. Moreover, we say that X is independent of $T \subset \{1, \ldots, N\}$ if $X = P_i X$ for all $i \in T$.

The following lemma collects basic algebraic properties of stochastic domination $\prec$. We shall use it tacitly throughout Sections 4–7. Roughly it says that, under some weak restrictions, $\prec$ satisfies all the usual algebraic properties of $\leqslant$ on $\mathbb{R}$.

LEMMA 3.3.   *(i) Suppose that $X(u, v) \prec \Psi(u, v)$ uniformly in $u \in U$ and $v \in V$. If $|V| \leqslant N^C$ for some constant C then*

$$\sum_{v \in V} X(u, v) \prec \sum_{v \in V} \Psi(u, v)$$

uniformly in u.

(ii) Suppose that $X_1(u) \prec \Psi_1(u)$ uniformly in u and $X_2(u) \prec \Psi_2(u)$ uniformly in u. Then

$$X_1(u)X_2(u) \prec \Psi_1(u)\Psi_2(u)$$

uniformly in u.

(iii) Suppose that $\Psi(u) \geqslant N^{-C}$ for all u and that for all p there is a constant C_p such that $\mathbb{E}|X(u)|^p \leqslant N^{C_p}$ for all u. Then, provided that $X(u) \prec \Psi(u)$ uniformly in u, we have

$$P_a X(u) \prec \Psi(u) \qquad and \qquad Q_a X(u) \prec \Psi(u)$$

uniformly in u and a.

PROOF. The claims (i) and (ii) follow from a simple union bound. The claim (iii) follows from Chebyshev's inequality, using a high-moment estimate combined with Jensen's inequality for partial expectation. We omit the details. $\qquad\qquad\square$

Note that if for any $\varepsilon > 0$ and $p \geqslant 1$ we have

$$\mathbb{E}|X|^p \leqslant N^\varepsilon \Psi^p \tag{3.1}$$

for large enough N (depending on ε and p) then $X \prec \Psi$ by Chebyshev's inequality. Moreover, if $X \leqslant \Psi$ almost surely, then $X \prec \Psi$. Hence $O_\prec(\Psi)$ describes a larger class of random variables than $O(\Psi)$.

We need the following bound on Λ.

LEMMA 3.4. *We have*

$$\Lambda(z) \prec \frac{1}{\sqrt{M\eta}} \tag{3.2}$$

uniformly for $z \in \mathbf{S}$.

Away from the spectral edges, i.e. for $z \in \mathbf{S}$, this bound was proved in Proposition 3.3 of [24]. In [24], the matrix entries x_{ij} were assumed to have at most subexponential tails (a stronger assumption than (2.11) for all p), but the proof of [24] extends trivially to our case. See [10] for a simplified and generalized alternative proof.

The following result collects some elementary facts about m.

LEMMA 3.5. *We have the identity*

$$1 - |m|^2 = \frac{\eta|m|^2}{\operatorname{Im} m}. \tag{3.3}$$

There is a constant $c > 0$ such that

$$c \leqslant |m| \leqslant 1 \tag{3.4}$$

on $z \in \mathbf{S}$. Furthermore,

$$|m|^2 = 1 - \eta\alpha + O(\eta^2) \tag{3.5}$$

for $z \in \mathbf{S}$, with α given in (1.3). We also have the bounds

$$\operatorname{Im} m \asymp 1, \qquad 1 - |m|^2 \asymp \eta \tag{3.6}$$

for $z \in \mathbf{S}$. (The implicit constants in the two latter estimates depend on κ.)

PROOF. The identity (3.3) follows by taking the imaginary part of (2.15). The estimate (3.4) was proved in [23], Lemma 4.2. From (2.16) we find $\operatorname{Im} m = 1/\alpha + O(\eta)$, from which (3.5) follows easily using (3.3). Finally, (3.6) follows from Lemma 4.2 in [23] combined with (3.3) and (3.4). $\qquad \square$

The following resolvent identities form the backbone of all of our proofs. They first appeared in [24, Lemmas 4.1 and 4.2] and [9, Lemma 6.10]. The idea behind them is that a resolvent entry G_{ij} depends strongly on the i-th and j-th columns of H, but weakly on all other columns. The first set of identities (called Family A) determine how to make a resolvent entry G_{ij} independent of an additional index $k \neq i, j$. The second set (Family B) identities express the dependence of a resolvent entry G_{ij} on the entries in the i-th or in the j-th column of H.

LEMMA 3.6 (RESOLVENT IDENTITIES; [8, LEMMA 3.5]). *For any Hermitian matrix H and $T \subset \{1, \ldots, N\}$ the following identities hold.*

Family A. *For $i, j, k \notin T$ and $k \neq i, j$, we have*

$$G_{ij}^{(T)} = G_{ij}^{(Tk)} + \frac{G_{ik}^{(T)} G_{kj}^{(T)}}{G_{kk}^{(T)}}, \qquad \frac{1}{G_{ii}^{(T)}} = \frac{1}{G_{ii}^{(Tk)}} - \frac{G_{ik}^{(T)} G_{ki}^{(T)}}{G_{ii}^{(T)} G_{ii}^{(Tk)} G_{kk}^{(T)}}. \tag{3.7}$$

Family B. *For $i, j \notin T$ satisfying $i \neq j$ we have*

$$G_{ij}^{(T)} = -G_{ii}^{(T)} \sum_{k}^{(Ti)} h_{ik} G_{kj}^{(Ti)} = -G_{jj}^{(T)} \sum_{k}^{(Tj)} G_{ik}^{(Tj)} h_{kj}. \tag{3.8}$$

DEFINITION 3.7. *The deterministic control parameter Ψ is* admissible *if*

$$M^{-1/2} \leqslant \Psi^{(N)}(z) \leqslant M^{-\gamma/2} \tag{3.9}$$

for all N and $z \in \mathbf{S}$. (Recall the parameter γ from (2.13).)

A typical example of an admissible control parameter is

$$\Psi(z) = \frac{1}{\sqrt{M\eta}}. \tag{3.10}$$

If Ψ is admissible then the lower bound in (3.9) together with (2.12) ensure that $h_{ij} \prec \Psi$.

The following lemma gives an expansion formula for the diagonal entries of G.

LEMMA 3.8. *Suppose that $\Lambda \prec \Psi$ for some admissible Ψ. Defining*

$$Z_i := \sum_{k,l}^{(i)} Q_i \big(h_{ik} G_{kl}^{(i)} h_{li} \big),$$

we have

$$G_{ii} = m + m^2 Z_i + O_{\prec}\big(\Psi^2 + M^{-1/2}\big), \qquad Z_i \prec \Psi. \tag{3.11}$$

PROOF. The claim is an immediate consequence of Equations (9.1) and (9.2) in [8]. (Related but less explicit formulas were also obtained in [25]). $\qquad \square$

3.1. Averaging of fluctuations. In this section we collect the necessary results from [8]. The following proposition is a special case of the Fluctuation Averaging Theorem of [8].

PROPOSITION 3.9. *Suppose that $\Lambda \prec \Psi$ for some admissible control parameter Ψ. Then*

$$\sum_{a}^{(\mu\nu)} s_{\rho a} G_{\mu a} G_{a\nu} \prec \Psi(\Psi + M^{-1/4})^2, \qquad \sum_{a}^{(\mu)} s_{\rho a} G_{\mu a} G_{a\mu}^* \prec \Psi^2, \tag{3.12}$$

and

$$\sum_{a}^{(\mu)} s_{\rho a} Q_a (G_{\mu a} G_{a\mu}) \prec \Psi^3, \qquad \sum_{a}^{(\mu)} s_{\rho a} Q_a (G_{\mu a} G_{a\mu}^*) \prec \Psi^2 \big(\Psi + M^{-1/4}\big)^2, \tag{3.13}$$

as well as

$$\sum_{a \neq b}^{(\mu)} s_{\rho a} s_{ab} G_{ba} G_{a\mu} G_{\mu b}^* \prec \Psi^2 \big(\Psi + M^{-1/4}\big)^2, \qquad \sum_{a \neq b}^{(\mu)} s_{\rho a} s_{ba} G_{ba} G_{a\mu} G_{\mu b}^* \prec \Psi^2 \big(\Psi + M^{-1/4}\big)^2. \tag{3.14}$$

16

PROOF. All of these estimates follow immediately from Theorem 4.8, Lemma B.1, and Proposition B.2 of [8], recalling that by assumption $\operatorname{Im} m \geqslant c_\kappa$ by (3.6). □

The important quantity on the right-hand sides of (3.12), (3.13) and (3.14) is Ψ. The additional factors $M^{-1/4}$ are a technical[1] nuisance, but their precise form will play some role in the large-η regime, where $M^{-1/4}$ is not negligible compared to Ψ.

To interpret these estimates, we note that each summand in (3.12), (3.13), and (3.14) has a naive size given by Ψ^k, where k is the number of off-diagonal resolvent entries in the summand. Without averaging, this naive size would be a sharp upper bound. In the second estimate in (3.12) the averaging does not improve the bound since $G_{\mu a} G_{a\mu}^* = |G_{\mu a}|^2$ is positive. In all other estimates, the monomial on the left-hand side either has a nontrivial phase or its expectation is zero thanks to Q_a. Proposition 3.9 asserts that in these cases the averaged quantity is smaller than its individual summands. Note that this averaging of fluctuations is effective even though the entries of G may be strongly correlated. How many additional factors of Ψ one gains depends on the structure of the left-hand side in a subtle way; see Theorem 4.8 of [8] for the precise statement. For the applications in this paper the second bound in (3.13) is especially important; here the averaging yields a gain of two extra factors of Ψ.

We remark that all these bounds also hold if the weight functions $s_{\rho a}$ are replaced with a more general weight function. The precise definition is given in Definition 4.4 of [8]. All the weights used in this paper satisfy Definition 4.4 of [8].

We also note that averaging in indices can be replaced by expectations. We shall need the following special case of Theorem 4.15 of [8].

PROPOSITION 3.10. *Suppose that $\Lambda \prec \Psi$ for some admissible control parameter Ψ. Then for $a \neq \mu, \nu$*

$$P_a(G_{\mu a} G_{a\nu}) \prec \Psi(\Psi + M^{-1/4})^2. \tag{3.15}$$

4. Self-consistent equation for T

After these preparations, we now move on to the main arguments of this paper. Throughout this section we work in the general d-dimensional setting of Section 2.1. In this section we derive a self-consistent equation for T, given in Theorem 4.1, whose error terms are controlled precisely using the fluctuation averaging from Proposition 3.9. In Section 5 we solve this self-consistent equation; the result is given in Proposition 5.1.

THEOREM 4.1. *Suppose that $\Lambda \prec \Psi$ for some admissible control parameter Ψ. Then we have*

$$T_{xy} = |m|^2 \sum_i s_{xi} T_{iy} + |m|^2 s_{xy} + O_\prec \left(\Psi^4 + \Psi^2 M^{-1/2} \right). \tag{4.1}$$

In matrix notation,

$$T = |m|^2 ST + |m|^2 S + \mathcal{E} \tag{4.2}$$

where the matrix entries of the error satisfy

$$\mathcal{E}_{xy} \prec \Psi^4 + \Psi^2 M^{-1/2}. \tag{4.3}$$

---

[1]This nuisance is necessary, however, and Proposition 3.9 would be false without the factors of $M^{-1/4}$. See [8, Remark 4.10].

The naive size of T_{xy} is of order Ψ^2. Notice that the error term in the self-consistent equation (4.2) is smaller by two orders. This improvement is essentially due to second estimate of (3.13).

REMARK 4.2. Instead of averaging in the first index of the resolvent in the definition of T (2.29), we could have averaged in the second, resulting in the quantity $T'_{xy} := \sum_j |G_{xj}|^2 s_{jy}$. Then T' satisfies the self-consistent equation

$$T' \;=\; |m|^2 T'S + |m|^2 S + \mathcal{E}',$$

where $\mathcal{E}'$ also satisfies (4.3).

Before the proof we mention that this result also gives a self-consistent equation for the two-sided averaged quantity

$$Y_{xz} \;:=\; (TS)_{xz} \;=\; \sum_{iy} s_{xi}|G_{iy}|^2 s_{yz} \;=\; (ST')_{xz}\,.$$

Taking the average $\sum_y s_{yz}$ of (4.1), we get the following corollary.

COROLLARY 4.3. *Suppose that $\Lambda \prec \Psi$ for some admissible control parameter Ψ. Then we have*

$$Y = |m|^2 SY + |m|^2 S^2 + \mathcal{E} \qquad and \qquad Y = |m|^2 YS + |m|^2 S^2 + \widetilde{\mathcal{E}}\,, \tag{4.4}$$

where $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ each satisfy (4.3).

The rest of this section is devoted to the proof of Theorem 4.1. We begin by writing

$$T_{xy} \;=\; \sum_i s_{xi}P_i|G_{iy}|^2 + \widetilde{T}_{xy}\,, \qquad \widetilde{T}_{xy} \;:=\; \sum_i s_{xi}Q_i|G_{iy}|^2\,. \tag{4.5}$$

Then by the second formula in (3.13), we have

$$\widetilde{T}_{xy} = O_{\prec}(\Psi^4 + \Psi^2 M^{-1/2})\,. \tag{4.6}$$

Notice that (3.13) applies only to the summands $i \neq y$ in (4.5). The estimate for the summand $i = y$ follows from

$$s_{xy}Q_y|G_{yy}|^2 \;=\; s_{xy}Q_y|G_{yy} - m|^2 + 2s_{xy}\operatorname{Re}\bar{m}Q_y(G_{yy} - m) \;=\; O_{\prec}(\Psi^2 M^{-1/2})\,,$$

where we used that $(G_{yy} - m) \prec \Psi$ (see (3.11)) and that Ψ is admissible, and in particular $\Psi M^{-1} \leqslant \Psi^2 M^{-1/2}$.

We shall compute $\sum_i s_{xi}P_i|G_{iy}|^2$ up to error terms of order Ψ^4. We have the following result.

LEMMA 4.4. *Suppose that $\Lambda \prec \Psi$ for some admissible control parameter Ψ. Then*

$$P_i|G_{iy}|^2 \;=\; |m|^2 P_i T_{iy} + |m|^2 \delta_{iy} + O_{\prec}(\Psi^4 + \Psi^2 M^{-1/2}) + \delta_{iy} O_{\prec}(\Psi^2 + M^{-1/2}) \tag{4.7}$$

and

$$\sum_i s_{xi}P_i|G_{iy}|^2 \;=\; |m|^2 \sum_i s_{xi}T_{iy} + |m|^2 s_{xy} + O_{\prec}(\Psi^4 + \Psi^2 M^{-1/2})\,. \tag{4.8}$$

(It is possible to improve the last error term in (4.7), but we shall not need this.) Before proving Lemma 4.4, we show how it implies Theorem 4.1.

