Fixed energy universality for generalized Wigner matrices

P. Bourgade

Cambridge University
bourgade@math.ias.edu

L. Erdős

Institute of Science and Technology Austria
lerdos@ist.ac.at

H.-T. Yau

Harvard University
htyau@math.harvard.edu

J. Yin

University of Wisconsin, Madison
jyin@math.wisc.edu

We prove the Wigner-Dyson-Mehta conjecture at fixed energy in the bulk of the spectrum for generalized symmetric and Hermitian Wigner matrices. Previous results concerning the universality of random matrices either require an averaging in the energy parameter or they hold only for Hermitian matrices if the energy parameter is fixed. We develop a homogenization theory of the Dyson Brownian motion and show that microscopic universality follows from mesoscopic statistics.

Keywords: Universality, Homogenization, Dyson Brownian motion.

The work of P. B. is partially supported by the NSF grant DMS-1208859. The work of L. E. is partially supported by ERC Advanced Grant, RANMAT 338804. The work of H.-T. Y. is partially supported by the NSF grant DMS-1307444 and the Simons investigator fellowship. The work of J. Y. is partially supported by NSF Grant DMS-1207961. The major part of this research was conducted when all authors were visiting IAS and were also supported by the NSF Grant DMS-1128255.

1 Introduction

E. Wigner discovered that energy levels of large quantum systems exhibit remarkably simple universality patterns. He introduced a fundamental model, the Wigner matrix ensemble, and postulated that the statistics of the eigenvalue gaps, i.e. differences of consecutive eigenvalues, depend only on the symmetry class of the model and are independent of the details of the ensemble. Although the central universal objects in Wigner's original work were the eigenvalue gap distributions, the subsequent developments showed that the correlation functions play a key role. In fact, a few years after Wigner's pioneering work, Gaudin, Mehta and Dyson computed explicitly the eigenvalue correlation functions for the Gaussian cases and expressed the eigenvalue gap distributions in terms of them. Later on, Mehta formalized a version of the (Wigner-Dyson-Mehta) universality conjecture in his seminal book [28] by stating that the appropriately rescaled correlation functions for any Wigner ensemble coincide with those for the Gaussian cases as N, the size of the matrix, tends to infinity. This holds for both real symmetric and complex Hermitian ensembles (Conjectures 1.2.1 and 1.2.2 in [28]). The topology of the convergence, however, was not specified explicitly.

One possible topology for the correlation functions is the pointwise convergence. But the convergence in this topology cannot hold for Wigner ensembles with discrete (e.g. Bernoulli) matrix elements so it could only be used for a certain subclass of Wigner matrices. Thus a reasonably strong topology suitable for the universality of the whole class of Wigner matrices is the vague convergence of the local correlation functions, rescaled around a fixed energy E; in short we will call it *fixed energy universality* (see Section 2 for the precise definitions). Certainly, instead of fixing the energy E, we can also take weak convergence in E or equivalently, taking some average in the energy. We will call universality in this weaker topology *averaged energy universality.* Finally, one can go back to Wigner's original point of view and ask for universality of the gap distributions.

The Wigner-Dyson-Mehta conjecture has been widely open until the recent work [10] where a general scheme to approach it was outlined and carried out for complex Hermitian matrices. The basic idea is first to establish a local version of the semicircle law and use it as an input to control the correlation function asymptotics in the Brezin-Hikami formula (which is related to Harish-Chandra/Itzykson-Zuber formula). This provides universality for the so-called Gaussian divisible models with a very small Gaussian component, or "noise" (previously, this universality was established by Johansson [25] when the noise is of order one). The last step is an approximation of a general Wigner ensemble by Gaussian divisible ones and this leads to the fixed energy universality for Hermitian Wigner matrices whose matrix elements have smooth distributions. The various restrictions on the laws of matrix elements were greatly relaxed in subsequent works [8, 11, 16, 32]. In particular, using the local semicircle law [12] as a main input, Tao-Vu [32] proved a comparison theorem which provides an approximation result for Wigner matrices satisfying a four moment matching condition. Finally, the conditions for tail distributions of the matrix elements were greatly relaxed in [8, 16, 33]. For a concise review on the recent progress on the universality for random matrices, see, e.g., [17].

For real symmetric matrices, no algebraic formula in the spirit of Brezin-Hikami is known. A completely new method based on relaxation of the Dyson Brownian Motion (DBM) to local equilibrium was developed in a series of papers [14, 15, 20]. This approach is very robust and applies to all symmetry classes of random matrices, including also sample covariance matrices and sparse matrices [8, 15], but it yields only the *average energy universality.* Although the energy averaging is on a very small scale, it so far cannot be completely removed with this method.

We now comment on a parallel development for the universality of the eigenvalue gaps which was Wigner's original interest. Correlation functions at a fixed energy E carry full information about the distribution of the eigenvalues near E. In particular, the Fredholm determinant and the Jimbo-Miwa-Mori-Sato formulae yield the probability that no eigenvalues appear in a neighborhood around E. The universality of the distribution of the gap with a *fixed label* (which we will call gap universality), e.g. the difference between, say, the $N/2$-th and $(N/2-1)$-th eigenvalues, however cannot be deduced rigorously from the fixed energy universality. Conversely, the gap universality does not imply the fixed energy universality either. The reason is that eigenvalues with a fixed label fluctuate on a scale larger than the mean eigenvalue spacing, so fixed energy and fixed label universalities are not equivalent. The gap universality was established in [18] via a

De Giorgi-Nash-Moser type Hölder regularity result for a discrete parabolic equation with time dependent random coefficients $B_{ij}(t) = (x_i(t) - x_j(t))^{-2}$ where $\mathbf{x}(t)$ is the DBM trajectory. The gap universality for the special case of Hermitian matrices satisfying the four moment matching condition was proved earlier in [31].

To summarize, the Wigner-Dyson-Mehta conjecture was completely resolved in the sense of averaged energy and fixed label gap universalities for both symmetric and Hermitian ensembles. In the sense of fixed energy universality, it was proved for the Hermitian matrices, but not for real symmetric ones. In the current paper, we settle this last remaining case of the Wigner-Dyson-Mehta conjecture by proving the universality of local correlation functions at any fixed energy E in the bulk spectrum for generalized Wigner matrices of any symmetry classes. Our theorem in particular implies the following three new results for real symmetric matrices (including the Bernoulli cases): (1) existence of the density of states on microscopic scales for generalized Wigner matrices, (2) the extension of the Jimbo-Miwa-Mori-Sato formula of the gap probability to generalized real Wigner matrices, (3) the precise distribution of the condition number or the smallest (in absolute value) eigenvalue of generalized Wigner matrices. Our proof also applies to the third symmetry class, the symplectic matrices, but we will focus on the real symmetric case as this is the most complicated case from the technical point of view.

The essence of the current work is a homogenization theory for the discrete parabolic equation with time dependent random coefficients $B_{ij}(t) = (x_i(t) - x_j(t))^{-2}$. By a rigidity property of the DBM trajectories, the random coefficients are close to deterministic ones, $B_{ij}(t) \approx (\gamma_i - \gamma_j)^{-2}$ if $|i - j| \gg 1$ (the typical locations γ_i are defined in (2.5)). The continuous version of the corresponding heat kernel is explicitly known; in fact locally it is given by $e^{-t|p|}(i, j)$ where $|p| = \sqrt{-\Delta}$. By coupling two DBM for two different initial conditions $\mathbf{x}(0)$ and $\mathbf{y}(0)$ (one for Wigner, one for a reference Gaussian ensemble), we show that after a sufficiently long time, the difference between $x_i(t)$ and $y_i(t)$ is given by the deterministic heat kernel acting on the difference of the initial data. Due to the scaling properties of the explicit heat kernel, this latter involves only mesoscopic linear statistics of the initial conditions which are more accessible than microscopic ones. Homogenization thus enables us to transfer mesoscopic statistics to microscopic ones. The main steps of the proof will be described in the next section in more details.

Convention. For two N-dependent positive quantities $a = a_N$, $b = b_N$ we say that a and b are comparable, $a \sim b$, if there exists a constant $C > 0$, independent of N, such that $C^{-1} \leqslant a/b \leqslant C$.

2 Main result and sketch of the proof

2.1 The model and the result. We consider the following class of random matrices.

Definition 2.1. *A generalized Wigner matrix H_N is a Hermitian or symmetric $N \times N$ matrix whose upper-triangular elements $h_{ij} = \overline{h_{ji}}$, $i \leqslant j$, are independent random variables with mean zero and variances $\sigma_{ij}^2 = \mathbb{E}(|h_{ij}|^2)$ that satisfy the following two conditions:*

(i) Normalization: for any $j \in [\![1, N]\!]$, $\sum_{i=1}^N \sigma_{ij}^2 = 1$.

(ii) Non-degeneracy: $\sigma_{ij}^2 \sim N^{-1}$ for all $i, j \in [\![1, N]\!]$.

In the Hermitian case, we furthermore assume that $\operatorname{Var}\Re(h_{ij}) \sim \operatorname{Var}\Im(h_{ij})$ for $i \neq j$ and that one of the following holds: (1) $\Re(h_{ij}), \Im(h_{ij})$ are independent, or (2) the law of h_{ij} is isotropic, i.e. $|h_{ij}|$ is independent of $\arg h_{ij}$, which is uniform on $(0, 2\pi)$.

We additionally assume that there exists $p > 0$ large but fixed such that

$$\sup_{i,j,N} \mathbb{E}\left((\sqrt{N}|h_{ij}|)^p \right) < \infty. \tag{2.1}$$

For example $p = 10$ is sufficient for our purpose, in this work we will not try to get the lowest possible exponent p, for the clarity of exposition. We denote by

$$x_1 \leqslant \ldots \leqslant x_N$$

the N random eigenvalues of a generalized Wigner matrix H_N. Let $\mu^{(N)}(\mathbf{u})$ be the associated probability distribution of the spectrum, where $\mathbf{u} = (u_1, \dots, u_N)$ is an element of the simplex $\Sigma = \{\mathbf{u} \,:\, u_1 \leqslant \dots \leqslant u_N\} \subset \mathbb{R}^N$. The universal limiting point process for random spectra will be uniquely characterized by the limits of the k-point correlation functions for $k = 1, 2, \dots$ as $N \to \infty$. These are defined[1] by

$$\rho_k^{(N)}(u_1, \dots, u_k) = \int_{\mathbb{R}^{N-k}} \widetilde{\mu}^{(N)}(\mathbf{u}) \mathrm{d}u_{k+1} \dots \mathrm{d}u_N, \tag{2.2}$$

where $\widetilde{\mu}^{(N)}$ is the symmetrized version of $\mu^{(N)}$, defined on $\mathbb{R}^N$ instead of the simplex: $\widetilde{\mu}^{(N)}(\mathbf{u}) = \frac{1}{N!}\mu^{(N)}(\mathbf{u}^{(\sigma)})$ where $\mathbf{u}^{(\sigma)} = (u_{\sigma(1)}, \dots, u_{\sigma(N)})$ with $u_{\sigma(1)} \leqslant \dots \leqslant u_{\sigma(N)}$. The limiting density ($k = 1$ point correlation function) of the eigenvalues is the Wigner semicircle law and it will be denoted

$$\mathrm{d}\varrho(x) = \varrho(x)\mathrm{d}x = \frac{1}{2\pi}\sqrt{(4 - x^2)_+}\mathrm{d}x.$$

In the fundamental particular case where H_N is a real symmetric matrix from the Gaussian Orthogonal Ensemble (GOE), the correlation functions are known to converge on microscopic scales, the limit being expressible as a determinant [6, 7, 28, 29]: for any $\mathbf{v} \in \mathbb{R}^k$ and $E \in (-2, 2)$, we have

$$\frac{1}{\varrho(E)^k}\rho_k^{(N,\mathrm{GOE})}\left(E + \frac{\mathbf{v}}{N\varrho(E)}\right) = \frac{1}{\varrho(E)^k}\rho_k^{(N)}\left(E + \frac{v_1}{N\varrho(E)}, \dots, E + \frac{v_k}{N\varrho(E)}\right) \to \rho_k^{(\mathrm{GOE})}(\mathbf{v}), \tag{2.3}$$

where this limit is independent of $E \in (-2, 2)$. For complex Hermitian matrices from the Gaussian Unitary Ensemble (GUE), the same statement holds with a different limit $\rho_k^{(\mathrm{GUE})}(\mathbf{v})$.

Bulk universality for generalized Wigner matrices was considered for various convergence types, notably the two following ones. We state them only in the symmetric case, the Hermitian setting being similar.

Fixed energy universality (in the bulk). For any $k \geqslant 1$, $F : \mathbb{R}^k \to \mathbb{R}$ continuous and compactly supported and for any $\kappa > 0$, we have, uniformly in $E \in [-2 + \kappa, 2 - \kappa]$,

$$\lim_{N\to\infty} \frac{1}{\varrho(E)^k} \int \mathrm{d}\mathbf{v} F(\mathbf{v}) \rho_k^{(N)}\left(E + \frac{\mathbf{v}}{N\varrho(E)}\right) = \int \mathrm{d}\mathbf{v} F(\mathbf{v}) \rho_k^{(\mathrm{GOE})}(\mathbf{v}). \tag{2.4}$$

Averaged energy universality (in the bulk). For any $k \geqslant 1$, $F : \mathbb{R}^k \to \mathbb{R}$ continuous and compactly supported, and for any $\varepsilon, \kappa > 0$, we have, uniformly in $E \in [-2 + \kappa, 2 - \kappa]$,

$$\lim_{N\to\infty} \frac{1}{\varrho(E)^k} \int_E^{E+s} \frac{\mathrm{d}x}{s} \int \mathrm{d}\mathbf{v} F(\mathbf{v}) \rho_k^{(N)}\left(x + \frac{\mathbf{v}}{N\varrho(E)}\right) \mathrm{d}\mathbf{v} = \int \mathrm{d}\mathbf{v} F(\mathbf{v}) \rho_k^{(\mathrm{GOE})}(\mathbf{v}), \qquad s := N^{-1+\varepsilon}.$$

Fixed energy universality obviously implies averaged energy universality. As mentioned in the introduction, fixed energy universality was proved for Hermitian matrices from the generalized Wigner ensemble. This required the use of the Brézin-Hikami-Johansson formula, a tool with no known analogue for symmetric matrices. General methods developed in the past five years, such as the local relaxation flow, allowed to prove universality for the symmetric class only in the sense of averaged energy. Our result establishes universality at fixed energy, with no need for any averaging, for the symmetric class. It also provides a new proof for the Hermitian class.

Theorem 2.2 (Universality at fixed energy)**.** *For symmetric or Hermitian matrices from the generalized Wigner ensemble satisfying (2.1), fixed energy universality holds in the bulk of the spectrum.*

The above theorem implies for example that the joint interval probabilities converge, more precisely, for disjoint intervals $I_1, \dots, I_\ell$, and integers $n_1, \dots, n_\ell \in \mathbb{N}$, the limit

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\{x_i \in E + \frac{I_j}{N\varrho(E)}\}\right| = n_j, 1 \leqslant j \leqslant \ell\right)$$

[1] Note that, strictly speaking, (2.2) only makes sense when $\mu^{(N)}$ has a density; in the general case, it needs to be interpreted in the distribution sense, i.e. after integrating with respect to a regular test function, as understood on the LHS of (2.4).

exists. It is independent of $E \in (-2, 2)$ and of the details of the distributions of the matrix entries, in particular they can be computed in the Gaussian ensemble where more explicit formulas are available. For example, the gap probability for Bernoulli random matrices converges on the microscopic scale:

$$\lim_{N\to\infty} \mathbb{P}\left(\{x_i \in E + \frac{[0,t]}{N\pi\varrho(E)}\} = \varnothing\right) = E_1(0,t),$$

where E_1 can be made explicit from the solution to a Painlevé equation of fifth type [23, 34].

Before giving the main ideas of the proof, we introduce the typical locations of eigenvalues with respect to the semicircular distribution: they are defined by

$$\int_{-2}^{\gamma_k} \mathrm{d}\varrho = \frac{k+\frac{1}{2}}{N}, \qquad k = 1, 2, \dots, N. \tag{2.5}$$

2.2 Sketch of the proof. We now outline the main steps towards the proof of Theorem 2.2, in the symmetric case. As mentioned in the introduction, it does not rely on improvements of existing methods such as the local relaxation flow. The Dyson Brownian motion plays again a key role in the following approach, but surprisingly our method requires understanding its behaviour for relatively large time, $t = N^{-\tau}$, for some small τ, instead of $t = N^{-1+\varepsilon}$ for small ε.

First step. Coupling and discrete integral operator. We run a coupled DBM with two different initial conditions, one from the Wigner ensemble we wish to study and one from a comparison Gaussian ensemble. At time 0, let $\mathbf{x} = \mathbf{x}(0)$ be the ordered spectrum of a generalized Wigner matrix, and let $\mathbf{y}(0)$ denote the eigenvalues of an independent GOE matrix. In the actual proof we have to start the coupling at a time $t_0 \sim N^{-\tau_0}$, $\tau_0 > \tau$ instead of time 0, but we neglect this technical issue in the current presentation. Consider the unique strong solutions for the following Dyson Brownian motion, more precisely its Ornstein-Uhlenbeck version:

$$\mathrm{d}x_\ell(t) = \sqrt{\frac{2}{N}}\mathrm{d}B_\ell(t) + \left(\frac{1}{N}\sum_{k\neq\ell}\frac{1}{x_\ell(t)-x_k(t)} - \frac{1}{2}x_\ell(t)\right)\mathrm{d}t, \tag{2.6}$$

$$\mathrm{d}y_\ell(t) = \sqrt{\frac{2}{N}}\mathrm{d}B_\ell(t) + \left(\frac{1}{N}\sum_{k\neq\ell}\frac{1}{y_\ell(t)-y_k(t)} - \frac{1}{2}y_\ell(t)\right)\mathrm{d}t.$$

Note that the underlying Brownian trajectories $(B_\ell)_{1\leqslant\ell\leqslant N}$ are the same. Then the normalized differences $\delta_\ell(t) := e^{t/2}(x_\ell(t) - y_\ell(t))$ satisfy an integral equation of parabolic type, namely

$$\partial_t \delta_\ell(t) = \sum_{k\neq\ell} b_{k\ell}(t)(\delta_k(t) - \delta_\ell(t)), \quad b_{k\ell}(t) = \frac{1}{N(x_\ell(t)-x_k(t))(y_\ell(t)-y_k(t))}. \tag{2.7}$$

Second step. Homogenization. We consider the following continuous analogue of (2.7):

$$\partial_t f_t = -K f_t, \qquad (Kf)(x) := \int_{-2}^{2} \frac{f(x)-f(y)}{(x-y)^2}\varrho(y)\mathrm{d}y. \tag{2.8}$$

A key step in our approach consists in proving that (2.8) gives a good approximation for (2.7). Indeed, if the initial conditions match in the sense that f_0 is smooth enough and $f_0(\gamma_k) = \delta_k(0)$, then for any $t = N^{-\tau}$, with a sufficiently small $\tau > 0$ there exists $\varepsilon > 0$ such that for any bulk index ℓ (i.e. $\ell \in [\![\alpha N, (1-\alpha)N]\!]$ for some small fixed $\alpha > 0$) we have

$$\delta_\ell(t) = (e^{-tK} f_0)_\ell + \mathrm{O}(N^{-1-\varepsilon}). \tag{2.9}$$

The above homogenization result holds for generic trajectories $\mathbf{x}(t)$, $\mathbf{y}(t)$. It relies on optimal rigidity estimates for these eigenvalues [21], a level repulsion bound similar to [13], and Hölder continuity for equations of type (2.7), obtained in [17].

Third step. The continuous heat kernel. The heat kernel for the equation (2.8) can be expressed by an explicit formula, see (3.22). For short times, i.e. τ close to 1 it almost coincides with $e^{-t|p|}$ where $|p| = \sqrt{-\Delta}$.

However, we will need τ close to 0, hence the effect of the curvature from the semicircle law cannot be neglected, and the explicit formula will be useful. This allows us to compute explicitly $e^{-tK}f_0$ and to rewrite (2.9) as follows. There exists $\varepsilon > 0$ such that, for a fixed E in the bulk and for any ℓ satisfying $|\gamma_\ell - E| < N^{-1+\varepsilon}$, we have

$$x_\ell(t) - y_\ell(t) = \widetilde{\zeta}_t^{\mathbf{x}} - \widetilde{\zeta}_t^{\mathbf{y}} + \mathrm{O}(N^{-1-\varepsilon}), \tag{2.10}$$

where

$$\widetilde{\zeta}_t^{\mathbf{x}} := \frac{1}{N}\sum_{k=1}^{N}\left(P_t(x_k(0)) - P_t(\gamma_k)\right), \tag{2.11}$$

is a smooth linear statistics of $\mathbf{x} = \mathbf{x}(0)$ on the *mesoscopic* scale $t = N^{-\tau} \ll 1$. Here P_t is an explicit function, the antiderivative of the heat kernel (3.22) (see (4.57)). We repeat the above steps with the initial condition $\mathbf{x}$ replaced by $\mathbf{z}$, the spectrum of another GOE independent of $\mathbf{x}$ and $\mathbf{y}$. In summary, we proved

$$x_\ell(t) = y_\ell(t) - \widetilde{\zeta}_t^{\mathbf{y}} + \widetilde{\zeta}_t^{\mathbf{x}} + \mathrm{O}(N^{-1+\varepsilon}), \quad z_\ell(t) = y_\ell(t) - \widetilde{\zeta}_t^{\mathbf{y}} + \widetilde{\zeta}_t^{\mathbf{z}} + \mathrm{O}(N^{-1+\varepsilon}). \tag{2.12}$$

From a technical perspective, the existence of a closed form for the kernel of K is essential in the above second and third steps.

Fourth step: Reformulation of universality through mesoscopic observables. For any continuous and compactly supported test function $Q : \mathbb{R}^k \to \mathbb{R}$ and $E \in (-2, 2)$, define

$$\mathcal{Q}(\mathbf{x}, E) := \sum_{i_1,\dots,i_k=1}^{N} Q(N(x_{i_1} - E), N(x_{i_2} - x_{i_1}), \dots, N(x_{i_k} - x_{i_1})).$$

Theorem 2.2 can be restated as

$$\mathbb{E}\mathcal{Q}(\mathbf{x}(0), E) = \mathbb{E}\mathcal{Q}(\mathbf{z}(0), E) + \mathrm{o}(1). \tag{2.13}$$

Let $\widehat{Q}$ denote the Fourier transform in the first variable. By a standard approximation argument, it is sufficient to prove (2.13) for any Q such that $\widehat{Q}$ is compactly supported, in $[-m, m]$, say. We will first prove that (2.13) holds for the corresponding DBM trajectories after some time $t = N^{-\tau}$, where τ will depend on m:

$$\mathbb{E}\mathcal{Q}(\mathbf{x}(t), E) = \mathbb{E}\mathcal{Q}(\mathbf{z}(t), E) + \mathrm{o}(1). \tag{2.14}$$

Using the representation (2.12), we easily see that (2.14) holds if we have

$$\mathbb{E}\mathcal{Q}(\mathbf{y}(t) - \widetilde{\zeta}_t^{\mathbf{y}}, E - \widetilde{\zeta}_t^{\mathbf{x}}) = \mathbb{E}\mathcal{Q}(\mathbf{y}(t) - \widetilde{\zeta}_t^{\mathbf{y}}, E - \widetilde{\zeta}_t^{\mathbf{z}}) + \mathrm{o}(1). \tag{2.15}$$

Note that $\mathbf{y}(t)$ and $\widetilde{\zeta}_t^{\mathbf{y}}$ are independent from $\widetilde{\zeta}_t^{\mathbf{x}}$ and $\widetilde{\zeta}_t^{\mathbf{z}}$. Moreover, (2.15) is simpler than the original microscopic universality problem, because $\widetilde{\zeta}_t^{\mathbf{x}}$ and $\widetilde{\zeta}_t^{\mathbf{z}}$ are mesoscopic observables. In the next two steps, we explain how (2.15) can be proved under the following, strange, compatibility assumption between the dynamics time and the Fourier support of the test function:

$$\tau \leqslant \frac{c}{m^2}. \tag{2.16}$$

Fifth step. Mesoscopic fluctuations for the Gaussian orthogonal ensemble. To justify (2.15), we first prove that the distribution of $\zeta_t^{\mathbf{z}} := N\widetilde{\zeta}_t^{\mathbf{z}}$ is very close to a Gaussian, with variance of order $\tau \log N$. This central limit theorem for the linear statistics of type (2.11) relies on the method initiated in [24]. For reasons apparent in the next step, we will need to control the the distribution of $\zeta_t^{\mathbf{z}}$ even beyond its natural scale $(\tau \log N)^{1/2}$. More precisely, we will prove that, for some fixed constants b and $c > 0$, we have

$$\widehat{\mu}_{\mathbf{z}}(\lambda) = \widehat{\mu}_{\mathbf{z},t}(\lambda) = \mathbb{E}\left(e^{-\mathrm{i}\lambda\zeta_t^{\mathbf{z}}}\right) = e^{-\frac{\lambda^2}{2}\tau \log N - \mathrm{i}\lambda b} + \mathrm{O}(N^{-c}). \tag{2.17}$$

For *macroscopic* linear statistics, corresponding to t independent of N in our notation, Johansson proved the central limit theorem in [24] by considering the logarithm of their Laplace transform. The proof of (2.17) involves additional technicalities because $\widehat{\mu}_{\mathbf{z}}$ may vanish, see Section 5. In particular we will need rigidity estimates from [4] to prove (2.17).

Sixth step. Reverse heat flow. For any fixed $a, h \in \mathbb{R}$, consider the functions

$$F(a) := \mathbb{E}\mathcal{Q}(\mathbf{y}(t) - \widetilde{\zeta}_t^{\mathbf{y}}, E - a), \quad F_h(a) := F(a-h) - F(a).$$

We can express the convolution of F_h with $\mu_{\mathbf{z}}$ as follows:

$$\begin{aligned}(F_h * \mu_{\mathbf{z}})(a) &= \mathbb{E}\mathcal{Q}(\mathbf{y}(t) - \widetilde{\zeta}_t^{\mathbf{y}}, E - a + h - \widetilde{\zeta}_t^{\mathbf{z}}) - \mathbb{E}\mathcal{Q}(\mathbf{y}(t) - \widetilde{\zeta}_t^{\mathbf{y}}, E - a - \widetilde{\zeta}_t^{\mathbf{z}}) \\ &= \mathbb{E}\mathcal{Q}(\mathbf{z}(t), E - a + h) - \mathbb{E}\mathcal{Q}(\mathbf{z}(t), E - a) + \mathrm{O}(N^{-c}) = \mathrm{O}(N^{-c}).\end{aligned} \tag{2.18}$$

for some $c > 0$. Here the second equality follows from the second formula in (2.12). The last step in (2.18) uses that on microscopic scales and with a high accuracy the distribution of the GOE spectrum is translation invariant; a fact that follows from an effective polynomial speed of the the convergence (2.3) uniformly in E in the bulk.

From the estimate (2.18) on $F_h * \mu_{\mathbf{z}}$ we bound F_h. This is a reverse heat flow type of question, because $\mu_{\mathbf{z}}$ is almost a Gaussian distribution, from the previous step. One can reverse the heat flow because (1) F_h is analytic, explaining our original Fourier support restriction on Q, and (2) the estimate (2.17) is precise enough. Namely, taking the Fourier transform in (2.18), we obtain

$$\widehat{F}_h(\lambda) = \mathrm{O}\left(N^{-c}\widehat{\mu}_{\mathbf{z}}(\lambda)^{-1}\right) = \mathrm{O}(N^{m^2\tau - c})$$

for any λ in the Fourier support of Q, where we used (2.17). This explains why we need (2.16) in order to prove that $\widehat{F}_h$, and then F_h, are o(1). We therefore obtained

$$\mathbb{E}\mathcal{Q}(\mathbf{y}(t) - \widetilde{\zeta}_t^{\mathbf{y}}, E - a + h) = \mathbb{E}\mathcal{Q}(\mathbf{y}(t) - \widetilde{\zeta}_t^{\mathbf{y}}, E - a) + \mathrm{o}(1)$$

uniformly in a, h. It is then elementary, by simple convolution, to prove (2.15) and therefore (2.14).

Seventh step. Green function comparison theorem. Finally, to obtain universality for the eigenvalues $\mathbf{x}(0)$ of the initial Wigner ensemble from the time evolved ones $\mathbf{x}(t)$, $t = N^{-\tau}$, we use a Green function comparison theorem, of type close to the one introduced in [20]: (2.14) implies (2.13).

3 Homogenization

From this section, we only consider symmetric matrices which is the most involved case since the level repulsion estimate requires an additional regularization. Section 3.1 would not be necessary for $\beta > 1$, the rest of the proof is insensitive to the value of β.

3.1 Regularized dynamics. Our goal is to estimate the coupled difference $\delta_\ell(t) = e^{t/2}(x_\ell(t) - y_\ell(t))$, which satisfies the dynamics (2.7). Notice that the singularity of the coefficient b_{jk} is not integrable and this will create serious difficulties in the analysis. We first perform a cutoff to tame this singularity which requires a level repulsion estimate. Since such estimate holds only for large enough time, we will perform the regularization only after an initial time $t_0 := N^{-\tau_0}/2$ with some $\tau_0 > 0$. We then show that the difference between the original and cutoff dynamics is negligible for times $t \in (t_0, 1)$. The estimates in this section are valid for any fixed τ_0. We assume $\tau_0 \leqslant 1$ which is the relevant regime.

