Optimal rigidity and maximum of the characteristic polynomial of Wigner matrices

Paul Bourgade
Courant Institute
New York University
bourgade@cims.nyu.edu

Patrick Lopatto
Brown University
Division of Applied Mathematics
patrick_lopatto@brown.edu

Ofer Zeitouni
Weizmann Institute of Science
and Courant Institute
ofer.zeitouni@weizmann.ac.il

We determine to leading order the maximum of the characteristic polynomial for Wigner matrices and β-ensembles. In the special case of Gaussian-divisible Wigner matrices, our method provides universality of the maximum up to tightness. These are the first universal results on the Fyodorov–Hiary–Keating conjectures for these models, and in particular answer the question of optimal rigidity for the spectrum of Wigner matrices.

Our proofs combine dynamical techniques for universality of eigenvalue statistics with ideas surrounding the maxima of log-correlated fields and Gaussian multiplicative chaos.

1 Introduction

1.1 Universality in the Fyodorov–Hiary–Keating program. In 2012, Fyodorov, Hiary, and Keating (FHK) initiated a new line of research on the connection between random matrix theory and the Riemann zeta function. Motivated by ideas from statistical mechanics, they conjectured that the extremal statistics for both characteristic polynomials of random matrices and the zeta function on the critical line are identical to those of logarithmically correlated fields [47]. Such fields arise whenever one superimposes randomness equally on all length scales, and are characterized by correlations proportional to the logarithm of the inverse distance between two points. The branching random walk and the two-dimensional Gaussian free field are paradigmatic examples.

The FHK conjecture states that for a Haar-distributed $N \times N$ unitary matrix U_N, the random variable X_N determined by the equality

$$\max_{|z|=1} \log|\det(z - \mathrm{U}_N)| = \log N - \tfrac{3}{4}\log\log N + X_N \tag{1.1}$$

converges to a variable X_∞ in distribution as $N \to \infty$, where X_∞ is distributed as the sum of two independent Gumbel random variables. Recently, there have been significant advances towards proving (1.1) and analogous results for certain other random matrix ensembles, as we discuss below. However, all previous results have been limited to specific models, which admit representations either as a determinantal point process or a sparse matrix model

In this paper, we consider the FHK conjecture for a far broader class of random matrices. We study the following generalization that encompasses real symmetric and complex Hermitian random matrices with independent entries (Wigner matrices), and systems of interacting particles at inverse temperature $\beta > 0$ and governed by a general potential $V : \mathbb{R} \to \mathbb{R}$ (β-ensembles). Under quite general conditions, the limiting spectrum of a Wigner matrix or β-ensemble is deterministic and supported on some compact interval $[A, B]$. We make this *one-cut* hypothesis in the statement below, although similar asymptotics should hold in the bulk in the complementary *multicut* case. Fix a small $\varepsilon > 0$ and set $I = [A + \varepsilon, B - \varepsilon]$. We let $\lambda_1 \leqslant \ldots \leqslant \lambda_N$ denote with the eigenvalues of the matrix or the particles of the β-ensemble, as appropriate, and set $\det(E) = \prod_{i=1}^{N}(E - \lambda_i)$.

Problem 1. *Consider any β-ensemble or any Wigner matrix; in the matrix case, set $\beta = 1$ if it is real symmetric or $\beta = 2$ if it is complex Hermitian. With the above conventions, show that*

$$\sqrt{\frac{\beta}{2}} \cdot \max_{E \in I} \big(\log|\det(E)| - \mathbb{E}\big[\log|\det(E)|\big]\big) = \log N - \tfrac{3}{4}\log\log N + Z_N, \ \textit{where} \ \lim_{N\to\infty} Z_N \overset{(\mathrm{d})}{=} Z_\infty \qquad (1.2)$$

for a random variable Z_∞ satisfying the tail decay asymptotic $c\,ye^{-2y} \leqslant \mathbb{P}(Z_\infty > y) \leqslant c^{-1}ye^{-2y}$ as $y \to \infty$, for some fixed $c > 0$. (The exact distribution of Z_∞ may depend on the matrix entries, or on V and β.)

Such a prediction was made in [49] for the Gaussian Unitary Ensemble (GUE). Both [49] and our paper focus on the bulk of the spectrum, since this corresponds to the original FHK setting (1.1).

The first contribution of this paper is to establish that the first order term in the conjecture (1.2) is correct, both for Wigner matrices and β-ensembles defined by a general class of potentials (Theorem 1.2 and Theorem 1.9). We also establish this conjecture up to tightness for the class of Gaussian-divisible Wigner matrices, in the sense that the maximum for these ensembles can be coupled with the maximum for the Gaussian Orthogonal Ensemble (GOE) up to an error of order 1.

Our first order result is new even for the GOE; previous studies were limited to the GUE [63]. Because the general models identified in Problem 1 are not integrable for $\beta \neq 2$, and do not admit a sparse representation using independent variables for non-quadratic V, the techniques previously used to prove FHK asymptotics are not applicable. Instead, we adopt dynamical ideas based around Dyson Brownian motion, which have not previously been applied to FHK asymptotics due the singular, non-local character of the relevant observable.

The method we develop for Problem 1 also leads to a sharp characterization of a fundamental property of random matrices, eigenvalue rigidity. This term refers to the observation that eigenvalues of such matrices behave as repelling particles, with interactions that suppress their fluctuations and trap them near deterministic locations. We fix a small constant $c > 0$ and consider the following problem.