PROOF OF THEOREM 4.1. Equation (4.1) is an immediate consequence of (4.8), (4.6), and (4.5). Hence Theorem 4.1 follows from Lemma 4.4. $\qquad\square$

18

PROOF OF LEMMA 4.4. Throughout the following we shall repeatedly need the simple estimate

$$\frac{1}{G_{ii}} = \frac{1}{m + O_{\prec}(\Psi)} = \frac{1}{m}\big(1 + O_{\prec}(\Psi)\big) = O_{\prec}(1), \tag{4.9}$$

where the first step follows from $\Lambda \prec \Psi$, the second from the fact that Ψ is admissible, and the last from (3.4). In particular, for $k \neq i, j$, from (3.7) we get the estimate

$$G_{ij}^{(k)} = G_{ij} - \frac{G_{ik}G_{kj}}{G_{kk}} = O_{\prec}(\delta_{ij} + \Psi + \Psi^2) = O_{\prec}(\delta_{ij} + \Psi). \tag{4.10}$$

We start the proof of Lemma 4.4 with the case $i \neq y$. Using (3.11) we get

$$|G_{iy}|^2 = |G_{ii}|^2 \sum_{k,l}^{(i)} h_{ik}G_{ky}^{(i)}G_{yl}^{(i)*}h_{li}$$

$$= |m|^2\Big|1 + mZ_i + O_{\prec}\big(\Psi^2 + M^{-1/2}\big)\Big|^2 \sum_{k,l}^{(i)} h_{ik}G_{ky}^{(i)}G_{yl}^{(i)*}h_{li}$$

$$= |m|^2\big(1 + 2\operatorname{Re}(mZ_i)\big) \sum_{k,l}^{(i)} h_{ik}G_{ky}^{(i)}G_{yl}^{(i)*}h_{li} + O_{\prec}(\Psi^4 + \Psi^2 M^{-1/2}), \tag{4.11}$$

where in the last step we used $Z_i \prec \Psi$ and the large deviation bound (see Lemma B.2)

$$\sum_k^{(i)} h_{ik}G_{ky}^{(i)} \prec \Psi. \tag{4.12}$$

We may now compute the contribution of the main term in (4.11) to $P_i|G_{iy}|^2$. Still assuming $i \neq y$, we find

$$P_i \sum_{k,l}^{(i)} h_{ik}G_{ky}^{(i)}G_{yl}^{(i)*}h_{li} = \sum_k^{(i)} s_{ik}\big|G_{ky}^{(i)}\big|^2$$

$$= \sum_k^{(i)} s_{ik}\Big|G_{ky} - \frac{G_{ki}G_{iy}}{G_{ii}}\Big|^2$$

$$= \sum_k^{(i)} s_{ik}\Big[|G_{ky}|^2 - 2\operatorname{Re}\Big(\overline{G}_{ky}\frac{G_{ki}G_{iy}}{G_{ii}}\Big)\Big] + O_{\prec}(\Psi^4)$$

$$= T_{iy} - 2\operatorname{Re}\sum_k^{(i)} s_{ik}\overline{G}_{ky}\frac{G_{ki}G_{iy}}{G_{ii}} + O_{\prec}\big(\Psi^4 + \Psi^2 M^{-1}\big). \tag{4.13}$$

In the third step we used (4.9), and in the last step we added the missing term $k = i$ to obtain T_{iy}; the

resulting error term is $O_\prec(\Psi^2 M^{-1})$ since $i \neq y$. Next, using (4.9) we get

$$\sum_k^{(i)} s_{ik}\overline{G}_{ky}\frac{G_{ki}G_{iy}}{G_{ii}} = \frac{1}{m}\sum_k^{(i)} s_{ik}G_{ki}G_{iy}G_{yk}^* + O_\prec(\Psi^4)$$

$$= \frac{1}{m}\sum_k^{(i)} s_{ik}G_{ki}G_{iy}G_{yk}^{(i)*} + O_\prec(\Psi^4), \qquad (4.14)$$

In the second step, using (3.7) and (4.9), we inserted an upper index i as a preparation to taking the partial expectation P_i. We obtain

$$P_i\sum_{k,l}^{(i)} h_{ik}G_{ky}^{(i)}G_{yl}^{(i)*}h_{li} = T_{iy} - 2\mathrm{Re}\sum_k^{(i)} s_{ik}\frac{1}{m}G_{ki}G_{iy}G_{yk}^{(i)*} + O_\prec\left(\Psi^4 + \Psi^2 M^{-1}\right). \qquad (4.15)$$

Now we take the partial expectation in i in (4.15). Using that

$$P_i\left(G_{ki}G_{iy}G_{yk}^{(i)*}\right) = G_{yk}^{(i)*}P_i\left(G_{ki}G_{iy}\right) \prec (\delta_{yk}+\Psi)\Psi\left(\Psi+M^{-1/4}\right)^2$$

by Proposition 3.10 and (4.10), we find that P_i applied to the second term in (4.15) results in a quantity $O_\prec\left(\Psi^2(\Psi+M^{-1/4})^2\right)$. Thus the contribution of main term in (4.11) to $P_i|G_{iy}|^2$ is

$$|m|^2 P_i\sum_{k,l}^{(i)} h_{ik}G_{ky}^{(i)}G_{yl}^{(i)*}h_{li} = |m|^2 P_i T_{iy} + O_\prec\left(\Psi^4 + \Psi^2 M^{-1/2}\right). \qquad (4.16)$$

Next, we look at the contribution of the term with a Z in (4.11):

$$|m|^2 P_i\left[\left(mZ_i + \overline{mZ_i}\right)\sum_{k,l}^{(i)} h_{ik}G_{ky}^{(i)}G_{yl}^{(i)*}h_{li}\right]$$

$$= |m|^2 P_i\left[\sum_{c,d}^{(i)} Q_i\left(h_{ic}\left(mG_{cd}^{(i)} + \overline{m}G_{cd}^{(i)*}\right)h_{di}\right)\sum_{k,l}^{(i)} h_{ik}G_{ky}^{(i)}G_{yl}^{(i)*}h_{li}\right]$$

$$= |m|^2\sum_{k,l}^{(i)} s_{ik}s_{il}\left(mG_{lk}^{(i)}G_{ky}^{(i)}G_{yl}^{(i)*} + \overline{m}G_{lk}^{(i)*}G_{ky}^{(i)}G_{yl}^{(i)*}\right) + O_\prec(\Psi^2 M^{-1} + M^{-2})$$

$$= |m|^2 2\mathrm{Re}\left(m\sum_{k,l}^{(i)} s_{ik}s_{il}G_{lk}^{(i)}G_{ky}^{(i)}G_{yl}^{(i)*}\right) + O_\prec(\Psi^2 M^{-1} + M^{-2}).$$

In the second step we assumed for simplicity that we are dealing with the complex Hermitian case (2.9); thus, thanks to Q_i, the only allowed pairing of the entries of H imposes $c = l$ and $k = d$. (In the real symmetric case (2.8), there is one other term, (4.18) below, which is estimated in the same way.) The error terms stem from the summands $c = d = l = k \neq y$ and $c = d = l = k = y$, respectively. Here we used (4.10) as well as the bound $\mathbb{E}|h_{ic}^4| \leqslant CM^{-1}s_{ic} \leqslant CM^{-2}$, an immediate consequence of (2.11) and (2.12). In the last step we switched the indices k and l.

Now we remove the upper indices at the expense of an error of size $O_\prec(\Psi^4)$, and then add back the exceptional summation index i as before. This gives

$$|m|^2 P_i \left(2\operatorname{Re}(mZ_i) \sum_{k,l}^{(i)} h_{ik} G_{ky}^{(i)} G_{yl}^{(i)*} h_{li} \right) = |m|^2 2\operatorname{Re}\left[m \sum_{k,l} s_{ik} s_{il} G_{lk} G_{ky} G_{yl}^* \right] + O_\prec(\Psi^4 + \Psi^2 M^{-1} + M^{-2})$$
$$= O_\prec(\Psi^4 + \Psi^2 M^{-1/2}), \tag{4.17}$$

where in the second step we used (3.14); the various cases of coinciding indices k, l, y are easily dealt with using the bound $M^{-1/2} \leqslant \Psi$.

As remarked above, in the real symmetric case (2.8) the pairing $c = k$, $d = l$ is also possible. This gives rise to the additional error term

$$\sum_{k,l} s_{ik} s_{il} G_{kl} G_{ky} G_{yl}^* \prec \Psi^4 + \Psi^2 M^{-1/2}, \tag{4.18}$$

where we used (3.14).

Combining (4.11), (4.16) and (4.17) yields

$$P_i |G_{iy}|^2 = |m|^2 P_i T_{iy} + O_\prec(\Psi^4 + \Psi^2 M^{-1/2}) \tag{4.19}$$

for $i \neq y$. This proves (4.7) for the case $i \neq y$.

If $i = y$ we compute

$$P_y |G_{yy}|^2 = |m|^2 + P_y |G_{yy} - m|^2 + 2P_y \operatorname{Re}[m(G_{yy} - m)] = |m|^2 + O_\prec(\Psi^2 + M^{-1/2}). \tag{4.20}$$

Here we used that $G_{yy} - m \prec \Psi$ and that $P_y(G_{yy} - m) \prec \Psi^2 + M^{-1/2}$ by (3.11). It is possible to compute this term to high order in Ψ, but we shall not need this.

For the proof of (4.8) we run almost the same argument as above but now we aim at removing all upper indices i. We first consider the summands $i \neq y$. From (4.15) we get

$$P_i \sum_{k,l}^{(i)} h_{ik} G_{ky}^{(i)} G_{yl}^{(i)*} h_{li} = T_{iy} - 2\operatorname{Re} \sum_k s_{ik} \frac{1}{m} G_{ki} G_{iy} G_{yk}^* + O_\prec(\Psi^4 + \Psi^2 M^{-1/2}), \tag{4.21}$$

where we removed the upper index i using (3.7), and included the summand $k = i$ at the expense of a negligible error term. Taking the average $\sum_i^{(y)} s_{xi}$ of the second term on the right-hand side yields

$$\sum_i^{(y)} \sum_k s_{xi} s_{ik} \frac{1}{m} G_{ki} G_{iy} G_{yk}^* = \sum_{i,k} s_{xi} s_{ik} \frac{1}{m} G_{ki} G_{iy} G_{yk}^* + O_\prec(\Psi^2 M^{-1} + M^{-2}) = O_\prec(\Psi^4 + \Psi^2 M^{-1/2}). \tag{4.22}$$

In the first step we just added the exceptional index $i = y$, and estimated the additional terms with $i = y$ using $s_{xy} s_{yk} \leqslant M^{-1} s_{yk} \leqslant M^{-2}$ as well as $G_{ky} G_{yy} G_{yk}^* \prec \delta_{ky} + \Psi^2$. In the second step we used (3.14). Note that the gain comes from the summation index i.

Thus the contribution of the main term of (4.11) to $\sum_i^{(y)} s_{xi} P_i |G_{iy}|^2$ is

$$|m|^2 \sum_i^{(y)} s_{xi} P_i \sum_{k,l}^{(i)} h_{ik} G_{ky}^{(i)} G_{yl}^{(i)*} h_{li} = |m|^2 \sum_i^{(y)} s_{xi} T_{iy} + O_\prec(\Psi^4 + \Psi^2 M^{-1/2}). \tag{4.23}$$

The contributions of the error terms in (4.11) to $\sum_{i \neq y} s_{xi} P_i |G_{iy}|^2$ are of order $O_{\prec}(\Psi^4 + \Psi^2 M^{-1/2})$; this is true even without averaging (see (4.17)). Thus we have

$$\sum_{i}^{(y)} s_{xi} P_i |G_{iy}|^2 = |m|^2 \sum_{i}^{(y)} s_{xi} T_{iy} + O_{\prec}(\Psi^4 + \Psi^2 M^{-1/2}). \tag{4.24}$$

Finally, we consider the case $i = y$. From (4.20) we get

$$s_{xy} P_y |G_{yy}|^2 = |m|^2 s_{xy} + O_{\prec}(M^{-1} \Psi^2) = |m|^2 s_{xy} + O_{\prec}(\Psi^4).$$

This formula provides the missing summands $i = y$ in (4.24) and hence yields (4.8). □

5. Solving the equation for T

Throughout this section we work in the general d-dimensional setting of Section 2.1. In this section we solve the self-consistent equation (4.2) to determine T. This involves inverting the matrix $1 - |m|^2 S$. The stability of the self-consistent equation (4.2) is provided by the spectral gap of S. In the regime of complete delocalization, this gap is larger (and hence more effective) if we restrict S to the subspace orthogonal to the vector $\mathbf{e} = N^{-1/2}(1, 1, \ldots, 1)^T$ (see the remarks after Lemma 5.2 for more details). Therefore, we deal with the span of $\mathbf{e}$ and its orthogonal complement separately. Define the rank-one projection

$$\Pi := \mathbf{e}\mathbf{e}^*.$$

Thus, the entries Π_{ij} of Π are all equal to $1/N$, and $S\Pi = \Pi S = \Pi$ since S is stochastic by (2.3). The complementary projection is denoted by $\overline{\Pi} := 1 - \Pi$.

We perform this splitting on T_{xy} only in the x coordinate, regarding y as fixed. Thus, we split

$$T_{xy} = \overline{T}_y + (T_{xy} - \overline{T}_y),$$

where we defined the averaged vector

$$\overline{T}_y := \frac{1}{N} \sum_x T_{xy} = \frac{1}{N} \sum_i |G_{iy}|^2 = \frac{1}{N\eta} \operatorname{Im} G_{yy};$$

here the last step follows easily by spectral decomposition of G. We can use the local semicircle law, Lemma 3.4, to get

$$\overline{T}_y = \frac{\operatorname{Im} m}{N\eta} \left[1 + O_{\prec}\left(\frac{1}{\sqrt{M\eta}} \right) \right]. \tag{5.1}$$

It is instructive to perform the same averaging with the deterministic profile Θ:

$$\frac{1}{N} \sum_x \Theta_{xy} = \left(\Pi \frac{|m|^2 S}{1 - |m|^2 S} \right)_{yy} = \frac{1}{N} \frac{|m|^2}{1 - |m|^2} = \frac{\operatorname{Im} m}{N\eta}, \tag{5.2}$$

where in the last step we used (3.3).

Having dealt with the component ΠT in (5.1), we devote the rest of this section to the component $\overline{\Pi} T$. The following proposition contains the main result of this section.

PROPOSITION 5.1. *Suppose that $\Lambda \prec \Psi$ for some admissible control parameter Ψ. Then we have for all y*

$$T_{xy} = \overline{T}_y + |m|^2 \left(\frac{S - \Pi}{1 - |m|^2 S} \right)_{xy} + \widetilde{\mathcal{E}}_{xy}, \tag{5.3}$$

where the error satisfies

$$\max_{x,y} |\widetilde{\mathcal{E}}_{xy}| \prec \frac{1}{\eta + \left(\frac{W}{L} \right)^2} \left(\Psi^4 + \Psi^2 M^{-1/2} \right). \tag{5.4}$$

The main tool in the proof of Proposition 5.1 is a control on the spectral gap of S on the space orthogonal to $\mathbf{e}$. In order to state it, we need the Euclidean matrix norm $\|A\|$ as well as the $\ell^\infty \to \ell^\infty$ norm of the matrix A,

$$\|A\|_\infty := \max_i \sum_j |A_{ij}|.$$

The following lemma shows that S has a spectral gap of order $(W/L)^2$ when restricted to the space orthogonal to $\mathbf{e}$. Its proof is postponed to the end of this section.

LEMMA 5.2. *We have the bounds*

$$\left\| \frac{1}{1 - |m|^2 S} \overline{\Pi} \right\| \leqslant \frac{C}{\eta + \left(\frac{W}{L} \right)^2} \tag{5.5}$$

and

$$\left\| \frac{1}{1 - |m|^2 S} \overline{\Pi} \right\|_\infty \leqslant \frac{C \log N}{\eta + \left(\frac{W}{L} \right)^2}. \tag{5.6}$$

In the regime of complete delocalization, $\eta \leqslant (W/L)^2$, the control on $(1 - |m|^2 S)^{-1}$ is stronger on the space orthogonal to $\mathbf{e}$. Indeed, in that regime the bound (5.5) is better than the trivial bound

$$\left\| \frac{1}{1 - |m|^2 S} \right\| = \frac{1}{1 - |m|^2} \leqslant \frac{C}{\eta} \tag{5.7}$$

from (3.6).