We recall the equation for the $\mathbf{x}(t)$ dynamics from (2.6) and for $t \in (t_0, 1)$, we define its regularized version as

$$\mathrm{d}\widehat{x}_j(t) = \sqrt{\frac{2}{N}}\mathrm{d}B_j(t) + \left(\frac{1}{N}\sum_{k \neq j} \frac{1}{x_j(t) - x_k(t) + \varepsilon_{jk}} - \frac{1}{2}\widehat{x}_j(t)\right)\mathrm{d}t \tag{3.1}$$

with $\varepsilon_{jk} = \varepsilon$ for $j > k$, $\varepsilon_{jk} := -\varepsilon$ for $j < k$, and we set $\widehat{x}_i(s) := x_i(s)$, for $s \leqslant t_0$. Notice that $\widehat{\mathbf{x}}(t)$ may not preserve the ordering, but we do not need this property. Let $q_i := N(x_i - \widehat{x}_i)$ denote the rescaled difference between the original dynamics and the regularized one. It satisfies the equation ($t > t_0$)

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \Omega_i - \frac{q_i}{2}, \qquad \text{with} \quad \Omega_i(t) := \sum_{j \neq i} \frac{\varepsilon_{ij}}{(x_i - x_j)(t)((x_i - x_j)(t) + \varepsilon_{ij})}. \tag{3.2}$$

Since $q_i(t_0) = 0$, we can solve this equation by

$$q_i(t) = \int_{t_0}^{t} e^{-(t-s)/2}\Omega_i(s)\mathrm{d}s, \qquad t \geqslant t_0. \tag{3.3}$$

Let $p > 2$ and p' be its conjugate exponent. We have

$$\mathbb{E}\sup_{t_0 \leqslant t \leqslant 1}\left|\int_{t_0}^{t}\Omega_i(s)\mathrm{d}s\right| \leqslant \sup_{t_0 \leqslant t \leqslant 1}\mathbb{E}|\Omega_i(t)| \leqslant \sup_{t_0 \leqslant t \leqslant 1}\sum_{j \neq i}\left(\mathbb{E}\frac{1}{|x_i - x_j|^{p'}(t)}\right)^{1/p'}\left(\mathbb{E}\frac{\varepsilon^p}{\left|(x_i - x_j)(t) + \varepsilon_{ij}\right|^p}\right)^{1/p}. \tag{3.4}$$

Recall the rigidity estimate from [21] asserting that for any $\xi, D > 0$, if for some large moment $p = p(\xi, D)$ (2.1) is satisfied, then there exists $C > 0$ such that

$$\mathbb{P}(\mathcal{G}_{\xi,\mathbf{x},i}) \geqslant 1 - CN^{-D}, \quad \text{with} \quad \mathcal{G}_{\xi,\mathbf{x},i} := \{|x_i(t) - \gamma_i| \leqslant N^{-2/3+\xi}(\widehat{i})^{-1/3}, 0 \leqslant t \leqslant 1\}, \tag{3.5}$$

where $\widehat{i} := \min(i, N + 1 - i)$ (the subscript $\mathbf{x}$ refers to the $\mathbf{x}(t)$ process). The original rigidity estimate in [21] was formulated for any fixed generalized Wigner ensemble, i.e. for a fixed t. A minor continuity argument in the time variable ensures that rigidity holds simultaneously for all times in a compact interval (see Lemma 9.3 of [18] for a similar argument). Note also that the original rigidity estimate from [21] assumes subexponential decay of the entries distribution, but this is easily weakened to the finite moment assumption (2.1) (see remark 2.4 in [9]).

Denote by $g_i(t) := x_{i+1}(t) - x_i(t)$ the gap at i-th location. A trivial estimate yields that

$$\left(\mathbb{E}\mathcal{G}_{\xi,\mathbf{x},i}\frac{\varepsilon^p}{(g_i(t) + \varepsilon)^p}\right)^{1/p} \leqslant \left(\mathbb{E}\mathcal{G}_{\xi,\mathbf{x},i}\frac{\varepsilon^2}{(g_i(t) + \varepsilon)^2}\right)^{1/p}$$

(with a slight abuse of notations we write $\mathcal{G}_{\xi,\mathbf{x},i}$ instead of its characteristic function within the expectation). Using the level repulsion estimate, i.e., Corollary B.2, we have for any $\xi > 0$ that

$$\left(\mathbb{E}\mathcal{G}_{\xi,\mathbf{x},i}\frac{\varepsilon^2}{(g_i(t) + \varepsilon)^2}\right)^{1/p} \leqslant C\varepsilon^{2/p}N^{2/p}N^{(C_1\tau+\xi)/p}|\log\varepsilon|^{1/p}, \qquad t = N^{-\tau} \in [t_0, 1] \tag{3.6}$$

where C_1 is the constant from Corollary B.2. We introduced the notation $t = N^{-\tau}$ and we will use t and τ in parallel, similarly to the notation $t_0 = N^{-\tau_0}/2$. The other factor in (3.4) is even easier to estimate and it gives

$$\left(\mathbb{E}\frac{1}{|x_i - x_j|^{p'}(t)}\right)^{1/p'} \leqslant N^{1/p'}N^{(C_0\tau+\xi)/p'}.$$

Choosing for example $\varepsilon = N^{-3C_0-100}$, $2 < p < 3$, and ξ small, we therefore proved that

$$\mathbb{P}(\mathcal{G}_{\mathbf{x}}) \geqslant 1 - CN^{-2}, \quad \text{with} \quad \mathcal{G}_{\mathbf{x}} := \{\sup_{t_0 \leqslant t \leqslant 1, i \in [\![1,N]\!]} |x_i - \widehat{x}_i|(t) \leqslant N^{-2C_0-50}\}. \tag{3.7}$$

Hence the trajectories of $\widehat{\mathbf{x}}$ and $\mathbf{x}$ are very close to each other.

We also regularize the $\mathbf{y}(t)$ dynamics, i.e. we have

$$\mathrm{d}y_j(t) = \sqrt{\frac{2}{N}}\mathrm{d}B_j(t) + \left(\frac{1}{N}\sum_{k\neq j}\frac{1}{y_j(t)-y_k(t)} - \frac{1}{2}y_j(t)\right)\mathrm{d}t, \tag{3.8}$$
$$\mathrm{d}\widehat{y}_j(t) = \sqrt{\frac{2}{N}}\mathrm{d}B_j(t) + \left(\frac{1}{N}\sum_{k\neq j}\frac{1}{y_j(t)-y_k(t)+\varepsilon_{jk}} - \frac{1}{2}\widehat{y}_\ell(t)\right)\mathrm{d}t.$$

with the same definition as previously for ε_{ij}, and $\widehat{y}_j(t) := y_j(t)$ for $t < t_0$. Note that B_j represents the same Brownian motion in each of the equations (2.6), (3.1) and (3.8). With a similar argument, we can assume that $\mathbf{y}$ is very close to $\widehat{\mathbf{y}}$ on another set $\mathcal{G}_{\mathbf{y}}$. We now define $\mathcal{G}_1 := \mathcal{G}_{\mathbf{x}} \cap \mathcal{G}_{\mathbf{y}}$.

Now we analyse the difference of the two cutoff dynamics. Setting $w_i := Ne^{t/2}(\widehat{x}_i - \widehat{y}_i)$, $\mathbf{w}$ satisfies an equation of the form $(t > t_0)$

$$\frac{\mathrm{d}w_i}{\mathrm{d}t} = \frac{1}{N}\sum_{j\neq i}\frac{w_j - w_i}{[(x_i-x_j)(t)+\varepsilon_{ij}][(y_i-y_j)(t)+\varepsilon_{ij}]} + \zeta_i \tag{3.9}$$

with an error term ζ_i satisfying

$$|\zeta_i(t)| \leqslant \sum_{j\neq i}\frac{|x_i-\widehat{x}_i| + |x_j-\widehat{x}_j| + |y_i-\widehat{y}_i| + |y_j-\widehat{y}_j|}{((x_i-x_j)(t)+\varepsilon_{ij})\,((y_i-y_j)(t)+\varepsilon_{ij})}.$$

With the level repulsion estimate as in (3.6), we have

$$\mathbb{E}\mathcal{G}_1|\zeta_i(t)| \leqslant CN^{-2C_0-50}\sum_{j\neq i}\left(\mathbb{E}\frac{1}{((x_i-x_j)(t)+\varepsilon_{ij})^2}\right)^{1/2}\left(\mathbb{E}\frac{1}{((y_i-y_j)(t)+\varepsilon_{ij})^2}\right)^{1/2} \leqslant CN^{-C_0-30}. \tag{3.10}$$

Therefore, there is a set $\mathcal{G}_2$ with $\mathbb{P}(\mathcal{G}_2) \geqslant 1 - CN^{-1}$ such that on this set we have

$$\sup_{t_0\leqslant t\leqslant 1}\max_i \int_{t_0}^{t}|\zeta_i(s)|\mathrm{d}s \leqslant N^{-C_0-20}. \tag{3.11}$$

We now show that ζ_i is negligible in the equation (3.9). This follows from the stability of the parabolic equation

$$\partial_t \mathbf{v}(t) = -\mathcal{B}(t)\mathbf{v}(t), \tag{3.12}$$

where $\mathcal{B}$ is the positive and positivity preserving matrix defined by

$$(\mathcal{B}(t)\mathbf{v})_i = \sum_{j=1}^{N} B_{ij}(t)(v_i - v_j), \tag{3.13}$$
$$B_{ij}(t) := \begin{cases} \frac{1}{N(x_i-x_j)(t)(y_i-y_j)(t)} & \text{if } t \leqslant t_0, \\ \frac{1}{N((x_i-x_j)(t)+\varepsilon_{ij})((y_i-y_j)(t)+\varepsilon_{ij})} & \text{if } t > t_0. \end{cases} \tag{3.14}$$

Indeed, suppose that $\mathbf{w}$ satisfies (3.9), i.e.

$$\partial_s \mathbf{w}(t) = -\mathcal{B}(t)\mathbf{w}(t) + \boldsymbol{\zeta}(t) \tag{3.15}$$

with the same initial data at time t_0, $\mathbf{v}(t_0) = \mathbf{w}(t_0)$, and $\boldsymbol{\zeta}$ satisfying the estimate (3.11). Then we have

$$\partial_t(\mathbf{w}-\mathbf{v})(t) = -\mathcal{B}(t)(\mathbf{w}-\mathbf{v})(t) + \boldsymbol{\zeta}(t) \tag{3.16}$$

with vanishing initial data. Let $U_{\mathcal{B}}(s,t)$ denote the semigroup associated with (3.12) from time s to time $t > s$, i.e.

$$\partial_t U_{\mathcal{B}}(s,t) = -\mathcal{B}(t) U_{\mathcal{B}}(s,t) \tag{3.17}$$

for any $t \geqslant s$ and $U_{\mathcal{B}}(s,s) = I$. By the Duhamel formula, we have

$$(\mathbf{w}-\mathbf{v})(t) = \int_{t_0}^t \mathcal{U}_{\mathcal{B}}(s,t)\boldsymbol{\zeta}(s)\mathrm{d}s.$$

Since $\mathcal{U}_{\mathcal{B}}$ is a contraction (in any L^p norm, in particular in L^∞), using (3.11) in $\mathcal{G}_2$ we have

$$\|(\mathbf{w}-\mathbf{v})(t)\|_\infty \leqslant N^{-C_0-20}, \qquad t \in [t_0, 1],$$

i.e., the effect of the perturbative term $\boldsymbol{\zeta}$ on the solution is negligible.

To summarize, we proved that the set $\mathcal{G} = \mathcal{G}_1 \cap \mathcal{G}_2$ satisfies $\mathbb{P}(\mathcal{G}) \geqslant 1 - CN^{-1}$ uniformly in $0 \leqslant \tau \leqslant 1$ and in this set $\mathcal{G}$ we have, for any i and $t \in (t_0, 1)$, that

$$N e^{t/2}(x_i(t) - y_i(t)) = v_i(t) + \mathrm{O}\left(N^{-1}\right) \tag{3.18}$$

where $\mathbf{v}$ satisfies (3.12) with initial condition $\mathbf{v}^0 = N(\mathbf{x}^0 - \mathbf{y}^0)$.

3.2 Continuous space operator. We now construct an operator in the continuum which approximates the discrete operator defined by $\mathcal{B}$. Recall the definition of the typical location γ_k from (2.5). If we replace x_i and y_i by γ_i and neglect the regularization ε, we have the following classical operator $\mathcal{U}$ on $\ell^2([\![1, N]\!])$:

$$(\mathcal{U}\mathbf{u})_j := \sum_{i \neq j} \frac{1}{N|\gamma_i - \gamma_j|^2}(u_j - u_i). \tag{3.19}$$

We now define an operator K acting on smooth functions on $[-2, 2]$ as

$$(Kf)(x) = \int_{-2}^2 \frac{f(x) - f(y)}{(x-y)^2}\mathrm{d}\varrho(y), \tag{3.20}$$

where the integral is in the principal value sense. Then K is the continuum limit of $\mathcal{U}$ in the sense that, for large N, $(\mathcal{U}\mathbf{u}^f)_j \approx (Kf)(\gamma_j)$, where $\mathbf{u}_j^f = f(\gamma_j)$. The following lemma provides an explicit formula for the evolution kernel e^{-tK}.

Lemma 3.1. *Let f be smooth with all derivatives uniformly bounded. For any $x, y \in (-2, 2)$, denote $x = 2\cos\theta$, $y = 2\cos\phi$ with $\theta, \phi \in (0, \pi)$. Then*

$$(e^{-tK}f)(x) = \int p_t(x,y) f(y)\mathrm{d}\varrho(y) \tag{3.21}$$

where the kernel is given by

$$p_t(x,y) := \frac{1-e^{-t}}{(1+e^{-t}-2e^{-t/2}\cos(\theta+\phi))(1+e^{-t}-2e^{-t/2}\cos(\theta-\phi))} = \frac{1-e^{-t}}{|e^{\mathrm{i}(\theta+\phi)} - e^{-t/2}|^2\,|e^{\mathrm{i}(\theta-\phi)} - e^{-t/2}|^2}. \tag{3.22}$$

Remark 1. The above formula is the same as the one in [1, page 462]. Lemma 3.1 shows that Biane's q-Ornstein Uhlenbeck generator coincides (for $q = 0$) with the convolution kernel (3.20).

Remark 2. If we neglect the curvature of the semicircle, i.e. γ_i's are equidistant on scale $1/N$, and formally extend the operator $\mathcal{U}$ to $\mathbb{Z}$, we obtain the following translation invariant operator $\mathcal{U}^\infty$ on $\ell^2(\mathbb{Z})$:

$$(\mathcal{U}^\infty \mathbf{u})_j := \sum_{i\in\mathbb{Z}\setminus\{j\}} \frac{N}{|i-j|^2}(u_j - u_i), \qquad \mathbf{u}\in\ell^2(\mathbb{Z}). \tag{3.23}$$

The Fourier transform of the kernel $1/k^2$ is given by $\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{1}{k^2}(1-e^{-ikp}) = \frac{\pi}{12}|p|$, i.e.

$$\widehat{(\mathcal{U}^\infty \mathbf{u})}(p) = c_0 N|p|\widehat{u}(p), \qquad c_0 := \frac{\pi}{12}, \quad p\in[-\pi,\pi],$$

where $\widehat{u}(p) := \sum_{k\in\mathbb{Z}} e^{-ipk}u_k$ and $u_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ipk}\widehat{u}(p)\mathrm{d}p$. Therefore the heat kernel of $\mathcal{U}^\infty$ can be computed by Fourier transform for any $t\geqslant 0$:

$$e^{-t\mathcal{U}^\infty}(0,k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-tc_0N|p|}e^{-ikp}\mathrm{d}p = \frac{1}{N}\frac{2c_0 t}{(tc_0)^2+(k/N)^2}\Big(1-(-1)^k e^{-\pi c_0 t}\Big). \tag{3.24}$$

The operator $\mathcal{U}^\infty$ is the discrete analogue of the operator $N\sqrt{-\Delta}$. The heat kernel $e^{-tN\sqrt{-\Delta}}$ is closely related to $p_t(x,y)$, but, compared with (3.22), there are substantial differences near the edges and also when t is large.

Proof of Lemma 3.1. Let U_n be the Chebishev polynomial of the second kind, defined by

$$U_N(\cos\theta) = \frac{\sin((n+1)\theta)}{\sin\theta}.$$

and we defined $P_n(x) = U_n(x/2)$. The proof relies on the following diagonalization: for any $n\geqslant 0$,

$$KP_n = \frac{n}{2}P_n. \tag{3.25}$$

For $n=0$, (3.25) is obvious. For $n=1$ this is the classical equilibrium relation

$$\int_{-2}^{2}\frac{\mathrm{d}\varrho(y)}{x-y} = \frac{x}{2}. \tag{3.26}$$

The following recursion relation is classical:

$$P_{n+1}(x) = xP_n(x) - P_{n-1}(x). \tag{3.27}$$

This yields, assuming that (3.25) holds up to the index n,

$$\begin{aligned}
KP_{n+1}(x) &= xKP_n(x) - KP_{n-1}(x) + \int_{-2}^{2}\frac{P_n(y)}{x-y}\mathrm{d}\varrho(y)\\
&= x\frac{n}{2}P_n(x) - \frac{n-1}{2}P_{n-1}(x) + P_n(x)\int_{-2}^{2}\frac{\mathrm{d}\varrho(y)}{x-y} + \int_{-2}^{2}\frac{P_n(y)-P_n(x)}{x-y}\mathrm{d}\varrho(y)\\
&= \frac{n+1}{2}xP_n(x) - \frac{n-1}{2}P_{n-1}(x) + \int_{-2}^{2}\frac{P_n(y)-P_n(x)}{x-y}\mathrm{d}\varrho(y)
\end{aligned}$$

where we used (3.26). Hence (3.25) will be proved with $n+1$ instead of n if

$$\int_{-2}^{2}\frac{P_n(x)-P_n(y)}{x-y}\mathrm{d}\varrho(y) = P_{n-1}(x) \tag{3.28}$$

holds. To prove (3.28) for any n, one can again proceed by induction. This formula is obviously true for $n=0,1$. Assuming it is true up to index n, with (3.27) we get

$$\int_{-2}^{2}\frac{P_{n+1}(x)-P_{n+1}(y)}{x-y}\mathrm{d}\varrho(y) = xP_{n-1}(x) + \int P_{n-1}(y)\mathrm{d}\varrho(y) - P_{n-2}(x) = P_n(x)$$

where we used that P_{n-1} is orthogonal to 1 with respect to the semicircle measure. This concludes the proof of (3.28) and therefore (3.25) for all n.

The conclusion of the lemma now easily follows: the kernel, defined through $(e^{-tK}f)(x) = \int p_t(x,y)f(y)\mathrm{d}\varrho(y)$, can be written in the eigenbasis as $p_t(x,y) = \sum_{n\geqslant 0} e^{-\frac{n}{2}t}P_n(x)P_n(y)$. Using the representation $P_n(2\cos\theta) = \sin((n+1)\theta)/\sin\theta$ and expanding the sin to get four geometric series concludes the proof. $\square$

We record some properties of the kernel (3.22) that easily follow from the explicit formula and from the asymptotics $\gamma_j + 2 \sim (j/N)^{2/3}$ for $j \leqslant N/2$,

$$p_t(\gamma_i,\gamma_j) \leqslant \frac{Ct}{t^2+(\gamma_i-\gamma_j)^2}, \qquad i \in [\![\alpha N, (1-\alpha)N]\!], \quad j \in [\![1,N]\!], \quad t \leqslant 1, \tag{3.29}$$

$$\sum_j p_t(\gamma_i,\gamma_j) \leqslant \sum_j \frac{Ct}{t^2+(\gamma_i-\gamma_j)^2} \leqslant C, \qquad i \in [\![\alpha N, (1-\alpha)N]\!], \quad t \leqslant 1, \tag{3.30}$$

$$|\partial_x p_t(\gamma_i, x)| \leqslant \frac{Ct|\gamma_i - x|}{(t^2+(\gamma_i-x)^2)^2} \qquad i \in [\![\alpha N, (1-\alpha)N]\!], \quad t \leqslant 1, \tag{3.31}$$

where the constant C depends only on the positive parameter $\alpha > 0$.

3.3 The homogenization result. For any $\delta \in \mathbb{R}$ and $E \in (-2,2)$ we define the index set

$$I(\delta) = I(E,\,\delta) := \{i:\ |\gamma_i - E| \leqslant N^{-1+\delta}\}. \tag{3.32}$$

The main result of this section is the following theorem.

Theorem 3.2. *Suppose $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are two DBM driven by the same Brownian motions (see (2.6)) and with initial data given by the spectra of two generalized Wigner matrices. There exist positive constants $\tau_0 \leqslant 1/4$, $\delta_1,\delta_2,\delta_3$ such that for any $t \in [2t_0, 1]$, with $t_0 := N^{-\tau_0}/2$, and $|E| < 2-\kappa$ with $\kappa > 0$ we have*

$$\mathbb{P}\left(\max_{i\in I(E,\delta_1)} \Big|Nx_i(t) - N\Big(y_i(t) + (\Psi_{t-t_0}\mathbf{x}(t_0))_i - (\Psi_{t-t_0}\mathbf{y}(t_0))_i\Big)\Big| \geqslant N^{-\delta_2}\right) \leqslant N^{-\delta_3}, \tag{3.33}$$

where Ψ_s is a linear operator defined by

$$(\Psi_s\mathbf{x})_i := e^{-s/2}\frac{1}{N}\sum_j p_s(\gamma_i,\gamma_j)x_j. \tag{3.34}$$

The main tool to prove Theorem 3.2 is a homogenization result. In order to state it, we first construct a partition of unity as follows. For any $j = 1,2,\ldots,N$ let

$$\tilde{g}_j := \min(\gamma_{j+1}-\gamma_j, \gamma_j - \gamma_{j-1}) \sim N^{-2/3}(\widehat{j})^{-1/3}$$

with the convention that $\gamma_{N+1} = \infty$, $\gamma_0 = -\infty$. For all even indices j define a smooth function ξ_j supported in $[\gamma_{j-1} + \tilde{g}_j/100, \gamma_{j+1} - \tilde{g}_j/100]$ and $\xi_j(x) = 1$ for $|x - \gamma_j| \leqslant \tilde{g}_j/100$ such that

$$\int_{\gamma_{j-1}}^{\gamma_j} \xi_j(x)\mathrm{d}\varrho(x) = \int_{\gamma_j}^{\gamma_{j+1}} \xi_j(x)\mathrm{d}\varrho(x) = \frac{1}{2N}. \tag{3.35}$$

For odd indices j, we define ξ_j by $\xi_j(x) = 1 - \xi_{j-1}(x)$ for $\gamma_{j-1} \leqslant x \leqslant \gamma_j$ and by $\xi_j(x) = 1 - \xi_{j+1}(x)$ for $\gamma_j \leqslant x \leqslant \gamma_{j+1}$. We thus have $\xi_j(x) + \xi_{j+1}(x) = 1$ for any j whenever $x \in [\gamma_j, \gamma_{j+1}]$. In particular, $\sum_j \xi_j(x) = 1$. Notice that by construction

$$\int \xi_j(x)\mathrm{d}\varrho(x) = \frac{1}{N}, \qquad \operatorname{supp}\xi_j \subset [\gamma_{j-1} + \tilde{g}_j/100, \gamma_{j+1} - \tilde{g}_j/100] \tag{3.36}$$

hold for all $j = 1, 2, \ldots, N$. For any discrete function (i.e. vector) $\mathbf{v} : i \to v_i$ define its continuous extension by

$$e_{\mathbf{v}}(x) := \sum_j \xi_j(x) v_j. \tag{3.37}$$

Notice that $e_{\mathbf{v}}(\gamma_i) = v_i$.

The main homogenization result is the following theorem. It is formulated for the parabolic equation (3.12) with general random coefficients $B_{ij}(t)$ under certain conditions. Later we will verify that rigidity and level repulsion for $\mathbf{x}(t)$ and $\mathbf{y}(t)$ imply that $B_{ij}(t)$ defined in (3.14) satisfy these conditions.

Theorem 3.3. *Fix three small positive constants, ξ, ρ, α. Choose $\tau_0 \in [0, \frac{1}{4}]$, and set $t_0 := N^{-\tau_0}/2$. Consider the equation (3.12) with time dependent random coefficients $B_{ij}(s)$ in the time interval $s \in [t_{in}, t_{end}]$ with $t_0 < t_{end} - t_{in} \leqslant C$. Denote by $\mathcal{F} = \mathcal{F}_\xi$ the event on which the following two bounds hold:*

$$\Big| B_{ij}(s) - \frac{1}{N(\gamma_i - \gamma_j)^2} \Big| \leqslant \frac{N^{-\frac{2}{3}+\xi}\big[(\widehat{i})^{-1/3} + (\widehat{j})^{-1/3}\big]}{N(\gamma_i - \gamma_j)^3}, \qquad \forall i, j, \ |i-j| \geqslant N^\xi, \quad \forall s \in [t_{in}, t_{end}], \tag{3.38}$$

$$B_{ij}(s) \geqslant \frac{N^{-\xi}}{N(\gamma_i - \gamma_j)^2}, \qquad \forall i, j, \quad \forall s \in [t_{in}, t_{end}]. \tag{3.39}$$

Furthermore we assume that

$$\max_{ij} \max_{s \in [t_{in}, t_{end}]} \mathbb{E}\left[\mathcal{F} |B_{ij}(s)|\right] \leqslant \frac{N^\rho}{N|\gamma_i - \gamma_j|^2}. \tag{3.40}$$

If ξ and ρ are small enough, then there are constants $c_4, c_5 > 0$, so that the following holds. For any fixed space-time point $(t, i) \in [t_{in} + t_0, t_{end}] \times [\![\alpha N, (1-\alpha)N]\!]$ there is an event $\mathcal{S} \subset \mathcal{F}$ with $\mathbb{P}(\mathcal{F} \setminus \mathcal{S}) \leqslant N^{-c_4}$ so that on $\mathcal{S}$ we have

$$\Big| (U_{\mathcal{B}}(t_{in}, t)\mathbf{v})_i - \int p_{t - t_{in}}(\gamma_i, y) e_{\mathbf{v}}(y) \mathrm{d}\varrho(y) \Big| \leqslant N^{-c_5} \|\mathbf{v}\|_\infty \tag{3.41}$$

for any vector $\mathbf{v} \in \mathbb{R}^N$ and for any sufficiently large $N \geqslant N_0(\alpha)$. Note that the set $\mathcal{S}$ depends on the choice (t, i), but the exponents c_4, c_5 do not.

Remark. Our proof can easily be extended to hold for any $\tau_0 < 1/3$, but then the smallness of ξ, ρ, c_4, c_5 will depend on how close τ_0 is to $1/3$. However, even the $1/3$ threshold for the exponent τ_0 is not optimal, it is due to various cutoffs that can be improved with more work. We do not pursue this direction since, for the purpose of this paper, only the small τ_0 regime is needed.

The following statement asserts that rigidity and level repulsion estimates on the DBM trajectories ensure that the conditions in Theorem 3.3 hold for B_{ij} given in (3.14) with a high probability provided that τ_0 is small.