PROOF OF PROPOSITION 5.1. Multiplying (4.2) by $\overline{\Pi}$ from the left yields

$$\overline{\Pi} T = |m|^2 S \overline{\Pi} T + |m|^2 (S - \Pi) + \overline{\Pi} \mathcal{E},$$

where we used that $S\Pi = \Pi S = \Pi$. Therefore

$$\overline{\Pi} T = |m|^2 \frac{S - \Pi}{1 - |m|^2 S} + \widetilde{\mathcal{E}}, \qquad \widetilde{\mathcal{E}} := \frac{1}{1 - |m|^2 S} \overline{\Pi} \mathcal{E}.$$

Note that $(\overline{\Pi} T)_{xy} = T_{xy} - \overline{T}_y$. Using (5.6) we therefore get (5.3) whose error term satisfies

$$\max_{x,y} |\widetilde{\mathcal{E}}_{xy}| \leqslant \left\| \frac{1}{1 - |m|^2 S} \overline{\Pi} \right\|_\infty \max_{x,y} |\mathcal{E}_{xy}| \prec \frac{1}{\eta + \left(\frac{W}{L} \right)^2} \left(\Psi^4 + \Psi^2 M^{-1/2} \right).$$

This completes the proof of (5.3) and (5.4). □

23

Next, we estimate $|G_{ij} - \delta_{ij}m|^2$ in terms of T_{ij}. In other words, we derive pointwise estimates on G_{ij} from estimates on the *averaged* quantity T_{xy}. This gives rise to an improved bound on Λ, which we may plug back into Proposition 5.1. Thus we get a self-improving scheme which may be iterated.

LEMMA 5.3. *Suppose that $\Lambda \prec \Psi$ with some admissible control parameter Ψ and $T_{ij} \prec \Omega_{ij}^2$ for a family of admissible control parameters Ω_{ij} indexed by a pair (i, j) (see Definition 3.7). Then*

$$\left| G_{ij} - \delta_{ij}m \right|^2 \; \prec \; \Omega_{ij}^2 + \Psi^4 + \delta_{ij} \sum_k \Omega_{ik}^2 s_{ki} \,. \tag{5.8}$$

(Here we write $\Omega_{ij}^2 := (\Omega_{ij})^2$.)

PROOF. We fix the index j throughout the proof. Let first $i \neq j$. Then (3.8) gives

$$G_{ij} \;=\; G_{ii} \sum_k^{(i)} h_{ik} G_{kj}^{(i)} \,. \tag{5.9}$$

We shall use the large deviation bounds from Theorem B.1 to estimate the sum. For that we shall need a bound on

$$\sum_k^{(i)} s_{ik} |G_{kj}^{(i)}|^2 \;=\; \sum_k^{(i)} s_{ik} \left(|G_{kj}|^2 + O_\prec\left(|G_{ki}G_{ij}|^2 \right) \right) \;=\; T_{ij} - s_{ii}|G_{ij}|^2 + O_\prec\left(\Psi^4 \right) \;\prec\; \Omega_{ij}^2 + \Psi^4 \,, \tag{5.10}$$

where in the first step we used (3.7) and (4.9). Since $G_{ii} \prec 1$, we get from (5.9) and Theorem B.1 (i) that

$$|G_{ij}|^2 \;\prec\; \Omega_{ij}^2 + \Psi^4 \,.$$

To estimate $G_{ii} - m$, we use (3.11) to get

$$|G_{ii} - m|^2 \;\leqslant\; C|Z_i|^2 + O_\prec(\Psi^4 + M^{-1}) \;\prec\; \sum_k \Omega_{ik}^2 s_{ki} + \Psi^4 + M^{-1} \,. \tag{5.11}$$

Here we used

$$|Z_i|^2 \;\leqslant\; \left| \sum_k^{(i)} \left(|h_{ik}|^2 - s_{ik} \right) G_{kk}^{(i)} \right|^2 + \left| \sum_{k \neq l}^{(i)} h_{ik} G_{kl}^{(i)} h_{li} \right|^2 \;\prec\; M^{-1} + \sum_k \Omega_{ik}^2 s_{ki} \,,$$

where in the second step we used Theorem B.1 (i) and (ii), with the bounds $G_{kk}^{(i)} \prec 1$ and

$$\sum_{k \neq l}^{(i)} s_{ik} |G_{kl}^{(i)}|^2 s_{li} \;\prec\; \sum_k \Omega_{ik}^2 s_{ki} + \Psi^4 \,.$$

This last estimate follows along the lines of (5.10), whereby the error terms resulting from the removal of the upper indices are estimated by Cauchy-Schwarz; we omit the details. Finally M^{-1} can be absorbed into $\sum_k \Omega_{ik}^2 s_{ki}$ by admissibility of Ω_{ij}. $\square$

We may now combine Proposition 5.1 and Lemma 5.3 in an iterative self-improving scheme, which results in an improved bound on Λ.

COROLLARY 5.4. *Suppose that $\Lambda \prec \Psi$ and $T_{ij} \prec \Omega^2$ for all i and j, where Ψ and Ω are admissible control parameters. Then*

$$\Lambda^2 \prec \Omega^2 \,. \tag{5.12}$$

PROOF. We apply Lemma 5.3 to the constant control parameter $\Omega_{ij} = \Omega$ for each i, j. Thus, suppose that $T_{ij} \prec \Omega^2$ for all i, j, Lemma 5.3 yields

$$\Lambda^2 \prec \Psi^2 \qquad \Longrightarrow \qquad \Lambda^2 \prec \Omega^2 + \Psi^4 \,.$$

Now we can iterate this estimate, $\Omega^2 + \Psi^4$ taking the role of Ψ^2 in controlling Λ^2. Thus after one iteration we get

$$\Lambda^2 \prec \Omega^2 + (\Omega^2 + \Psi^4)^2 \prec \Omega^2 + \Psi^8 \,.$$

After k iterations we get $\Lambda^2 \prec \Omega^2 + \Psi^{2^k}$. Since Ω and Ψ are admissible, we have $\Psi^{2^k} \prec \Omega^2$ for $k \sim |\log \gamma|$. This completes the proof. □

What remains is the proof of Lemma 5.2, which relies on Fourier transformation. We introduce the dual lattice of $\mathbb{T} \equiv \mathbb{T}_L^d$,

$$P \equiv P_L^d := \frac{2\pi}{L} \mathbb{T}_L^d \,.$$

For $p \in \mathbb{R}^d$ define

$$\widehat{S}(p) := \sum_{x \in \mathbb{T}} \mathrm{e}^{-\mathrm{i}p \cdot x} s_{x0} = \sum_{x \in \mathbb{T}} \mathrm{e}^{-\mathrm{i}p \cdot x} \frac{1}{Z_{L,W}} f\left(\frac{x}{W}\right) \,. \tag{5.13}$$

In particular, if $p \in P$ then $\widehat{S}(p)$ is the discrete Fourier transform of s_{x0}. Since s_{xy} is translation invariant and L-periodic, we get for all $x, y \in \mathbb{T}$ that

$$s_{xy} = s_{[x-y]_L 0} = \frac{1}{N} \sum_{p \in P} \mathrm{e}^{\mathrm{i}p \cdot (x-y)} \widehat{S}(p) \,.$$

PROOF OF LEMMA 5.2. First we show that for large enough L the Euclidean matrix norm satisfies

$$\|S\overline{\Pi}\| \leqslant 1 - c_1 \left(\frac{W}{L}\right)^2 \tag{5.14}$$

with some positive constant c_1 depending on the profile f. Since the matrix entries s_{ij} are translation invariant (see (2.2)), it is sufficient to compute its Fourier transform as defined in (5.13). Using the property $\widehat{Su}(p) = \widehat{S}(p)\widehat{u}(p)$, the fact that $\widehat{\Pi}(p) = \delta_{p0}$, and Plancherel's identity, we find

$$\|S\overline{\Pi}\| = \max\{|\widehat{S}(p)| : p \in P \setminus \{0\}\} \leqslant 1 - c_1\left(\frac{W}{L}\right)^2 \,. \tag{5.15}$$

The last step follows easily from $\widehat{S}(p) \geqslant -1 + \delta$ (recall (2.4)) and the representation

$$1 - \widehat{S}(p) = \sum_{x \in \mathbb{T}} (1 - \cos(p \cdot x)) \frac{1}{Z_{L,W}} f\left(\frac{x}{W}\right) \geqslant c \sum_{x \in \mathbb{T}} (p \cdot x)^2 \frac{1}{Z_{L,W}} f\left(\frac{x}{W}\right) \geqslant c_1 \left(\frac{W}{L}\right)^2 \,,$$

where in the second step we used $|p \cdot x| \leqslant \pi$, and in the last step $|p| \geqslant 2\pi/L$.

From (3.6) we get $1 - |m|^2 \geqslant c\eta$, which, combined with (5.15), yields

$$|m|^2 \|S\overline{\Pi}\| \;\leqslant\; 1 - c\left(\frac{W}{L}\right)^2 - c\eta \,. \tag{5.16}$$

Thus we get

$$\left\|\frac{1}{1 - |m|^2 S}\,\overline{\Pi}\right\| \;\leqslant\; \sum_{k=0}^{\infty} |m|^{2k} \|S\overline{\Pi}\|^k \;\leqslant\; \frac{C}{\eta + \left(\frac{W}{L}\right)^2} \,.$$

This is (5.5).

In order to prove (5.6), we first observe that $\|S\|_\infty \leqslant 1$ as follows from the estimate

$$\max_i |(S\mathbf{v})_i| \;\leqslant\; \max_i \left|\sum_x s_{ix} v_x\right| \;\leqslant\; \max_x |v_x| \,,$$

where $\mathbf{v} = (v_i)$ is an arbitrary vector. Thus, for any vector $\mathbf{v}$ satisfying $\langle \mathbf{v}, \mathbf{e} \rangle = 0$ any $k_0 \in \mathbb{N}$ we have

$$\left\|\frac{1}{1 - |m|^2 S}\,\mathbf{v}\right\|_\infty \;\leqslant\; \sum_{k=0}^{k_0 - 1} |m|^{2k} \|S^k \mathbf{v}\|_\infty + \sum_{k=k_0}^{\infty} |m|^{2k} \|S^k \mathbf{v}\|_2$$

$$\leqslant\; k_0 \|\mathbf{v}\|_\infty + \sum_{k=k_0}^{\infty} |m|^{2k} \|S\overline{\Pi}\|^k \|\mathbf{v}\|_2$$

$$\leqslant\; k_0 \|\mathbf{v}\|_\infty + \sqrt{N}\, \frac{|m|^{2k_0} \|S\overline{\Pi}\|^{k_0}}{1 - |m|^2 \|S\overline{\Pi}\|}\|\mathbf{v}\|_\infty \,,$$

where we used the bound $\|\mathbf{v}\|_\infty \leqslant \|\mathbf{v}\|_2 \leqslant \sqrt{N}\|\mathbf{v}\|_\infty$ and (3.4). Choosing $k_0 = C(\log N)[\eta + (W/L)^2]^{-1}$ with a sufficiently large constant C, we obtain (5.6) exactly as above using the bound (5.16). This completes the proof of Lemma 5.2. $\qquad\square$

6. Delocalization bounds

In this section we prove our main results – Theorems 2.2, 2.5, and 2.11. We return to the one-dimensional case, $d = 1$, and continue to assume (2.17). In particular, we write N instead of L. The simple extension to higher dimensions is given in Section 8.

6.1. Delocalization without profile: proof of Theorem 2.2. Suppose that $\Lambda \prec \Psi$ for some admissible control parameter Ψ. Then (5.3) together with (5.4), (5.1), and (2.38) yield

$$T_{ij} \;\prec\; \frac{1}{N\eta} + \frac{1}{W\sqrt{\eta}} + \frac{1}{\eta + \left(\frac{W}{N}\right)^2}\left(\Psi^4 + \Psi^2 W^{-1/2}\right) \;\prec\; \Phi^2 + \frac{1}{\eta + \left(\frac{W}{N}\right)^2}\left(\Psi^4 + \Psi^2 W^{-1/2}\right) \,. \tag{6.1}$$

Recalling Corollary 5.4, we have therefore proved

$$\Lambda^2 \;\prec\; \Psi^2 \qquad \Longrightarrow \qquad \Lambda^2 \;\prec\; \Phi^2 + \frac{N^2}{W^2}\left(\Psi^4 + \Psi^2 W^{-1/2}\right) \,, \tag{6.2}$$

i.e. the upper bound $\Lambda^2 \prec \Psi^2$ can be replaced with the stronger bound (6.2).

We can now iterate (6.2), exactly as in the proof of Corollary 5.4. We start the iteration with $\Psi_0 :=$ $(W\eta)^{-1/2}$; see Lemma 3.4. Explicitly, the iteration reads

$$\Psi_{k+1}^2 := \Phi^2 + \frac{N^2}{W^2}\left(\Psi_k^4 + \Psi_k^2 W^{-1/2}\right).$$

From (6.2) and Lemma 3.4 we get that $\Lambda^2 \prec \Psi_k$ for any fixed k.

In order perform the iteration, we require

$$\frac{N^2}{W^2}\Psi_0^2 \ll 1 \qquad \text{and} \qquad \frac{N^2}{W^2}W^{-1/2} \ll 1. \tag{6.3}$$

Thus we get the conditions $N \ll W^{5/4}$ and $\eta \gg N^2/W^3$. (Here we used (2.1)). Satisfying these two conditions is the reason we need to impose the restriction on W in Theorem 2.2, Corollary 2.4, and Theorem 2.5. Using (6.3) and the fact that Φ is by definition admissible, it is now easy to see that there is a finite constant k, which depends on the implicit constants c in $\ll$ and $\gg$ above, such that $\Psi_k^2 \leqslant C\Phi^2$. This concludes the proof of Theorem 2.2.

6.2. Delocalization with profile: proof of Theorem 2.5. By assumption we have $(W/N)^2 \leqslant \eta \leqslant 1$, so that in particular $\Phi^2 = W^{-1}\eta^{-1/2} =: \Psi^2$. Note that this Ψ is admissible. From (2.20) we get $\Lambda \prec \Psi$. Now observe that $\frac{\operatorname{Im} m}{N\eta} = \Pi_{xy}\frac{\operatorname{Im} m}{\eta}$ for all x and y, as well as

$$\frac{\operatorname{Im} m}{\eta}\Pi + |m|^2\frac{S - \Pi}{1 - |m^2|S} = |m|^2\frac{S}{1 - |m^2|S}$$

by (3.3) and the property $\Pi S = S\Pi = \Pi$. Thus (5.3) together with (5.4) and (5.1) implies the first estimate of (2.30), since in the regime $\eta \geqslant (W/N)^2$ and $W^{5/4} \gg N$ the error term (5.4) is bounded by

$$\frac{1}{\eta}\left(\Psi^4 + \Psi^2 W^{-1/2}\right) \leqslant \frac{C}{N\eta}.$$

The second estimate of (2.30) follows from the first one and (4.7).

Next, (2.31) follows by using (2.33) in (2.30).

Finally, using Lemma 5.3 with $\Omega_{ij}^2 = \Upsilon_{ij}$ and $\Psi := W^{-1/2}\eta^{-1/4}$, we obtain

$$\left|G_{ij} - \delta_{ij}m\right|^2 \prec \Upsilon_{ij} + \Psi^4 + \delta_{ij}\sum_k \Upsilon_{ik}s_{ki} \prec \Upsilon_{ij}. \tag{6.4}$$

Here we used that Ψ^4 can be absorbed into $(N\eta)^{-1} \leqslant \Upsilon_{ij}$ and in the last summation $\sum_k \Upsilon_{ik}s_{ki}$ can be absorbed into $\Upsilon_{ii} \geqslant \frac{C}{W\sqrt{\eta}}$. This proves (2.32), and hence concludes the proof of Theorem 2.5.

6.3. Delocalization with a small mean-field component: proof of Theorem 2.11. In order to prove Theorem 2.11 we repeat the arguments from the previous sections almost to the letter. The self-consistent equations from Theorem 4.1 remains unchanged except that S_ε replaces S in (4.2). The key observation is that, on the subspace orthogonal to $\mathbf{e}$, the lower bound on $1 - |m|^2S_\varepsilon$ is better than that on $1 - |m|^2S$. Indeed, using (3.6) we get

$$(1 - |m|^2S_\varepsilon)\overline{\Pi} = 1 - |m|^2(1 - \varepsilon)S\overline{\Pi} \geqslant 1 - (1 - c\eta)(1 - \varepsilon)S\overline{\Pi} \geqslant c(\eta + \varepsilon)$$

27

with some positive constant c. This implies that (5.5) and (5.6) can be improved to

$$\left\| \frac{1}{1 - |m|^2 S_\varepsilon} \overline{\Pi} \right\| \leqslant \frac{C}{\eta + \varepsilon + \left(\frac{W}{N}\right)^2} \,, \qquad \left\| \frac{1}{1 - |m|^2 S_\varepsilon} \overline{\Pi} \right\|_\infty \leqslant \frac{C \log N}{\eta + \varepsilon + \left(\frac{W}{N}\right)^2} \,. \tag{6.5}$$

Suppose now that $\Lambda \prec \Psi$ for some admissible control parameter Ψ. Then the statement of Proposition 5.1 is modified to

$$T_{xy} = \overline{T}_y + |m|^2 \left(\frac{S_\varepsilon - \Pi}{1 - |m|^2 S_\varepsilon} \right)_{xy} + \widetilde{\mathcal{E}}_{xy} \,, \tag{6.6}$$

where the error term satisfies

$$\max_{x,y} |\widetilde{\mathcal{E}}_{xy}| \prec \frac{1}{\eta + \varepsilon + \left(\frac{W}{N}\right)^2} \left(\Psi^4 + \Psi^2 W^{-1/2} \right) \,. \tag{6.7}$$

Notice that the Fourier transforms of S and S_ε (defined by (5.13)) satisfy

$$\widehat{S}_\varepsilon(p) = (1 - \varepsilon)\widehat{S}(p) + \varepsilon \delta_{p0}. \tag{6.8}$$

Thus we have

$$\left(\frac{S_\varepsilon}{1 - |m|^2 S_\varepsilon} \right)_{xy} = \frac{1}{N} \sum_{p \in P: p \neq 0} \mathrm{e}^{ip(x-y)} \frac{(1 - \varepsilon)\widehat{S}(p)}{1 - |m|^2 (1 - \varepsilon)\widehat{S}(p)} + \frac{\mathrm{Im}\, m}{|m|^2 N \eta}$$

$$= \frac{1}{N} \sum_{p \in P} \mathrm{e}^{ip(x-y)} \frac{(1 - \varepsilon)\widehat{S}(p)}{1 - |m|^2 (1 - \varepsilon)\widehat{S}(p)} + \frac{\mathrm{Im}\, m}{|m|^2 N \eta} + O\left(\frac{1}{(\eta + \varepsilon)N} \right) \,. \tag{6.9}$$

Here we treated the zero mode $p = 0$ separately; it is given by

$$\frac{1}{N} \sum_x \left(\frac{S_\varepsilon}{1 - |m|^2 S_\varepsilon} \right)_{xy} = \left(\Pi \frac{S_\varepsilon}{1 - |m|^2 S_\varepsilon} \right)_{xy} = \frac{1}{N(1 - |m|^2)} = \frac{\mathrm{Im}\, m}{|m|^2 N \eta} \,,$$

where in the last step we used (3.3). The error term in (6.9) is estimated using a similar calculation.

Notice that the coefficient of $\widehat{S}(p)$ in the denominator of (6.9) is now $|m|^2(1-\varepsilon) = 1 - \varepsilon - (1-\varepsilon)\alpha\eta + O(\eta^2)$, where we used (3.5). The results and the proof of Proposition 2.8 remain unchanged when S is replaced with S_ε, except that $\alpha\eta$ must be replaced with $(1 - \varepsilon)\alpha\eta + \varepsilon$ on the right-hand side of (2.36), and the whole expression is multiplied by an additional factor $(1 - \varepsilon)$. Moreover, instead of (2.38), we now have

$$\max_{xy} \left(\frac{S_\varepsilon}{1 - |m|^2 S_\varepsilon} \right)_{xy} \asymp \frac{1}{N\eta} + \frac{1}{W\sqrt{\eta + \varepsilon}} \,. \tag{6.10}$$

Recall the definition (2.20) of Φ_ε. Following the proof of Theorem 2.2, instead of (6.2) we now obtain

$$\Lambda^2 \prec \Psi^2 \qquad \Longrightarrow \qquad \Lambda^2 \prec \Phi_\varepsilon^2 + \frac{1}{\eta + \varepsilon + \left(\frac{W}{N}\right)^2} \left(\Psi^4 + \Psi^2 W^{-1/2} \right) \,. \tag{6.11}$$

As in Section 6.1, we can iterate (6.11) under the conditions

$$\frac{1}{W\eta} \ll \eta + \varepsilon + \left(\frac{W}{N}\right)^2 \,, \qquad W^{-1/2} \ll \eta + \varepsilon + \left(\frac{W}{N}\right)^2 \,. \tag{6.12}$$

28

(Note that the a priori estimate $(W\eta)^{-1}$ is still determined by W despite the small mean-field component. In Lemma 3.4 it is given by $(M\eta)^{-1/2}$ where $M = (\max_{ij} s_{ij})^{-1} \sim (\varepsilon N^{-1} + W^{-1})^{-1} \sim W$.) The first condition of (6.12) holds if

$$\eta(\eta + \varepsilon) \;\gg\; W^{-1}\,, \tag{6.13}$$

and the second holds if either

$$\eta + \varepsilon \;\gg\; W^{-1/2} \tag{6.14}$$

or

$$N \;\ll\; W^{5/4}\,. \tag{6.15}$$

This concludes the proof of part (i).

In order to get complete delocalization of the resolvent, i.e. $\Lambda^2 \prec (N\eta)^{-1}$, we require $\Lambda \prec \Phi_\varepsilon^2$ as well as

$$W\sqrt{\eta + \varepsilon} \;\geqslant\; N\eta\,, \tag{6.16}$$

which ensures that $\Phi_\varepsilon = (N\eta)^{-1}$. Hence we get complete delocalization of the resolvent provided that (6.13), (6.14), and (6.16) hold. This concludes the proof of part (ii).

If $\varepsilon \gg (N/W^2)^{2/3}$ then there exists an η such that the assumptions of part (ii) are met. Hence part (ii) and Proposition 7.1 yields part (iii). This concludes the proof of Theorem 2.11.

7. Complete delocalization of eigenvectors

In this section we derive a delocalization result for the eigenvectors of H, using the complete delocalization $\Lambda^2 \prec (N\eta)^{-1}$ as input. We denote the eigenvalues of H by $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_N$, and the associated normalized eigenvectors by $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N$. We use the notation $\mathbf{u}_\alpha = (u_\alpha(x))_{x=1}^N$. We shall only consider eigenvectors associated with eigenvalues lying in the interval

$$I \;:=\; [-2 + \kappa, 2 - \kappa]\,,$$

where $\kappa > 0$ is fixed.

For $\ell \equiv \ell(L)$ define the characteristic function $P_{x,\ell}$ projecting onto the complement of the ℓ-neighbourhood of x,

$$P_{x,\ell}(y) \;:=\; \mathbf{1}(|y - x| \geqslant \ell)\,.$$

Let $\varepsilon > 0$ and define the random subset of eigenvector indices through

$$\mathcal{A}_{\varepsilon,\ell} \;:=\; \left\{\alpha : \lambda_\alpha \in I\,,\; \sum_x |u_\alpha(x)|\,\|P_{x,\ell}\,\mathbf{u}_\alpha\| \leqslant \varepsilon\right\},$$

which indexes the set of eigenvectors localized on scale ℓ up to an error ε; see Remark 7.2 below for more details on its interpretation.

PROPOSITION 7.1 (COMPLETE DELOCALIZATION OF EIGENVECTORS). *Suppose that $\Lambda \prec \Psi$ for some admissible control parameter Ψ. Let $\eta \equiv \eta_N$ be a sequence satisfying $M^{-1+\gamma} \leqslant \eta \ll 1$. Suppose that*

$$\sup_{E \in I} |G_{xy}(E + \mathrm{i}\eta)|^2 \;\prec\; \frac{1}{N\eta} + \delta_{xy}\,. \tag{7.1}$$

Let $\ell \ll N$. Then we have for any $\varepsilon > 0$

$$\frac{|\mathcal{A}_{\varepsilon,\ell}|}{N} \;\leqslant\; C\sqrt{\varepsilon} + O_{\prec}(N^{-c}) \,.$$

REMARK 7.2. The set $\mathcal{A}_{\varepsilon,\ell}$ contains, in particular, all indices associated with eigenvectors that are exponentially localized in balls of radius $O(\ell)$. In fact, exactly as in [7, Corollary 3.4], Proposition 7.1 implies that the fraction of eigenvectors subexponentially localized on scales ℓ vanishes with high probability for large N.

PROOF OF PROPOSITION 7.1. As usual, we omit the spectral parameter $z = E + \mathrm{i}\eta$, where $E \in I$ is arbitrary and η is the parameter given in the statement of Proposition 7.1. By assumption on Λ, we have for all x

$$\frac{\eta}{\operatorname{Im} m} \sum_y |G_{yx}|^2 \;=\; \frac{\operatorname{Im} G_{xx}}{\operatorname{Im} m} \;=\; 1 + O_{\prec}(\Psi) \,, \tag{7.2}$$

uniformly in $E \in I$, where in the first step we used the spectral decomposition of G. Thus, for all x, the map $y \mapsto \frac{\eta}{\operatorname{Im} m}|G_{yx}|^2$ is approximately a probability distribution on $\{1, \dots, N\}$. Roughly, (7.1) states that this probability distribution is supported on the order of N sites of $\{1, \dots, N\}$. More precisely, (7.1) yields (introducing the standard basis vector δ_x defined by $(\delta_x)(y) := \delta_{xy}$), for any fixed x,

$$
\begin{aligned}
\frac{\eta}{\operatorname{Im} m} \left\| P_{x,\ell}\, G\, \delta_x \right\|^2 \;&=\; \frac{\eta}{\operatorname{Im} m} \sum_y \mathbf{1}(|y - x| \geqslant \ell)\, |G_{yx}|^2 \\
&=\; \frac{\eta}{\operatorname{Im} m} \sum_y |G_{yx}|^2 - \frac{\eta}{\operatorname{Im} m} \sum_y \mathbf{1}(|y - x| < \ell)\, |G_{yx}|^2 \\
&=\; 1 + O_{\prec}(\Psi) + O_{\prec}\!\left[\frac{\eta}{\operatorname{Im} m}\!\left(\frac{N^{1-c}}{N\eta} + 1 \right) \right] \\
&=\; 1 + O_{\prec}(N^{-c}) \,,
\end{aligned}
\tag{7.3}
$$

uniformly in $E \in I$. Here in the third step we used (7.2) and (7.1), and in the last step the upper bound $\eta \leqslant M^{-c}$ and the fact that Ψ is admissible.

In order to obtain a statement about the eigenvectors, we do a spectral decomposition $G = \sum_\alpha \frac{\mathbf{u}_\alpha \mathbf{u}_\alpha^*}{\lambda_\alpha - z}$, which yields for arbitrary $\zeta > 0$

$$
\begin{aligned}
\frac{\eta}{\operatorname{Im} m} \left\| P_{x,\ell}\, G\, \delta_x \right\|^2 \;&=\; \frac{\eta}{\operatorname{Im} m} \left\| \sum_\alpha \frac{1}{\lambda_\alpha - z}\, \overline{u}_\alpha(x)\, P_{x,\ell}\mathbf{u}_\alpha \right\|^2 \\
&\leqslant\; \frac{\eta}{\operatorname{Im} m}\left(1 + \frac{1}{\zeta}\right) \left\| \sum_{\alpha \in \mathcal{A}_{\varepsilon,\ell}} \frac{1}{\lambda_\alpha - z}\, \overline{u}_\alpha(x)\, P_{x,\ell}\mathbf{u}_\alpha \right\|^2 + \frac{\eta}{\operatorname{Im} m}(1 + \zeta) \left\| \sum_{\alpha \in \mathcal{A}_{\varepsilon,\ell}^c} \frac{1}{\lambda_\alpha - z}\, \overline{u}_\alpha(x)\, P_{x,\ell}\mathbf{u}_\alpha \right\|^2 , \quad (7.4)
\end{aligned}
$$

where we introduced the complement set $\mathcal{A}_{\varepsilon,\ell}^c := \{1, \dots, N\} \setminus \mathcal{A}_{\varepsilon,\ell}$. In order to estimate the first term on

the right-hand side of (7.4), we write

$$\frac{\eta}{\operatorname{Im} m}\left\|\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}}\frac{1}{\lambda_\alpha-z}\,\overline{u}_\alpha(x)\,P_{x,\ell}\mathbf{u}_\alpha\right\|^2 \leqslant \frac{\eta}{\operatorname{Im} m}\left\|\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}}\frac{1}{\lambda_\alpha-z}\,\overline{u}_\alpha(x)\,\mathbf{u}_\alpha\right\|^2$$

$$\leqslant \frac{\eta}{\operatorname{Im} m}\sum_\alpha\frac{|u_\alpha(x)|^2}{|\lambda_\alpha-z|^2}$$

$$= \frac{\operatorname{Im} G_{xx}}{\operatorname{Im} m} = 1+O_\prec(\Psi)\,.$$

Therefore we may estimate the left-hand side by its square root to get the bound

$$\frac{\eta}{\operatorname{Im} m}\left\|\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}}\frac{1}{\lambda_\alpha-z}\,\overline{u}_\alpha(x)\,P_{x,\ell}\mathbf{u}_\alpha\right\|^2 \leqslant (1+O_\prec(\Psi))\sqrt{\frac{\eta}{\operatorname{Im} m}}\left\|\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}}\frac{1}{\lambda_\alpha-z}\,\overline{u}_\alpha(x)\,P_{x,\ell}\mathbf{u}_\alpha\right\|$$

$$\leqslant \big(C+O_\prec(\Psi)\big)\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}}\sqrt{\frac{\eta}{(E-\lambda_\alpha)^2+\eta^2}}\,|u_\alpha(x)|\,\|P_{x,\ell}\mathbf{u}_\alpha\|\,,\qquad(7.5)$$

where in the last step we used (3.6).