Theorem 3.4. *There exist positive constants $c_4, c_5 > 0$ and $\tau_0 > 0$ such that the following holds. Fix $\alpha \in (0,1)$, set $t_0 := N^{-\tau_0}/2$. Consider the equation (3.12) with coefficients $B_{ij}(s)$ given by two coupled DBM $\mathbf{x}(s)$ and $\mathbf{y}(s)$, $s \in [0,1]$, as defined in (3.14). Then for any space-time point $(t, i) \in [2t_0, 1] \times [\![\alpha N, (1-\alpha)N]\!]$ there exists a set $\mathcal{S} = \mathcal{S}(t, i)$ in the joint probability space of the coupled DBM's $\mathbf{x}(s)$ and $\mathbf{y}(s)$, with $\mathbb{P}(\mathcal{S}) \geqslant 1 - N^{-c_4}$, such that on the set $\mathcal{S}$ (3.41) holds for any $\mathbf{v} \in \mathbb{R}^N$ and $N \geqslant N_0(\alpha, \tau_0)$.*

Proof. Choose $t_{in} := t_0$, $t_{end} := 1$ in Theorem 3.3. The estimates (3.38) and (3.39) directly follow from the rigidity bound (3.5) on the set

$$\widetilde{\mathcal{F}}_\xi := \bigcap_i \mathcal{G}_{\xi, \mathbf{x}, i} \cap \mathcal{G}_{\xi, \mathbf{y}, i},$$

thus $\mathcal{F}_\xi \supset \widetilde{\mathcal{F}}_\xi$ (here we used the fact that the parameter ε in the definition of B_{ij} is much smaller than the rigidity threshold $N^{-1+\xi}$). From (3.5) $\widetilde{\mathcal{F}}_\xi$ has a very high probability, $\mathbb{P}(\widetilde{\mathcal{F}}_\xi) \geqslant 1 - N^{-D}$ for any D (note that $\mathcal{F}$ and $\widetilde{\mathcal{F}}$ are independent of τ_0). For (3.40) we claim that

$$\mathbb{E}\mathcal{F}_\xi |B_{jk}(s)| \leqslant \frac{N^{C_0 \tau_0 + 3\xi}}{N|\gamma_j - \gamma_k|^2}$$

holds for any $\xi > 0$. Indeed, for $|j-k| \geqslant N^{\xi}$ this follows form the rigidity estimates. For $|j-k| \leqslant N^{\xi}$ with $j < k$ one may estimate $B_{jk}(s) \leqslant B_{j,j+1}(s)$ and then use a Schwarz inequality similar to (3.10). Finally, applying Corollary B.2 as in (3.6), we get

$$\mathbb{E}B_{j,j+1}(s) \leqslant N^{C_0\tau+\xi} N^{1/3} (\widehat{j})^{2/3} |\log \varepsilon| \leqslant \frac{N^{C_0\tau_0+2\xi}}{N|\gamma_j - \gamma_{j+1}|^2} \leqslant \frac{N^{C_0\tau_0+3\xi}}{N|\gamma_j - \gamma_k|^2}.$$

Setting $\rho = C_0\tau_0 + 3\xi$, we verified (3.40). Choosing τ_0 and ξ sufficiently small, we can apply Theorem 3.3 to conclude (3.41). For the probability of $\mathcal{S}$ we have $\mathbb{P}(\mathcal{S}) \geqslant 1 - N^{-D} - N^{-c_4}$ which satisfies the required bound by reducing c_4 a bit. □

Proof of Theorem 3.2. Pick positive constants c_4, c_5, τ_0 sufficiently small so that Theorem 3.4 applies. Without loss of generality we may assume that $\tau_0, c_4, c_5 \leqslant 1/100$ and $\tau_0 \leqslant c_5/100$. Recall the notation $t_0 = N^{-\tau_0}/2$. For brevity we write $\mathbf{x} := \mathbf{x}(t_0)$ and $\mathbf{y} := \mathbf{y}(t_0)$. We would like to apply Theorem 3.4 for the vector $\mathbf{v}$ of the form there is such a factor: x and y evolve by OU, but v does not.

$$v_j := Ne^{t_0/2}(x_j - y_j) \lesssim N^{\xi}(\widehat{j}/N)^{-1/3}, \tag{3.42}$$

but then $\|\mathbf{v}\|_\infty \sim N^{1/3+\xi}$ in (3.41) would be too large, as the edge indices contribute. So we have to perform a cutoff and use (3.41) only for the bulk indices and use an $L^1 \to L^\infty$ heat kernel bound to control the contribution near the edge. We therefore rewrite $\mathbf{v} = \mathbf{w} + \mathbf{u}$ where $w_j := v_j$ if $N^{1-\nu} \leqslant j \leqslant N - N^{1-\nu}$ and $w_j := 0$ otherwise, for some exponent $\nu > 0$ chosen later. Equation (3.41) with initial condition $\mathbf{w}$ yields

$$\Big|(U_{\mathcal{B}}(t_0,t)\mathbf{w})_i - \int p_{t-t_0}(\gamma_i, y) e_{\mathbf{w}}(y) \mathrm{d}\varrho(y)\Big| \leqslant N^{-c_5}\|\mathbf{w}\|_\infty \leqslant N^{-c_5+\frac{\nu}{3}+\xi} \tag{3.43}$$

on the set $\mathcal{S}(t,i)$ for any $i \in I(E,\delta_1)$. Using the definition of $e_{\mathbf{w}}$, from (3.31), (3.30) and (3.36) we have

$$\begin{aligned}
&\Big|\int p_{t-t_0}(\gamma_i, y) e_{\mathbf{w}}(y)\mathrm{d}\varrho(y) - \frac{1}{N}\sum_j p_{t-t_0}(\gamma_i,\gamma_j) w_j\Big| \\
&\qquad \leqslant \sum_j |w_j| \Big| \int p_{t-t_0}(\gamma_i, y)\xi_j(y)\mathrm{d}\varrho(y) - \frac{1}{N} p_{t-t_0}(\gamma_i,\gamma_j)\Big| \\
&\qquad \leqslant \sum_j |w_j| \int |p_{t-t_0}(\gamma_i,y) - p_{t-t_0}(\gamma_i,\gamma_j)|\xi_j(y)\mathrm{d}\varrho(y) \\
&\qquad \leqslant CN^{-2+\frac{2\nu}{3}+\xi} \sum_j \frac{t|\gamma_i - \gamma_j|}{(t^2 + (\gamma_i-\gamma_j)^2)^2} \leqslant Ct^{-1}N^{-1+\frac{2\nu}{3}+\xi}, \quad i \in I(E,\delta_1), \quad t \geqslant 2t_0.
\end{aligned}$$

since $|y - \gamma_j| \leqslant CN^{-1+\nu/3}$ and $|w_j| \leqslant N^{\frac{\nu}{3}+\xi}$ on the support of ξ_j with $\widehat{j} \geqslant N^{1-\nu}$. Moreover, from (3.29) and using that $u_j \neq 0$ only for $\widehat{j} \leqslant N^{1-\nu}$,

$$\Big|\frac{1}{N}\sum_j p_{t-t_0}(\gamma_i,\gamma_j)u_j\Big| \leqslant \frac{Ct}{N}\sum_j |u_j| \leqslant CtN^{-\frac{2}{3}\nu+\xi}, \quad i \in I(E,\delta_1), \quad t \geqslant 2t_0. \tag{3.44}$$

Together with (3.43) this gives that

$$\Big|(U_{\mathcal{B}}(t_0,t)\mathbf{w})_i - \frac{1}{N}\sum_j p_{t-t_0}(\gamma_i,\gamma_j)v_j\Big| \leqslant C\left(N^{-c_5+\frac{\nu}{3}+\xi} + t^{-1}N^{-1+\frac{2\nu}{3}+\xi} + tN^{-\frac{2}{3}\nu+\xi}\right) \tag{3.45}$$

on the set $\mathcal{S}(t,i)$ for any $t \geqslant 2t_0$, $i \in I(E,\delta_1)$.

Moreover, thanks to the following Proposition 3.5 (in our application $b = N^{-\xi}$, we use (3.39) and we shift the initial time from 0 to t_0) we have

$$\|U_{\mathcal{B}}(t_0,t)\mathbf{u} - \bar{u}\|_\infty \leqslant CN^{3\xi}t^{-3}N^{-1}\sum_j |u_j - \bar{u}|, \qquad t \geqslant 2t_0,$$

where $\bar{u} := N^{-1} \sum_j u_j$ and thus $|\bar{u}| \leqslant CN^{-2\nu/3+\xi}$. Hence we have proved that

$$\|U_{\mathcal{B}}(t_0, t)\mathbf{u}\|_\infty \leqslant CN^{-\frac{2}{3}\nu+\xi} + t^{-3}N^{-\frac{2}{3}\nu+4\xi}. \tag{3.46}$$

Combining this with (3.45), choosing $\nu = c_5$, $\xi = c_5/100$ and recalling $\tau_0 \leqslant c_5/100$, we have proved that

$$\Big|(U_{\mathcal{B}}(t_0, t)\mathbf{v})_i - N^{-1} \sum_j p_{t-t_0}(\gamma_i, \gamma_j)v_j\Big| \leqslant CN^{-c_5/2}, \tag{3.47}$$

on the set $\mathcal{S}(t, i)$ for any $i \in I(E, \delta_1)$ and any t with $1 \geqslant t \geqslant 2t_0$.

Finally, we need to guarantee that (3.47) holds for all $i \in I(E, \delta_1)$ simultaneously, i.e. we take the intersection

$$\mathcal{S}(t) := \bigcap_{i \in I(E,\delta_1)} \mathcal{S}(t, i).$$

The cardinality of $I(E, \delta_1)$ is bounded by CN^{δ_1} and $\mathbb{P}(\mathcal{S}(t, i)) \geqslant 1-N^{-c_4}$, so by choosing $\delta_1 < c_4$, we obtain that $\mathbb{P}(\mathcal{S}(t)) \geqslant 1 - \frac{1}{2}N^{-c_4}$. Now we choose $\delta_2 < c_5/2$ and $\delta_3 < c_4$ and together with (3.18), we conclude the proof of Theorem 3.2. □

For any $\mathbf{u} \in \mathbb{R}^N$ we define the ℓ^p norms as

$$\|\mathbf{u}\|_p := \big(\frac{1}{N} \sum_{i=1}^N |u_i|^p\big)^{1/p}.$$

The following decay estimate extends Proposition 10.4 of [4]. Notice that the convention of ℓ^p norm in this paper differs from that used in [4] by a normalization factor N^{-1}.

Proposition 3.5. *Suppose that the coefficients of the equation* (3.12) *satisfy for some constant b that*

$$B_{jk}(s) \geqslant \frac{b}{N(\gamma_j - \gamma_k)^2}, \quad 0 \leqslant s \leqslant \sigma. \tag{3.48}$$

Then for any $\mathbf{u}$ with $\sum_j u_j = 0$ we have the decay estimate

$$\|U_{\mathcal{B}}(0, s)\mathbf{u}\|_\infty \leqslant C(sb)^{-3}\|\mathbf{u}\|_1, \quad 0 \leqslant s \leqslant \sigma. \tag{3.49}$$

Proof. We first prove the same inequality for the operator K, i.e., for any mean zero function f that

$$\|e^{-2tK} f\|_\infty \leqslant \frac{C}{t^3}\|f\|_{\varrho,1}, \tag{3.50}$$

where

$$\|f\|_{\varrho,p} := \Big(\int |f(x)|^p \mathrm{d}\varrho(x)\Big)^{1/p}. \tag{3.51}$$

Recall Corollary 4 of [1] (see also (10.19) of [4]) asserting that there is a constant C so that

$$\|f\|^2_{\varrho,3} \leqslant C \left(\|f\|^2_{\varrho,2} + \langle f, Kf\rangle_\varrho\right). \tag{3.52}$$

By the explicit diagonalization of K, (3.25), the spectral gap of K is equal to 1/2. Hence for $\int f \mathrm{d}\varrho = 0$, we have $\|f\|^2_{\varrho,2} \leqslant 2\langle f, Kf\rangle_\varrho$ and thus

$$\|f\|^2_{\varrho,3} \leqslant C\langle f, Kf\rangle_\varrho. \tag{3.53}$$

We shall drop the subscript ϱ in the following argument. Suppose f_t solves the equation

$$\partial_t f_t = -Kf_t \tag{3.54}$$

and the initial data has zero mean, i.e., $\int f_0 \mathrm{d}\varrho = 0$. Then we have

$$\partial_s \|f_s\|_2^2 = -\langle f_s, K f_s \rangle \leqslant -C \|f_s\|_2^{\frac{8}{3}} \|f_s\|_1^{-\frac{2}{3}},$$

where we have used (3.53) and the Hölder inequality

$$\|f\|_3^2 \geqslant \|f\|_2^{\frac{8}{3}} \|f\|_1^{-\frac{2}{3}}.$$

Since $\|f_s\|_1$ is non-increasing, we can integrate this inequality to have

$$\|f_t\|_2 \leqslant \frac{C}{t^{3/2}} \|f_0\|_1.$$

For any g with $\int g \mathrm{d}\varrho = 0$, we have

$$|\langle g, e^{-tK} f_t \rangle| = |\langle e^{-tK} g, f_t \rangle| \leqslant \left\| e^{-tK} g \right\|_2 \|f_t\|_2 \leqslant \frac{C}{t^3} \|g\|_1 \|f_0\|_1. \tag{3.55}$$

Since $\int e^{-tK} f_t \mathrm{d}\varrho = \int f_0 \mathrm{d}\varrho = 0$ by assumption, the mean zero condition of g can be removed and we have thus proved (3.50).

We can now follow the similar argument to prove (3.49). After a time rescaling, we can assume that $b = 1$. The key ingredient in the previous argument is the Sobolev inequality (3.52). Now we will need a discrete version. This can be achieved by extending a discrete function to the continuum with a simple interpolation procedure. This idea was used in [4] and we will not repeat it here. Once a discrete version of (3.52) is proved, the rest of the proof is identical to the one in the continuum. Thus we have proved (3.49). □

3.4 Proof of Theorem 3.3. Without loss of generality, we can assume that $t_{in} = 0$ by a simple time shift. For simplicity, we also set $t_{end} = 1$, as the actual value of t_{end} influences only irrelevant constant prefactors. By definition (3.37) and the equation (3.12), we have

$$\partial_t e_{\mathbf{v}(t)} = -\mathcal{R}_t \mathbf{v}(t), \tag{3.56}$$

where

$$(\mathcal{R}_t \mathbf{v})(x) := \sum_{j,k=1}^{N} \xi_j(x)(v_j - v_k) B_{kj}(t) \tag{3.57}$$

takes the vector $\mathbf{v}$ to the function $\mathcal{R}_t \mathbf{v}$ for any fixed t. Suppose that $f = f(t,x)$ is a solution to the continuum equation

$$\partial_t f(x) = -(Kf)(x), \tag{3.58}$$

where K is defined in (3.20). Then we have

$$\partial_t (e_{\mathbf{v}(t)} - f(t)) = -K(e_{\mathbf{v}}(t) - f(t)) + [K e_{\mathbf{v}(t)} - \mathcal{R}_t \mathbf{v}(t)]. \tag{3.59}$$

We will need to solve this equation from time $t_{in} = 0$ to t. We will take the initial condition at time $t_{in} = 0$ to be $\mathbf{v}$ for the discrete equation and $f(0) = e_{\mathbf{v}}$ for the continuous one.

By the Duhamel formula, we have

$$e_{\mathbf{v}(t)} - f(t) = \int_0^t e^{-(t-s)K} [K e_{\mathbf{v}(s)} - \mathcal{R}_s \mathbf{v}(s)] \mathrm{d}s = \Phi - \Omega, \tag{3.60}$$

where the functions Φ, Ω are given by

$$\Omega(z) := \int_0^t \mathrm{d}s \int d\varrho(x) p_{t-s}(z,x) \int \mathrm{d}\varrho(y) \frac{e_{\mathbf{v}}(s,y) - e_{\mathbf{v}}(s,x)}{|x-y|^2}, \tag{3.61}$$

$$\Phi(z) := \int_0^t \mathrm{d}s \int \int \mathrm{d}\varrho(x) \mathrm{d}\varrho(y) \sum_{j,k} p_{t-s}(z,x) \xi_j(x)(v_k(s) - v_j(s)) B_{kj}(s). \tag{3.62}$$

We have used that $\int \mathrm{d}\varrho(y) = 1$ in (3.62). We will need these functions for $z := \gamma_i$ in order to obtain (3.41), but we will keep the shorter z notation. Note that $e_{\mathbf{v}(t)}(\gamma_i) = v_i(t) = (U_{\mathcal{B}}(0,t)\mathbf{v})_i$, thus the left hand side of (3.41) is $\Phi(z) - \Omega(z)$ with $z = \gamma_i$.

Step 1: Cutoff of long range part. We first cutoff the contributions to Ω, Φ when $|x-y| \geqslant \ell$ for some $N^{-2/3} \ll \ell \ll 1$ to be fixed later on. In this regime, we will need to use cancellation between Ω and Φ. We start with the following definition that for any subset D in $\mathbb{R}^2 \times \mathbb{R}$ define

$$\Omega_D := \int_0^t \mathrm{d}s \int\int \mathbf{1}_D(x,y,s)\mathrm{d}\varrho(x)\mathrm{d}\varrho(y) p_{t-s}(z,x)\frac{e_{\mathbf{v}}(s,y) - e_{\mathbf{v}}(s,x)}{|x-y|^2}. \tag{3.63}$$

If D is symmetric under $x \leftrightarrow y$, then we have

$$\Omega_D = \frac{1}{2}\int_0^t \mathrm{d}s \int\int \mathbf{1}_D(x,y,t)\mathrm{d}\varrho(x)\mathrm{d}\varrho(y)[p_{t-s}(z,x) - p_{t-s}(z,y)]\frac{e_{\mathbf{v}}(s,y) - e_{\mathbf{v}}(s,x)}{|x-y|^2}. \tag{3.64}$$

Similarly we can define

$$\Phi_D := \int_0^t \mathrm{d}s \int\int \mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_D(x,y,t)\sum_{j,k} p_{t-s}(z,x)\xi_j(x)(v_k(s) - v_j(s))B_{kj}(s), \tag{3.65}$$

and for symmetric D we have

$$\Phi_D = \frac{1}{2}\int_0^t \mathrm{d}s \int\int \mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_D(x,y,t)\sum_{j,k}[p_{t-s}(z,x)\xi_j(x) - p_{t-s}(z,y)\xi_k(y)](v_k(s) - v_j(s))B_{kj}(s). \tag{3.66}$$

Let

$$\widehat{A}_\ell := \{(x,y) : |x-y| \geqslant \ell\}, \quad D_\ell := \widehat{A}_\ell \times [0,t]. \tag{3.67}$$

Using (3.62) and (3.61), we can decompose the error term $\Phi_{D_\ell} - \Omega_{D_\ell}$ in this region into $\Phi^1_{D_\ell} - \Omega^1_{D_\ell} + \Phi^2_{D_\ell} - \Omega^2_{D_\ell}$, where

$$\Phi^1_{D_\ell} - \Omega^1_{D_\ell} := -\int_0^t \mathrm{d}s \int\int \mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_{\widehat{A}_\ell}(x,y)p_{t-s}(z,x)\sum_j \xi_j(x)v_j(s)\Big[\sum_k B_{jk}(s) - \frac{1}{|x-y|^2}\Big]. \tag{3.68}$$

The second term is

$$\Phi^2_{D_\ell} - \Omega^2_{D_\ell} = \int_0^t \mathrm{d}s \int\int \mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_{\widehat{A}_\ell}(x,y)p_{t-s}(z,x)\sum_k v_k(s)\Big[\sum_j \xi_j(x)B_{jk}(s) - \xi_k(y)\frac{1}{|x-y|^2}\Big]. \tag{3.69}$$

Since $\sum_j \xi_j(x) = 1$, we can replace $\xi_k(y)\frac{1}{|x-y|^2}$ by $\sum_j \xi_j(x)\xi_k(y)\frac{1}{|x-y|^2}$. Recall the normalization condition $\int \mathrm{d}\varrho(y)\xi_k(y) = 1/N$ (3.36). If we could neglect the factor $\mathbf{1}_{\widehat{A}_\ell}(x,y)$ in this normalization, we had

$$\Phi^2_{D_\ell} - \Omega^2_{D_\ell} \approx \int_0^t \mathrm{d}s \int\int \mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_{\widehat{A}_\ell}(x,y)p_{t-s}(z,x)\sum_{k,j} v_k(s)\xi_j(x)\xi_k(y)\Big[NB_{jk}(s) - \frac{1}{|x-y|^2}\Big]. \tag{3.70}$$

Indeed, the difference between the two sides of (3.70) is

$$\int_0^t \mathrm{d}s \int \mathrm{d}\varrho(x)p_{t-s}(z,x)\sum_{k,j} v_k(s)\xi_j(x)B_{jk}(s)\int_{|x-y|\geqslant\ell}\mathrm{d}\varrho(y)\Big[1 - \xi_k(y)N\Big]. \tag{3.71}$$

Notice that $|\gamma_j - x| \leqslant \widetilde{g}_j$ from the support property of $\xi_j(x)$ and that the last integral is zero unless the support of $\xi_k(y)$ overlaps with one of the boundaries $y = x \pm \ell$ of the integration regime. Thus (3.71) can be bounded from above by

$$\begin{aligned}
\int_0^t \mathrm{d}s \int\int_{|x-y|\geqslant \ell} &\mathrm{d}\varrho(x)\mathrm{d}\varrho(y)p_{t-s}(z,x) \sum_{k,j,||\gamma_j-\gamma_k|-\ell|\leqslant 4\widetilde{g}_j+4\widetilde{g}_k} |v_k(s)|\xi_j(x)\xi_k(y)NB_{jk}(s) \\
&\leqslant C\|\mathbf{v}\|_\infty \ell^{-2}\int_0^t \mathrm{d}s \int\int_{|x-y|\geqslant \ell} \mathrm{d}\varrho(x)\mathrm{d}\varrho(y)p_{t-s}(z,x) \sum_{k,j,||\gamma_j-\gamma_k|-\ell|\leqslant 4\widetilde{g}_j+4\widetilde{g}_k} \xi_j(x)\xi_k(y) \\
&\leqslant C\|\mathbf{v}\|_\infty \ell^{-2}\int_0^t \mathrm{d}s \int\int \mathbf{1}\Big(\big||x-y|-\ell\big| \leqslant CN^{-1}[\varrho(x)^{-1}+\varrho(y)^{-1}]\Big)\mathrm{d}\varrho(x)\mathrm{d}\varrho(y)p_{t-s}(z,x) \\
&\leqslant C\|\mathbf{v}\|_\infty \ell^{-2}N^{-1}t,
\end{aligned} \tag{3.72}$$

where in the first step we used that under the constraints on the summations, we have $NB_{jk}\leqslant C\ell^{-2}$ from (3.38) and from the fact that $\widetilde{g}_j,\widetilde{g}_k \ll \ell$. In the second step we translated the constraint on the indices j,k to a constraint on x,y using that $\widetilde{g}_j \sim \varrho^{-1}(\gamma_j)$, and finally we integrated out y, x and s in this order. We also used the contraction property $\|\mathbf{v}(s)\|_\infty \leqslant \|\mathbf{v}\|_\infty$.

For the term on the r.h.s. of (3.70), we will use the coordinate system $x=2\cos\Theta(x)$ with $0\leqslant\Theta(x)\leqslant\pi$. From the estimate (3.38), for $|x-y|\geqslant \ell \gg N^{-2/3}$ we have

$$\sum_{k,j}\xi_j(x)\xi_k(y)\Big|NB_{jk}(s)-\frac{1}{|x-y|^2}\Big| \leqslant N^{-1+\xi}\frac{1}{|x-y|^3}[(\sin\Theta(x))^{-1}+(\sin\Theta(y))^{-1}]. \tag{3.73}$$

Together with $\|\mathbf{v}(t)\|_\infty \leqslant \|\mathbf{v}\|_\infty$, we have

$$\begin{aligned}
\Phi^2_{D_\ell}-\Omega^2_{D_\ell} &\leqslant N^{-1+\xi}\,\|\mathbf{v}\|_\infty \int_0^t \mathrm{d}s\int\int_{|x-y|\geqslant\ell}\mathrm{d}\varrho(x)\mathrm{d}\varrho(y)p_{t-s}(z,x)\frac{1}{|x-y|^3}[(\sin\Theta(x))^{-1}+(\sin\Theta(y))^{-1}] \\
&\leqslant \frac{t|\log\ell|}{\ell^2}N^{-1+\xi}\,\|\mathbf{v}\|_\infty\,.
\end{aligned} \tag{3.74}$$

In the last step we used that z is away from the edge, so in the regime where x is near the edge and $[\sin\Theta(x)]^{-1}\sim\varrho(x)^{-1}$ becomes singular, we know that $p_{t-s}(z,x)\leqslant C$ from (3.22).

The estimate of the $\Phi^1_{D_\ell}-\Omega^1_{D_\ell}$ term is similar. We write the $\mathrm{d}\varrho(y)$ integration in (3.68) as

$$\begin{aligned}
\int_{|x-y|\geqslant\ell}&\mathrm{d}\varrho(y)\Big[\sum_k B_{jk}(s)-\frac{1}{|x-y|^2}\Big] \\
&=\int_{|x-y|\geqslant\ell}\mathrm{d}\varrho(y)\sum_k\xi_k(y)\Big[NB_{jk}(s)-\frac{1}{|x-y|^2}\Big]+\sum_k B_{jk}(s)\int_{|x-y|\geqslant\ell}\mathrm{d}\varrho(y)(1-N\xi_k(y))
\end{aligned}$$

The first term can be estimated exactly the r.h.s. of (3.70) (the only difference is $v_j(s)$ in (3.68) instead of $v_k(s)$ in (3.70) but these factors are estimated by $\|\mathbf{v}\|_\infty$ anyway). The second term is analogous to (3.71). This completes the estimate of the regime $|x-y|\geqslant\ell$.

From now on, we will work on the complement of $\widehat{A}_\ell$, i.e. in the regime $|x-y|\leqslant\ell$. We will not use cancellation between Φ and Ω and will estimate them separately by splitting the integrals into further subregions. As the estimates for Φ and Ω are similar, we will work out only one of them in every region.

Step 2: Time region away from the final time t via the energy bound. In this step, we estimate the contribution to the integrals (3.61), (3.62) for times $s\in[0,t-t_1]$ with some $t_1\ll t$. The main idea to deal with this regime is to use energy bound for the dynamics (3.12) and the regularity of the continuous evolution kernel $p_{t-s}(x,y)$.

We start with a general estimate to show how energy bound is used to control Ω_D. For a set $D\subset\mathbb{R}^2\times[0,t]$, symmetric under $x \leftrightarrow y$, using the Schwarz inequality, we have

$$\Omega_D = \frac{1}{2}\int_0^t \mathrm{d}s \int\int \mathbf{1}_D(x,y,t)\mathrm{d}\varrho(x)\mathrm{d}\varrho(y)[p_{t-s}(z,x)-p_{t-s}(z,y)]\frac{e_{\mathbf{v}}(s,y)-e_{\mathbf{v}}(s,x)}{|x-y|^2} \leqslant \sqrt{W_1^D W_2^D}, \quad (3.75)$$

$$W_1^D := \frac{1}{2}\int_0^t \mathrm{d}s \int\int \mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_D[p_{t-s}(z,x)-p_{t-s}(z,y)]^2\frac{1}{|x-y|^2}, \quad (3.76)$$

$$W_2^D := \frac{1}{2}\int_0^t \mathrm{d}s \int\int \mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_D\frac{[e_{\mathbf{v}}(y)-e_{\mathbf{v}}(x)]^2}{|x-y|^2}. \quad (3.77)$$

We start with the second term W_2^D. For $\gamma_j \leqslant x \leqslant \gamma_{j+1}$ and $\gamma_k \leqslant y \leqslant \gamma_{k+1}$, from the construction of ξ's, we have

$$\begin{aligned} e_{\mathbf{v}}(x) - e_{\mathbf{v}}(y) &= \xi_j(x)v_j + \xi_{j+1}(x)v_{j+1} - \xi_k(y)v_k - \xi_{k+1}(y)v_{k+1} \\ &= (1-\xi_{j+1}(x))v_j + \xi_{j+1}(x)v_{j+1} - (1-\xi_{k+1}(y))v_k - \xi_{k+1}(y)v_{k+1} \\ &= v_j - v_k + \xi_{j+1}(x)(v_{j+1}-v_j) - \xi_{k+1}(y)(v_{k+1}-v_k). \end{aligned} \quad (3.78)$$

In particular, when $j=k$, we have

$$|e_{\mathbf{v}}(x)-e_{\mathbf{v}}(y)| = \Big|\xi_{j+1}(x)-\xi_{j+1}(y)\Big||v_{j+1}-v_j| \leqslant C\widetilde{g}_{j+1}^{-1}|x-y||v_{j+1}-v_j| \leqslant \frac{C|x-y||v_{j+1}-v_j|}{|\gamma_{j+1}-\gamma_j|^2}.$$

For neighboring indices, i.e. when $k=j+1$,

$$e_{\mathbf{v}}(x)-e_{\mathbf{v}}(y) = [\xi_{j+1}(x)-1](v_{j+1}-v_j) - \xi_{j+2}(y)(v_{j+2}-v_{j+1}). \quad (3.79)$$

Notice that when $|x-y| \leqslant \widetilde{g}_{j+1}/400$, we have $|x-\gamma_{j+1}| \leqslant \widetilde{g}_{j+1}/200$, $|y-\gamma_{j+1}| \leqslant \widetilde{g}_{j+1}/200$ and by definition of the ξ's we have $\xi_{j+1}(x)=1$, $\xi_{j+2}(y)=0$, so $e_{\mathbf{v}}(x)-e_{\mathbf{v}}(y)=0$. Therefore for any set D, we can bound W_2^D by

$$\begin{aligned} W_2^D &= \sum_{j,k}\int_0^t \int_{\gamma_j}^{\gamma_{j+1}}\mathrm{d}\varrho(x)\int_{\gamma_k}^{\gamma_{k+1}}\mathrm{d}\varrho(y)\frac{[e_{\mathbf{v}}(x)-e_{\mathbf{v}}(y)]^2}{|x-y|^2}\mathrm{d}s \\ &\leqslant C\sum_{j\neq k}\int_0^t \frac{[v_j(s)-v_k(s)]^2}{N^2|\gamma_j-\gamma_k|^2}\mathrm{d}s + C\sum_j \int_0^t \frac{[v_j(s)-v_{j+1}(s)]^2}{N^2|\gamma_j-\gamma_{j+1}|^2}\mathrm{d}s \\ &\leqslant N^\xi \frac{1}{N}\sum_{j,k}\int_0^t [v_j(s)-v_k(s)]^2 B_{kj}(s)\mathrm{d}s \\ &= N^\xi\frac{1}{N}\sum_j \left[v_j^2(0)-v_j^2(t)\right] \leqslant N^\xi\|\mathbf{v}\|_2^2 \leqslant N^\xi\|\mathbf{v}\|_\infty^2. \end{aligned} \quad (3.80)$$

In (3.80) we have used (3.39). The last step is the energy estimate that can be obtained by integrating the time derivative $\partial_s\|\mathbf{v}(s)\|_2^2$.

Choose $t_1 = tN^{-2\mathfrak{a}}$ with some $\mathfrak{a}>0$ to be fixed later and define $D_1 := \{|x-y|\leqslant \ell\}\times[0,t-t_1]$ (Here D_1 is a new set, not to be confused with D_ℓ defined earlier). For z in the bulk, we can use the explicit formula of p_t (3.22) so that

$$W_1^{D_1} = \int_0^{t-t_1}\mathrm{d}s\int_{|x-y|\leqslant\ell}\mathrm{d}\varrho(x)\mathrm{d}\varrho(y)[p_{t-s}(z,x)-p_{t-s}(z,y)]^2\frac{1}{|x-y|^2} \leqslant \frac{\ell}{t_1^2}. \quad (3.81)$$

Together with (3.75) and (3.80), we have proved that $|\Omega_{D_1}| \leqslant CN^\xi\frac{\sqrt{\ell}}{t_1}\|\mathbf{v}\|_\infty$. Similarly, we can bound Φ_{D_1} and obtain

$$|\Phi_{D_1}| + |\Omega_{D_1}| \leqslant CN^\xi\frac{\sqrt{\ell}}{t_1}\|\mathbf{v}\|_\infty. \quad (3.82)$$

Step 3: Time region near the final time t via the Hölder regularity. In this step and the next one we consider the final time region $s \in [t-t_1, t]$. Notice that we will not use the smoothness of the continuous kernel $p_{t-s}(x,y)$ which depends on $t-s$ and becomes singular when s is close to t. Instead, in Step 3 we consider the regime in (3.66) where x (hence also y) is not too far from the fixed reference point z. In this case we will use the the Hölder regularity of the solution to the equation (3.12). In Step 4, we look at the complement regime, when x and y are far from z, and we can use the large distance decay of the kernel p_t.

We first recall this basic Hölder estimate from [18]. We will need this result in the following form and in Appendix A we will explain how this particular version follows from the general statement in [18].

Lemma 3.6. *For any $t \geqslant t_0 = N^{-\tau_0}/2$ and a small constant $0 < \mathfrak{a} < 1-\tau_0$ fixed, we set $\ell_1 = tN^{-\mathfrak{a}}$. For any real z with $|z| < 2$ define*

$$\Xi_z(\ell_1) := \big\{(j,k) : 1 \leqslant j,k \leqslant N, |\gamma_j - z| \leqslant \ell_1,\ |\gamma_k - z| \leqslant \ell_1\big\}. \tag{3.83}$$

Consider the equation (3.12) with coefficients (3.14) satisfying (3.38)–(3.40). If the exponent $\rho > 0$ in (3.40) is sufficiently small, depending on $\mathfrak{a}$, then there exists a set $\mathcal{G} \subset [t - tN^{-\mathfrak{a}}, t]$ of "good times" with Lebesgue measure

$$|[t - tN^{-\mathfrak{a}}, t] \setminus \mathcal{G}| \leqslant (tN^{-\mathfrak{a}})^{1/4} N^{-3/4}, \tag{3.84}$$

and a set $\mathcal{R}_{z,t}$ in the probability space with

$$\mathbb{P}(\mathcal{R}_{z,t}) \geqslant 1 - N^{-\rho} \tag{3.85}$$

such that in the set $\mathcal{R}_{z,t}$ the following oscillation estimate holds: for any time $s \in \mathcal{G}$ and indices $j,k \in \Xi_z(\ell_1)$ we have

$$\big|v_j(s) - v_k(s)\big| \leqslant N^{-\mathfrak{q}\mathfrak{a}} \|\mathbf{v}\|_\infty. \tag{3.86}$$

Here the exponent $\mathfrak{q}$ is a positive constant independent of any parameters.

If (3.86) holds, we say that Hölder regularity holds at the space time point (z,t).

For any $t_1 \ll \ell_1 \ll t$ and $N^{-2/3} \ll \ell \ll \ell_1$, denote by

$$A_{\ell,\ell_1} := \{(x,y) : |x-y| \leqslant \ell;\ |z-x| \leqslant \ell_1 \text{ and } |z-y| \leqslant \ell_1\} \tag{3.87}$$

and consider Φ_D from (3.65) with $D_{\ell,\ell_1} := A_{\ell,\ell_1} \times [t-t_1, t]$. With a similar estimate on the boundary terms of the set A_{ℓ,ℓ_1} as in (3.72), one obtains

$$\Phi_{D_{\ell,\ell_1}} = \widetilde{\Phi}_{D_{\ell,\ell_1}} + O\Big(\|\mathbf{v}\|_\infty \ell^{-2} N^{-2/3} t_1\Big), \tag{3.88}$$

where

$$\widetilde{\Phi}_{D_{\ell,\ell_1}} := \frac{1}{2}\int_{t-t_1}^{t} \int \mathrm{d}\varrho(x)\mathrm{d}\varrho(y) \mathbf{1}_{A_{\ell,\ell_1}}(x,y)[p_{t-s}(z,x) - p_{t-s}(z,y)] \sum_{j,k} \xi_j(x)\xi_k(y) N B_{kj}(s)(v_k(s) - v_j(s))\mathrm{d}s. \tag{3.89}$$

Notice that the characteristic function on x,y puts a constraint on the indices j,k via the support properties of ξ's, in particular $(j,k) \in \Xi_z(\ell_1)$. From Lemma 3.6 there is a set $\mathcal{G}$ of "good times" and an event $\mathcal{R}_{z,t}$ such that the Hölder estimate (3.86) holds in the intersection of $\mathcal{R}_{z,t}$ and the event $\mathcal{F}$ defined in Theorem 3.3. Thus there is a positive constant $\mathfrak{q} > 0$ such that for j,k with $\xi_j(x)\xi_k(y)\mathbf{1}_{A_{\ell,\ell_1}}(x,y) \neq 0$ and $t - t_1 \leqslant s \leqslant t$ we have

$$\mathbb{E}\mathbf{1}(\mathcal{G})\mathbf{1}(\mathcal{F} \cap \mathcal{R}_{z,t})|B_{jk}(s)|\big|v_j(s) - v_k(s)\big| \leqslant N^{-\mathfrak{q}\mathfrak{a}} \|\mathbf{v}\|_\infty \mathbb{E}\mathcal{F}|B_{jk}(s)|. \tag{3.90}$$

Using the estimate (3.40), we have

$$\mathbb{E}\mathbf{1}(\mathcal{G})\mathbf{1}(\mathcal{F} \cap \mathcal{R}_{z,t})|B_{jk}(s)|\big|v_j(s) - v_k(s)\big| \leqslant N^{\rho} \frac{1}{N|\gamma_j - \gamma_k|^2} N^{-\mathfrak{q}\mathfrak{a}} \|\mathbf{v}\|_\infty. \tag{3.91}$$

We can use

$$\frac{1}{|\gamma_j-\gamma_k|^2}\leqslant\frac{C}{|x-y|^2+N^{-2}},\tag{3.92}$$

whenever $\xi_j(x)\xi_k(y)\neq 0$ and $j\neq k$. By splitting the time integration into good and bad times, we can bound the expectation of $\widetilde{\Phi}_{D_{\ell,\ell_1}}$ by

$$\begin{aligned}\mathbb{E}\mathbf{1}(\mathcal{F}\cap\mathcal{R}_{z,t})\widetilde{\Phi}_{D_{\ell,\ell_1}}&\leqslant N^{\rho}N^{-\mathfrak{q}\mathfrak{a}}\|\mathbf{v}\|_\infty\int_{t-t_1}^{t}\mathrm{d}s\int\mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_{A_{\ell,\ell_1}}(x,y)\frac{|p_{t-s}(z,x)-p_{t-s}(z,y)|}{|x-y|^2+N^{-2}}+N^{-\mathfrak{a}+\rho}\|\mathbf{v}\|_\infty\\&\leqslant N^{\rho}N^{-\mathfrak{q}\mathfrak{a}}\log N\|\mathbf{v}\|_\infty,\end{aligned}\tag{3.93}$$

where the second term comes from the "bad" times s after using the estimate (3.84) and estimating $|v_k(s)-v_j(s)|\leqslant 2\|\mathbf{v}\|_\infty$ in (3.89).

Step 4: Time region near the final time t via the decay of the kernel $p_{t-s}(x,y)$. We now consider the contribution from the region

$$\tilde{A}_{\ell,\ell_1}:=\{(x,y):|x-y|\leqslant\ell;|z-x|>\ell_1\text{ or }\quad|z-y|>\ell_1\},\qquad\widetilde{D}_{\ell,\ell_1}:=\tilde{A}_{\ell,\ell_1}\times[t-t_1,t],\tag{3.94}$$

i.e. estimate $\Phi_{\widetilde{D}_{\ell,\ell_1}}$, see (3.66). As in (3.88), it is sufficient to consider the more symmetrized version

$$\widetilde{\Phi}_{\widetilde{D}_{\ell,\ell_1}}:=\frac12\int_{t-t_1}^{t}\int\mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_{\tilde{A}_{\ell,\ell_1}}(x,y)[p_{t-s}(z,x)-p_{t-s}(z,y)]\sum_{j,k}\xi_j(x)\xi_k(y)NB_{kj}(s)(v_k(s)-v_j(s))\mathrm{d}s$$

with a common factor $\xi_j(x)\xi_k(y)$. Using $\ell\ll\ell_1$, we see that both $|z-x|$ and $|z-y|$ are bounded from below by $\ell_1/2$, so the p_{t-s} kernels are not singular. By (3.40) and $\|\mathbf{v}(t)\|_\infty\leqslant\|\mathbf{v}\|_\infty$, we have

$$\mathbb{E}\mathcal{F}|B_{jk}(s)|\big|v_j(s)-v_k(s)\big|\leqslant N^{\rho}\frac{1}{N|\gamma_j-\gamma_k|^2}\|\mathbf{v}\|_\infty.\tag{3.95}$$

Using (3.92), we can thus bound the expectation of $\widetilde{\Phi}_{\tilde{D}_{\ell,\ell_1}}$ by

$$\begin{aligned}\mathbb{E}\mathcal{F}\widetilde{\Phi}_{\tilde{D}_{\ell,\ell_1}}&\leqslant N^{\rho}\|\mathbf{v}\|_\infty\int_{t-t_1}^{t}\mathrm{d}s\int\mathrm{d}\varrho(x)\mathrm{d}\varrho(y)\mathbf{1}_{\tilde{A}_{\ell,\ell_1}}(x,y)|p_{t-s}(z,x)-p_{t-s}(z,y)|\frac{1}{|x-y|^2+N^{-2}}\\&\leqslant N^{\rho}\frac{t_1^2}{\ell_1^2}\|\mathbf{v}\|_\infty.\end{aligned}\tag{3.96}$$

Step 5: The conclusion. Collecting all error terms from (3.72), (3.74), (3.82), (3.93), (3.96), and neglecting irrelevant logarithmic factors, we have

$$\mathbb{E}\mathbf{1}(\mathcal{F}\cap\mathcal{R}_{z,t})|\Phi-\Omega|\leqslant\Big[\frac{t}{\ell^2}N^{-1+\xi}+\frac{\sqrt{\ell}}{t_1}+N^{\rho}N^{-\mathfrak{q}\mathfrak{a}}+N^{\rho}\frac{t_1^2}{\ell_1^2}\Big]\|\mathbf{v}\|_\infty.\tag{3.97}$$

Recall the choices $\ell_1=tN^{-\mathfrak{a}}$, $t_1=tN^{-2\mathfrak{a}}$, we have

$$\mathbb{E}\mathbf{1}(\mathcal{F}\cap\mathcal{R}_{z,t})|\Phi-\Omega|\leqslant N^{\xi+\rho}\Big[\frac{t}{\ell^2}N^{-1}+\frac{\sqrt{\ell}}{t}N^{2\mathfrak{a}}+N^{-\mathfrak{q}\mathfrak{a}}\Big]\|\mathbf{v}\|_\infty.\tag{3.98}$$

Choosing $\ell=t^2N^{-5\mathfrak{a}}$ so that the second term is small, we have

$$\mathbb{E}\mathbf{1}(\mathcal{F}\cap\mathcal{R}_{z,t})|\Phi-\Omega|\leqslant N^{\xi+\rho}\Big[t^{-3}N^{-1+10\mathfrak{a}}+N^{-\mathfrak{a}\mathfrak{q}}\Big]\|\mathbf{v}\|_\infty.\tag{3.99}$$

Hence for $t\geqslant t_0=N^{-\tau_0}/2$ with $\tau_0\leqslant 1/4$, one can choose sufficiently small positive exponents $\xi,\rho,\mathfrak{a}$, so that $\mathbb{E}\mathbf{1}(\mathcal{F}\cap\mathcal{R}_{z,t})|\Phi-\Omega|\leqslant N^{-c}\|\mathbf{v}\|_\infty$ with some positive $c>0$. We can choose $c\leqslant\rho$. After a Markov inequality and using (3.85), we see that $|\Phi-\Omega|\leqslant N^{-c/2}\|\mathbf{v}\|_\infty$ on an event $\mathcal{S}$ with probability larger than $1-N^{-c/2}$ This completes the proof of Theorem 3.3. □

4 Proof of the universality at fixed energy

In this section, we prove our main result Theorem 2.2. The key ingredient of the proof is Lemma 4.1 below, asserting that local eigenvalue statistics of DBM for sufficiently large but still of order o(1) times converges to those of GOE. In order to state this lemma, we first introduce some notations.

The trajectory $(\mathbf{x}(t))_{t\geqslant 0}$will always denote Dyson Brownian motion dynamics, on the simplex $x_1(t) \leqslant \ldots \leqslant x_N(t)$, with initial condition given by eigenvalues of a generalized Wigner matrix. See (2.6). The processes $(\mathbf{y}(t))_{t\geqslant 0}$, $(\mathbf{z}(t))_{t\geqslant 0}$ follow the same dynamics on the simplex, with different, independent, initial conditions, given by the spectrum of a GOE. Omission of the time parameter means initial condition: $x_k = x_k(0)$.

For any $k \in \mathbb{N}$ and any smooth function $O : \mathbb{R}^k \to \mathbb{R}$, we denote the $W^{2,\infty}(\mathbb{R}^k)$ Sobolev norm by

$$\|O\|_{W^{2,\infty}} = \sum_{\sum_j \alpha_j \leqslant 2} \Big\| \prod_{j=1}^{k} \partial_{x_j}^{\alpha_j} O(\mathbf{x}) \Big\|_\infty . \tag{4.1}$$

We will consider test functions $O \in W^{2,\infty}(\mathbb{R}^k)$ that are compactly supported in $[-L, L]^k$ for some $L > 0$.

For an initial Wigner matrix H_0 we define the Ornstein-Uhlenbeck matrix flow as the solution of the SDE

$$\mathrm{d}H_t = \frac{\mathrm{d}B_t}{\sqrt{N}} - \frac{H_t}{2}\mathrm{d}t, \qquad H_{t=0} = H_0, \tag{4.2}$$

where B_t is a matrix of standard real or complex Brownian motions in the same symmetry class as H_0. The distribution of H_t coincides with

$$H_t \overset{d}{\sim} e^{-t/2} H_0 + (1 - e^{-t})^{1/2} H^G, \tag{4.3}$$

where H^G is a standard GOE matrix, independent of H_0. Recall the well-known fact that the law of the solution $\mathbf{x}(t)$ to the DBM (2.6) is the same as that of the eigenvalues of H_t provided that the law of the initial data for (2.6) is given by the eigenvalues of H_0. Recall the definition of the correlation functions $\rho_k^{(N)}$ from Section 2 and define the rescaled correlation functions around a fixed energy E by

$$\rho_{k,E}^{(N,\,\mathrm{resc})}(\mathbf{v}) := \frac{1}{\varrho(E)^k} \rho_k^{(N)} \left(E + \frac{\mathbf{v}}{N\varrho(E)} \right). \tag{4.4}$$

We will use $\rho_{k,E,t}^{(N,\,\mathrm{resc})}(\mathbf{v})$ for the rescaled correlation functions of the eigenvalues of H_t.

Lemma 4.1. *For a fixed $k \in \mathbb{N}$ and $L > 0$, let $O \in W^{2,\infty}(\mathbb{R}^k)$ be a test function supported in $[-L, L]^k$. Suppose that H_0 satisfies all the assumptions in Definition 2.1 and* (2.1)*. For a fixed positive number τ we set $t = N^{-\tau}$. Fix any $\kappa > 0$. Then for any $|E| \leqslant 2 - \kappa$ we have*

$$\left| \int \mathrm{d}\mathbf{v} O(\mathbf{v}) \rho_{k,E,t}^{(N,\,\mathrm{resc})}(\mathbf{v}) - \int \mathrm{d}\mathbf{v} O(\mathbf{v}) \rho_k^{(\mathrm{GOE})}(\mathbf{v}) \right| \leqslant C \|O\|_{W^{2,\infty}} \tau^{1/2} \tag{4.5}$$

holds for any small enough $\tau \leqslant \tau_0(\kappa)$ any sufficiently large $N \geqslant N_0(\tau, \kappa)$. Here the constant C depends only on L and κ.

Throughout this section we use the relation $t = N^{-\tau}$ between t and τ, and we will use both letters in parallel. In order to extend the universality result from Wigner ensembles H_t with a Gaussian component of size of order $t = N^{-\tau}$ to all Wigner ensembles, we follow the standard approach via the following Green function comparison theorem.

Lemma 4.2. *Consider two $N \times N$ generalized Wigner matrices, $H^{(v)}$ and $H^{(w)}$ with matrix elements h_{ij} given by the random variables $N^{-1/2} v_{ij}$ and $N^{-1/2} w_{ij}$, respectively, and satisfying the assumptions in Definition 2.1 and the moment condition* (2.1)*. We assume that the first four moments of v_{ij} and w_{ij} satisfy, for some $\delta > 0$, that*

$$\left| \mathbb{E}(\Re\, v_{ij})^a (\Im\, v_{ij})^b - \mathbb{E}(\Re\, w_{ij})^a (\Im\, w_{ij})^b \right| \leqslant N^{-\delta - 2 + (a+b)/2}, \qquad 1 \leqslant a + b \leqslant 4. \tag{4.6}$$

Let $\rho_k^{(N,v)}$ and $\rho_k^{(N,w)}$ be the $k-$point correlation functions of the eigenvalues w.r.t. the probability law of the matrix $H^{(v)}$ and $H^{(w)}$, respectively. Then for any test function O and any $|E| \leqslant 2-\kappa$ we have

$$\lim_{N\to\infty} \int \mathrm{d}\mathbf{v} O(\mathbf{v}) \left(\rho_{k,E}^{(N,\,v,\,\mathrm{resc})}(\mathbf{v}) - \rho_{k,E}^{(N,\,w,\,\mathrm{resc})}(\mathbf{v}) \right) = 0. \tag{4.7}$$

Proof. Recall [19, Lemma 3.4], where it was proved that for any real random variable θ such that

$$\mathbb{E}\,\theta = 0, \quad \mathbb{E}\,\theta^2 = 1, \quad \mathbb{E}|\theta|^4 \leqslant C,$$

and small $t > 0$, there exists random variable $\widetilde{\theta} = \widetilde{\theta}(\theta, t)$ and an independent, standard normal random variable $X \sim \mathcal{N}(0,1)$ such that

(i) $\widetilde{\theta}$ has subexponential decay;

(ii) the first three moments of $e^{-t/2}\widetilde{\theta} + (1-e^{-t})^{1/2}X$ equal to those of θ;

(iii) the difference between the fourth moment of $e^{-t/2}\widetilde{\theta} + (1-e^{-t})^{1/2}X$ and θ is $\mathrm{O}(t)$.

Inspecting the proof in [19], one can easily show that (i) can be strengthened to require that $\widetilde{\theta}$ has a Gaussian decay. Moreover, one can easily extend this result to complex random variables θ if (1) $\Re(\theta), \Im(\theta)$ are independent, or (2) the law of θ is isotropic, i.e. $|\theta|$ is independent of $\arg\theta$, which is uniform on $(0, 2\pi)$. In this case there exists a complex random variable $\widetilde{\theta}$ satisfying the corresponding condition (1) or (2) and each item (i)-(iii).

We apply this result to each entry of H. Therefore, there exists a generalized Wigner matrix $\widetilde{H}$, satisfying the assumptions in Definition 2.1 and (2.1) such that if we define

$$\widetilde{H}_t \;=\; e^{-t/2}\widetilde{H} + (1-e^{-t})^{1/2} H^G,$$

then the first four moments of the matrix entries of $\widetilde{H}_t$ almost match those of H in the following sense:

$$\mathbb{E}\left[\Re(\widetilde{H}_t)_{ij}\right]^a \left[\Im(\widetilde{H}_t)_{ij}\right]^b = \mathbb{E}\left[\Re H_{ij}\right]^a \left[\Im H_{ij}\right]^b, \qquad 0 \leqslant a, b \leqslant 3, \quad 1 \leqslant a+b \leqslant 3,$$
$$\left|\mathbb{E}\left[\Re(\widetilde{H}_t)_{ij}\right]^a \left[\Im(\widetilde{H}_t)_{ij}\right]^b - \mathbb{E}\left[\Re H_{ij}\right]^a \left[\Im H_{ij}\right]^b\right| \leqslant CN^{-2}t, \qquad a+b = 4.$$

Furthermore, $\widetilde{H}_t$ satisfies the assumptions in Definition 2.1 and the decay condition (2.1). Applying Lemma 4.2 with the choice $H^{(v)} = H$ and $H^{(w)} = \widetilde{H}_t$, $t = N^{-\tau}$ and $\delta := \tau$, we obtain that the correlation functions of H asymptotically match those of $\widetilde{H}_t$, i.e.,

$$\lim_{N\to\infty} \int \mathrm{d}\mathbf{v} O(\mathbf{v}) \left(\rho_{k,E}^{(N,\,\mathrm{resc})}(\mathbf{v}) - \widetilde{\rho}_{k,E,t}^{(N,\,\mathrm{resc})}(\mathbf{v}) \right) = 0 \tag{4.8}$$

for any test function O. Now we can apply (4.5) with $\widetilde{H}$ and $\widetilde{H}_t$ playing the role of H_0 and H_t, respectively, since $\widetilde{H}$ satisfies the assumption in Definition 2.1 and (2.1). We obtain that the correlation functions of $\widetilde{H}_t$ asymptotically match those of H^G:

$$\limsup_{N\to\infty} \left| \int \mathrm{d}\mathbf{v} O(\mathbf{v}) \widetilde{\rho}_{k,E,t}^{(N,\,\mathrm{resc})}(\mathbf{v}) - \int \mathrm{d}\mathbf{v} O(\mathbf{v}) \rho_k^{(\mathrm{GOE})}(\mathbf{v}) \right| \leqslant C\|O\|_{W^{2,\infty}} \tau^{1/2}. \tag{4.9}$$

Combining (4.8) and (4.9), and letting $\tau \to 0$ after the $N \to \infty$ limit, we obtain that

$$\lim_{N\to\infty} \int \mathrm{d}\mathbf{v} O(\mathbf{v}) \left(\rho_{k,E}^{(N,\,\mathrm{resc})}(\mathbf{v}) - \rho_k^{(\mathrm{GOE})}(\mathbf{v}) \right) = 0 \tag{4.10}$$

holds for compactly supported test functions $O \in W^{2,\infty}$. To extend this result to a general continuous function O supported in $[-L, L]^k$, we use a simple approximation. For any $\varepsilon > 0$, there exist $W^{2,\infty}$ functions $O_{+,\varepsilon}$ and $O_{-,\varepsilon}$, supported in $[-L-\varepsilon, L+\varepsilon]^k$, such that

$$O_{+,\varepsilon} \geqslant O \geqslant O_{-,\varepsilon}, \quad \|O_{+,\varepsilon} - O_{-,\varepsilon}\|_\infty \leqslant 2\varepsilon.$$

Applying (4.10) to $O_{\pm,\varepsilon}$, we obtain

$$\limsup_{N\to\infty}\int \mathrm{d}\mathbf{v} O(\mathbf{v})\rho_{k,E}^{(N,\,\mathrm{resc})}(\mathbf{v}) \leqslant \int \mathrm{d}\mathbf{v} O_{+,\varepsilon}(\mathbf{v})\rho_k^{(\mathrm{GOE})}(\mathbf{v}).$$

and similar lower bound for lim inf. Together with the fact that $\rho_k^{(\mathrm{GOE})}$ is bounded, it implies that (4.10) holds for any continuous, compactly supported observable, which completes the proof of Theorem 2.2. □

4.1 Reduction to observables with compact Fourier support. This section presents an approximation argument: we show that universality for a special class of test functions can be extended to $W^{2,k}$ test functions as required in Lemma 4.1. After a change of variables, we will work with test functions that have a compact support in the Fourier space in the energy variable. Universality for such test functions is stated in Lemma 4.4 below and will be proven in the subsequent Section 4.2.

First we will need the following precise estimates on the correlation functions of GOE, which were proved in [30, Theorem 3] and [5].

Lemma 4.3. *As in* (2.3)*,*

$$\frac{1}{\varrho(E)^k}\rho_k^{(N,\mathrm{GOE})}\left(E+\frac{\mathbf{v}}{N\varrho(E)}\right)=\rho_k^{(\mathrm{GOE})}(\mathbf{v})+O(N^{-1/2}), \tag{4.11}$$

uniformly holds for $(\mathbf{v},E)$ *in any fixed compact subset of* $\mathbb{R}^k\times(-2,2)$ *(for matrices from the GUE, the same statement holds with a different limit* $\rho_k^{(\mathrm{GUE})}(\mathbf{v})$*).*

Let $\mathbf{x}^G$ be the vector of ordered eigenvalues of H^G and let $\mathbf{x}(t)$ be the eigenvalues of H_t in (4.2). Simply rescaling the variables in O with $\varrho(E)$, the above lemma shows that for the proof of (4.5) it is sufficient to prove that

$$\left|\mathbb{E}\sum_{i_1,i_2\cdots i_k=1}^{N} O\left(\left\{N(x_{i_j}(t)-E)\right\}_{j=1}^k\right)-\mathbb{E}\sum_{i_1,i_2\cdots i_k=1}^{N} O\left(\left\{N(x_{i_j}^G-E)\right\}_{j=1}^k\right)\right|\leqslant C\tau^{1/2} \tag{4.12}$$

holds for any compactly supported $O\in W^{2,\infty}(\mathbb{R}^k)$.

For brevity, we assume that O has only two arguments, i.e., $k=2$; the general case is proven analogously. Furthermore, with a change of variables $(a,b)\to(a,b-a)$, we use the test function of the form

$$Q(N(x_i-E),N(x_j-x_i)) \quad \text{instead of} \quad O(N(x_i-E),N(x_j-E)). \tag{4.13}$$

The new test function Q is still compactly supported and lies in $W^{2,\infty}$; its advantage is that it depends on E only through its first variable. Therefore, under the assumption of Lemma 4.1, it is sufficient to prove that for small enough τ,

$$\left|\mathbb{E}\sum_{i,j=1}^{N} Q\Big(N(x_i(t)-E),N(x_j(t)-x_i(t))\Big)-\mathbb{E}\sum_{i,j=1}^{N} Q\Big(N(x_i^G-E),N(x_j^G-x_i^G)\Big)\right|\leqslant C\|Q\|_{W^{2,\infty}}\tau^{1/2} \tag{4.14}$$

holds for sufficiently large N and with C depending only on L and κ. For simplicity, we define

$$\mathcal{Q}(\mathbf{x},E):=\sum_{i,j=1}^{N} Q\Big(N(x_i-E),N(x_j-x_i)\Big), \quad E\in\mathbb{R}. \tag{4.15}$$

Let $\widehat{Q}(p,y)$ be the Fourier transform of Q w.r.t. the first argument, i.e.,

$$\widehat{Q}(p,y)=\int_{\mathbb{R}} Q(x,y)e^{-ipx}\mathrm{d}x. \tag{4.16}$$

In this section, hat always denotes a partial Fourier transform, i.e. Fourier transform only in the first variable. The Fourier-space variables will be denoted by p. We will also say that $\widehat{Q}(p,y) \in W^{2,\infty}$ if $\widehat{Q}$ as a function of p, y is in the Sobolev space.

The following lemma, proven in Section 4.2, states that that universality for large time holds for observables whose Fourier transforms have compact support.

Lemma 4.4. *Let $Q : \mathbb{R}^2 \to \mathbb{R}$ be a function such that $\widehat{Q} \in W^{2,\infty}$, and*

$$\operatorname{supp} \widehat{Q} \subset [-m, m] \times [-L, L] \tag{4.17}$$

for some fixed m, $L \in \mathbb{N}$. There exists a constant δ_0 independent of m and L such that for any $\kappa > 0$

$$\mathbb{E}\, \mathcal{Q}(\mathbf{x}(t), E) - \mathbb{E}\, \mathcal{Q}(\mathbf{x}^G, E) = \mathrm{O}(N^{-\tau}), \qquad t = N^{-\tau} \tag{4.18}$$

holds uniformly for $|E| \leqslant 2 - \kappa$ and $\tau \leqslant \frac{\delta_0}{m^2+1}$.