Similarly, we may estimate the second term of (7.4) using

$$\left\|\sum_{\alpha\in\mathcal{A}^c_{\varepsilon,\ell}}\frac{1}{\lambda_\alpha-z}\,\overline{u}_\alpha(x)\,P_{x,\ell}\mathbf{u}_\alpha\right\|^2 \leqslant \left\|\sum_{\alpha\in\mathcal{A}^c_{\varepsilon,\ell}}\frac{1}{\lambda_\alpha-z}\,\overline{u}_\alpha(x)\,\mathbf{u}_\alpha\right\|^2 = \sum_{\alpha\in\mathcal{A}^c_{\varepsilon,\ell}}\frac{|u_\alpha(x)|^2}{|\lambda_\alpha-z|^2}\,.\qquad(7.6)$$

Combining (7.4) with (7.3), (7.5), and (7.6), we get

$$1+O_\prec(N^{-c}) \leqslant \big(C+O_\prec(N^{-c})\big)\Big(1+\frac{1}{\zeta}\Big)\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}}\sqrt{\frac{\eta}{(E-\lambda_\alpha)^2+\eta^2}}\,|u_\alpha(x)|\,\|P_{x,\ell}\mathbf{u}_\alpha\|$$

$$+ (1+\zeta)\sum_{\alpha\in\mathcal{A}^c_{\varepsilon,\ell}}\frac{\eta\,|u_\alpha(x)|^2}{\operatorname{Im} m\,|\lambda_\alpha-z|^2}\,,$$

uniformly for $E\in I$. Now taking the average $|I|^{-1}\int_I\mathrm{d}E$ and using Jensen's inequality, we find

$$1+O_\prec(N^{-c}) \leqslant \big(C+O_\prec(N^{-c})\big)\Big(1+\frac{1}{\zeta}\Big)\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}}|u_\alpha(x)|\,\|P_{x,\ell}\mathbf{u}_\alpha\|$$

$$+ (1+\zeta)\frac{1}{|I|}\int_I\mathrm{d}E\sum_{\alpha\in\mathcal{A}^c_{\varepsilon,\ell}}\frac{\eta\,|u_\alpha(x)|^2}{\operatorname{Im} m\,|\lambda_\alpha-E-\mathrm{i}\eta|^2}\,,$$

where we used that

$$\int_I\mathrm{d}E\,\frac{\eta}{(E-\lambda_\alpha)^2+\eta^2} \geqslant \frac{\pi}{2}\qquad(7.7)$$

for $\alpha\in\mathcal{A}_{\varepsilon,\ell}$. Averaging over x, i.e. taking $N^{-1}\sum_x$, yields

$$1+O_\prec(N^{-c}) \leqslant \big(C+O_\prec(N^{-c})\big)\Big(1+\frac{1}{\zeta}\Big)\varepsilon + (1+\zeta)\frac{1}{|I|}\int\mathrm{d}E\,\frac{1}{N}\sum_{\alpha\in\mathcal{A}^c_{\varepsilon,\ell}}\frac{\eta}{\operatorname{Im} m\,|\lambda_\alpha-E-\mathrm{i}\eta|^2}\,,\qquad(7.8)$$

where we used the definition of $\mathcal{A}_{\varepsilon,\ell}$. We may estimate the integral as

$$\frac{1}{|I|}\int \mathrm{d}E\,\frac{1}{N}\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}^c}\frac{\eta}{\operatorname{Im}m\,|\lambda_\alpha-E-\mathrm{i}\eta|^2} \;\leqslant\; \frac{1}{|I|}\int\mathrm{d}E\,\frac{1}{N}\frac{\operatorname{Im}\operatorname{Tr}G}{\operatorname{Im}m} \;=\; 1+O_\prec(\Psi)\,.$$

Setting $\zeta=\sqrt{\varepsilon}$ in (7.8) therefore yields

$$\frac{1}{|I|}\int\mathrm{d}E\,\frac{1}{N}\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}^c}\frac{\eta}{\operatorname{Im}m\,|\lambda_\alpha-E-\mathrm{i}\eta|^2} \;\geqslant\; 1-\big(C+O_\prec(N^{-c})\big)\sqrt{\varepsilon}-O_\prec(N^{-c}) \;\geqslant\; 1-C\sqrt{\varepsilon}-O_\prec(N^{-c})\,.$$

Using

$$\frac{1}{|I|}\int\mathrm{d}E\,\frac{1}{N}\sum_{\alpha}\frac{\eta}{\operatorname{Im}m\,|\lambda_\alpha-E-\mathrm{i}\eta|^2} \;=\; \frac{1}{|I|}\int\mathrm{d}E\,\frac{1}{N}\frac{\operatorname{Im}\operatorname{Tr}G}{\operatorname{Im}m} \;=\; 1+O_\prec(\Psi)$$

we therefore get

$$\frac{1}{|I|}\int\mathrm{d}E\,\frac{1}{N}\sum_{\alpha\in\mathcal{A}_{\varepsilon,\ell}}\frac{\eta}{\operatorname{Im}m\,|\lambda_\alpha-E-\mathrm{i}\eta|^2} \;\leqslant\; C\sqrt{\varepsilon}+O_\prec(N^{-c})\,.$$

The claim now follows from (7.7). $\qquad\square$

8. Extension to higher dimensions and a slowly decaying band

In this section we extend Theorem 2.2, Corollary 2.4, and Theorem 2.5 in two directions: higher dimensions d and a slowly decaying band.

The multidimensional analogues of the slowly decaying profile are left to the reader, as is the formulation of these extensions if a small mean-field component is added to the band matrix. All these results can be obtained in a straightforward manner following the proofs for the one-dimensional case with a rapidly decaying f.

8.1. Higher dimensions. Fix $d=1,2,3,\ldots$, and recall that $N=L^d$ and $M\asymp W^d$. Throughout this section we continue to assume (2.17). The following lemma gives the sharp upper bound on the size of Θ_{xy} defined by the formula (2.24), where x and y take on values in the d-dimensional torus $\mathbb{T}_L^d\equiv\mathbb{T}$.

LEMMA 8.1. *Let $d=2,3,\ldots$ and assume (2.17). Then there is a constant C such that*

$$\Theta_{xy} \;\leqslant\; C\max\left\{\frac{1}{M},\frac{1}{N\eta}\right\}$$

for all x and y.

PROOF. See Appendix A. $\qquad\square$

In order to state the precise form of the profile Θ_{xy}, we define the covariance matrix $D\equiv D_W$ through

$$D_{ij} \;:=\; \frac{1}{2}\sum_{x\in\mathbb{T}}\frac{x_i x_j}{W^2}s_{x0}\,. \tag{8.1}$$

We have
$$D \;=\; D_\infty + O(W^{-1}) \qquad \text{where} \qquad (D_\infty)_{ij} \;:=\; \frac{1}{2} \int_{\mathbb{R}^d} x_i x_j f(x) \, \mathrm{d}x \,.$$

Since $D_\infty > 0$ we get $D \geqslant c > 0$ uniformly in W.

Next, we define the d-dimensional Yukawa potential
$$V(x) \;:=\; \int_{\mathbb{R}^d} \frac{\mathrm{d}q}{(2\pi)^d} \frac{\mathrm{e}^{\mathrm{i}q \cdot x}}{1 + q^2} \qquad (x \in \mathbb{R}^d) \,,$$

where the integral is to be understood as the Fourier transform of a tempered distribution. For $d = 1$ we have $V(x) = \frac{1}{2}\mathrm{e}^{-|x|}$ and for $d = 3$ we have $V(x) = \frac{1}{4\pi|x|}\mathrm{e}^{-|x|}$. Generally, in $d \geqslant 3$ dimensions V has a singularity of type $|x|^{2-d}$ at the origin, and in $d = 2$ dimensions a singularity of type $\log|x|$; see e.g. [29, Theorem 6.23] for more details. The leading-order behaviour of the profile Θ_{x0} is given by

$$\theta_x \;:=\; \frac{|m|^2}{N} \sum_{p \in (\frac{2\pi}{L}\mathbb{Z})^d} \mathrm{e}^{\mathrm{i}p \cdot x} \frac{\chi(Wp)}{\alpha\eta + W^2(p \cdot Dp)} \;=\; \frac{|m|^2 (\alpha\eta)^{d/2-1}}{W^d \sqrt{\det D}} \sum_{k \in \mathbb{Z}^d} \left(V * \varphi_{\sqrt{\alpha\eta}}\right)\left(\frac{\sqrt{\alpha\eta}}{W}D^{-1/2}(x + kL)\right),$$
(8.2)

where χ is a smooth function satisfying $\chi(q) = 1$ for $|q| \leqslant 1/2$ and $\chi(q) = 0$ for $|q| \geqslant 1$, φ is a Schwartz function satisfying $\int \varphi = 1$, and $\varphi_t(x) := t^{-d}\varphi(x/t)$. (In fact, $\varphi(D^{-1/2}x)$ is the Fourier transform of $\chi(q)$.) The second step of (8.2) follows by Poisson summation; see Appendix A and in particular (A.14) for more details. The following lemma gives the precise error bounds in the approximation (8.2).

LEMMA 8.2. *Let $d = 1, 2, 3, \ldots$ and assume (2.17). Then*

$$\Theta_{xy} \;=\; \theta_{x-y} + O_K\left(\frac{1}{W^d}\left\langle \frac{x}{W} \right\rangle^{-K} + \frac{\eta^{d/2}}{W^d}\left\langle \frac{\sqrt{\eta}x}{W} \right\rangle^{-K}\right).$$
(8.3)

PROOF. See Appendix A. □

REMARK 8.3. The convolution in (8.2) smooths out the Yukawa potential on the scale $x \approx W$. The error terms in (8.3) are negligible compared to the main term θ_x in the regime $W \ll |x| \leqslant CW\eta^{-1/2}$. Therefore the approximation θ is meaningful from the profile scale $W\eta^{-1/2}$ down to the band scale W. The actual choice of the function χ in (8.2) is immaterial in the relevant regime $|x| \gg W$, as long as χ is equal to one in a neighbourhood of the origin.

Next, we state the counterparts of Theorem 2.2, Corollary 2.4, and Theorem 2.5 in the higher-dimensional setting. Their proofs are trivial modifications of the proofs of their one-dimensional counterparts, using Lemmas 8.1 and 8.2.

THEOREM 8.4 (IMPROVED LOCAL SEMICIRCLE LAW). *Let $d = 2, 3, \ldots$ and assume (2.17). Suppose moreover that $L \ll W^{1+d/4}$ and $\eta \gg L^2/W^{d+2}$. Then we have*

$$\Lambda^2 \;\prec\; \max\left\{\frac{1}{M}, \frac{1}{N\eta}\right\}$$
(8.4)

for $z \in \mathbf{S}$.

COROLLARY 8.5 (EIGENVECTOR DELOCALIZATION). *Let $d = 2, 3, \dots$ and assume (2.17). If $L \ll W^{1+d/4}$ then the eigenvectors of H are completely delocalized in the sense of Proposition 7.1.*

THEOREM 8.6 (DIFFUSION PROFILE). *Let $d = 2, 3, \dots$ and assume (2.17). Suppose that $L \ll W^{1+d/4}$ and $(W/L)^2 \leqslant \eta \leqslant 1$. Then*

$$|T_{xy} - \Theta_{xy}| \prec \frac{1}{N\eta} + \frac{1}{M^{3/2}\eta}, \qquad \left| P_x |G_{xy}|^2 - \delta_{xy}|m|^2 - |m|^2 \Theta_{xy} \right| \prec \frac{1}{N\eta} + \frac{1}{M^{3/2}\eta} + \frac{\delta_{xy}}{\sqrt{M}}. \qquad (8.5)$$

Moreover, the analogues of (2.31) and (2.32) hold with

$$\Upsilon_{xy} \equiv \Upsilon_{xy}^{(K)} := \frac{\eta^{d/2-1}}{W^d} \left(V * \varphi_{\sqrt{\eta}} \right) \left(\frac{\sqrt{\alpha\eta}}{W} D^{-1/2}(x - y) \right) + \frac{1}{W^d} \left\langle \frac{x - y}{W} \right\rangle^{-K} + \frac{\eta^{d/2}}{W^d} \left\langle \frac{\sqrt{\eta}(x - y)}{W} \right\rangle^{-K},$$

where K is an arbitrary, fixed, positive integer.

8.2. Slowly decaying band. In this section we make the following assumption on the band shape. Suppose that $d = 1$ and f is smooth and symmetric, and satisfies

$$f(x) = \frac{h(x)}{|x|^{1+\beta}} \qquad (8.6)$$

for some fixed $\beta \in (0, 2)$. Here h is a symmetric function satisfying

$$|h(x) - h_0| \leqslant C\langle x \rangle^{-3} \qquad (8.7)$$

for some fixed $h_0 > 0$. Note that by definition f is smooth and symmetric, so that $h(x) = O(|x|^{1+\beta})$ near the origin.

In order to avoid technical issues arising from the periodicity of S, we cut off the tail of f at scales $x \approx N$. Thus we set

$$s_{xy} := \frac{1}{Z} f\left(\frac{[x - y]_N}{W} \right) \sigma\left(\frac{[x - y]_N}{N} \right);$$

here σ is a smooth, symmetric bump function satisfying $\sigma(x) = 1$ for $|x| \leqslant a$ and $\sigma(x) = 0$ for $|x| \geqslant b$, where $0 < a < b < 1/2$. As usual, Z is a normalization constant.

The following lemma is the analogue of Lemma 5.2. Its proof is similar to that of Lemma 5.2; the key input is Lemma A.2 (iii).

LEMMA 8.7. *Suppose that $d = 1$ and that (8.6) and (8.7) hold. Then*

$$\left\| \frac{1}{1 - |m|^2 S} \overline{\Pi} \right\|_\infty \leqslant \frac{C \log N}{\eta + \left(\frac{W}{N} \right)^\beta}. \qquad (8.8)$$

Next, we give the sharp upper bound on the peak of the profile.

LEMMA 8.8. *Suppose that $d = 1$ and that (8.6) and (8.7) hold. Then*

$$\Theta_{xy} \leqslant C \max\left\{ \frac{\eta^{1/\beta - 1} + 1}{W}, \frac{1}{N\eta} \right\}.$$

PROOF. See Appendix A. □

In order to describe the asymptotic shape of the profile, we define

$$B := h_0 \frac{W}{Z} \int_{\mathbb{R}} \mathrm{d}u \, \frac{1 - \cos u}{|u|^{1+\beta}} = h_0 \int_{\mathbb{R}} \mathrm{d}u \, \frac{1 - \cos u}{|u|^{1+\beta}} + O\left(\frac{1}{W} + \left(\frac{W}{N}\right)^\beta\right),$$

which plays a role similar to the unrenormalized diffusion constant D from (2.26). Moreover, define the function

$$V(x) := \int_{\mathbb{R}} \frac{\mathrm{d}q}{2\pi} \frac{\mathrm{e}^{\mathrm{i}qx}}{1 + |q|^\beta},$$

which is bounded for $\beta > 1$. It is easy to check that for $\beta > 1$ and $|x| \geqslant 1$ we have

$$V(x) = C_\beta |x|^{-1-\beta} + O(|x|^{-2-\beta}) \tag{8.9}$$

with an explicitly computable constant $C_\beta > 0$.