We now prove Lemma 4.1 assuming Lemma 4.4 holds. The first step is to approximate a compactly supported observable $Q(x, y) \in W^{2,\infty}$ by an observable $Q(x, y)$ whose Fourier transform $\widehat{Q}(p, y)$ is compactly supported as required in Lemma 4.4. The following lemma provides an effective control on this approximation.

Lemma 4.5. *Let $q \in W^{2,\infty}(\mathbb{R})$ be a symmetric cutoff function, supported on $[-1, 1]$ such that $q(p) = 1$ for $|p| \leqslant 1/2$, and $q'(p) \leqslant 0$ for $p > 0$. For $Q \in W^{2,\infty}$ and* $\operatorname{supp} Q \in [-L, L]^2$*, define Q_m via its partial Fourier transform as*

$$\widehat{Q_m}(p, y) := \widehat{Q}(p, y) q_m(p), \qquad q_m(p) := q(p/m). \tag{4.19}$$

Then there exists a constant C, depending only on L and q, such that for any $m \in \mathbb{N}$, $(x, y) \in \mathbb{R}^2$, we have

$$|(Q_m - Q)(x, y)| \leqslant \frac{C\|Q\|_{W^{2,\infty}}}{(1 + x^2)} m^{-1}, \tag{4.20}$$

$$\|\widehat{Q_m}\|_{W^{2,\infty}} \leqslant C\|Q\|_{W^{2,\infty}}. \tag{4.21}$$

Proof of Lemma 4.5. We have

$$x^n (Q_m - Q)(x, y) = \int_{\mathbb{R}} \frac{(i\partial_p)^n}{2\pi} \left[\left(\widehat{Q_m} - \widehat{Q} \right)(p, y) \right] e^{ipx} \mathrm{d}p, \tag{4.22}$$

$$p^{n_1} (\partial_p)^{n_2} \widehat{Q}(p, y) = \int_{\mathbb{R}} (-i)^{n_1+n_2} (\partial_x)^{n_1} \left[x^{n_2} Q(x, y) \right] e^{-ipx} \mathrm{d}x. \tag{4.23}$$

Using (4.23) with $n_1 = 0, 2$, since Q is compactly supported in $[-L, L]^2$, we have

$$|(\partial_p)^n \widehat{Q}(p, y)| \leqslant C_{n,L} (1 + |p|^2)^{-1} \|Q\|_{W^{2,\infty}}. \tag{4.24}$$

Similarly, using (4.22) we have

$$|x^n (Q_m - Q)(x, y)| \leqslant C_n \int \sum_{n_1+n_2=n} |(\partial_p)^{n_1} (q_m - 1)(p)| \left| (\partial_p)^{n_2} \widehat{Q}(p, y) \right| \mathrm{d}p.$$

By definition, if $n_1 \neq 0$ then $|(\partial_p)^{n_1} (q_m - 1)| \leqslant C m^{-1}$. If $n_1 = 0$ then $|(\partial_p)^{n_1} (q_m - 1)|$ is supported in $\{p : |p| \geqslant m/2\}$. Together with (4.24), we have

$$|x^n (Q_m - Q)(x, y)| \leqslant C_{n,L} \|Q\|_{W^{2,\infty}} \ m^{-1}.$$

Choosing $n = 0$ and 2, we complete the proof of (4.20). The (4.21) can be easily derived from the definition of Q_m and (4.23) (with $n_1 = 0$). □

Lemma 4.5 provides an approximation for any smooth observables with compact support by observables with compact support in the Fourier space. On the other hand, to estimate the error resulting from this approximation, we will need the following corollary which gives an effective bound on the density of $\mathbf{x}(t)$, the eigenvalues of H_t, at the local scale $1/N$.

Corollary 4.6. *Define*

$$\#(\mathbf{x}, E, s_1, s_2) := \left|\left\{(i,j) \in \mathbb{N}^2 : |x_i - E| \leqslant \frac{s_1}{N}, \quad |x_i - x_j| \leqslant \frac{s_2}{N}\right\}\right|. \tag{4.25}$$

For any fixed $L \geqslant 1$ and $\kappa > 0$, with the δ_0 in Lemma 4.4, there exists constant $C > 0$ such that for $\tau \leqslant \delta_0/2$

$$\limsup_{N\to\infty} \max_{|E|\leqslant 2-\kappa} \mathbb{E}\,\#(\mathbf{x}(t), E, 1, L) \leqslant C, \qquad t = N^{-\tau}, \quad \tau \leqslant \delta_0/2. \tag{4.26}$$

Proof. Let $g, h \in W^{2,\infty}(\mathbb{R})$ be two real functions such that $\mathrm{supp}\,(\widehat{g}) = [-1,1]$, $\mathrm{supp}(h) = [-2L, 2L]$. We assume that $\min_{|x|\leqslant a} g(x) \geqslant b$ for some $0 < a \leqslant 1$ and $b > 0$. Furthermore, we assume that $1 \geqslant h(x) \geqslant 0$ for any $x \in \mathbb{R}$ and $h(x) = 1$ for $|x| \leqslant L$. Define

$$Q(x,y) := b^{-2} g^2(x) h(y). \tag{4.27}$$

Since $\widehat{g^2} = \widehat{g} \star \widehat{g}$, it is clear that Q satisfies the assumption in Lemma 4.4 with $m = 2$, and L being replaced by $2L$. Then applying (4.18) to Q defined in (4.27), and using that $\mathbb{E}\,\mathcal{Q}(\mathbf{x}^G, E)$ is bounded from Lemma 4.3, we have that

$$\limsup_{N\to\infty} \mathbb{E}\,\mathcal{Q}(\mathbf{x}(t), E) \leqslant C \tag{4.28}$$

holds uniformly for any $|E| \leqslant 2 - \kappa$. From the definition of Q in (4.27), we have

$$\min_{x,y\in\mathbb{R}} Q(x,y) \geqslant 0 \quad \text{and} \quad Q(x,y) \geqslant 1, \quad (x,y) \in [-a,a] \times [L,L].$$

Then (4.28) implies

$$\limsup_{N\to\infty} \max_{|E|\leqslant 2-\kappa} \mathbb{E}\,\#(\mathbf{x}(t), E, a, L) \leqslant C$$

for some constant C. Hence (4.26) also holds, since $a \sim 1$, which completes the proof of Corollary 4.6. $\square$

We now have all the ingredients to complete the proof of Lemma 4.1.

Proof of Lemma 4.1. For any compactly supported $Q \in W^{2,\infty}(\mathbb{R}^2)$, we construct Q_m as in (4.19). The definition of Q_m, and (4.21) guarantee that Q_m satisfy the assumption of Lemma 4.4. Then applying Lemma 4.4 for Q_m, we obtain for any fixed $m \in N$ that

$$\mathbb{E}\,\mathcal{Q}_m(\mathbf{x}(t), E) = \mathbb{E}\,\mathcal{Q}_m(\mathbf{x}^G, E) + \mathrm{O}(N^{-\tau}), \qquad t = N^{-\tau}, \qquad \tau \leqslant \frac{\delta_0}{m^2+1}, \tag{4.29}$$

where δ_0 is from (4.18). On the other hand, we will show below that if $\tau \leqslant \delta_0/2$, then

$$|\mathbb{E}\,\mathcal{Q}(\mathbf{x}(t), E) - \mathbb{E}\,\mathcal{Q}_m(\mathbf{x}(t), E)| \leqslant C\,\|Q\|_{W^{2,\infty}} m^{-1} \tag{4.30}$$

holds for some C independent of m and τ and and large enough N. Notice that (4.30) also holds if we replace the $\mathbf{x}(t)$ with $\mathbf{x}^G$, since $\mathbf{x}(t) \overset{d}{\sim} \mathbf{x}^G$ if $\mathbf{x}(0) \overset{d}{\sim} \mathbf{x}^G$. Combining (4.29) and (4.30), and choosing $m = \delta\tau^{-1/2}$ with a small $\delta \leqslant \delta_0$, we obtain (4.14), i.e., (4.12) in the case $k = 2$. One can easily extend the above proof to the general k case. Together with Lemma 4.3, it implies the desired result (4.5).

Hence it only remains to prove (4.30). Using (4.20), we have

$$\max_y \max_{x:|x-n|\leqslant 1} |(Q_m - Q)(x,y)| \leqslant \frac{C\|Q\|_{W^{2,\infty}}}{1+n^2} m^{-1}.$$

With the definition of $\#(\mathbf{x}(t), E+\frac{n}{N}, 1, L)$ in (4.25), it implies that

$$|\mathbb{E}\,\mathcal{Q}(\mathbf{x}(t),E)-\mathbb{E}\,\mathcal{Q}_m(\mathbf{x}(t),E)|\leqslant \|Q\|_{W^{2,\infty}}\frac{1}{m}\sum_n \frac{C}{1+n^2}\mathbb{E}\,\#(\mathbf{x}(t),E+\frac{n}{N},1,L). \tag{4.31}$$

For $n\leqslant N^{1/2}$, $\mathbb{E}\,\#(\mathbf{x}(t),E+\frac{n}{N},1,L)$ can be bounded by (4.26) (after replacing κ with $\kappa/2$). For $n\geqslant N^{1/2}$ we can use the trivial bound

$$\mathbb{E}\,\#(\mathbf{x}(t),E+\frac{n}{N},1,L)\leqslant N^{\xi}$$

for any $\xi>0$ that directly follows from the rigidity of eigenvalues of H_t if $N\geqslant N_0(\xi)$ is large enough. Inserting these bounds into (4.31), we obtain (4.30) and complete the proof of Lemma 4.1. □

4.2 Universality for test functions with compact Fourier support: Proof of Lemma 4.4. Recall that E satisfies $|E|\leqslant 2-\kappa$. All the constants in the following proof depend on κ, but we will not carry this dependence explicitly in the notation. For any nonnegative integer α introduce the notation

$$Q^{(\alpha)}(x,y):=(\partial_x)^\alpha Q(x,y).$$

Recall

$$x^{n_1}(\partial_x)^{n_2}Q(x,y)=\int\frac{i^{n_1+n_2}}{2\pi}(\partial_p)^{n_1}\left[p^{n_2}\widehat{Q}(p,y)\right]e^{ipx}\mathrm{d}p.$$

As in (4.20), with assumption (4.17) and $\widehat{Q}\in W^{2,\infty}$ we obtain that there exists some constant C such that for any $\alpha\in\mathbb{Z}_{\geqslant 0}$, and $y\in\mathbb{R}$

$$\left|Q^{(\alpha)}(x,y)\right|\leqslant Cd_\alpha(1+x^2)^{-1},\qquad d_\alpha:=\left(m^2+\alpha^2\right)m^{\alpha-1}\|\widehat{Q}\|_{W^{2,\infty}}. \tag{4.32}$$

The multiindex α used in this section has nothing to do with the threshold α to indicate indices away from the edge, see e.g., (3.29).

The main input to prove Lemma 4.4 is the homogenization result, Theorem 3.2, stating that for any $\tau<\tau_0$ with a sufficiently small τ_0, two coupled DBMs (2.6) driven by the same Brownian motions satisfy the estimate

$$Nx_i(t)-N\Big(y_i(t)+(\Psi_{t-t_0}\mathbf{x}(t_0))_i-(\Psi_{t-t_0}\mathbf{y}(t_0))_i\Big)=O(N^{-\delta_2}),\qquad t=N^{-\tau}\geqslant 2t_0=N^{-\tau_0}, \tag{4.33}$$

for all $i\in I(\delta_1)$ with probability bigger than $1-N^{-\delta_3}$. We recall from (3.34) that $\Psi_t\mathbf{x}$ is given by $(\Psi_t\mathbf{x})_i=N^{-1}\sum_k p_t(\gamma_i,\gamma_k)x_k$, and $\delta_1,\delta_2,\delta_3$ are small positive exponents.

In our application we choose $\mathbf{y}(t_0)$ to be distributed by μ_G, i.e, the eigenvalue distribution of a Gaussian matrix ensemble. Since $\mathbf{x}(t_0)$'s are the eigenvalues of a generalized Wigner matrix, we denote their distribution by μ_W. The joint distribution of the coupled DBM processes $\{\mathbf{x}(s)\}_{0\leqslant s\leqslant t}$ and $\{\mathbf{y}(s)\}_{0\leqslant s\leqslant t}$, as defined in (2.6), is given by $\mu_W\otimes\mu_G\otimes\mu_B$, where $\mu_B=\mu\left(\{B_\ell(t)\}_{1\leqslant\ell\leqslant N,0\leqslant s\leqslant t}\right)$ is the measure of the independent Brownian motions. For simplicity, for expectation w.r.t. $\mu_W\otimes\mu_G\otimes\mu_B$, we just use $\mathbb{E}$. For the expectation of functionals f of $\mathbf{x}(t)$, we will sometimes use $\mathbb{E}^{\mu_W}f(\mathbf{x}(t))$ instead of $\mathbb{E}^{\mu_W\otimes\mu_B}f(\mathbf{x}(t))$ and similarly we use $\mathbb{E}^{\mu_G}f(\mathbf{y}(t))$ for functionals of $\mathbf{y}(t)$.

Below, we apply the homogenization result (4.33) to $Q^{(\alpha)}$. Recall the definition of $I(\delta)$ from (3.32).

Lemma 4.7. *For $\mathbf{x}\in\mathbb{R}^N$ and $s\geqslant 0$, define*

$$\xi_s^{\mathbf{x}}:=\xi_s^{\mathbf{x}}(E):=N(\Psi_s\mathbf{x})_{i_0}-N(\Psi_s\boldsymbol{\gamma})_{i_0},\quad i_0:=\min\{i:\ \gamma_i\geqslant E\}. \tag{4.34}$$

(The notation $\xi_s^{\mathbf{x}}$ should not be confused with the rigidity exponent ξ.) Recalling τ_0 provided by Theorem 3.2, there exists $\delta_Q\leqslant\tau_0$ such that for any $0<\delta\leqslant\delta_Q$ and any $0<\tau\leqslant\delta/5$ we have (with the usual $t=N^{-\tau}$, $t_0=N^{-\tau_0}/2$ conventions)

$$\max_{|E'-E|\leqslant N^{4\tau-1}}\left|\mathbb{E}^{\mu_W}\ \mathcal{Q}^{(\alpha)}(\mathbf{x}(t),E')-\sum_{i,j\in I(\delta)}\mathbb{E}\,Q^{(\alpha)}\Big(N(y_i(t)-E')+\xi_{t-t_0}^{\mathbf{x}(t_0)}-\xi_{t-t_0}^{\mathbf{y}(t_0)},\quad N(y_j(t)-y_i(t))\Big)\right|$$
$$\leqslant 3N^{-\delta/2}d_{\alpha+1} \tag{4.35}$$

for large enough $N \geqslant N_0$ where N_0 is independent of α, the order of derivatives.

Notice that although Lemma 4.4 is formulated at a fixed energy E, for its proof we will need to understand $\mathbb{E}\, \mathcal{Q}^{(\alpha)}(\mathbf{x}(t), E')$ for nearby energies E' as well, which explains the introduction of E' in (4.35).

Proof. First we show that the summation over all indices i, j in the definition of $\mathcal{Q}$ (4.15) can be restricted to the interval $I(\delta)$. This directly follows from the rigidity of eigenvalues $\mathbf{x}(t)$ and from the bound (4.32): there exists some $\delta_c > 0$ (here we use subscript c for cutoff) such that for $0 < \delta \leqslant \delta_c$ and $\tau \leqslant \delta/5$, we have

$$\max_{|E'-E|\leqslant N^{-1+4\tau}} \left| \mathcal{Q}^{(\alpha)}(\mathbf{x}(t), E') - \sum_{i,j\in I(\delta)} Q^{(\alpha)}\Big(N(x_i(t)-E'), \quad N(x_j(t)-x_i(t))\Big) \right| \leqslant d_\alpha N^{-\delta/2} \tag{4.36}$$

holds with probability greater than $1 - N^{-10}$ for large enough N independent of α.

With Theorem 3.2, and from the derivative estimate from (3.31), one can easily check that there exists some constants δ_h ("h" stands for homogenization), and τ_0 such that (3.33) holds for any $0 < \tau \leqslant \tau_0$ and δ_1, δ_2, $\delta_3 \leqslant \delta_h$ and we also have

$$|p_t(\gamma_i, \gamma_j) - p_t(\gamma_{i+1}, \gamma_j)| \leqslant N^{-3\delta_h}, \qquad \forall\, i \in I(\delta_h), \quad 1 \leqslant j \leqslant N. \tag{4.37}$$

Using (4.37) and the rigidity of eigenvalues, we know that for any $0 < \tau < \tau_0$,

$$\mathbb{P}\left(\max_{i,j\in I(\delta_h)} |(\Psi_{t-t_0}(\mathbf{x}(t_0)-\boldsymbol{\gamma}))_i - (\Psi_{t-t_0}(\mathbf{x}(t_0)-\boldsymbol{\gamma}))_j| \geqslant N^{-1-\delta_h} \right) \leqslant N^{-10}. \tag{4.38}$$

Define $\mathcal{Q}^{(\alpha)}(\mathbf{x}, E)$ as in (4.15) with $Q^{(\alpha)}$ replacing Q. Combining (3.33), (4.38) and (4.36), and using rigidity and (4.32), we obtain that for any δ: $0 < \delta \leqslant \delta_Q := \min(\delta_c, \delta_h)/3$ and $0 < \tau < \min(\delta/5, \tau_0)$,

$$\max_{|E'-E|\leqslant N^{4\tau-1}} \left| \mathcal{Q}^{(\alpha)}(\mathbf{x}(t), E') - \sum_{i,j\in I(\delta)} Q^{(\alpha)}\Big(N(y_i(t)-E') + \xi_{t-t_0}^{\mathbf{x}(t_0)} - \xi_{t-t_0}^{\mathbf{y}(t_0)}, \quad N(y_j(t)-y_i(t))\Big) \right| \leqslant 2N^{-\delta/2} d_{\alpha+1} \tag{4.39}$$

holds with probability larger than $1 - 2N^{-\delta}$ for large enough N independent of α. Here for the second variable of $Q^{(\alpha)}$ we first use (3.33), for any $i, j \in I(\delta)$,

$$\begin{aligned} x_j(t) - x_i(t) =& y_j(t) - y_i(t) + (\Psi_{t-t_0}\mathbf{x})_j - (\Psi_{t-t_0}\mathbf{x})_i - (\Psi_{t-t_0}\mathbf{y})_j + (\Psi_{t-t_0}\mathbf{y})_i + O(N^{-1-\delta}) \\ =& y_j(t) - y_i(t) + (\Psi_{t-t_0}(\mathbf{x}-\boldsymbol{\gamma}))_j - (\Psi_{t-t_0}(\mathbf{x}-\boldsymbol{\gamma}))_i - (\Psi_{t-t_0}(\mathbf{y}-\boldsymbol{\gamma}))_j + (\Psi_{t-t_0}(\mathbf{y}-\boldsymbol{\gamma}))_i + O(N^{-1-\delta}) \\ =& y_j(t) - y_i(t) + O(N^{-1-\delta}), \end{aligned}$$

with the shorthand writing $\mathbf{x} = \mathbf{x}(t_0)$, $\mathbf{y} = \mathbf{y}(t_0)$, where in the second step we smuggled in the γ's and in the last step we used (4.38). Similar argument applies to the first variable of $Q^{(\alpha)}$.

On the complement event of probability at most $2N^{-\delta}$ but still on the event where the rigidity holds, we use that for any fixed τ, δ and $\xi > 0$,

$$\left| \mathcal{Q}^{(\alpha)}(\mathbf{x}(t), E') \right| + \left| \sum_{i,j\in I(\delta)} Q^{(\alpha)}\Big(N(y_i(t)-E') + \xi_{t-t_0}^{\mathbf{x}(t_0)} - \xi_{t-t_0}^{\mathbf{y}(t_0)}, \quad N(y_j(t)-y_i(t))\Big) \right| \leqslant d_\alpha N^{\xi}$$

holds for all $E' : |E' - E| \leqslant N^{4\tau-1}$. Finally, on the event where the rigidity does not hold, we can estimate Q^α by maximum norm; the contribution of this event is still negligible in the expectation. Together with (4.39), we obtain (4.35) and complete the proof of Lemma 4.7. ☐

To understand the second term in (4.35), we define a (non-random) function F as follows:

$$F(a) := \mathbb{E}^{\mu_G} \sum_{i,j\in I(\delta_Q)} Q\Big(N(y_i(t)-E) + a - \xi_{t-t_0}^{\mathbf{y}(t_0)}, \quad N(y_j(t)-y_i(t))\Big), \tag{4.40}$$

and we always assume $\delta_Q \leqslant 10^{-4}$. We can now rewrite (4.35) as follows: for $\tau < \min(\delta/5, \tau_0)$

$$\max_{|h|\leqslant N^{4\tau}} \left|\mathbb{E}^{\mu_W}\ \mathcal{Q}^{(\alpha)}(\mathbf{x}(t), E+h/N) - \mathbb{E}^{\mu_W}\ F^{(\alpha)}(\xi^{\mathbf{x}(t_0)}_{t-t_0} - h)\right| \leqslant 3N^{-\delta_Q/2} d_{\alpha+1}. \tag{4.41}$$

We do not have a direct understanding of $\xi^{\mathbf{x}(t_0)}_{t-t_0}$; although it concerns local statistics on a relatively large mesoscopic scale $t - t_0 \sim N^{-\tau}$, but in (4.41) we would need it with a precision that cannot be obtained from the available local semicircle laws for Wigner matrices. The key observation is that F is essentially a constant function, so the actual distribution of $\xi^{\mathbf{x}(t_0)}_{t-t_0}$ does not matter. The following lemma formalizes the statement that F is essentially a constant:

Lemma 4.8. *With the choice* $\delta_0 := \min(\delta_Q/3, \tau_0)$ *and* $\tau < \frac{\delta_0}{m^2+1}$, *we have*

$$F(a) - F(0) = O(d_2 N^{-\tau}), \quad \forall\, a:\ |a| \leqslant N^{4\tau}. \tag{4.42}$$

We first prove Lemma 4.4 assuming that Lemma 4.8 holds and then we will prove Lemma 4.8 in the next Section 4.3. Using rigidity for $\mathbf{x}$ and the fact that

$$p_t(\gamma_i, \gamma_j) \leqslant \frac{Ct}{t^2 + (\gamma_i - \gamma_j)^2}, \quad i \in I(\delta_Q), \quad 1 \leqslant j \leqslant N,$$

(see (3.29)), we obtain that for any $0 < \tau < \tau_0$ and $\xi > 0$,

$$\mathbb{P}\left(\max_{i\in I(\delta_Q)} |(\Psi_{t-t_0}\mathbf{x})_i - (\Psi_{t-t_0}\boldsymbol{\gamma})_i| \geqslant N^{-1+\xi}\right) \leqslant N^{-10}. \tag{4.43}$$

Choosing ξ small enough in (4.43), we have

$$|\xi^{\mathbf{x}(t_0)}_{t-t_0}| \leqslant N^{\tau}. \tag{4.44}$$

Hence for $\tau < \frac{\delta_0}{m^2+1}$, and $|E' - E| \leqslant N^{4\tau-1}$ we have

$$\mathbb{E}^{\mu_W}\ \mathcal{Q}(\mathbf{x}(t), E') = \mathbb{E}^{\mu_W}\ F(\xi^{\mathbf{x}(t_0)}_{t-t_0}) + O(N^{-\tau} d_1) = F(0) + O(N^{-\tau} d_2). \tag{4.45}$$

In the first step we used (4.41) and in the second we used (4.42) and (4.44). This implies that, up to a negligible error, the left side of the last equation is independent of the specific initial Wigner ensemble μ_W, in particular, it is the same as for the Gaussian ensemble, i.e. μ_G. Since in the Gaussian case, we have $\mathbf{x}(t) \overset{d}{\sim} \mathbf{x}^G$, this proves (4.18) and completes the proof of Lemma 4.4. □

4.3 Constantness of F: Proof of Lemma 4.8. Notice that F is defined exclusively by the Gaussian ensemble, so the proof of Lemma 4.8 will be a Gaussian calculation where additional tools are available.

Step 1: Apriori bounds on F. For convenience, we define

$$F_h(a) := F(a-h) - F(a). \tag{4.46}$$

By definition, for the α-th derivative of $F(a)$, we have

$$F^{(\alpha)}(a) := \mathbb{E} \sum_{i,j\in I(\delta_Q)} Q^{(\alpha)}\Big(N(y_i(t)-E) + a - \xi^{\mathbf{y}(t_0)}_{t-t_0}, \quad N(y_j(t)-y_i(t))\Big), \quad F^{(\alpha)}_h(a) := F^{(\alpha)}(a-h) - F^{(\alpha)}(a). \tag{4.47}$$

It follows from (4.41) that for $\tau < \min(\delta_Q/5, \tau_0)$

$$\left|\mathbb{E}^{\mu_W}\mathcal{Q}^{(\alpha)}\big(\mathbf{x}(t), E - rN^{-1}\big) - \mathbb{E}^{\mu_W}\mathcal{Q}^{(\alpha)}\big(\mathbf{x}(t), E + hN^{-1} - rN^{-1}\big) - \mathbb{E}^{\mu_W} F^{(\alpha)}_h(\xi^{\mathbf{x}(t_0)}_{t-t_0} + r)\right| \leqslant 6 d_{\alpha+1} N^{-\delta_Q/2} \tag{4.48}$$

uniformly holds for $\alpha \geqslant 0$ and $|h|$, $|r| \leqslant N^{4\tau}$.

The following lemma provides an a priori bound on the derivatives of F.

Lemma 4.9. *With F defined in* (4.40)*, for any positive ξ, and $\tau < \min(\delta_Q/5, \tau_0)$*

$$\|F^{(\alpha)}\|_\infty \leqslant d_\alpha N^\xi \tag{4.49}$$

holds uniformly for $\alpha \geqslant 0$. Furthermore, uniformly for $\alpha \geqslant 0$ we have

$$\mathbf{1}(|a| \geqslant N^{2\delta_Q})|F^{(\alpha)}(a)| \leqslant |a|^{-2} d_\alpha N^\xi. \tag{4.50}$$

Proof. For any fixed $a > 0$, we define a subset of the probability space $\Omega_a := \Omega_{\xi,\delta_Q,a}$. If $|a| \leqslant 3N^2$ then Ω_a is the event such that

$$\max_{i \in I(\delta_Q)} |\gamma_i - y_i(t)| \leqslant N^{-1+\xi}, \qquad |\xi^{\mathbf{y}(t_0)}_{t-t_0}| \leqslant N^{\delta_Q}, \quad \text{and} \quad \max_i |y_i(t)| \leqslant N^2$$

hold. If $|a| \geqslant 3N^2$, then let the event Ω_a be the set on which

$$\max_i |y_i(t)| \leqslant \frac{|a|}{3N}$$

holds. Note that in the second case, the upper bound of y_i's implies $|\xi^{\mathbf{y}(t_0)}_{t-t_0}| \leqslant |a|/3$. Since $\sum_i y_i^2(t) = \operatorname{Tr} H(t)^2 = \sum_{ij} |h_{ij}|^2 \sim N^{-1}\chi^2_{N^2}$, where $\chi^2_{N^2}$ is chi-square distribution with N^2 degrees of freedom, one can easily check that $\chi^2_{N^2}$ is smaller than $2N^2$ with a very high probability, and the probability density decay faster than polynomials. Together with rigidity, and (4.43), it implies that

$$\mathbb{P}(\Omega_a^c) \leqslant \min(N^{-10}, N^{-5}a^{-2}). \tag{4.51}$$

By the definition of Ω_a and (4.32), for any $\alpha \geqslant 0$, $\xi > 0$ and $|a| \leqslant 3N^2$, we have

$$\begin{aligned} &\left|\mathbb{E}\mathbf{1}(\Omega_a) \sum_{i,j \in I(\delta_Q)} Q^{(\alpha)}\Big(N(y_i(t) - E) + a - \xi^{\mathbf{y}(t_0)}_{t-t_0}, N(y_j(t) - y_i(t))\Big)\right| \\ &\qquad\qquad \leqslant \max_{b,b' \in \mathbb{R}} \sup_{\omega \in \Omega_a} \sum_{i \in I(\delta_Q)} \left|Q^{(\alpha)}\left(N(y_i(t) - E) + b,\ b'\right)\right| N^{2\xi} \leqslant\ d_\alpha N^{4\xi}. \end{aligned} \tag{4.52}$$

One can easily obtain the same bound for $|a| \geqslant 3N^2$, since in that case $|N(y_i(t) - E) + a - \xi^{\mathbf{y}(t_0)}_{t-t_0}| \geqslant |a|/10$ on the event Ω_a.