PROPOSITION 8.9. *Suppose that $d = 1$ and that (8.6) and (8.7) hold for some $\beta > 1$. Suppose moreover that*

$$\left(\frac{W}{N}\right)^\beta \ll \eta \ll 1. \tag{8.10}$$

Then there is a constant $c > 0$, depending on the implicit exponents in (8.10), such that for $x \in \mathbb{T}$ we have

$$\Theta_{x0} = \frac{|m|^2}{W\alpha\eta}\left(\frac{\alpha\eta}{B}\right)^{1/\beta} V\left[\left(\frac{\alpha\eta}{B}\right)^{1/\beta} \frac{x}{W}\right] + O\left(\frac{\eta^{1/\beta-1}}{W^{1+c}}\right) \tag{8.11}$$

PROOF. See Appendix A. □

The matrix Θ is the resolvent of a *superdiffusive* operator, whose symbol in Fourier space is $B|Wp|^\beta$. Thus, under the identification $t = \eta^{-1}$ from Remark 2.7, we find that the associated dynamics scales according to $x \sim Wt^{1/\beta}$ instead of the diffusive scaling $x \sim Wt^{1/2}$.

We may now state the counterparts of Theorem 2.2, Corollary 2.4, and Theorem 2.5 for the slowly decaying band. Their proofs are trivial modifications of those for the strongly decaying band, using Lemmas 8.7 and 8.8.

THEOREM 8.10 (IMPROVED LOCAL SEMICIRCLE LAW). *Suppose that $d = 1$ and that (8.6) and (8.7) hold. Suppose moreover that $N \ll W^{1+1/2\beta}$ and $\eta \gg (N/W)^\beta/W$. Then we have*

$$\Lambda^2 \prec \max\left\{\frac{\eta^{1/\beta-1} + 1}{W}, \frac{1}{N\eta}\right\} \tag{8.12}$$

for $z \in \mathbf{S}$.

COROLLARY 8.11 (EIGENVECTOR DELOCALIZATION). *Suppose that $d = 1$ and that (8.6) and (8.7) hold. If $N \ll W^{1+1/2\beta}$ then the eigenvectors of H are completely delocalized in the sense of Proposition 7.1.*

THEOREM 8.12 (DIFFUSION PROFILE). *Suppose that $d = 1$ and that (8.6) and (8.7) hold for some $\beta \geqslant 1$. Suppose that $N \ll W^{1+1/2\beta}$ and $(W/N)^\beta \leqslant \eta \leqslant 1$. Then*

$$|T_{xy} - \Theta_{xy}| \prec \frac{1}{N\eta}, \qquad \left||P_x|G_{xy}|^2 - \delta_{xy}|m|^2 - |m|^2\Theta_{xy}\right| \prec \frac{1}{N\eta} + \frac{\delta_{xy}}{\sqrt{W}}. \tag{8.13}$$

Moreover, the analogues of (2.31) and (2.32) hold.

A. The deterministic profile

In this appendix we establish bounds and asymptotics for the deterministic profile Θ_{xy}.

A.1. Proof of Proposition 2.8. We assume $d = 1$ and abbreviate $\mathbb{T} \equiv \mathbb{T}_N^1$ as well as $P \equiv P_N^1$. We assume without loss of generality that $x \in \mathbb{T}$.

First we notice that by the symmetry of f and (2.17), the Fourier transform $\widehat{f}(q) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}qx} f(x) \, \mathrm{d}x$ is a smooth function with a Taylor expansion

$$\widehat{f}(q) \; = \; 1 - D_\infty q^2 + q^4 g(q) \qquad (q \in \mathbb{R}), \tag{A.1}$$

where g is a bounded smooth function satisfying $g(q) = g(-q)$. Clearly, $\widehat{f}$ is real and $\|\widehat{f}\|_\infty \leqslant 1$. Moreover, we claim that for any $\varepsilon > 0$ there exists an $\varepsilon' > 0$ such that

$$\widehat{f}(q) \; \leqslant \; 1 - \varepsilon' \qquad \text{if} \qquad |q| \; \geqslant \; \varepsilon \,; \tag{A.2}$$

indeed, this follows easily from the identity

$$1 - \widehat{f}(q) \; = \; \int_{\mathbb{R}} \big(1 - \cos(qx)\big) f(x) \, \mathrm{d}x \,.$$

Next, define the lattice

$$Q \; := \; WP \; = \; \frac{2\pi W}{N} \mathbb{T} \,.$$

For $q \in [-\pi W, \pi W]$ define

$$\widehat{S}_W(q) \; := \; \widehat{S}(q/W) \,.$$

Thus we get

$$\left(\frac{S}{1 - |m|^2 S} \right)_{x0} \; = \; \frac{1}{N} \sum_{p \in P} \mathrm{e}^{\mathrm{i}px} \frac{\widehat{S}(p)}{1 - |m|^2 \widehat{S}(p)} \; = \; \frac{1}{N} \sum_{q \in Q} \mathrm{e}^{\mathrm{i}qx/W} \frac{\widehat{S}_W(q)}{1 - |m|^2 \widehat{S}_W(q)} \tag{A.3}$$

for $x \in \mathbb{T}$.

As a guide for intuition, we have $\widehat{S}_W(q) \approx \widehat{f}(q)$, as can be seen from

$$\widehat{S}_W(q) \; = \; \sum_{x \in \mathbb{T}} \mathrm{e}^{-\mathrm{i}xq/W} \frac{1}{Z_{N,W}} f\left(\frac{x}{W}\right) \; \approx \; \int_{-N/2}^{N/2} \mathrm{e}^{-\mathrm{i}xq/W} f\left(\frac{x}{W}\right) \frac{1}{W} \, \mathrm{d}x \; = \; \int_{-N/2W}^{N/2W} \mathrm{e}^{-\mathrm{i}qy} f(y) \, \mathrm{d}y \; \approx \; \widehat{f}(q) \,. \tag{A.4}$$

Thus, our proof consists in controlling the error in the approximation

$$\left(\frac{S}{1 - |m|^2 S} \right)_{x0} \; \approx \; \frac{1}{W} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}qx/W} \frac{\widehat{f}(q)}{1 - |m|^2 \widehat{f}(q)} \, \mathrm{d}q \,.$$

As a first step, we establish basic properties of $\widehat{S}_W$ that are analogous to (A.2) and (A.1).

LEMMA A.1. *The function $\widehat{S}_W$ is smooth with uniformly bounded derivatives, real, and symmetric with $|\widehat{S}_W(q)| \leqslant 1$ and $\widehat{S}_W(0) = 1$. Moreover, it has the following properties.*

(i) *For any $\varepsilon > 0$ there exists an $\varepsilon' > 0$ such that*

$$\widehat{S}_W(q) \leqslant 1 - \varepsilon' \qquad if \qquad |q| \geqslant \varepsilon \tag{A.5}$$

for large enough W (depending on ε).

(ii) *For any $k, K \in \mathbb{N}$ there is a constant $C_{k,K}$ such that*

$$|\partial_q^k \widehat{S}_W(q)| \leqslant \frac{C_{k,K}}{\langle q \rangle^K} + O(N^{-1}), \qquad (q \in [-\pi W, \pi W]). \tag{A.6}$$

(iii) *There exists a smooth function g_W whose derivatives are bounded uniformly in W such that*

$$\widehat{S}_W(q) = 1 - Dq^2 + q^4 g_W(q), \qquad (q \in [-\pi W, \pi W]). \tag{A.7}$$

PROOF. The proof of (i) is similar to that of (A.2).

To prove (ii), we use summation by parts combined with (2.17). Let the integers N_- and N_+ denote the end points of $\mathbb{T} = [N/2, N/2)$, i.e. $\mathbb{T} = [N_-, N_+]$. Then we find

$$
\begin{aligned}
\widehat{S}_W(q) &= \frac{1}{Z_{N,W}} \sum_{x=N_-}^{N_+} \mathrm{e}^{-iqx/W} f\left(\frac{x}{W}\right) \\
&= \frac{1}{Z_{N,W}} \sum_{x=N_-}^{N_+-1} \sum_{y=N_-}^{x} \mathrm{e}^{-iqy/W} \left[f\left(\frac{x}{W}\right) - f\left(\frac{x+1}{W}\right) \right] + \frac{1}{Z_{N,W}} f\left(\frac{N_+}{W}\right) \sum_{y=N_-}^{N_+} \mathrm{e}^{-iqy/W} \\
&= \frac{1}{Z_{N,W}} \sum_{x=N_-}^{N_+-1} \frac{\mathrm{e}^{-iqN_-/W} - \mathrm{e}^{-iq(x+1)/W}}{1 - \mathrm{e}^{-iq/W}} \left[f\left(\frac{x}{W}\right) - f\left(\frac{x+1}{W}\right) \right] + O_K\left(\frac{N}{W} \left(\frac{W}{N}\right)^K\right) \\
&\leqslant \frac{C_K}{W} \sum_x \frac{1}{|1 - \mathrm{e}^{-iq/W}|} \frac{1}{W} \left\langle \frac{x}{W} \right\rangle^{-K} + C_K \left(\frac{W}{N}\right)^{K-1} \\
&\leqslant \frac{C}{|q|} + O(N^{-1}).
\end{aligned} \tag{A.8}
$$

Here we used that

$$\frac{1}{|1 - \mathrm{e}^{-iq/W}|} \leqslant \frac{CW}{|q|}$$

(recall that $|q| \leqslant \pi W$) and that

$$\left| f\left(\frac{x}{W}\right) - f\left(\frac{x+1}{W}\right) \right| \leqslant \frac{C_K}{W} \left\langle \frac{x}{W} \right\rangle^{-K}$$

to estimate the main term, and (2.1) to estimate the error term with $K \sim 1/\delta$. We also have the trivial bound on $|\widehat{S}_W| \leqslant 1$. Thus we have

$$|\widehat{S}_W(q)| \leqslant \frac{C}{\langle q \rangle} + O(N^{-1}).$$

We can iterate the above argument for the main term in (A.8), thus obtaining higher order divided differences of f. Since f is smooth and decays rapidly, (A.6) follows for $k = 0$. The proof for $k > 0$ is analogous.

In order to prove (iii), we write

$$g_W(q) := \frac{1 - Dq^2 - \widehat{S}_W(q)}{q^4} = \frac{1}{Z_{N,W}} \sum_{x \in \mathbb{T}} \frac{1 - (qx/W)^2/2 - \cos(qx/W)}{(qx/W)^4} \left(\frac{x}{W}\right)^4 f\left(\frac{x}{W}\right).$$

Now (iii) follows from the fact that the function $h(q) := \left(1 - q^2/2 - \cos(q)\right)q^{-4}$ is smooth and its derivatives are bounded. $\qquad\square$

Having proved Lemma A.1, we may now complete the proof of Proposition 2.8. Fix a small constant $\varepsilon > 0$ and introduce a partition of unity $\chi + \overline{\chi} = 1$ on $\mathbb{R}$ with smooth functions such that $\chi(q) = 1$ for $|q| \leqslant \varepsilon$ and $\chi(q) = 0$ for $|q| \geqslant 2\varepsilon$. Since $\widehat{S}_W(q) \leqslant 1$ and $|m|^2 \leqslant 1 - c\eta$ (see (3.6)), we have

$$1 - |m|^2 \widehat{S}_W(q) \geqslant 1 - |m|^2 \geqslant c\eta$$

with some positive constant c. By (A.5) we have on the support of $\overline{\chi}$

$$1 - |m|^2 \widehat{S}_W(q) \geqslant \varepsilon' \tag{A.9}$$

for some ε' depending on ε. Then from (A.3) we have

$$\left(\frac{S}{1 - |m|^2 S}\right)_{x0} = \frac{1}{N} \sum_{q \in Q} e^{iqx/W} \frac{\widehat{S}_W(q)\chi(q)}{1 - |m|^2 \widehat{S}_W(q)} + \frac{1}{N} \sum_{q \in Q} e^{iqx/W} \frac{\widehat{S}_W(q)\overline{\chi}(q)}{1 - |m|^2 \widehat{S}_W(q)}$$

$$= \frac{1}{N} \sum_{q \in Q} e^{iqx/W} \frac{\widehat{S}_W(q)\chi(q)}{1 - |m|^2 \widehat{S}_W(q)} + O_K\left(\frac{1}{W}\left\langle\frac{x}{W}\right\rangle^{-K}\right). \tag{A.10}$$

Here we used that the function

$$R(q) = \frac{\widehat{S}_W(q)\overline{\chi}(q)}{1 - |m|^2 \widehat{S}_W(q)},$$

extended to the whole real line, is smooth and its derivatives are bounded uniformly in N and W (by (A.9)). (These bounds may of course depend on ε). Moreover, $R(q) = O_K(\langle q\rangle^{-K})$ for any K; see (A.6). By summation by parts, as in (A.8), we find that for such a function we have

$$\frac{1}{N} \sum_{q \in Q} e^{iqx/W} R(q) = O_K\left(\frac{1}{W}\left\langle\frac{x}{W}\right\rangle^{-K}\right) \tag{A.11}$$

for any K.

Now we consider the first term in (A.10). For the following we use $A_i(q, \eta, N, W)$ with $i = 1, 2, 3, \ldots$ to denote functions that are smooth in q and whose q-derivatives are uniformly bounded in q, η, W, and N. Using the Taylor expansion (A.7) and (3.5), we have (omitting the arguments for brevity)

$$1 - |m|^2 \widehat{S}_W = \alpha\eta + Dq^2 + A_1 q^4 + A_2 \eta q^2 + A_3 \eta^2.$$

This gives (again omitting the arguments)

$$\left(\frac{\widehat{S}_W}{1 - |m|^2 \widehat{S}_W} - \frac{1}{\alpha\eta + Dq^2}\right)\chi = \frac{A_4 q^4 + A_5 \eta q^2 + A_6 \eta^2}{\left(\alpha\eta + Dq^2 + A_1 q^4 + A_2 \eta q^2 + A_3 \eta^2\right)\left(\alpha\eta + Dq^2\right)}\chi$$

$$= \frac{A_4 r^4 + A_5 r^2 + A_6}{\left(\alpha + Dr^2 + A_1 \eta r^4 + A_2 \eta r^2 + A_3 \eta\right)\left(\alpha + Dr^2\right)}\chi$$

$$=: F_{N,W,\eta}(r), \tag{A.12}$$

where we introduced the new variable $r := \eta^{-1/2}q$. By definition, $A_1, \ldots, A_6$ and their q-derivatives are uniformly bounded. Since $D \geqslant c > 0$ and $r \leqslant \varepsilon\eta^{-1/2}$ on the support of χ, we find that for small enough ε the denominator of the second line of (A.12) is bounded away from zero, uniformly in r, η, W, and N. We therefore conclude that $F_{N,W,\eta}$ is smooth and its derivatives (in the variable r) are uniformly bounded.

Using summation by parts, exactly as in (A.8), we get

$$\frac{1}{N}\sum_{q \in Q} e^{iqx/W} F_{N,W,\eta}(\eta^{-1/2}q) \leqslant \frac{C_K}{W}\left\langle\frac{\sqrt{\eta}x}{W}\right\rangle^{-K}. \tag{A.13}$$

Here we used that the sum on the left-hand side ranges over a set of size $O(N/W)$ due to the factor χ in the definition of $F_{N,W,\eta}$. Therefore (A.12) and (A.13) imply that the first term of (A.10) is given by

$$\frac{1}{N}\sum_{q \in Q} e^{iqx/W}\frac{\widehat{S}_W(q)\chi(q)}{1 - |m|^2 \widehat{S}_W(q)} = \frac{1}{N}\sum_{q \in Q} e^{iqx/W}\frac{\chi(q)}{\alpha\eta + Dq^2} + O_K\left(\frac{1}{W}\left\langle\frac{\sqrt{\eta}x}{W}\right\rangle^{-K}\right). \tag{A.14}$$

Notice that the error term in (A.11) is smaller than in (A.14). Next, we remove the factor χ from the main term, exactly as in (A.11). Plugging this into (A.10) yields

$$\left(\frac{S}{1 - |m|^2 S}\right)_{x0} = \frac{1}{N}\sum_{q \in Q} e^{iqx/W}\frac{1}{\alpha\eta + Dq^2} + O_K\left(\frac{1}{W}\left\langle\frac{\sqrt{\eta}x}{W}\right\rangle^{-K}\right). \tag{A.15}$$

We can extend the summation in the main term

$$\frac{1}{N}\sum_{q \in Q} e^{iqx/W}\frac{1}{\alpha\eta + Dq^2} = \frac{1}{N}\sum_{q \in \frac{2\pi W}{N}\mathbb{Z}} e^{iqx/W}\frac{1}{\alpha\eta + Dq^2} + O\left(\frac{1}{W}\int_{\mathbb{R}}\frac{\mathbf{1}(|q| \geqslant \pi W)}{\alpha\eta + Dq^2}\,dq\right),$$

where the error term on the right-hand side is of order $O(W^{-2})$. Thus we have

$$\left(\frac{S}{1 - |m|^2 S}\right)_{x0} = \frac{1}{N}\sum_{q \in \frac{2\pi W}{N}\mathbb{Z}} e^{iqx/W}\frac{1}{\alpha\eta + Dq^2} + O_K\left(\frac{1}{W}\left\langle\frac{\sqrt{\eta}x}{W}\right\rangle^{-K}\right) + O\left(\frac{1}{W^2}\right).$$

The main term can be computed by the Poisson summation formula

$$\frac{1}{\pi}\sum_{n \in \mathbb{Z}}\frac{e^{inx}}{a^2 + n^2} = \frac{1}{a}\sum_{k \in \mathbb{Z}} e^{-a|x + 2\pi k|},$$

where $a > 0$. Thus,

$$\frac{1}{N}\sum_{q\in\frac{2\pi W}{N}\mathbb{Z}}\mathrm{e}^{\mathrm{i}qx/W}\frac{1}{\alpha\eta+Dq^2} = \frac{1}{2W\sqrt{D\alpha\eta}}\sum_{k\in\mathbb{Z}}\exp\left[-\frac{\sqrt{\alpha\eta}}{W\sqrt{D}}|x+kN|\right].$$

This concludes the proof of (2.37).

In order to prove (2.38), it suffices to analyse the asymptotics of the expression

$$R := \frac{1}{W\sqrt{\eta}}\sum_{k\in\mathbb{Z}}\mathrm{e}^{-\sqrt{\eta}\frac{N}{W}k}.$$

We consider two cases. If $\eta \geqslant \left(\frac{W}{N}\right)^2$ then $R \asymp \frac{1}{W\sqrt{\eta}}$. On the other hand, if $\eta \leqslant \left(\frac{W}{N}\right)^2$ we use an integral approximation to get

$$R = \frac{1}{N\eta}\sqrt{\eta}\frac{N}{W}\sum_k \mathrm{e}^{-\sqrt{\eta}\frac{N}{W}k} \asymp \frac{1}{N\eta}.$$

This concludes the proof of (2.38), and hence of Proposition 2.8.

A.2. Higher dimensions: proofs of Lemmas 8.1 and 8.2.

PROOF OF LEMMA 8.1. We follow the argument from the proof of Proposition 2.8 in the previous section, and merely sketch the differences. We use the d-dimensional lattices

$$\mathbb{T} \equiv \mathbb{T}_L^d, \qquad Q := \frac{2\pi W}{L}\mathbb{T}.$$

Exactly as in (A.3), we get

$$\left(\frac{S}{1-|m|^2S}\right)_{x0} = \frac{1}{N}\sum_{q\in Q}\mathrm{e}^{\mathrm{i}q\cdot x/W}\frac{\widehat{S}_W(q)}{1-|m|^2\widehat{S}_W(q)} \qquad \text{where} \qquad \widehat{S}_W(q) := \sum_{x\in\mathbb{T}}\mathrm{e}^{-\mathrm{i}q\cdot x/W}s_{x0}. \qquad \text{(A.16)}$$

Next, the basic properties of $\widehat{S}_W$ listed in Lemma A.1, and their proofs, carry over verbatim to the higher-dimensional setting. Now (A.7) reads $\widehat{S}_W(q) = 1 - (q \cdot Dq)(1 + A_2(q))$ where $A_2(q) = O(|q|^2)$ uniformly in W. Let χ be a smooth bump function on $\mathbb{R}^d$, as in the proof of Proposition 2.8. As in (A.10), we find

$$\left(\frac{S}{1-|m|^2S}\right)_{x0} = \frac{1}{N}\sum_q\mathrm{e}^{\mathrm{i}q\cdot x/W}\frac{\widehat{S}_W(q)\chi(q)}{1-|m|^2\widehat{S}_W(q)} + O_K\left(\frac{1}{W^d}\left\langle\frac{x}{W}\right\rangle^{-K}\right)$$

for arbitrary $K \in \mathbb{N}$. Here, and for the rest of this proof, the summation in q ranges over the lattice $\left(\frac{2\pi W}{N}\mathbb{Z}\right)^d$. We split

$$\left(\frac{S}{1-|m|^2S}\right)_{x0} = \frac{1}{N}\sum_q\mathrm{e}^{\mathrm{i}q\cdot x/W}\frac{\chi(q)}{\alpha\eta+q\cdot Dq} + \frac{1}{N}\sum_q\mathrm{e}^{\mathrm{i}q\cdot x/W}R(q)\chi(q) + O_K\left(\frac{1}{W^d}\left\langle\frac{x}{W}\right\rangle^{-K}\right), \qquad \text{(A.17)}$$

where

$$R(q) := \frac{\widehat{S}_W(q)}{1-|m|^2\widehat{S}_W(q)} - \frac{1}{\alpha\eta+q\cdot Dq}.$$

40

Note that, unlike in the proof of Proposition 2.8, we keep the cutoff function χ in the main term since the function $(\eta + q \cdot Dq)^{-1}$ is not integrable in higher dimensions.

The main term of (A.17) can be computed using Poisson summation:

$$\frac{1}{N} \sum_{q \in (\frac{2\pi W}{N} \mathbb{Z})^d} e^{iq \cdot x/W} \frac{\chi(q)}{\alpha\eta + qDq} = \frac{(\alpha\eta)^{d/2-1}}{W^d \sqrt{\det D}} \sum_{k \in \mathbb{Z}^d} (V * \varphi_{\sqrt{\alpha\eta}}) \left(\frac{\sqrt{\alpha\eta}}{W} D^{-1/2}(x - y + kL) \right). \tag{A.18}$$

Using that $V(x) \asymp |x|^{2-d}$ near the origin, we find $\|V * \varphi_{\sqrt{\alpha\eta}}\|_\infty \leqslant CW^{-d}$. By treating the two cases $\eta \leqslant \left(\frac{W}{N}\right)^2$ and $\eta \geqslant \left(\frac{W}{N}\right)^2$ separately, we find exactly as in the last paragraph of the proof of Proposition 2.8 that (A.18) is bounded by $CW^{-d} + C(N\eta)^{-1}$.

What remains therefore is the estimate of the error term containing R in (A.17). To that end, we write

$$R(q) = \frac{B_4 + \eta B_2 + \eta^2 B_0}{\left(\alpha\eta + (q \cdot Dq)(1 + A_2) + \eta A_2' + \eta^2 A_0\right)(\alpha\eta + q \cdot Dq)}, \tag{A.19}$$

where $B_0, B_2, B_4, A_0, A_2, A_2'$ are smooth and bounded functions of q, each of order $O(|q|^i)$ near the origin uniformly in W and η, where i denotes the subscript of the corresponding function. Using the change of variables $q = \sqrt{\eta} r$ it is now easy to see that the error term containing R in (A.17) is bounded by CW^{-d}. This concludes the proof of Lemma 8.1. $\qquad\square$

PROOF OF LEMMA 8.2. We need a more precise bound on the error term of (A.17) than the bound CW^{-d} from the proof of Lemma 8.1. In fact, we claim that

$$\frac{1}{N} \sum_q e^{iqx/W} R(q)\chi(q) = O_K\left(\frac{1}{W^d} \left\langle \frac{x}{W} \right\rangle^{-K} + \frac{\eta^{d/2}}{W^d} \left\langle \frac{\sqrt{\eta}x}{W} \right\rangle^{-K} \right). \tag{A.20}$$

The proof of (A.20) is a rather laborious exercise in Taylor expansion whose details we omit. The basic strategy is similar to the analysis of (A.12), except that we expand $\widehat{S}_W$ up to order $d/2 + 2$ (instead of 4). This completes the proof of Lemma 8.2. $\qquad\square$

A.3. Slowly decaying band: proof of Lemma 8.8 and Proposition 8.9. We begin by proving the following auxiliary result, which gives the relevant asymptotics of $\widehat{S}_W$. For $q \neq 0$ define

$$b(q) := h_0 \frac{W}{Z} \int_{\mathbb{R}} du \, \frac{1 - \cos u}{|u|^{1+\beta}} \sigma\left(\frac{Wu}{qN}\right) = B + O\left[\left(\frac{W}{qN}\right)^\beta\right]. \tag{A.21}$$

We also set $b(0) := 0$, so that b is continuous.

LEMMA A.2. *Suppose that $d = 1$ and that (8.6) and (8.7) hold. Then the following are true.*

(i) *For any $K \in \mathbb{N}$ there exists a constant C_K such that for $|q| \geqslant 1$ we have*

$$|\widehat{S}_W(q)| \leqslant \frac{C_K}{|q|^K}.$$

(ii) *For $|q| \leqslant 1$ we have*

$$\widehat{S}_W(q) = 1 - b(q)|q|^\beta + O(q^2) \tag{A.22}$$

uniformly in N and W.

(iii) There is a constant c_1 such that

$$\|S\overline{\Pi}\| \;\leqslant\; 1 - c_1 \left(\frac{W}{N}\right)^{\beta}.$$

PROOF. Part (i) is proved similarly to (A.6), using summation by parts.

In order to prove part (ii), we write

$$1 - \widehat{S}_W(q) \;=\; \frac{1}{Z} \sum_{x \in \frac{1}{W}\mathbb{Z}} \big(1 - \cos(qx)\big) \frac{h(x)}{|x|^{1+\beta}} \sigma\left(\frac{Wx}{N}\right). \tag{A.23}$$

Let χ be a smooth, symmetric bump function satisfying $\chi(x) = 1$ for $|x| \leqslant 1$ and $\chi(x) = 0$ for $|x| \geqslant 2$. Write $\overline{\chi} := 1 - \chi$. We introduce the splitting

$$h \;=\; \chi h + \overline{\chi}(h - h_0) + h_0 \overline{\chi}$$

on the right-hand side of (A.23). It is easy to check that the two first terms give a contribution of order $O(q^2)$. The last term of the splitting gives rise to

$$\frac{h_0}{Z} \sum_{x \in \frac{1}{W}\mathbb{Z}} \big(1 - \cos(qx)\big) \frac{\overline{\chi}(x)}{|x|^{1+\beta}} \sigma\left(\frac{Wx}{N}\right) \;=\; h_0 \frac{W}{Z} \int_{\mathbb{R}} \big(1 - \cos(qx)\big) \frac{\overline{\chi}(x)}{|x|^{1+\beta}} \sigma\left(\frac{Wx}{N}\right) \mathrm{d}x + O\left(\frac{q^2}{W^2} + \frac{1}{N^2}\right),$$

$$\tag{A.24}$$

where the last step follows from a mid-point Riemann sum approximation. Now a change of variables $u = qx$ easily yields (A.22).

Part (iii) follows from part (ii) using an argument similar to (5.15). □

PROOF OF LEMMA 8.8. The claim follows from the bound

$$\Theta_{xy} \;\leqslant\; \frac{C}{W} \frac{W}{N} \sum_{q \in Q} \frac{\widehat{S}_W(q)}{1 - |m|^2 \widehat{S}_W(q)} \;\leqslant\; \frac{C}{N\eta} + \frac{C}{W} \int_{\mathbb{R}} \mathrm{d}q \, \frac{\widehat{S}_W(q) \mathbf{1}(|q| \geqslant \eta^{1/\beta})}{1 - |m|^2 \widehat{S}_W(q)} \;\leqslant\; \frac{C}{N\eta} + \frac{C(\eta^{1/\beta - 1} + 1)}{W},$$

where the first term is the contribution of the low modes $|q| \leqslant \eta^{1/\beta}$ and the second term the contribution of the high modes $|q| \geqslant \eta^{1/\beta}$, which may be replaced with an integral and estimated using Lemma A.2. We omit the details. □

PROOF OF PROPOSITION 8.9. We proceed similarly to the proof of Proposition 2.8. We choose a cutoff scale ε, and denote by χ the bump function from the proof of Lemma A.2. The scale ε satisfies $\eta^{1/\beta} \ll \varepsilon \ll 1$, and will be chosen by optimizing at the end of the proof.

We use the expansions (A.22) and (3.5). Thus we find, as in the proof of Proposition 2.8,

$$\left(\frac{S}{1 - |m|^2 S}\right)_{x0} \;=\; \frac{1}{N} \sum_{q \in Q} \mathrm{e}^{\mathrm{i}qx/W} \frac{\chi(q/\varepsilon)}{\alpha\eta + B|q|^{\beta}} + \frac{1}{N} \sum_{q \in Q} \mathrm{e}^{\mathrm{i}qx/W} R(q)\chi(q/\varepsilon) + O\left(\frac{\varepsilon^{1-\beta}}{W}\right), \tag{A.25}$$

where χ is a smooth bump function as in the proof of Proposition 2.8 and

$$R(q) \;:=\; \frac{\widehat{S}_W(q)}{1 - |m|^2 \widehat{S}_W(q)} - \frac{1}{\alpha\eta + B|q|^{\beta}} \;=\; \frac{(B-b)|q|^{\beta} + O\big(\eta^2 + \eta|q|^{\beta} + q^2\big)}{\big(\alpha\eta + b|q|^{\beta} + O(\eta^2 + \eta|q|^{\beta} + q^2)\big)\big(\alpha\eta + B|q|^{\beta}\big)}.$$

Note that for $q \in Q \setminus \{0\}$ we have $b(q) \geqslant c$. Using (A.21) we may therefore estimate, as in (A.12), to get

$$\frac{1}{N} \sum_{q \in Q} |R(q)| \chi(q/\varepsilon) \leqslant \frac{C}{W}\left(\log N \, \eta^{1/\beta - 2} \left(\frac{W}{N}\right)^{\beta} + 1 + \varepsilon^{2-\beta} \eta^{1/\beta - 1} \right).$$

Next, the bump function in the main term of (A.25) may be easily removed, and the summation in q extended to the whole lattice $\frac{2\pi W}{N}\mathbb{Z}$, at the expense of an error of order $O(\varepsilon^{1-\beta}/W)$. Putting everything together, we get

$$\left(\frac{S}{1 - |m|^2 S}\right)_{x0} = \frac{1}{N} \sum_{q \in \frac{2\pi W}{N}\mathbb{Z}} \mathrm{e}^{\mathrm{i}qx/W} \frac{1}{\alpha\eta + B|q|^{\beta}} + \frac{\eta^{1/\beta - 1}}{W} O\big(W^{-c} + \varepsilon^{2-\beta} + \varepsilon^{1-\beta}\eta^{1-1/\beta}\big)$$

for some $c > 0$, where we used (8.10). Setting $\varepsilon := \eta^{1-1/\beta}$ and Poisson summation yields

$$\Theta_{x0} = \frac{|m|^2}{W\alpha\eta}\left(\frac{\alpha\eta}{B}\right)^{1/\beta} \sum_{k \in \mathbb{Z}} V\left[\left(\frac{\alpha\eta}{B}\right)^{1/\beta} \frac{x + kN}{W}\right] + O\left(\frac{\eta^{1/\beta - 1}}{W^{1+c}}\right).$$

Now (8.11) follows by noting that by (8.9), under the assumption (8.10), only the term $k = 0$ is of leading order. $\qquad\square$

B. Multilinear large deviation estimates

In this appendix we give a generalization of the large deviation estimate of Corollary B.3 [24]. The proof is simpler and the statement is formulated under the assumption (2.11) instead of the stronger subexponential decay assumption. Moreover, since the current proof does not rely on the Burkholder inequality, it is trivially generalizable to arbitrary multilinear estimates.