On the other hand, the contribution from Ω^c to $F^{(\alpha)}(a)$ is negligible thanks to (4.51). Hence together with (4.52), we obtain (4.49). Similarly, with (4.32) and $|\xi^{\mathbf{y}(t_0)}_{t-t_0}| \leqslant N^{\delta_Q}$, we have (4.50). □

Step 2. Estimating F with a Gaussian convolution. In order to show that $F_h(a)$ is negligible, we first prove that its convolution with a Gaussian kernel is small (and in Step 3 below we remove this convolution). This is formulated in Lemma 4.10 below. We cannot prove this result directly, but we can show that $\mathbb{E}F_h(X)$ is small, where X is a random variable close to a Gaussian. The key is to choose the random variable X appropriately: it will be the mesoscopic statistics $\xi^{\mathbf{x}}_{t-t_0}$ defined in (4.34) but applied to the case where $\mathbf{x}$ is distributed by GOE. On one hand, by going back to the homogenization result, we show that $\mathbb{E}F_h(\xi^{\mathbf{x}}_{t-t_0})$ is small, this will be formulated in (4.56) below. On the other hand, by using the Gaussian fluctuation of mesoscopic eigenvalue statistics, we show that $\xi^{\mathbf{x}}_{t-t_0}$ is close to a Gaussian random variable, this will follow from the combination of (4.59) and (4.64) below. Now we explain these two ingredients in detail.

The homogenization results in the form (4.41) and (4.48) hold any Wigner ensemble $\mathbf{x}(0)$. In particular, they also hold for the case $\mu_W = \mu_G$. To avoid confusion with the other Gaussian ensemble denoted by $\mathbf{y}$ earlier, when taking μ_W to be μ_G we denote the eigenvalues by $\mathbf{z}$ instead of $\mathbf{x}$ in this argument. Since for any $t > 0$, the probability measure of $\mathbf{z}(t)$ is also μ_G, then for any $|E| \leqslant 2 - \kappa/2$ (for brevity we write $\rho_2^{(N,G)}$

instead of $\rho_2^{(N,\mathrm{GOE})}$ and similarly for the limiting correlation functions)

$$\begin{aligned}\mathbb{E}^{\mu_G}\mathcal{Q}^{(\alpha)}\big(\mathbf{z}(t),E\big)&=\int Q^{(\alpha)}\left(u_1,u_2-u_1\right)\rho_2^{(N,G)}(E+\frac{\mathbf{u}}{N})\mathrm{d}\mathbf{u},\quad \mathbf{u}=(u_1,u_2)\\ &=\sum_{n\in\mathbb{Z}}\int_{|u_1|\leqslant\frac12}Q^{(\alpha)}\left(n+u_1,u_2-u_1\right)\rho_2^{(N,G)}(E_n+\frac{\mathbf{u}}{N})\mathrm{d}\mathbf{u},\quad E_n=E+nN^{-1}\\ &=\sum_{|n|\leqslant N^{1/2}}\int_{|u_1|\leqslant\frac12,|u_2-u_1|\leqslant L}Q^{(\alpha)}\left(n+u_1,u_2-u_1\right)\rho_2^{(N,G)}(E_n+\frac{\mathbf{u}}{N})\mathrm{d}\mathbf{u}+O(d_\alpha N^{-1/2+2\xi})\end{aligned}\tag{4.53}$$

for any fixed $\xi>0$. Here for the last line, we used (4.32) and rigidity of eigenvalues. It follows from Lemma 4.3 (with choosing the compact set $\{\mathbf{u}:|u_1|\leqslant\frac12,|u_2-u_1|\leqslant L\}\times\{x:|x|\leqslant 2-\kappa/3\}$) that the last line of (4.53) equals

$$\sum_{|n|\leqslant N^{1/2}}\int_{|u_1|\leqslant\frac12,|u_2-u_1|\leqslant L}Q^{(\alpha)}\left(n+u_1,u_2-u_1\right)\varrho(E_n)^2\rho_2^{(G)}\big(\varrho(E_n)\mathbf{u}\big)\mathrm{d}\mathbf{u}+O(d_\alpha N^{-1/2+2\xi})\tag{4.54}$$

where C depends on κ and L. For $|h|$, $|r|\leqslant N^{1/2}$, we define the notations

$$E^*=E_0-rN^{-1},\quad E^{**}=E_0+hN^{-1}-rN^{-1},\quad E_n^*=E^*+n,\quad E_n^{**}=E^{**}+n.$$

It is well known from the explicit formula that $\rho_2^{(G)}(\mathbf{v})$ is uniformly smooth on any compact support. Then

$$\varrho(E_n^*)^2\rho_2^{(G)}(\varrho(E_n^*)\mathbf{u})-\varrho(E_n^{**})^2\rho_2^{(G)}(\varrho(E_n^{**})\mathbf{u})=O(d_\alpha N^{-1/2}).$$

Together with (4.54) and (4.53) we obtain that

$$\Big|\mathbb{E}^{\mu_G}\mathcal{Q}^{(\alpha)}\big(\mathbf{z}(t),E-rN^{-1}\big)-\mathbb{E}^{\mu_G}\mathcal{Q}^{(\alpha)}\big(\mathbf{z}(t),E+hN^{-1}-rN^{-1}\big)\Big|\leqslant Cd_\alpha N^{-1/2+\xi}\leqslant d_\alpha N^{-1/3}\tag{4.55}$$

uniformly holds for $\alpha\geqslant 0$ and $|h|$, $|r|\leqslant N^{4\tau}$.

We remark that one can also prove (4.55) directly from (4.54) without using the smoothness of $\rho_2^{(G)}(\mathbf{v})$ but using a version of (4.32) for $\partial_y Q^{(\alpha)}$. It requires $\widehat{Q}\in W^{3,\infty}$, so it can be implemented by increasing the regularity condition from $W^{2,\infty}$ to $W^{3,\infty}$ from the beginning of the proof. Therefore, with (4.48) applied to μ_G instead of μ_W, we have

$$\Big|\mathbb{E}^{\mu_G}F_h^{(\alpha)}(\xi_{t-t_0}^{\mathbf{z}(t_0)}+r)\Big|\leqslant 7d_{\alpha+1}N^{-\delta_Q/2}\tag{4.56}$$

for any $0<\tau<\min(\delta_Q/5,\tau_0)$.

The next ingredient is to show that $\xi_{t-t_0}^{\mathbf{z}(t_0)}$ is close to a Gaussian random variable. Recall $\xi_{t-t_0}^{\mathbf{z}(t_0)}$ is defined as

$$\xi_{t-t_0}^{\mathbf{z}(t_0)}=\sum_k p_{t-t_0}(\gamma_{i_0},\gamma_k)\left(z_k(t_0)-\gamma_k\right),\quad i_0:=\min\{i:\ \gamma_i\geqslant E\}.$$

The kernel $p_s(x,y)$ is originally defined on $[-2,2]^2$; we now extend it linearly to a larger set in the second variable so that it remains a differentiable function. For $|\gamma|\geqslant 2$, we simply define $p_s(\gamma_{i_0},\gamma)$ such that $\partial_\gamma p_s(\gamma_{i_0},\gamma)=\partial_\gamma p_s(\gamma_{i_0},\pm 2)$. We also define $P_s:\mathbb{R}\to\mathbb{R}$ as a function such that

$$P_s(\gamma):=\int_{\gamma_{i_0}}^{\gamma}p_s(\gamma_{i_0},x)\mathrm{d}x\qquad\text{for}\quad|\gamma|\leqslant 3\tag{4.57}$$

and $\operatorname{supp}P_s=[-4,4]$ and $|P_s''(\gamma)|\leqslant C$ for $2\leqslant|\gamma|\leqslant 4$. With lemma lem:diagonalization on p_s, it is easy to check that for $i_0:i_0\sim N$, and $N-i_0\sim N$, and $s\ll 1$,

$$\|P_s\|_\infty\leqslant C,\quad {P'}_s(\gamma)\leqslant\frac{Cs}{s^2+(\gamma_{i_0}-\gamma)^2},\quad {P''}_s(\gamma)\leqslant\frac{Cs|\gamma_{i_0}-\gamma|}{s^4+(\gamma_{i_0}-\gamma)^4},\quad\gamma\in\mathbb{R}.\tag{4.58}$$

Then with (4.58), rigidity of eigenvalues $\mathbf{z}(t_0)$ and mean value theorem, for any $\xi > 0$, we have

$$\mathbb{P}\left(\left|\xi_s^{\mathbf{z}(t_0)} - \zeta_s^{\mathbf{z}(t_0)}\right| \geqslant N^{-1+\xi}\right) \leqslant N^{-10}, \quad \zeta_s^{\mathbf{z}(t_0)} := \sum_j \Big[P_s\big(z_j(t_0)\big) - P_s(\gamma_j)\Big], \qquad \forall s \ll 1. \tag{4.59}$$

Combining (4.59), (4.49) and (4.56), then we obtain that for any fixed $\tau < \min(\delta_Q/5, \tau_0)$

$$\left|\mathbb{E}^{\mu_G} F_h^{(\alpha)}(\zeta_{t-t_0}^{\mathbf{z}(t_0)} + r)\right| \leqslant 8 d_{\alpha+1} N^{-\delta_Q/2} \tag{4.60}$$

uniformly holds for $\alpha \geqslant 0$ and $|h|$, $|r| \leqslant N^{4\tau}$.

The characteristic function of linear statistics of $z_j(t_0)$ in the form

$$\sum_j P_{t-t_0}\big(z_j(t_0)\big) - \int_{-2}^{2} P_{t-t_0}(s)\varrho(s)\mathrm{d}s \tag{4.61}$$

will be analyzed in Section 5 in details. The main result (Theorem 5.4) states that this linear statistics is asymptotically Gaussian with parameters (expectation and variance) expressed as certain functionals of P_{t-t_0}. These functionals are somewhat complicated and will be defined later right above (5.8). With (4.58), a simple calculation gives that their values on P_{t-t_0} are given by

$$\sigma^2(P_{t-t_0}) = \tau \log N + \mathrm{o}(\log N), \quad \delta(P_{t-t_0}) = O(1), \quad \varepsilon(P_{t-t_0}) = O(N^{2\tau}), \qquad t - t_0 \approx N^{-\tau}.$$

With these values, Theorem 5.4 states that

$$\mathbb{E}^{\mu_G} \exp\left(i\lambda \zeta_{t-t_0}^{\mathbf{z}(t_0)}\right) = e^{-\frac{\lambda^2}{2}\sigma(P_{t-t_0})^2 + \mathrm{i}\lambda\left(\delta(P_{t-t_0}) + \widetilde{\delta}(P_{t-t_0})\right)} + \mathrm{O}\left(N^{-1/100}\right), \tag{4.62}$$

for $|\lambda| \leqslant (2\tau)^{-1/2}$, where $\widetilde{\delta}(P_s)$ is defined as

$$\widetilde{\delta}(P_s) := \sum_j P_s(\gamma_j) - \int_{-2}^{2} \varrho(u) P_s(u)\mathrm{d}u$$

to account for the difference between (4.61) and the definition of $\zeta_{t-t_0}^{\mathbf{z}(t_0)}$ in (4.59). By a Riemann sum approximation, one can easily obtain $\widetilde{\delta}(P_{t-t_0}) = O(1)$. Theorem 5.4 concerns only the small λ regime; but Lemma 5.6 complements it in the regime $(2\tau)^{-1/2} \leqslant |\lambda| \leqslant N^{1/10}$ with a crude estimate of order $N^{-1/100}$. Note that in this regime and for small τ the first term in the r.h.s. of (4.62) is smaller than $N^{-1/100}$, so (4.62) holds throughout the regime $|\lambda| \leqslant N^{1/10}$.

We now define ζ as a new Gaussian random variable with expectation $\delta(P_{t-t_0}) + \widetilde{\delta}(P_{t-t_0})$ and variance $\sigma(P_{t-t_0})$:

$$\zeta \sim \mathcal{N}\Big(\delta(P_{t-t_0}) + \widetilde{\delta}(P_{t-t_0}),\ \sigma(P_{t-t_0})\Big). \tag{4.63}$$

Using (4.62), the distribution of ζ is close to that of $\zeta_{t-t_0}^{\mathbf{z}(t_0)}$ in the following way;

$$\mathbb{E}^{\mu_G} \exp\left(i\lambda\, \zeta_{t-t_0}^{\mathbf{z}(t_0)}\right) = \mathbb{E}\exp\left(i\lambda\,\zeta\right) + \mathrm{O}\left(N^{-1/100}\right) \tag{4.64}$$

for $|\lambda| \leqslant N^{1/10}$. Finally, we use (4.64) to replace $\zeta_{t-t_0}^{\mathbf{z}(t_0)}$ with ζ in the bound (4.60). This gives the following main result of Step 2.

Lemma 4.10. *Define ζ as in* (4.63) *with $t = N^{-\tau}$ and $\tau < \min(\delta_Q/5, \tau_0)$. Then*

$$\left|\mathbb{E} F_h^{(\alpha)}(\zeta + r)\right| \leqslant C d_{\alpha+2} N^{-\delta_Q/2} \tag{4.65}$$

uniformly holds for $\alpha \geqslant 0$ and $|h|$, $|r| \leqslant N^{4\tau}$, where C is independent of α, h and r.

Proof. We define

$$F_{h,r,\alpha}(a) := F_h^{(\alpha)}(r-a).$$

It follows from (4.64) that

$$\left|\mathbb{E}F_h^{(\alpha)}(\zeta+r)-\mathbb{E}F_h^{(\alpha)}(\zeta_{t-t_0}^{\mathbf{z}(t_0)}+r)\right| \leqslant CN^{-1/100}\int_{|p|\leqslant N^{1/10}}\left|(\widehat{F_{h,r,\alpha}})(p)\right|\mathrm{d}p+\int_{|p|\geqslant N^{1/10}}\left|(\widehat{F_{h,r,\alpha}})(p)\right|\mathrm{d}p. \tag{4.66}$$

For the last term, using Lemma 4.9, we have

$$\left|(\widehat{F_{h,r,\alpha}})(p)\right| = \frac{1}{|p|^2}\left|\int\left(F_{h,r,\alpha}\right)''(a)e^{-ipa}\mathrm{d}a\right| \leqslant Cd_{\alpha+2}N^{3\delta_Q}|p|^{-2}.$$

Similarly, we have $\|\widehat{F_{h,r,\alpha}}\|_\infty \leqslant Cd_{\alpha+2}N^{3\delta_Q}$. Then

$$\|(\widehat{F_{h,r,\alpha}})\|_1 \leqslant Cd_{\alpha+2}N^{3\delta_Q}, \qquad \int_{|p|\geqslant N^{1/10}}\left|(\widehat{F_{h,r,\alpha}})(p)\right|\mathrm{d}p \leqslant Cd_{\alpha+2}N^{-1/10+3\delta_Q}. \tag{4.67}$$

Together with (4.66) and (4.60), we obtain (4.65) and complete the proof of Lemma 4.10. ☐

Step 3: Removal of the Gaussian convolution. The expectation w.r.t. the Gaussian variable ζ in (4.65) can be viewed as evolving the standard heat equation on the function F_h and its derivatives up to times given by the variance $\sigma=\sigma(P_{t-t_0})$. We will thus show that from the estimates on the heat evolution on F_h given in (4.65) we have effective estimates on the function F_h. This is similar to backward uniqueness of the heat equation for analytic functions, supplemented by precise bounds. This step is the reason why we need to consider test functions with compact Fourier support in Section 4.2.

Recall $F^{(\alpha)}$ are uniformly bounded in (4.49). Together with (4.32), we can define:

$$U(r,t) := \sum_{\alpha=0}^{\infty}\frac{s^\alpha}{\alpha!}F_h^{(2\alpha)}(r), \qquad r,s\in\mathbb{R}.$$

With (4.49) and the estimate $d_\alpha \leqslant C_m^\alpha$ from (4.32), this power series is convergent and termwise differentiable in both variables arbitrary many times. It is easy to check that

$$\partial_s U(r,s) = \partial_r^2 U(r,s).$$

Thus $U(r,s)$ is the solution of the heat equation with an initial condition $U(r,0)=F_h(r)$ that is analytic in a strip around the real axis. Therefore the usual semigroup property extends to negative times as well and for any $\mu,\sigma>0$ we have

$$F_h(r)=U(r,0)=\int U\left(r-r',-\sigma^2\right)\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(r')^2}{\sigma^2}}\mathrm{d}r' = \int U\left(r-r'+\mu,-\sigma^2\right)\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(r'-\mu)^2}{\sigma^2}}\mathrm{d}r'.$$

Recall ζ defined in (4.63). Choosing $\mu=\delta(P_{t-t_0})+\widetilde{\delta}(P_{t-t_0})$ and $\sigma=\sigma(P_{t-t_0})$, we obtain

$$F_h(r)=\sum_{\alpha=1}^{\infty}\frac{\Big(-\sigma^2(P_{t-t_0})\Big)^\alpha}{\alpha!}\mathbb{E}F_h^{(2\alpha)}\big(\zeta+\delta(P_{t-t_0})+\widetilde{\delta}(P_{t-t_0})\big), \quad r\in\mathbb{R}.$$

Using (4.65) and (4.32), for $\tau<\min(\delta_Q/5,\tau_0)$, we obtain that

$$|F_h(r)| \leqslant Cd_2N^{m^2\tau-\delta_Q/2} \tag{4.68}$$

uniformly holds for $|h|$, $|r|\leqslant N^{4\tau}$. Let $\delta_0:=\min(\delta_Q/3,\tau_0)$. Inserting (4.68) into (4.48) with $r=\alpha=0$, we obtain (4.42) and complete the proof of Lemma 4.8.

5 Mesoscopic fluctuations for Gaussian ensembles

This section follows Johansson's method [24] to prove Gaussian fluctuations of linear statistics at any mesoscopic scale $N^{-1+\varepsilon}$. An important ingredient is the optimal rigidity of the eigenvalues obtained in [2–4], allowing the choice of any $\varepsilon > 0$. Moreover, while limiting Gaussian behaviour of linear statistics is obtained in [24] by characterizing the Laplace transform, in this section we choose to work with the Fourier transform, for the sake of better estimates on the speed of convergence. This implies technical complications: the partition function may vanish.

Consider the probability measure

$$\mathrm{d}\mu(\mathbf{y}) := \frac{1}{Z} \prod_{1 \leqslant k < \ell \leqslant N} |y_k - y_\ell|^\beta e^{-\beta \frac{N}{4} \sum_{k=1}^N y_k^2} \mathrm{d}\mathbf{y} \tag{5.1}$$

on the simplex $y_1 < \cdots < y_N$. For a given function $f : \mathbb{R} \to \mathbb{R}$ we consider the general linear statistics

$$S_N(f) := \sum_{k=1}^N f(y_k) - N \int f(s)\varrho(s)\mathrm{d}s,$$

and we are interested in the Fourier transform

$$Z(\lambda) := Z_N(\lambda) = \mathbb{E}_\mu(e^{\mathrm{i}\lambda S_N(f)}).$$

We will need the following complex measure, modification of the GOE: assuming $Z(\lambda) \neq 0$, we define

$$\mathrm{d}\mu^\lambda(\mathbf{y}) := \frac{e^{\mathrm{i}\lambda S_N(f)}}{Z(\lambda)} \mathrm{d}\mu(\mathbf{y}).$$

The following lemma about the total variation of μ^λ is elementary.

Lemma 5.1. *If $Z(\lambda) \neq 0$, for any measurable A we have $|\mu^\lambda|(A) \leqslant \frac{\mu(A)}{|Z(\lambda)|}$.*

We will use the following rigidity estimate, proved for a wide class of β-ensembles including the quadratic beta ensemble in [4]. We use the notation $\widehat{k} = \min(k, N+1-k)$.

Lemma 5.2. *For any $\xi > 0$ there exists $c > 0$ such that for any $N \geqslant 1$ and $k \in [\![1, N]\!]$ we have*

$$\mu\left(|y_k - \gamma_k| > N^{-\frac{2}{3}+\xi}(\widehat{k})^{-\frac{1}{3}}\right) \leqslant e^{-N^c}.$$

As an easy consequence of Lemmas 5.1 and 5.2, rigidity estimates for μ yield rigidity estimates for μ^λ, at the expense of a factor $Z(\lambda)^{-1}$. It also gives estimates on the 1-point function and variances for the measure μ^λ. We recall the definition of the correlation functions from (2.2), in particular the 1-point function satisfies

$$N \int h(s)\varrho_1^{(N,\lambda)}(s)\mathrm{d}s = \mathbb{E}_{\mu^\lambda}\left(\sum_k h(y_k)\right) = \int \sum_k h(y_k)\mathrm{d}\mu^\lambda(\mathbf{y}),$$

for any continuous bounded test-function h. We also define the complex variance by $\mathrm{Var}^{\mu^\lambda}(X) = \mathbb{E}^{\mu^\lambda}\left(X^2\right) - \left(\mathbb{E}^{\mu^\lambda} X\right)^2$. We introduce the notation for the Stieltjes transform of the empirical measure, and its expectation w.r.t. μ^λ, by

$$s_N(z) := \frac{1}{N}\sum_k \frac{1}{z - y_k}, \qquad m_{N,\lambda}(z) := \mathbb{E}^{\mu^\lambda}(s_N(z)).$$

We will also use the following notation for the Stieltjes transform of the semicircle distribution :

$$m(z) := \int \frac{\varrho(s)}{z-s}\mathrm{d}s = \frac{z - \sqrt{z^2-4}}{2},$$

where the square root is chosen so that m is holomorphic on $[-2,2]^{\mathrm{c}}$ and $m(z) \to 0$ as $|z| \to \infty$.

Lemma 5.3. *Assume that $Z(\lambda)\neq 0$. For any $\xi>0$ there exists $c>0$ such that for any $N\geqslant 1$ and $k\in[\![1,N]\!]$ we have*

$$|\mu^\lambda|\left(|y_k-\gamma_k|>N^{-\frac{2}{3}+\xi}(\widehat{k})^{-\frac{1}{3}}\right)\leqslant\frac{e^{-N^c}}{|Z(\lambda)|}. \tag{5.2}$$

As a consequence, the following estimates hold: for fixed $\xi>0$, for any $0<|\eta|<1$ (remember $z=E+\mathrm{i}\eta$), $N\geqslant 1$, and $f\in\mathscr{C}^2(\mathbb{R})$ we have

$$\int_{\mathbb{R}}\frac{f'(s)}{z-s}\varrho_1^{(N,\lambda)}(s)\mathrm{d}s-\int_{\mathbb{R}}\frac{f'(s)}{z-s}\varrho(s)\mathrm{d}s=\frac{N^{-1+\xi}}{|Z(\lambda)|}\,\mathrm{O}\left(\int\frac{|f''(s)|}{|z-s|}\mathrm{d}s+\int\frac{|f'(s)|}{|z-s|^2}\mathrm{d}s+\frac{e^{-N^c}}{\eta^2}(\|f'\|_\infty+\|f''\|_\infty)\right), \tag{5.3}$$

$$m_{N,\lambda}(z)-m(z)=\mathrm{O}\left(\frac{N^{-1+\xi}}{|\eta Z(\lambda)|}\right), \tag{5.4}$$

$$m'_{N,\lambda}(z)-m'(z)=\mathrm{O}\left(\frac{N^{-1+\xi}}{\eta^2|Z(\lambda)|}\right), \tag{5.5}$$

$$\mathrm{Var}^{\mu^\lambda}\left(\frac{1}{N}\sum_k\frac{1}{z-y_k}\right)=\mathrm{O}\left(\frac{N^{-2+2\xi}}{\eta^2|Z(\lambda)|^2}\right). \tag{5.6}$$

Proof. The rigidity estimate (5.2) is immediate from Lemmas 5.1 and 5.2. For the proof of (5.3), we first write the left hand side of (5.3) as

$$\frac{1}{N}\sum_{k=1}^N\mathbb{E}^{\mu^\lambda}\left(\int_{\gamma_k}^{y_k}\partial_s\left(\frac{f'(s)}{z-s}\right)\mathrm{d}s\right)+\sum_{k=1}^N\int_{\gamma_{k-1}}^{\gamma_k}\mathrm{d}s\varrho(s)\int_s^{\gamma_k}\partial_u\left(\frac{f'(u)}{z-u}\right)\mathrm{d}u. \tag{5.7}$$

Let $\gamma(u):=\max\{\gamma_k:\gamma_k\leqslant u\}$. The second sum above is easily bounded by

$$\int_{-2}^2\left|\partial_u\frac{f'(u)}{z-u}\right|\mathrm{d}u\int_{\gamma(u)}^u\varrho(s)\mathrm{d}s=\mathrm{O}\left(\frac{1}{N}\int\left(\frac{|f'(s)|}{|z-s|^2}+\frac{|f''(s)|}{|z-s|}\right)\mathrm{d}s\right).$$

To bound the first term in (5.7), we first denote $A=\{\forall k\in[\![1,N]\!],\ |y_k-\gamma_k|<N^{-\frac{2}{3}+\xi}(\widehat{k})^{-\frac{1}{3}}\}$. Thanks to (5.2),

$$\frac{1}{N}\sum_{k=1}^N\mathbb{E}^{\mu^\lambda}\left(\mathbb{1}_{A^c}\int_{\gamma_k}^{y_k}\partial_s\left(\frac{f'(s)}{z-s}\right)\mathrm{d}s\right)=\mathrm{O}\left(\frac{e^{-N^c}}{|Z(\lambda)|\eta^2}(\|f'\|_\infty+\|f''\|_\infty)\right).$$

In the event A, we have

$$\frac{1}{N}\sum_{k=1}^N\mathbb{E}^{\mu^\lambda}\left(\mathbb{1}_A\int_{\gamma_k}^{y_k}\partial_s\left(\frac{f'(s)}{z-s}\right)\mathrm{d}s\right)=\mathrm{O}\left(\frac{1}{|Z(\lambda)|N}\int\left|\partial_s\frac{f'(s)}{z-s}\right|\sum_{k=1}^N\mathbb{1}_{s\in I_k}\mathrm{d}s\right)$$

where $I_k=\{s:|s-\gamma_k|\leqslant N^{-\frac{2}{3}+\xi}(\widehat{k})^{-\frac{1}{3}}\}$. This concludes the proof of (5.3) by noting that for any fixed s we have $|\{k:s\in I_k\}|\leqslant N^\xi$.

The bounds (5.4), (5.5) and (5.6) can be proved the same way, by discussing the cases A and A^{c}. For example, the considered variance can be written

$$\mathbb{E}^{\mu^\lambda}\left(\left(\frac{1}{N}\sum_k\frac{1}{z-y_k}-\frac{1}{z-\gamma_k}\right)^2\right)-\left(\mathbb{E}^{\mu^\lambda}\frac{1}{N}\sum_k\left(\frac{1}{z-y_k}-\frac{1}{z-\gamma_k}\right)\right)^2.$$

The first term can be bounded as previously and yields an error of size $\frac{N^{-2+2\xi}}{|Z(\lambda)|\eta^2}$ (in A). The second one yields the higher order error $\frac{N^{-2+2\xi}}{|Z(\lambda)|^2\eta^2}$, concluding the proof. $\square$

For the following theorem, we need the notations

$$\kappa(s) := \max\{N^{-2/3}, \min(|s-2|, |s+2|)\},$$
$$\mathrm{d}\nu(s) := \frac{1}{2}\left(\delta_{s-2} + \delta_{s+2} - \frac{1}{2\pi}\frac{\mathrm{d}s}{\sqrt{4-s^2}}\right).$$

Theorem 5.4. *Let f be a (N-dependent) real function of class $\mathscr{C}^2$ such that, for any N, we have $\|f\|_\infty < C$, $\|f'\|_\infty, \|f''\|_\infty \leqslant N^C$, $\int |f'| < C$. Let*

$$\sigma(f)^2 := \frac{1}{2\pi^2\beta}\iint_{(-2,2)^2}\left(\frac{f(x)-f(y)}{x-y}\right)^2 \frac{4-xy}{\sqrt{4-x^2}\sqrt{4-y^2}}\mathrm{d}x\mathrm{d}y,$$
$$\delta(f) := \left(\frac{2}{\beta}-1\right)\int f(s)\mathrm{d}\nu(s),$$
$$\varepsilon(f) := \left(1 + \int |f''(s)|\kappa(s)^{-\frac{1}{2}}\mathrm{d}s\right)^2.$$

For any fixed $\xi > 0$, uniformly in the set

$$\left\{\lambda \,:\, \lambda^2\sigma(f)^2 < \left|\log|N^{-1+3\xi}\varepsilon(f)|\right|\right\} \cap \{|\lambda| < N^\xi\} \tag{5.8}$$

we have

$$Z(\lambda) = \mathbb{E}^\mu\left(e^{\mathrm{i}\lambda S_N(f)}\right) = e^{-\frac{\lambda^2}{2}\sigma(f)^2 + \mathrm{i}\lambda\delta(f)} + \mathrm{O}\left(N^{-1+3\xi}\varepsilon(f)\right).$$

Proof. The main tools for the proof of this theorem are the loop equation (5.9) and the Helffer-Sjöstrand formula to go from the Stieltjes transform to any test function. To derive proper asymptotics in the loop equation, an important input is the optimal rigidity and its consequences, Lemma 5.3.

We begin with $\frac{\mathrm{d}}{\mathrm{d}\lambda}\log Z(\lambda) = \mathbb{E}^{\mu^\lambda}(\mathrm{i}S_N(f))$, and therefore want to estimate expectation of general linear statistics for the measure μ^λ. We begin with the expectation of the Stieltjes transform.

First step: analysis of the loop equation. The loop equation is a well-known algebraic identity for the expectation of the empirical measure. In our case it takes the following form:

$$(m_{N,\lambda}(z) - m(z))^2 - \sqrt{z^2-4}\,(m_{N,\lambda}(z) - m(z)) + \mathrm{i}\frac{\lambda}{\beta N}\int_{\mathbb{R}}\frac{f'(s)}{z-s}\varrho_1^{(N,\lambda)}(s)\mathrm{d}s$$
$$- \frac{1}{N}\left(\frac{2}{\beta}-1\right)m'_{N,\lambda}(z) - \mathrm{Var}^{\mu^{(\lambda)}}(s_N(z)) = 0. \tag{5.9}$$

Note that, when compared to the loop equation initiated in [24] (written in a form closer to (5.9) in [4] Section 6.2), we only consider the special case of quadratic external potential, hence extra simplifications occur. From the estimates from Lemma 5.3 and our assumptions for the theorem, the loop equation (5.9) implies that uniformly in $\eta > N^{-1+\xi}$ we have

$$X_N(z)^2 - b(z)X_N(z) + c_N(z) = \mathrm{O}(\omega_N(z)), \tag{5.10}$$
$$X_N(z) = m_{N,\lambda}(z) - m(z),$$
$$b(z) = \sqrt{z^2-4},$$
$$c_N(z) = \mathrm{i}\frac{\lambda}{\beta N}\int_{\mathbb{R}}\frac{f'(s)}{z-s}\varrho(s)\mathrm{d}s - \frac{1}{N}\left(\frac{2}{\beta}-1\right)m'(z),$$
$$\omega_N(z) = \frac{N^{-2+2\xi}}{|Z(\lambda)|^2}\left(\frac{1}{\eta}\int|f''| + \frac{1}{\eta^2}\right).$$

Let

$$\Omega_N := \{z = E + \mathrm{i}\eta : N^{\xi}\min(N^{-2/3}, N^{-1}\kappa(E)^{-1/2}) \leqslant |\eta| \leqslant 3, |E| \leqslant 4\}.$$

A simple analysis exercise shows that

$$\sup_{s\in[-2,2]} \frac{\varrho(s)}{|z-s|} \leqslant C\eta^{-1}\max(\eta, \kappa(E))^{1/2}.$$

Using this estimate together with $\int |f'| < C$ and $|\lambda| \leqslant N^{\xi}$, we have $|b(z)|^2 > cN^{\xi}|c_N(z)|$ for any $z \in \Omega_N$. We consider two cases to identify the relevant root of (5.10).

(i) If $|b(z)|^2 > N^{\xi}\omega_N(z)$, by monotonicity we also have $|b(z')|^2 > N^{\xi}\omega_N(z')$ for any $z' = E + \mathrm{i}\eta', |\eta'| > |\eta|$. Moreover, using (5.4), together with $|Z(\lambda)| > N^{-1+\frac{3\xi}{2}}$ (obtained from our assumption $|b(z)|^2 > N^{\xi}\omega_N(z)$), we have $m_{N,\lambda}(z) - m(z) \to 0$ when $|\eta|$ is of order 1.

All together, by continuity we proved that in this case, for any $z \in \Omega_N$,

$$m_{N,\lambda}(z) - m(z) = c_N(z)/b(z) + \mathrm{O}(\omega_N(z)/b(z)).$$

(ii) Assume $|b(z)|^2 \leqslant N^{\xi}\omega_N(z)$ (in particular $|\omega_N(z)| > |c_N(z)|$). Any solution of (5.10) satisfies

$$|X_N(z)| \leqslant C\max(|b(z)|, \sqrt{|c_N(z)| + |\omega_N(z)|}) \leqslant CN^{\xi/2}\sqrt{\omega_N(z)} \leqslant CN^{\xi}|\omega_N(z)/b(z)|.$$

In all cases, we therefore proved that uniformly in Ω_N we have

$$m_{N,\lambda}(z) - m(z) = \frac{c_N(z)}{b(z)} + \mathrm{O}\left(N^{\xi}\frac{\omega_N(z)}{|b(z)|}\right). \tag{5.11}$$

Second step, integration. Let $\widetilde{f}$ coincide with f on $(-3,3)$, such that $\widetilde{f}(x) = 0$ for $|x| > 4$ and $\|(f-\widetilde{f})^{(\ell)}\|_\infty < C$ for $\ell = 0, 1, 2$. From (5.2) we have

$$\mathbb{E}^{\mu^\lambda}(\mathrm{i}S_N(f)) = \mathbb{E}^{\mu^\lambda}\left(\mathrm{i}S_N(\widetilde{f})\right) + \mathrm{O}\left(\frac{e^{-N^c}}{|Z(\lambda)|}\right).$$

Let $\chi : \mathbb{R} \to \mathbb{R}^+$ be a smooth symmetric function such that $\chi(x) = 1$ for $x \in (-1,1)$ and $\chi(x) = 0$ for $|x| > 2$. By the Helffer-Sjöstrand formula [22] we have

$$\mathbb{E}^{\mu^\lambda}\left(\mathrm{i}S_N(\widetilde{f})\right) = -\frac{\mathrm{i}}{2\pi}\iint_{\mathbb{R}^2}\left(\mathrm{i}y\widetilde{f}''(x)\chi(y) + \mathrm{i}(\widetilde{f}(x) + \mathrm{i}y\widetilde{f}'(x))\chi'(y)\right)N(m_{N,\lambda}(x+\mathrm{i}y) - m(x+\mathrm{i}y))\mathrm{d}x\mathrm{d}y.$$

We now bound some error terms.

(i) Using the estimate (5.4), we have (note that $\widetilde{f}(x)\chi'(y)$ and $\widetilde{f}'(x))\chi'(y)$ both vanish for $z = x + \mathrm{i}y \notin \Omega_N$, and $f''(x)\chi(y) = f''(x)$ when $z \notin \Omega_N$)

$$\begin{aligned}&\iint_{\Omega_N^{\mathrm{c}}}\left(\mathrm{i}y\widetilde{f}''(x)\chi(y) + \mathrm{i}(\widetilde{f}(x) + \mathrm{i}y\widetilde{f}'(x))\chi'(y)\right)N(m_{N,\lambda}(x+\mathrm{i}y) - m(x+\mathrm{i}y))\mathrm{d}x\mathrm{d}y\\ &= \iint_{\Omega_N^{\mathrm{c}}}\mathrm{i}y\widetilde{f}''(x)\chi(y)N(m_{N,\lambda}(x+\mathrm{i}y) - m(x+\mathrm{i}y))\mathrm{d}x\mathrm{d}y = \mathrm{O}\left(\frac{N^{2\xi}}{N|Z(\lambda)|}\int|\widetilde{f}''(x)|\kappa(x)^{-1/2}\mathrm{d}x\right).\end{aligned}$$

(ii) A simple calculation yields (note that $|b(x+\mathrm{i}y)| > c$ when $\chi'(y) \neq 0$)

$$\iint_{\Omega_N}(|\widetilde{f}(x)| + |y\widetilde{f}'(x)|)|\chi'(y)|N\frac{|\omega_N(z)|}{|b(z)|}\mathrm{d}x\mathrm{d}y = \mathrm{O}\left(\frac{N^{-1+2\xi}}{|Z(\lambda)|^2}\left(1 + \int|f''|\right)\right).$$

Moreover,

$$\iint_{\Omega_N}|\widetilde{f}''(x)|y\chi(y)N\frac{\omega_N(z)}{|b(z)|}\mathrm{d}x\mathrm{d}y = \mathrm{O}\left(\frac{N^{-1+2\xi}}{|Z(\lambda)|^2}\left(1 + \int|f''|\right)\int|\widetilde{f}''(x)|\kappa(x)^{-1/2}\mathrm{d}x\right).$$

(iii) Finally, thanks to the easy estimate $|c_N(z)| \leqslant C(|\lambda|/(Ny)+1/(Ny))$, we have

$$\iint_{\Omega_N^c} \left(\mathrm{i}y\widetilde{f}''(x)\chi(y)+\mathrm{i}(\widetilde{f}(x)+\mathrm{i}y\widetilde{f}'(x))\chi'(y)\right) N\frac{c_N(z)}{b(z)}\mathrm{d}x\mathrm{d}y = N^{-1+\xi}\,\mathrm{O}\left(\int |\widetilde{f}''(x)|\kappa(x)^{-1/2}\mathrm{d}x\right).$$

Let

$$\widetilde{\sigma}(f)^2 := -\frac{1}{2\pi\beta}\iint_{\mathbb{R}^2}\left(\mathrm{i}y\widetilde{f}''(x)\chi(y)+\mathrm{i}(\widetilde{f}(x)+\mathrm{i}y\widetilde{f}'(x))\chi'(y)\right) b(z)^{-1}\left(\int \frac{f'(s)}{z-s}\varrho(s)\mathrm{d}s\right)\mathrm{d}x\mathrm{d}y,$$
$$\widetilde{\delta}(f) := \left(\frac{2}{\beta}-1\right)\frac{1}{2\pi}\iint_{\mathbb{R}^2}\left(\mathrm{i}y\widetilde{f}''(x)\chi(y)+\mathrm{i}(\widetilde{f}(x)+\mathrm{i}y\widetilde{f}'(x))\chi'(y)\right) b(z)^{-1}m'(z)\mathrm{d}x\mathrm{d}y.$$

Using (i), (ii) and (iii) all together, we proved that

$$\mathbb{E}^{\mu^\lambda}\left(\mathrm{i}S_N(f)\right) = -\lambda\widetilde{\sigma}(f)^2+\mathrm{i}\widetilde{\delta}(f)+\mathrm{O}\left(\frac{N^{-1+2\xi}}{|Z(\lambda)|^2}\varepsilon(f)\right).$$

Let $g(\lambda) = e^{\lambda^2\widetilde{\sigma}(f)^2-2\mathrm{i}\widetilde{\delta}(f)\lambda}Z(\lambda)^2$. The above equation implies $g'(\lambda) = e^{\lambda^2\widetilde{\sigma}(f)^2-2\mathrm{i}\widetilde{\delta}(f)\lambda}\,\mathrm{O}(N^{-1+2\xi}\varepsilon(f))$, so $g(\lambda) = 1+e^{\lambda^2\widetilde{\sigma}(f)^2}\,\mathrm{O}(N^{-1+3\xi}\varepsilon(f))$. On our set (5.8), by continuity in λ this implies

$$Z(\lambda) = e^{-\frac{\lambda^2}{2}\widetilde{\sigma}(f)^2+\mathrm{i}\lambda\widetilde{\delta}(f)}+\mathrm{O}\left(N^{-1+3\xi}\varepsilon(f)\right). \tag{5.12}$$

We now want to prove $\widetilde{\sigma}(f)^2 = \sigma(f)^2$ and $\widetilde{\delta}(f) = \delta(f)$. If f is fixed independent of N, (5.12) proves that $S_N(f)$ converges to a Gaussian random variable with variance $\widetilde{\sigma}(f)^2$ and shift $\widetilde{\delta}(f)$. Thanks to [24, Theorem 2.4] we can identify this shift: we know that $S_N(f)$ converges to a Gaussian with shift $\delta(f)$. Thanks to [26, Theorem 2], we can identify the variance: for $\beta = 1$, $S_N(f)$ converges to a Gaussian with variance $\sigma(f)^2$. This implies the identity $\widetilde{\sigma}(f)^2 = \sigma(f)^2$ and $\widetilde{\delta}(f) = \delta(f)$ for any f, and concludes the proof. □

Remark 5.5. *In the previous theorem, the error term $\varepsilon(f)$ is quadratic in $\int|f''|$, which is sufficient for our purpose, as we will apply it for f fluctuating at the mesoscopic scale $N^{-\tau}$ for some small τ.*

If one is interested in the mesoscopic statistics at scale $N^{-1+\varepsilon}$ for some small ε and the support of f is of order 1, the above reasoning fails. On the other hand, if f is supported in the bulk, with support size $(\int|f''|)^{-1}$, then by taking in the previous reasoning χ a cutoff function on scale $(\int|f''|)^{-1}$ one obtains an error linear in $\int|f''|$ instead of quadratic, which is sufficient to prove Gaussianity of $S_N(f)$ at this very small mesoscopic scale.

Assuming $\varepsilon(f)$ has size N^θ for some $\theta\in(0,1)$, Theorem 5.4 gives a very accurate control of $Z(\lambda)$ in the regime $|\lambda| \leqslant c(\theta)(\log N)^{1/2}/\sigma(f)$. The purpose of the following lemma is to get a rough polynomial bound on Z in the regime $|\lambda| > c(\theta)(\log N)^{1/2}/\sigma(f)$.

Lemma 5.6. *Let f be a (N-dependent) real function of class $\mathscr{C}^2$ such that, for any N, we have $\|f\|_\infty < C$, $\|f'\|_\infty, \|f''\|_\infty \leqslant N^C$, $\int|f'| < C$.*

Assume that $\varepsilon(f) \leqslant N^{1/2}$, $c \leqslant (\log N)^{1/2}/\sigma(f)$, and $\sigma(f) > c$. Then for any $|\lambda| \in [(\log N)^{1/2}/\sigma(f), N^{1/10}]$ we have

$$|Z(\lambda)| \leqslant C\ N^{-1/100}. \tag{5.13}$$

Proof. Without loss of generality, we can assume $\lambda > 0$. Note that for $\lambda = (\log N)^{1/2}/\sigma(f)$, from Theorem 5.4 we have $|Z(\lambda)| \leqslant N^{-1/100}$, so we only need to prove the following statement: if $\lambda\in[(\log N)^{1/2}/\sigma(f), N^{1/10}]$ and $|Z(\lambda)| > N^{-1/100}$ then

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\Re\log Z(\lambda) < 0. \tag{5.14}$$

To prove the above statement, we begin as in the proof of Theorem 5.4 with $\frac{\mathrm{d}}{\mathrm{d}\lambda}\Re\log Z(\lambda) = \Re\mathbb{E}^{\mu^\lambda}(\mathrm{i}S_N(f))$. If we repeat exactly the proof of Theorem 5.4 except that we substitute Ω_N with

$$\Omega_{N,\lambda} := \{z = E+\mathrm{i}\eta : \lambda\min(N^{-2/3}, N^{-1}\kappa(E)^{-1/2}) \leqslant |\eta| \leqslant 3, |E| \leqslant 4\}.$$

Then the following variant of (5.11) holds uniformly in $\Omega_{N,\lambda}$:

$$m_{N,\lambda}(z) - m(z) = \frac{c_N(z)}{b(z)} + \mathrm{O}\left(\lambda \frac{\omega_N(z)}{|b(z)|}\right).$$

This allows us to reproduce all error estimates (i), (ii) and (iii) in the integration step, always replacing Ω_N with $\Omega_{N,\lambda}$. We end up with

$$\Re \mathbb{E}^{\mu^\lambda}(\mathrm{i}S_N(f)) = -\lambda \widetilde{\sigma}(f) + \mathrm{O}\left(\frac{N^{-1}\lambda^2}{|Z(\lambda)|^2}\varepsilon(f)\right).$$

From our strong assumptions $\lambda \leqslant N^{1/10}$, $|Z(\lambda)| > N^{-1/100}$ and $\varepsilon(f) < N^{1/2}$ the above term is positive for large enough N. This concles the proof of (5.14) and the lemma. □

Appendix A Hölder regularity

We now explain the proof of Lemma 3.6, i.e., the Hölder regularity for (3.12). It directly follows from Theorem 10.3 of [18] after checking the conditions. We recall that the setup of [18] was the discrete equation

$$\partial_s \mathbf{v}(s) = -\mathcal{A}(s)\mathbf{v}(s), \qquad \mathcal{A}(s) = \mathcal{B}(s) + \mathcal{W}(s), \tag{A.1}$$

in a finite $I \subset [\![1, N]\!]$ of size $|I| = K$ and in a time interval $s \in [0, \sigma]$. Here $\mathcal{W}(t)$ is an diagonal operator given by $(\mathcal{W}(t)\mathbf{v})_i = W_i(t)v_i$. We will apply this result for $I = [\![1, N]\!]$, i.e. $K = N$. The time interval is $[0, \sigma] := [0, t]$. The key assumption on the coefficients B_{jk} is the following strong regularity condition (Definition 9.7 in [18]). We remind the reader that, compared with the scalings of this paper, the time in [18] is rescaled by a factor N while the coefficient B_{jk} is rescaled by a factor $1/N$. The microscopic coordinates used in [18] are chosen so that the eigenvalue spacing is of order one and the time to equilibrium is of order N. In this paper, all scalings are dictated by the original scalings of the DBM, so the following setup uses the scaling convention in this paper.

Theorem 10.3 in [18] had two conditions, called $(\mathbf{C1})_\rho$ and $(\mathbf{C2})_\xi$. The first condition is the following concept of strong regularity:

Definition A.1. *The equation*

$$\partial_t \mathbf{v}(t) = -\mathcal{B}(t)\mathbf{v}(t) \tag{A.2}$$

is called **regular** *at the space-time point (z, σ) with exponent ρ, if*

$$\sup_{0 \leqslant s \leqslant \sigma} \sup_{1 \leqslant M \leqslant N} \frac{1}{1/N + |s - \sigma|} \left| \int_s^\sigma \frac{1}{M} \sum_{i \in I\,:\,|i-z| \leqslant M} \sum_{j \in I\,:\,|j-z| \leqslant M} B_{ij}(s')\mathrm{d}s' \right| \leqslant N^{1+\rho}. \tag{A.3}$$

Furthermore, the equation is called **strongly regular** *at the space-time point (z, σ) with exponent ρ if it is regular at all points $\{z\} \times \{\sigma\Xi + \sigma\}$, where*

$$\Xi = \left\{ -2^{-m}(1 + 2^{-k}) \;:\; 0 \leqslant k, m \leqslant C \log N \right\}.$$

Strong regularity (A.3) at (z, t) with exponent 2ρ follows from (3.40) on a set $\mathcal{R}_{z,t}$ of probability at least $1 - CN^{-\rho}(\log N)^4$. Without the double supremum in (A.3) this would clearly follow from the Markov inequality and the cardinality $|\Xi| \leqslant C(\log N)^2$. However, the suprema over all s and M can be replaced by suprema over a dyadic choice of $s = 2^{-a}\sigma$, $M = 2^b$, with intgers $a, b \leqslant C \log N$, explaining the additional logarithmic factors.

The other condition, $(\mathbf{C2})_\xi$, expresses various a priori bounds on B_{ij} that follow from (3.38) and (3.39). More precisely, we need for any $0 \leqslant s \leqslant t$

$$B_{ij}(s) \geqslant \frac{N^{-\xi}}{N|i-j|^2}, \text{ for any } i, j \text{ with } \widehat{i}, \widehat{j} \geqslant cN, \tag{A.4}$$

$$\frac{\mathbf{1}(\min\{\widehat{i}, \widehat{j}\} \geqslant cN)}{CN|i-j|^2} \leqslant B_{ij}(s) \leqslant \frac{C}{N|i-j|^2}, \text{ for any } |i-j| \geqslant C'N^\xi \tag{A.5}$$

with some constants C, C', c, where recall that $\widehat{i} = \min\{i, N+1-i\}$ denotes the distance from the edge. Finally, in [18] the diagonal operator is assumed to satisfy

$$W_i(s) \leqslant \frac{N^\xi}{N\widehat{i}}, \quad \text{if} \quad \widehat{i} \geqslant N^\xi, \tag{A.6}$$

but in our application the diagonal operator is not present. Having verified these conditions (with a possible modified value of ρ), Lemma 3.6 directly follows from Theorem 10.3 of [18]. □

Appendix B Level repulsion estimate

The following level repulsion estimate is adapted from [13]. The main differences are:

(i) it is given for symmetric matrices instead of Hermitian;

(ii) we consider the generalized Wigner class instead of Wigner;

(iii) the matrix entries are smooth on scale $N^{-\tau/2}$ instead of 1.

We closely follow the method from [13], where the Hermitian case was given in details. Since the adjustment of the proof to the symmetric case requires technical changes, for the convenience of the reader, we will give the main steps of the proof and explain the modifications.

Proposition B.1. *Let H_N be a symmetric generalized Wigner matrix satisfying (2.1), and G_N a $N \times N$ GOE matrix. For any $t > 0$ we denote $\mu_1(t) \leqslant \ldots \leqslant \mu_N(t)$ the eigenvalues of $\sqrt{1-t}H_N + \sqrt{t}G_N$. Define the set*

$$\mathcal{G}_\xi = \left\{|\mu_i - \gamma_i| \leqslant N^{-2/3+\xi}(\hat{i})^{-1/3} \text{ for all } \ i \in [\![1, N]\!]\right\}. \tag{B.1}$$

For any fixed κ there exists $C_1 > 0$ such that for any $k \geqslant 1$, $\tau, \xi > 0$, there exists $C_2 > 0$ such that for any $N \in \mathbb{N}$, $E \in (-2+\kappa, 2-\kappa)$, $t \in [N^{-\tau}, 1]$ and $\varepsilon > 0$ we have

$$\mathbb{P}\left(\{|\{\mu_i(t) \in [E, E+\varepsilon/N]\}| \geqslant k\} \cap \mathcal{G}_\xi\right) \leqslant C_2 N^{2k\xi + C_1 k^2 \tau} \varepsilon^{\frac{k(k+1)}{2}}.$$

Compared to [13, Theorem 3.5], the above Wegner estimate bound has extra N^ξ factors, because our proof does not use subgaussian decay of the matrix entries (we only assume condition (2.1) instead). The same comment applies to the following corollary.

Corollary B.2. *Assume the same conditions as Proposition B.1.*

For any fixed $\alpha > 0$ there exists $C_1 > 0$ such that for any $\tau, \xi > 0$, there exists $C_2 > 0$ sucht that for any $N \in \mathbb{N}$, $i \in [\![\alpha N, (1-\alpha)N]\!]$, $t \in [N^{-\tau}, 1]$ and $\varepsilon > 0$ we have

$$\mathbb{P}\left(\left\{|\mu_{i+1}(t) - \mu_i(t)| \leqslant \frac{\varepsilon}{N}\right\} \cap \mathcal{G}_\xi\right) \leqslant C_2 N^{2k\xi + C_1 \tau} \varepsilon^2.$$

Proof. For any $j \in \mathbb{Z}$, define $E_j = \gamma_i + j\frac{\varepsilon}{N}$. We then have the events inclusion

$$\left\{|\mu_{i+1}(t) - \mu_i(t)| \leqslant \frac{\varepsilon}{N}\right\} \cap \mathcal{G}_\xi \subset \bigcup_{|j| \leqslant \frac{N^\xi}{\varepsilon}+1} \left(\left\{\left|\left\{\mu_\ell(t) \in \left[E_j, E_j + \frac{2\varepsilon}{N}\right]\right\}\right| \geqslant 2\right\} \cap \mathcal{G}_\xi\right).$$

The union bound together with Proposition B.1 applied with $k = 2$ allow to conclude. □

The above Corollary B.2 actually holds for eigenvalues up to the edge (with the exponents εN^{-1} and $N^{-1+\delta}$ being replaced by $\varepsilon N^{-2/3}(\widehat{i})^{-1/3}$, $N^{-2/3+\delta}(\widehat{i})^{-1/3}$, respectively). The proof requiring just formal changes, we will only present the bulk case here, for notational simplicity.

To prepare the proof of Proposition B.1, we need the following lemmas. In particular, Proposition B.1 will require a regularity assumption of type (B.2) for the matrix entries. Note that this condition was weakened in [27] to $\int (f'/f)^4 f < \infty$ (where f is the density of real and imaginary parts of the matrix entries), but we will not need this improvement.

Lemma B.3. *Let $H=(h_{ij})_{1\leqslant i,j\leqslant N}$ be a symmetric generalized Wigner matrix satisfying* (2.1) *and $\tau>0$.*

We denote $f=e^{-g}$ the probability density of $\sqrt{1-t}\sqrt{N}h_{ij}+\sqrt{t}\mathcal{N}$, where $t\in[N^{-\tau},1]$ and $\mathcal{N}$ is a standard Gaussian independent from H. Then there exists $C>0$ such that for any $a\geqslant 1$ there exists $c_a>0$ such that uniformly in $N,i,j,s\in\mathbb{R}$, we have

$$|\widehat{f}(s)|\leqslant c_a\frac{N^{Ca\tau}}{(1+s^2)^a},\quad |\widehat{fg''}(s)|\leqslant c_a\frac{N^{Ca\tau}}{(1+s^2)^a}. \tag{B.2}$$

Proof. The first inequality is elementary:

$$\left|\int e^{\mathrm{i}sx}f(x)\mathrm{d}x\right|=|\mathbb{E}(e^{\mathrm{i}s\sqrt{1-t}\sqrt{N}h_{ij}})||\mathbb{E}(e^{\mathrm{i}s\sqrt{t}\mathcal{N}})|\leqslant e^{-s^2\frac{t}{2}}\leqslant c_a\frac{t^{-a}}{(1+s^2)^a}.$$

For the second one, we have $fg''=f'^2/f-f''$ and

$$\left|\int e^{\mathrm{i}sx}f''(x)\mathrm{d}x\right|\leqslant s^2\left|\int e^{\mathrm{i}sx}f(x)\mathrm{d}x\right|\leqslant s^2c_{a+1}\frac{t^{-(a+1)}}{(1+s^2)^{a+1}}\leqslant c_{a+1}\frac{t^{-2a}}{(1+s^2)^a},$$

so we only need to bound (without loss of generality we can assume a is an integer)

$$\left|\int e^{\mathrm{i}sx}\frac{f'(x)^2}{f(x)}\mathrm{d}x\right|\leqslant\frac{2^a}{(1+s^2)^a}\left|\int e^{\mathrm{i}sx}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^{2a}\frac{f'(x)^2}{f(x)}\mathrm{d}x\right|\mathbb{1}_{|s|\geqslant1}+\left|\int\frac{f'(x)^2}{f(x)}\mathrm{d}x\right|\mathbb{1}_{|s|\leqslant1}. \tag{B.3}$$

Let ν be the distribution of $\sqrt{1-t}\sqrt{N}h_{ij}$. Then for any $K>0$ we have

$$\begin{aligned}|f'(x)|=\left|\frac{1}{\sqrt{2\pi t}}\int\nu(\mathrm{d}u)\frac{x-u}{t}e^{-\frac{(x-u)^2}{2t}}\right|&\leqslant C\ Kt^{-1}|f(x)|+C\ t^{-3/2}e^{-\frac{K^2}{2t}}\int\nu(\mathrm{d}u)|x-u|\\&\leqslant CxN^{C\tau}|f(x)|+CN^{C\tau}e^{-\frac{x^2}{2t}},\end{aligned}$$

where we chose $K=x$ and used (2.1) so that ν has finite first moment. Moreover, we obviously have $|f(x)|>cN^{-C\tau}e^{-\frac{x^2}{2t}}$, so we can easily bound the second term on the right hand side of (B.3):

$$\left|\int\frac{f'(x)^2}{f(x)}\mathrm{d}x\right|\leqslant C\left|\int|f'(x)|(1+|x|)N^{C\tau}\mathrm{d}x\right|\leqslant CN^{C\tau}.$$

For the first term of the right hand side in (B.3), and expansion of the $2a$-th derivative of this ratio and the same cut argument by $K=x$ yields

$$\int\left|\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^{2a}\frac{f'(x)^2}{f(x)}\right|\mathrm{d}x\leqslant CN^{C\tau a},$$

which concludes the proof. □

Lemma B.4. *Fix $p\in\mathbb{N}^*$ and $N\geqslant p+3$. Let $\mathbf{u}_1,\dots,\mathbf{u}_{N-1}$be an orthonormal basis in $\mathbb{R}^N$, and set $\xi_\alpha=|\mathbf{b}\cdot\mathbf{u}_\alpha|^2$, where the components of $\mathbf{b}$ are independent centered real random variables with density $f=e^{-g}$ satisfying* $\operatorname{Var}b_i\sim1$*, the decay (2.1) and the density smoothness assumption (B.2), uniformly in N and $i\in[\![1,N-1]\!]$.*

Let $\alpha_1,\dots,\alpha_p,\beta_1,\beta_2,\beta_3$ be distinct indices in $[\![1,N-1]\!]$. Let $c_j>0$, $j\in[\![1,p]\!]$, $d_\alpha\in\mathbb{R}$ for all $1\leqslant\alpha\leqslant N-1$, $d_{\beta_1},d_{\beta_2},d_{\beta_3}>0$.