Throughout the following we consider random variables X satisfying

$$\mathbb{E}X = 0, \qquad \mathbb{E}|X|^2 = 1, \qquad \|X\|_p \leqslant \mu_p \tag{B.1}$$

for all p with some μ_p. Here we set $\|X\|_p := \big(\mathbb{E}|X|^p\big)^{1/p}$.

THEOREM B.1 (LARGE DEVIATION BOUNDS). *Let $\big(X_i^{(N)}\big)$, $\big(Y_i^{(N)}\big)$, $\big(a_{ij}^{(N)}\big)$, and $\big(b_i^{(N)}\big)$ be independent families of random variables, where $N \in \mathbb{N}$ and $i, j = 1, \ldots, N$. Suppose that all entries $X_i^{(N)}$ and $Y_i^{(N)}$ are independent and satisfy (B.1).*

(i) Suppose that $\big(\sum_i |b_i|^2\big)^{1/2} \prec \Psi$. Then $\sum_i b_i X_i \prec \Psi$.

(ii) Suppose that $\big(\sum_{i \neq j} |a_{ij}|^2\big)^{1/2} \prec \Psi$. Then $\sum_{i \neq j} a_{ij} X_i X_j \prec \Psi$.

(iii) Suppose that $\big(\sum_{i,j} |a_{ij}|^2\big)^{1/2} \prec \Psi$. Then $\sum_{i,j} a_{ij} X_i Y_j \prec \Psi$.

If all of the above random variables depend on an index u and the hypotheses of (i) – (iii) are uniform in u, then so are the conclusions.

The rest of this appendix is devoted to the proof of Theorem B.1. Our proof in fact generalizes trivially to arbitrary multilinear estimates for quantities of the form $\sum_{i_1,\ldots,i_k}^* a_{i_1\ldots i_k}(u)X_{i_1}(u)\cdots X_{i_k}(u)$, where the star indicates that the summation indices are constrained to be distinct.

We first recall the following version of the Marcinkiewicz-Zygmund inequality.

LEMMA B.2. *Let $X_1,\ldots,X_N$ be a family of independent random variables each satisfying* (B.1) *and suppose that the family (b_i) is deterministic. Then*

$$\left\|\sum_i b_i X_i\right\|_p \;\leqslant\; (Cp)^{1/2}\mu_p\Big(\sum_i |b_i|^2\Big)^{1/2} \tag{B.2}$$

PROOF. The proof is a simple application of Jensen's inequality. Writing $B^2 := \sum_j |b_i|^2$, we get, by the classical Marcinkiewicz-Zygmund inequality [39] in the first line, that

$$\begin{aligned}
\left\|\sum_i b_i X_i\right\|_p^p &\;\leqslant\; (Cp)^{p/2}\left\|\Big(\sum_i |b_i|^2|X_i|^2\Big)^{1/2}\right\|_p^p \\
&\;=\; (Cp)^{p/2}B^p\,\mathbb{E}\Big[\Big(\sum_i \frac{|b_i|^2}{B^2}|X_i|^2\Big)^{p/2}\Big] \\
&\;\leqslant\; (Cp)^{p/2}B^p\,\mathbb{E}\Big[\sum_i \frac{|b_i|^2}{B^2}|X_i|^p\Big] \\
&\;\leqslant\; (Cp)^{p/2}B^p\mu_p^p\,.
\end{aligned}$$

$\square$

Next, we prove the following intermediate result.

LEMMA B.3. *Let $X_1,\ldots,X_N,Y_1,\ldots,Y_N$ be independent random variables each satisfying* (B.1)*, and suppose that the family (a_{ij}) is deterministic. Then for all $p \geqslant 2$ we have*

$$\left\|\sum_{i,j} a_{ij}X_i Y_j\right\|_p \;\leqslant\; Cp\,\mu_p^2\Big(\sum_{i,j}|a_{ij}|^2\Big)^{1/2}\,.$$

PROOF. Write

$$\sum_{i,j} a_{ij}X_i Y_j \;=\; \sum_j b_j Y_j\,, \qquad b_j \;:=\; \sum_i a_{ij}X_i\,.$$

Note that (b_j) and (Y_j) are independent families. By conditioning on the family (b_j), we therefore get from Lemma B.2 and the triangle inequality that

$$\left\|\sum_j b_j Y_j\right\|_p \;\leqslant\; (Cp)^{1/2}\mu_p\left\|\sum_j |b_j|^2\right\|_{p/2}^{1/2} \;\leqslant\; (Cp)^{1/2}\mu_p\Big(\sum_j \|b_j\|_p^2\Big)^{1/2}\,.$$

Using Lemma B.2 again, we have

$$\|b_j\|_p \;\leqslant\; (Cp)^{1/2}\mu_p\Big(\sum_i |a_{ij}|^2\Big)^{1/2}\,.$$

This concludes the proof. $\square$

LEMMA B.4. *Let $X_1, \ldots, X_N$ be independent random variables each satisfying* (B.1), *and suppose that the family (a_{ij}) is deterministic. Then we have*

$$\left\| \sum_{i \neq j} a_{ij} X_i X_j \right\|_p \leqslant Cp \, \mu_p^2 \left(\sum_{i \neq j} |a_{ij}|^2 \right)^{1/2}.$$

PROOF. The proof relies on the identity (valid for $i \neq j$)

$$1 = \frac{1}{Z_N} \sum_{I \sqcup J = \mathbb{N}_N} \mathbf{1}(i \in I) \mathbf{1}(j \in J), \tag{B.3}$$

where the sum ranges over all partitions of $\mathbb{N}_N = \{1, \ldots, N\}$ into two sets I and J, and $Z_N := 2^{N-2}$ is independent of i and j. Moreover, we have

$$\sum_{I \sqcup J = \mathbb{N}_N} 1 = 2^N - 2, \tag{B.4}$$

where the sum ranges over nonempty subsets I and J. Now we may estimate

$$\left\| \sum_{i \neq j} a_{ij} X_i X_j \right\|_p \leqslant \frac{1}{Z_N} \sum_{I \sqcup J = \mathbb{N}_N} \left\| \sum_{i \in I} \sum_{j \in J} a_{ij} X_i X_j \right\|_p \leqslant \frac{1}{Z_N} \sum_{I \sqcup J = \mathbb{N}_N} Cp \, \mu_p^2 \left(\sum_{i \neq j} |a_{ij}|^2 \right)^{1/2},$$

where we used that, for any partition $I \sqcup J = \mathbb{N}_N$, the families $(X_i)_{i \in I}$ and $(X_j)_{j \in J}$ are independent, and hence the Lemma B.3 is applicable. The claim now follows from (B.4). $\qquad \square$

As remarked above, the proof of Lemma B.4 may be easily extended to multilinear expressions of the form $\sum_{i_1, \ldots, i_k}^* a_{i_1 \ldots i_k} X_{i_1} \cdots X_{i_k}$.

We may now complete the proof of Theorem B.1.

PROOF OF THEOREM B.1. The proof is a simple application of Chebyshev's inequality. Part (i) follows from Lemma B.2, part (ii) from Lemma B.4, and part (iii) from Lemma B.3. We give the details for part (iii).

For $\varepsilon > 0$ and $D > 0$ we have

$$\mathbb{P}\left[\left| \sum_{i \neq j} a_{ij} X_i X_j \right| \geqslant N^\varepsilon \Psi \right] \leqslant \mathbb{P}\left[\left| \sum_{i \neq j} a_{ij} X_i X_j \right| \geqslant N^\varepsilon \Psi, \, \left(\sum_{i \neq j} |a_{ij}|^2 \right)^{1/2} \leqslant N^{\varepsilon/2} \Psi \right]$$

$$+ \mathbb{P}\left[\left(\sum_{i \neq j} |a_{ij}|^2 \right)^{1/2} \geqslant N^{\varepsilon/2} \Psi \right]$$

$$\leqslant \mathbb{P}\left[\left| \sum_{i \neq j} a_{ij} X_i X_j \right| \geqslant N^{\varepsilon/2} \left(\sum_{i \neq j} |a_{ij}|^2 \right)^{1/2} \right] + N^{-D-1}$$

$$\leqslant \left(\frac{Cp\mu_p^2}{N^{\varepsilon/2}} \right)^p + N^{-D-1}$$

for arbitrary D. In the second step we used the definition of $\left(\sum_{i \neq j} |a_{ij}|^2 \right)^{1/2} \prec \Psi$ with parameters $\varepsilon/2$ and $D + 1$. In the last step we used Lemma B.4 by conditioning on (a_{ij}). Given ε and D, there is a large enough

p such that the first term on the last line is bounded by N^{-D-1}. Since ε and D were arbitrary, the proof is complete.

The claimed uniformity in u in the case that a_{ij} and X_i depend on an index u also follows from the above estimate. □

References

[1] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, *Scaling theory of localization: Absence of quantum diffusion in two dimensions*, Phys. Rev. Lett. **42** (1979), 673–676.

[2] M. Aizenman and S. Molchanov, *Localization at large disorder and at extreme energies: an elementary derivation*, Comm. Math. Phys. **157** (1993), 245–278.

[3] P. W. Anderson, *Absence of diffusion in certain random lattices*, Phys. Rev. **109** (1958), 1492.

[4] M. Disertori, H. Pinson, and T. Spencer, *Density of states for random band matrices*, Comm. Math. Phys. **232** (2002), 83–124.

[5] A. Elgart, *Lifshitz tails and localization in the three-dimensional Anderson model*, Duke Math. J. **146** (2009), no. 2, 331–360.

[6] L. Erdős and A. Knowles, *Quantum diffusion and delocalization for band matrices with general distribution*, Ann. H. Poincaré **12** (2011), 1227–1319.

[7] ———, *Quantum diffusion and eigenfunction delocalization in a random band matrix model*, Comm. Math. Phys. **303** (2011), 509–554.

[8] L. Erdős, A. Knowles, and H.-T. Yau, *Averaging fluctuations in resolvents of random band matrices*, to appear in Ann. H. Poincaré. Preprint arXiv:1205.5664.

[9] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős-Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues*, Comm. Math. Phys. **314** (2012), 587–640.

[10] ———, *The local semicircle law for a general class of random matrices*, Electron. J. Probab **18** (2013), 1–58.

[11] ———, *Spectral statistics of Erdős-Rényi graphs I: Local semicircle law*, Ann. Prob. **41** (2013), 2279–2375.

[12] L. Erdős, S. Péché, J.A. Ramirez, B. Schlein, and H.-T. Yau, *Bulk universality for Wigner matrices*, Comm. Pure Appl. Math. **63** (2010), 895–925.

[13] L. Erdős, J. Ramirez, B. Schlein, T. Tao, V. Vu, and H.-T. Yau, *Bulk universality for Wigner hermitian matrices with subexponential decay*, Math. Res. Lett. **17** (2010), 667–674.

[14] L. Erdős, J. Ramirez, B. Schlein, and H.-T. Yau, *Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation*, Electr. J. Prob. **15** (2010), 526–604.

[15] L. Erdős, M. Salmhofer, and H.-T. Yau, *Quantum diffusion for the Anderson model in the scaling limit*, Ann. H. Poincaré **8** (2007), 621–685.

[16] _____, *Quantum diffusion of the random Schrödinger evolution in the scaling limit II. the recollision diagrams*, Comm. Math. Phys. **271** (2007), 1–53.

[17] _____, *Quantum diffusion of the random Schrödinger evolution in the scaling limit*, Acta Math. **200** (2008), 211–277.

[18] L. Erdős, B. Schlein, and H.-T. Yau, *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. **287** (2009), 641–655.

[19] _____, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Ann. Prob. **37** (2009), 815–852.

[20] _____, *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Not. **2010** (2009), 436–479.

[21] _____, *Universality of random matrices and local relaxation flow*, Invent. Math. **185** (2011), no. 1, 75–119.

[22] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin, *The local relaxation flow approach to universality of the local statistics of random matrices*, Ann. Inst. Henri Poincaré (B) **48** (2012), 1–46.

[23] L. Erdős, H.-T. Yau, and J. Yin, *Universality for generalized Wigner matrices with Bernoulli distribution*, J. Combinatorics **1** (2011), no. 2, 15–85.

[24] _____, *Bulk universality for generalized Wigner matrices*, Prob. Theor. Rel. Fields **154** (2012), 341–407.

[25] _____, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math **229** (2012), 1435–1515.

[26] O. N. Feldheim and S. Sodin, *A universality result for the smallest eigenvalues of certain sample covariance matrices*, Geom. Funct. Anal. **20** (2010), 88–123.

[27] J. Fröhlich and T. Spencer, *Absence of diffusion in the Anderson tight binding model for large disorder or low energy*, Comm. Math. Phys. **88** (1983), 151–184.

[28] Y. V. Fyodorov and A. D. Mirlin, *Scaling properties of localization in random band matrices: a σ-model approach*, Phys. Rev. Lett. **67** (1991), 2405–2409.

[29] E. Lieb and M. Loss, *Analysis*, second ed., American Mathematical Society, 2001.

[30] 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** (1996), no. 1, 3221–3230.

[31] S. A. Molchanov, L. A. Pastur, and A. M. Khorunzhii, *Limiting eigenvalue distribution for band random matrices*, Theor. Math. Phys. **90** (1992), 108–118.

[32] N. S. Pillai and J. Yin, *Universality of covariance matrices*, Preprint arXiv:1110.2501.

[33] J. Schenker, *Eigenvector localization for random band matrices with power law band width*, Comm. Math. Phys. **290** (2009), 1065–1097.

[34] S. Sodin, *The spectral edge of some random band matrices*, Ann. Math. **172** (2010), no. 3, 2223–2251.

[35] _____, *An estimate for the average spectral measure of random band matrices*, J. Stat. Phys. **144** (2011), no. 1, 46–59.

[36] T. Spencer, *SUSY statistical mechanics and random band matrices*, lecture notes.

[37] _____, *Lifshitz tails and localization*, Preprint (1993).

[38] _____, *Random banded and sparse matrices (Chapter 23)*, "Oxford Handbook of Random Matrix Theory" edited by G. Akemann, J. Baik, and P. Di Francesco (2011).

[39] D. Stroock, *Probability theory, and analytic view*, Cambridge University Press, 1999.

[40] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. Math. **62** (1955), 548–564.