(i) For any $r\in\left(1,\frac{p}{2}+1\right)$, there exists a constant $C_{r,p}<\infty$ such that

$$\mathbb{E}_{\mathbf{b}}\left(\left(\sum_{j=1}^{p}c_j\xi_{\alpha_j}\right)^2+\left(E-\sum_{\alpha=1}^{N-1}d_\alpha\xi_\alpha\right)^2\right)^{-\frac{r}{2}}\leqslant C_{r,p}\frac{N^{2C(r-1)\tau}}{\left(\prod_{j=1}^{p}c_j^{1/2}\right)^{\frac{2(r-1)}{p}}\min(d_{\beta_1},d_{\beta_2},d_{\beta_3})}. \tag{B.4}$$

(ii) For any $r \in \left(\frac{p+1}{2}, \frac{p}{2}+1\right)$, there exists a constant $C_{r,p} < \infty$ such that

$$\mathbb{E}_{\mathbf{b}} \left(\left(\sum_{j=1}^{p} c_j \xi_{\alpha_j} \right)^2 + \left(E - \sum_{\alpha=1}^{N-1} d_\alpha \xi_\alpha \right)^2 \right)^{-\frac{r}{2}} \leqslant C_{r,p} \frac{N^{Cp\tau}}{\left(\prod_{j=1}^{p-1} c_j^{1/2}\right) c_p^{r-\frac{p+1}{2}} \min(d_{\beta_1}, d_{\beta_2}, d_{\beta_3})}. \tag{B.5}$$

(iii) For any $r \in \left(1, \frac{p}{2}\right)$, there exists a constant $C_{r,p} < \infty$ such that

$$\mathbb{E}_{\mathbf{b}} \left(\sum_{j=1}^{p} c_j \xi_{\alpha_j} \right)^{-r} \leqslant C_{r,p} \frac{N^{C(r-1)\tau}}{(\min_j c_j^{1/2})^r}. \tag{B.6}$$

Proof. We closely follow the method of Lemma 8.2 in [13]. The main differences in the estimates (due to considering real instead of complex random variables) are the exponents $c_j^{1/2}$ in the upper bounds (instead of c_j when the b_j's take complex values), and the fact that we need to consider 3 variables $d_{\beta_1}\xi_{\beta_1}, d_{\beta_2}\xi_{\beta_2}, d_{\beta_3}\xi_{\beta_3}$ for convergence purpose (instead of 2 when the b_j's take complex values). Moreover, the extra error terms $N^{C(r-1)\tau}$ are of course due to our smoothness scale.

Let O be the orthogonal matrix with columns $\mathbf{u}_1, \dots, \mathbf{u}_{N-1}$, $\mathbf{x} = O^* \mathbf{b}$, $\mathrm{d}\mu(\mathbf{x}) = e^{-\Phi(\mathbf{x})} \prod_{\alpha=1}^{N-1} \mathrm{d}x_\alpha$, with $\Phi(\mathbf{x}) = \sum_{\ell=1}^{N-1} g((O\mathbf{x})_\ell)$, $F(t) = \int_{-\infty}^{t} \left((\sum_{j=1}^{p} c_j x_{\alpha_j}^2)^2 + s^2 \right)^{-r/2}$ and $D = x_{\beta_1}\partial_{x_{\beta_1}} + x_{\beta_2}\partial_{x_{\beta_2}} + x_{\beta_3}\partial_{x_{\beta_3}}$. Then the analogue of [13, equation (8.20)] is

$$I := \mathbb{E}_{\mathbf{b}} \left(\left(\sum_{j=1}^{p} c_j \xi_{\alpha_j} \right)^2 + \left(E - \sum_{\alpha=1}^{N-1} d_\alpha \xi_\alpha \right)^2 \right)^{-\frac{r}{2}} = \frac{1}{2} \int \mathrm{d}\mu(\mathbf{x}) \frac{F(E - \sum_{\alpha=1}^{N-1} d_\alpha x_\alpha^2)}{d_{\beta_1} x_{\beta_1}^2 + d_{\beta_2} x_{\beta_2}^2 + d_{\beta_3} x_{\beta_3}^2} (1 - D\Phi(\mathbf{x})).$$

We then can follow [13, equations (8.21), (8.23)] and bound $I \leqslant (A_1 + A_2 + A_3 + |B_1| + |B_2| + |B_3|)/2$ where

$$\begin{aligned}
A_1 &:= \int \mathrm{d}\mu(\mathbf{x}) \frac{\mathbb{1}_{\sum_{j=1}^{p} c_j x_{\alpha_j}^2 \leqslant \kappa}}{(\sum_{j=1}^{p} c_j x_{\alpha_j}^2)^{r-1}}, \\
A_2 &:= \frac{1}{\kappa^{r-1}} \int \mathrm{d}\mu(\mathbf{x}) \frac{1}{x_{\beta_1}^2 + x_{\beta_2}^2 + x_{\beta_3}^2}, \\
A_3 &:= \int \mathrm{d}\mu(\mathbf{x}) \frac{\mathbb{1}_{\sum_{j=1}^{p} c_j x_{\alpha_j}^2 \leqslant \kappa} \mathbb{1}_{x_{\beta_1}^2 + x_{\beta_2}^2 + x_{\beta_3}^2 \leqslant 1}}{(\sum_{j=1}^{p} c_j x_{\alpha_j}^2)^{r-1} (x_{\beta_1}^2 + x_{\beta_2}^2 + x_{\beta_3}^2)^2}, \\
B_k &:= \int \mathrm{d}\mu(\mathbf{x}) \frac{1}{(\sum_{j=1}^{p} c_j x_{\alpha_j}^2)^{r-1}} (\partial_{x_{\beta_k}} \Phi)^2, \ k = 1, 2, 3.
\end{aligned}$$

To prove (i), we first bound A_3. For this, let

$$\widetilde{f}(\mathbf{x}) := \frac{\mathbb{1}_{\sum_{k=1}^{p} c_j x_j^2 \leqslant \kappa} \mathbb{1}_{x_{p+1}^2 + x_{p+2}^2 + x_{p+3}^2 \leqslant 1}}{(\sum_{j=1}^{p} c_j x_j^2)^{r-1} (x_{p+1}^2 + x_{p+2}^2 + x_{p+3}^2)}.$$

The integral of $\widetilde{f}$ over $x_{p+1}, x_{p+2}, x_{p+3}$ is finite (for this we need at least 3 such terms), and changing the other variables $c_j^{1/2} x_j \to x_j$ and using $r < \frac{p}{2} + 1$ we have $\|f\|_1 \leqslant C_{r,p} \kappa^{\frac{p}{2}+1-r} / \prod_{j=1}^{p} c_j^{1/2}$. The reasoning of [13, equation (8.26)], with the first equation of (B.2) as an input, gives

$$A_3 \leqslant C_{r,p} \|\widetilde{f}\|_1 N^{Cp\tau} \leqslant C_{r,p} \frac{\kappa^{\frac{p}{2}+1-r}}{\prod_{j=1}^{p} c_j^{1/2}} N^{Cp\tau}.$$

The terms A_1 can be controlled in the same way and $A_2 \leqslant C\kappa^{r-1}$. The bound on B_1, B_2, B_3 amounts to the same estimate as A_1, A_2, A_3, thanks to the representation analogue to [13, equation (8.28)], and it requires the second estimate in (B.2). We therefore obtained

$$I \leqslant C_{r,p}\left(\frac{\kappa^{\frac{p}{2}+1-r}}{\prod_{j=1}^{p} c_j^{1/2}} N^{Cp\tau} + \frac{1}{\kappa^{r-1}}\right).$$

Optimization over κ concludes the proof of (i).

For (ii), we bound $I \leqslant (A_4 + A_5 + A_6 + A_7 + |B_1| + |B_2| + |B_3|)/2$ where

$$\begin{aligned}
A_4 &:= \int \mathrm{d}\mu(\mathbf{x}) \frac{\mathbb{1}_{x_{\alpha_p}^2 \leqslant 1}}{(\sum_{j=1}^{p-1} c_j x_{\alpha_j}^2 + c_p x_{\alpha_p}^2)^{r-1}},\\
A_5 &:= \int \mathrm{d}\mu(\mathbf{x}) \frac{1}{(\sum_{j=1}^{p-1} c_j x_{\alpha_j}^2 + c_p)^{r-1}},\\
A_6 &:= \int \mathrm{d}\mu(\mathbf{x}) \frac{\mathbb{1}_{x_{\beta_1}^2 + x_{\beta_2}^2 + x_{\beta_3}^2 \leqslant 1}}{(\sum_{j=1}^{p-1} c_j x_{\alpha_j}^2 + c_p)^{r-1}(x_{\beta_1}^2 + x_{\beta_2}^2 + x_{\beta_3}^2)},\\
A_7 &:= \int \mathrm{d}\mu(\mathbf{x}) \frac{\mathbb{1}_{x_{\alpha_p}^2 \leqslant 1}\mathbb{1}_{x_{\beta_1}^2 + x_{\beta_2}^2 + x_{\beta_3}^2 \leqslant 1}}{(\sum_{j=1}^{p-1} c_j x_{\alpha_j}^2 + c_p x_{\alpha_p}^2)^{r-1}(x_{\beta_1}^2 + x_{\beta_2}^2 + x_{\beta_3}^2)}.
\end{aligned}$$

To bound the term A_7, we now introduce the function

$$\widetilde{f}(x_1, \dots, x_{p+2}) := \frac{\mathbb{1}_{x_p^2 \leqslant 1}\mathbb{1}_{x_{p+1}^2 + x_{p+2}^2 + x_{p+3}^2 \leqslant 1}}{(\sum_{j=1}^{p-1} c_j x_j^2 + c_p x_p^2)^{r-1}(x_{p+1}^2 + x_{p+2}^2 + x_{p+3}^2)}.$$

Again, the integral of $\widetilde{f}$ over $x_{\beta_1}, x_{\beta_2}, x_{\beta_3}$ is finite. By changing the variable $c_j^{1/2} x_j \to x_j$ for $j \in [\![1, p-1]\!]$ we get

$$\|f\|_1 \leqslant \frac{1}{(\prod_{j=1}^{p-1} c_j^{1/2}) c_p^{r-\frac{p}{2}-\frac{1}{2}}} \int_{|x_p|<1} \frac{1}{(\sum_{j=1}^{p-1} x_i^2 + x_p^2)^{r-1}} \mathrm{d}x_1 \dots \mathrm{d}x_p \leqslant C_{r,p} \frac{1}{(\prod_{j=1}^{p-1} c_j^{1/2}) c_p^{r-\frac{p}{2}-\frac{1}{2}}},$$

where we used that the above integral is finite when $r \in (\frac{p+1}{2}, \frac{p}{2}+1)$: after changing variables, it is also of order

$$\int_{0<x<1, y>0} \frac{x^{-\frac{1}{2}} y^{\frac{p-1}{2}-1}}{(y+x)^{r-1}} \mathrm{d}x\mathrm{d}y \leqslant \int_{0<x<1} x^{-\frac{1}{2}} \int_{a>x} a^{\frac{p-1}{2}-r} \mathrm{d}a\mathrm{d}x \leqslant \int_0^1 x^{\frac{p}{2}-r} \mathrm{d}x < \infty.$$

Proceeding as in [13, equation (8.28)] we conclude that

$$A_7 \leqslant C_{r,p} \frac{N^{Cp\tau}}{(\prod_{j=1}^{p-1} c_j^{1/2}) c_p^{r-\frac{p}{2}-\frac{1}{2}}}.$$

The terms A_4, A_5, A_6 can be bounded in the same way. Similarly, B_1, B_2 and B_3 can be bounded by the previous reasoning after using the analogue of representation [13, equation (8.28)] in the real context. This concludes the proof of (ii). Finally, the proof of (iii) is elementary. □

Proof of Proposition B.1. We follow the method from [13, Sections 8 and 9]. The required preliminary results from [13] are listed below, as well as their substitute used in our proof, so that we can accomodate the weaker condition (2.1) instead of the subgaussian decay of the matrix entries.

(1) The localization result [13, Theorem 3.1] states in particular the following. Writing $\mathcal{N}_\eta(E) = \mathcal{N}_I$ the number of eigenvalues in $I = [E - \eta/2, E + \eta/2]$, then for any $\delta > 0$ one has

$$\mathbb{P}\left(\left|\frac{\mathcal{N}_{\eta^*}(E)}{N\eta^*} - \varrho(E)\right| \geqslant \delta\right) \leqslant Ce^{-c\delta\sqrt{N\eta^*}}. \tag{B.7}$$

Equation (B.7) states that in windows of scale η the fluctuations of the number of eigenvalues is of order $\sqrt{N\eta}$. Instead, we will use the rigidity result (remember the notation (B.1)), for any (small) $\xi > 0$ and (large) $D > 0$ we have

$$\mathbb{P}\left(\mathcal{G}_\xi\right) \geqslant 1 - N^{-D}. \tag{B.8}$$

The above rigidity estimate was proved in [21] assuming subexponential decay of the entries distribution, but this is easily weakened to the finite moment assumption (2.1) (see remark 2.4 in [9]). Compared to (B.7), the above bound on fluctuations of eigenvalues is better for mesoscopic scales ($\eta \sim N^{-1+c}$ for small fixed c) but it becomes worse at the microscopic scale ($\eta \sim N^{-1}$), explaining the final extra N^ξ factor in our statement of Proposition B.1.

(2) The tail distribution of the eigenvalue gap [13, Theorem 3.3]: denoting by μ_α the largest eigenvalue greater than E, there are constants $C, c > 0$ such that uniformly in $N, K \geqslant 0$ and E in the bulk of the spectrum we have

$$\mathbb{P}\left(\mu_{\alpha+1} - E \geqslant \frac{K}{N}, \alpha \leqslant N-1\right) \leqslant Ce^{-c\sqrt{K}}. \tag{B.9}$$

Again, the above result assumes subgaussian decay of the entries, in this paper it will therefore be substituted by (B.8).

(3) The analogue of [13, Theorem 3.4] requires smoothness of the entries. We therefore now assume the $\mu_i(t)$'s are as in Proposition B.1, so that they satisfy the density bounds (B.2). The average density of states becomes, in our context: denoting $I = [E - \varepsilon/(2N), E + \varepsilon/(2N)]$, there exists $C > 0$ such that uniformly in $0 \leqslant \varepsilon \leqslant 1$, we have

$$\mathbb{P}(\{\mathcal{N}_I \geqslant 1\} \cap \mathcal{G}_\xi) \leqslant CN^{\xi + C\tau}\varepsilon. \tag{B.10}$$

For the proof, we denote $(\lambda_\alpha^{(j)})_\alpha$ for the eigenvalues of the minor obtained from H by removing the j-th row and column, $(u_\alpha^{(j)})_\alpha$ the eigenvectors, and $\xi_\alpha^{(j)} = |\mathbf{b}^{(j)} \cdot u_\alpha^{(j)}|^2$ where $\mathbf{b}^{(j)} = \sqrt{N}(h_{j2}, \dots, h_{j,N})$.

The proof of (B.10) is the same as [13, Theorem 3.4], except that: (i) one needs to replace the definition [13, (8.3)] by $\Delta = N(\lambda_{\gamma+3}^{(1)} - E)$ by $N(\lambda_{\gamma+4}^{(1)} - E)$, because the analogue (B.4) of [13, (8.12)] requires three indexes d_β in the real case instead of two for complex entries, for convergence reasons; (ii) the error term has a factor $N^{C\tau}$ due to the deteriorated smoothness (B.2) and its consequeces in (B.4), (B.6); (iii) the rigidity input (B.7) and (B.9) used in [13] are replaced by (B.8), explaining the above extra N^ξ factor in the Wegner estimate (B.10).

Thanks to these preliminary results (1), (2), (3), the analogue of [13, Theorem 3.5], Proposition B.1, can be proved as follows. First, the inequality [13, (9.2)] still holds:

$$\mathcal{N}_I \leqslant \frac{C\varepsilon}{N} \sum_{j=1}^{N} \left(\left(\eta + \frac{\eta}{N} \sum_{\alpha=1}^{N-1} \frac{\xi_\alpha^{(j)}}{(\lambda_\alpha^{(j)} - E)^2 + \eta^2} \right)^2 + \left(E - h_{jj} + \frac{1}{N} \sum_{\alpha=1}^{N-1} \frac{(\lambda_\alpha^{(j)} - E)\xi_\alpha^{(j)}}{(\lambda_\alpha^{(j)} - E)^2 + \eta^2} \right)^2 \right)^{-\frac{1}{2}}. \tag{B.11}$$

We follow [13] and denote

$$\begin{aligned} d_\alpha^{(j)} &:= \frac{N(\lambda_\alpha^{(j)} - E)}{N^2(\lambda_\alpha^{(j)} - E)^2 + \varepsilon^2},\ c_\alpha^{(j)} = \frac{\varepsilon}{N^2(\lambda_\alpha^{(j)} - E)^2 + \varepsilon^2}, \\ \mu_{\gamma(N)} &:= \min\left\{\mu_\alpha : \mu_\alpha - E \geqslant \frac{\varepsilon}{N}\right\},\ \Delta_d^{(\mu)} = N(\mu_{\gamma(N)+d-1} - E). \end{aligned} \tag{B.12}$$

In the following, $\Delta_d^{(\mu)}$ is always well defined because we will always consider $d = \mathrm{O}(1)$ as $N \to \infty$: in the set $\mathcal{G}_\xi$ there are always many more than $d-1$ eigenvalues above $E + \varepsilon/N$.

Note that the proof of (B.10) actually gives a bit more, i.e. the analogue of [13, Corollary 8.1], which is the first step in the following induction (B.14): for any $M, d \geqslant 1$, we have

$$\mathbb{E}\left(\mathbb{1}_{\mathcal{N}_I \geqslant 1} (\Delta_d^{(\mu)})^M \mathbb{1}_{\mathcal{G}_\xi}\right) \leqslant CN^{M\xi + C\tau}\varepsilon. \tag{B.13}$$

To bound $\mathbb{P}(\mathcal{N}_I \geqslant k, \mathcal{G}_\xi)$, we introduce the more general quantity

$$I_N^{(\mu)}(M,k,\ell) := \mathbb{E}(\mathbb{1}_{\mathcal{N}_I^{(\mu)} \geqslant k} (\Delta_\ell^{(\mu)})^M \mathbb{1}_{\mathcal{G}_\xi}).$$

We will prove that

$$I_N^{(\mu)}(M,k,\ell) \leqslant C_k N^{Ck\tau} \varepsilon^k \max_{1\leqslant j\leqslant N} I_{N-1}^{(j)}(M+2,k-1,\ell+1). \tag{B.14}$$

By induction over k, together with the initial condition (B.13), this will conclude the proof, noting that $1+\sum_{j=2}^k j = \frac{k(k+1)}{2}$. To prove (B.14), thanks to (B.11) for any $r \geqslant 1$ we have

$$\begin{aligned} I_N^{(\mu)}(M,k,\ell) &\leqslant C_{k,s}\varepsilon^r \max_{1\leqslant j\leqslant N} \mathbb{E} \frac{\mathbb{1}_{\mathcal{N}_I^{(j)} \geqslant k-1} (\Delta_\ell^{(\mu)})^M \mathbb{1}_{\mathcal{G}_\xi}}{\left(\left(\sum_{\alpha=1}^{N-1} c_\alpha^{(j)} \xi_\alpha^{(j)}\right)^2 + \left(E - h_{jj} + \sum_{\alpha=1}^{N-1} d_\alpha^{(j)} \xi_\alpha^{(j)}\right)^2\right)^{\frac{r}{2}}} \\ &\leqslant C_{k,s}\varepsilon^r \max_{1\leqslant j\leqslant N} \mathbb{E} \frac{\mathbb{1}_{\mathcal{N}_I^{(j)} \geqslant k-1} (\Delta_{\ell+1}^{(\lambda^{(j)})})^M \mathbb{1}_{\mathcal{G}_\xi}}{\left(\left(\sum_{\alpha=1}^{N-1} c_\alpha^{(j)} \xi_\alpha^{(j)}\right)^2 + \left(E - h_{jj} + \sum_{\alpha=1}^{N-1} d_\alpha^{(j)} \xi_\alpha^{(j)}\right)^2\right)^{\frac{r}{2}}}, \end{aligned}$$

where we used Markov's inequality and convexity of $x \mapsto x^r$ in the first inequality, and interlacing in the second. We used the definition (B.12), applied to the eigenvalues of the minor, $\lambda^{(j)} = (\lambda_1^{(j)}, \dots, \lambda_{N-1}^{(j)})$, instead of μ. Quantities of type $\Delta_\ell^{(\lambda^{(j)})}$ are well defined becuse $\ell = \mathrm{O}(1)$ as $N \to \infty$ and in the set $\mathcal{G}_\xi$ there are always many more than ℓ eigenvalues above $E + \varepsilon/N$, by interlacing.We therefore have $I_N(M,k,\ell) \leqslant C_{k,s}(\max_{1\leqslant j\leqslant N} A_j + \max_{1\leqslant j\leqslant N} B_j)$ where

$$A_j := \varepsilon^r \mathbb{E} \frac{\mathbb{1}_{\mathcal{N}_I^{(j)} \geqslant k+2} (\Delta_{\ell+1}^{(\lambda^{(j)})})^M \mathbb{1}_{\mathcal{G}_\xi}}{\left(\left(\sum_{\alpha=1}^{N-1} c_\alpha^{(j)} \xi_\alpha^{(j)}\right)^2 + \left(E - h_{jj} + \sum_{\alpha=1}^{N-1} d_\alpha^{(j)} \xi_\alpha^{(j)}\right)^2\right)^{\frac{r}{2}}},$$

$$B_j := \varepsilon^r \mathbb{E} \frac{\mathbb{1}_{k-1 \leqslant \mathcal{N}_I^{(j)} \leqslant k+1} (\Delta_{\ell+1}^{(\lambda^{(j)})})^M \mathbb{1}_{\mathcal{G}_\xi}}{\left(\left(\sum_{\alpha=1}^{N-1} c_\alpha^{(j)} \xi_\alpha^{(j)}\right)^2 + \left(E - h_{jj} + \sum_{\alpha=1}^{N-1} d_\alpha^{(j)} \xi_\alpha^{(j)}\right)^2\right)^{\frac{r}{2}}}.$$

To bound A_j, denoting $\lambda_{\alpha_1}^{(j)}, \dots, \lambda_{\alpha_{k+2}}^{(j)}$ the first $k+2$ eigenvalues in I_η, we have $c_{\alpha_i} \geqslant 1/(2\varepsilon)$, so (B.6) gives, by a reasoning identical to [13, (9.8)], $A_j \leqslant C_{k,s}\varepsilon^{2r} I_{N-1}^{(j)}(M,k-1,\ell+1)$, provided $r \in (1, \frac{k+2}{2})$ (in particular one can get the exponent ε^k).

To bound the main term B_j, let $\alpha_1, \dots, \alpha_{k-1}$ be indices so that $\lambda_{\alpha_i}^{(j)} \in I_\eta$, $1 \leqslant j \leqslant k-1$. As there are less than $k+2$ eigenvalues in I_η, we can assume that for N large enough there are four eigenvalues at distance greater than $\varepsilon/(2N)$ from E, on its right for example. Let $\lambda_{\alpha_k} = \min\{\lambda_\alpha : \lambda_\alpha > E + \frac{\varepsilon}{2N}\}$. We also denote $\lambda_{\beta_1}^{(j)} \leqslant \lambda_{\beta_2}^{(j)} \leqslant \lambda_{\beta_3}^{(j)}$ the eigenvalues immediately on the right of $\lambda_{\alpha_k}^{(j)}$, and $\Delta = \Delta_4^{(\lambda^{(j)})} = N(\lambda_{\beta_3}^{(j)} - E)$. Then, analogously to [13, (9.10)], we have

$$B_j \leqslant \varepsilon^r C_{k,s} \mathbb{E}_{\lambda^{(j)},h_{jj}} \left(\mathbb{1}_{\mathcal{N}_I^{(j)} \geqslant k-1} (\Delta_{\ell+1}^{(\lambda^{(j)})})^M \mathbb{1}_{\mathcal{G}_\xi} \ \mathbb{E}_{\mathbf{b}^{(j)}} \left(\left(\sum_{i=1}^{k-1} \frac{\xi_{\alpha_i}^{(j)}}{\varepsilon} + \frac{\varepsilon}{\Delta^2} \xi_{\alpha_k}^{(j)} \right)^2 + \left(E - h_{jj} + \sum_{\alpha=1}^{N-1} d_\alpha^{(j)} \xi_\alpha^{(j)} \right)^2 \right)^{-\frac{r}{2}} \right)$$

We use (B.5) with $p = k$, $c_j = \varepsilon^{-1}$, $1 \leqslant j \leqslant p-1$, $c_p = \varepsilon\Delta^{-2}$, $\min(d_{\beta_1}^{(j)}, d_{\beta_2}^{(j)}, d_{\beta_3}^{(j)}) \geqslant 1/(2\Delta)$, $r \in (\frac{k+1}{2}, \frac{k}{2}+1)$:

$$B_j \leqslant C_{k,r}\varepsilon^r \frac{1}{(\prod_{i=1}^{k-1} \varepsilon^{-1/2})\varepsilon^{r-\frac{k+1}{2}}} N^{Ck\tau} I_{N-1}^{(j)}(M+2,k-1,\ell+1) \leqslant \varepsilon^k N^{Ck\tau} I_{N-1}^{(j)}(M+2,k-1,\ell+1).$$

This concludes the proof. □

References

[1] P Biane, *Free hypercontractivity*, Comm. Math. Phys. **184** (1997), no. 2, 457–474.

[2] P. Bourgade, L. Erdős, and H.-T. Yau, *Universality of general β-ensembles*, Duke Math. J. **163** (2014), no. 6, 1127–1190.

[3] ———, *Bulk universality of general β-ensembles with non-convex potential*, J. Math. Phys. **53** (2012), no. 9, 095221, 19.

[4] ———, *Edge universality of β-ensembles*, to appear in Comm. Math. Phys. (2013).

[5] P. Deift and D. Gioev, *Random matrix theory: invariant ensembles and universality*, Courant Lecture Notes in Mathematics, vol. 18, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2009. MR2514781 (2011f:60008)

[6] F. Dyson, *Statistical theory of the energy levels of complex systems. III*, J. Mathematical Phys. **3** (1962), 166–175.

[7] ———, *Correlations between eigenvalues of a random matrix*, Comm. Math. Phys. **19** (1970), 235–250.

[8] 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), no. 3, 587–640.

[9] ———, *The local semicircle law for a general class of random matrices*, Elect. J. Prob. **18** (2013), no. 59, 1–58.

[10] L. Erdős, S. Péché, J. A. Ramírez, B. Schlein, and H.-T. Yau, *Bulk universality for Wigner matrices*, Comm. Pure Appl. Math. **63** (2010), no. 7, 895–925.

[11] L. Erdős, J. Ramírez, B. Schlein, T. Tao, V. Vu, and H.-T. Yau, *Bulk universality for Wigner Hermitian matrices with subexponential decay*, Math. Res. Lett. **17** (2010), no. 4, 667–674.

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

[13] ———, *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Not. IMRN **3** (2010), 436–479.

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

[15] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin, *The local relaxation flow approach to universality of the local statistics for random matrices*, Ann. Inst. Henri Poincaré Probab. Stat. **48** (2012), no. 1, 1–46.

[16] L. Erdős and H.-T. Yau, *A comment on the Wigner-Dyson-Mehta bulk universality conjecture for Wigner matrices*, Electron. J. Probab. **17** (2011), no. 28, 5.

[17] ———, *Universality of local spectral statistics of random matrices*, Bull. Amer. Math. Soc. (N.S.) **49** (2012), no. 3, 377–414.

[18] ———, *Gap universality of generalized Wigner and beta ensembles*, preprint, arxiv:1211.3786 (2012).

[19] L. Erdős, H.-T. Yau, and J. Yin, *Universality for generalized Wigner matrices with Bernoulli distribution*, J. Comb. **2** (2011), no. 1, 15–81.

[20] ———, *Bulk universality for generalized Wigner matrices*, Probab. Theory Related Fields **154** (2012), no. 1-2, 341–407.

[21] ———, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math. **229** (2012), no. 3, 1435–1515.

[22] B. Helffer and J. Sjöstrand, *Équation de Schrödinger avec champ magnétique et equation de Harper*, Schrödinger operators, Lecture notes in Physics **345** (1989), 118–197.

[23] M. Jimbo, T. Miwa, Y. Môri, and M. Sato, *Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent*, Phys. D **1** (1980), no. 1, 80–158.

[24] K. Johansson, *On fluctuations of eigenvalues of random Hermitian matrices*, Duke Math. J. **91** (1998), no. 1, 151–204.

[25] ———, *Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices*, Comm. Math. Phys. **215** (2001), no. 3, 683–705.

[26] A. Lytova and L. Pastur, *Central limit theorem for linear eigenvalue statistics of the Wigner and the sample covariance random matrices*, Metrika **69** (2009), no. 2-3, 153–172.

[27] A. Maltsev and B. Schlein, *A Wegner Estimate for Wigner Matrices*, Entropy and the Quantum II. Arizona School of Analysis with Applications. R. Sims, D. Ueltschi Editors, American Mathematical Society (2011).

[28] M. Mehta, *Random matrices*, 2nd ed., Academic Press, Inc., Boston, MA, 1991. MR1083764 (92f:82002)

[29] M. Mehta and M. Gaudin, *On the density of eigenvalues of a random matrix*, Nuclear Phys. **18** (1960), 420–427.

[30] M. Shcherbina, *Orthogonal and Symplectic Matrix Models: Universality and Other Properties*, Comm. Math. Phys. **307** (2011), no. 3, 761–790.

[31] T. Tao, *The asymptotic distribution of a single eigenvalue gap of a Wigner matrix*, Probab. Theory Related Fields **157** (2010), no. 1-2, 81–106.

[32] T. Tao and V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math. **206** (2011), no. 1.

[33] ———, *The Wigner-Dyson-Mehta bulk universality conjecture for Wigner matrices*, Electronic J. Probab. **16** (2011), 2104–2121.

[34] C. Tracy and H. Widom, *On orthogonal and symplectic matrix ensembles*, Comm. Math. Phys. **177** (1996), no. 3, 727–754.