Problem 2. *For general self-adjoint random matrices or β-ensembles, how large is* $\max_{cN \leqslant i \leqslant (1-c)N} \big|\lambda_i - \mathbb{E}[\lambda_i]\big|$?

This can be understood as asking for either of the following two things:

(i) An estimate giving the exact size of the maximum on a set of *high probability*, i.e. $1 - \mathrm{o}(1)$.

(ii) A bound that captures the correct order of this maximum with *overwhelming probability*, i.e. $1 - \mathrm{O}(N^{-D})$ for any $D > 0$.

The second contribution of this paper is to answer both versions of this question. For (i), we identify the size of the maximum, including the correct constant prefactor, for Wigner matrices (the first part of Theorem 1.8) and β-ensembles (Corollary 1.10). These are the first optimal rigidity results for matrix ensembles that are not unitary invariant. Previous works in this direction relied on reducing the rigidity question to one about a Riemann–Hilbert problem; such a translation is only possible for integrable ensembles [33, 37]. For (ii), we obtain the Gaussian decay of the distribution of $\lambda_i - \mathbb{E}[\lambda_i]$ well beyond the fluctuations regime, in the second part of Theorem 1.8. This solves the longstanding question of rigidity on the scale $(\log N)/N$ with overwhelming probability.

Our results are obtained by a novel combination of methods coming from the study of universality for random matrices (in particular, heat flow, coupling and homogenization), with ideas coming from the theory of logarithmically correlated fields. We now give precise statements of our main results, and defer a complete survey of the existing literature to Section 1.3 below.

1.2 ***Results.*** We begin with the definition of Wigner matrices.

Definition 1.1. A Wigner matrix $H = H(N)$ is a real symmetric or complex Hermitian $N \times N$ matrix whose upper-triangular elements $\{H_{ij}\}_{i\leqslant j}$ are independent random variables with mean zero and variances $\mathbb{E}\left[|H_{ij}|^2\right] = N^{-1}$ for all $i \neq j$, and $\mathbb{E}\left[|H_{ii}|^2\right] = CN^{-1}$ for all i, where $C > 0$ is a constant. We have $H_{ij} = \overline{H}_{ji}$ for $i > j$, and in the case that H is complex Hermitian, we suppose that the variables $\{\operatorname{Im} H_{ij}\}_{i\geqslant j}, \{\operatorname{Re} H_{ij})\}_{i\geqslant j}$ are all independent and satisfy $\mathbb{E}[(\operatorname{Im} H_{ij})^2] = \mathbb{E}[(\operatorname{Re} H_{ij})^2]$ for all $i \neq j$. Further, there exists a constant $c > 0$ such that, for all $i, j \in [\![1, N]\!]$ and $x > 0$,

$$\mathbb{P}\left(|\sqrt{N} H_{ij}| > x\right) \leqslant c^{-1} \exp\left(-x^c\right). \tag{1.3}$$

Moreover, a symmetric Wigner matrix is called Gaussian-divisible if it has the same distribution as $\sqrt{1-\varepsilon^2}H + \varepsilon G$, where H is a Wigner matrix as defined above, independent of the GOE matrix G. Here $\varepsilon \in (0, 1)$ does not depend on N.

We recall that the empirical spectral density of a Wigner matrix converges to the semicircle law as $N \to \infty$, see e.g. [2]. This distribution has density

$$\rho_{\rm sc}(x) = \frac{\sqrt{(4-x^2)_+}}{2\pi}, \tag{1.4}$$

where $(x)_+ = \max(x, 0)$. We consider the principal branch of the logarithm, extended to the negative real numbers by continuity from above, given by $\log(re^{\mathrm{i}\theta}) = \log(r) + \mathrm{i}\theta$ for any $r > 0$ and $\theta \in (-\pi, \pi]$. As is usual, we define z^α by $\exp(\alpha \log(z))$. In particular for real λ and E we have $\operatorname{Re}\log(E-\lambda) = \log|E-\lambda|$ and $\operatorname{Im}\log(E-\lambda) = \pi \mathbb{1}_{\lambda>E}$ Given a probability measure ν with bounded density and a matrix H with eigenvalues $\lambda_1 \leqslant \ldots \leqslant \lambda_N$, for real E we also define

$$L_N(E) = \sum_{j=1}^{N} \log\left(E - \lambda_j\right) - N \int_{\mathbb{R}} \log\left(E - x\right) \mathrm{d}\nu(x), \tag{1.5}$$

which is the logarithm of the characteristic polynomial up to a centering shift. The following theorem is our main result on the maximum of the characteristic polynomial for Wigner matrices.

Theorem 1.2. *Let H be a symmetric Wigner matrix as in Definition 1.1 and set $d\nu(x) = \rho_{\rm sc}(x)\,\mathrm{d}x$ in* (1.5)*. Then for any $\varepsilon, \kappa > 0$ we have*

$$\mathbb{P}\left(\sup_{|E|<2-\kappa} \frac{\operatorname{Re} L_N(E)}{\sqrt{2}\log N} \in [1-\varepsilon, 1+\varepsilon]\right) = 1 - \mathrm{o}(1),$$
$$\mathbb{P}\left(\sup_{|E|<2-\kappa} \frac{\operatorname{Im} L_N(E)}{\sqrt{2}\log N} \in [1-\varepsilon, 1+\varepsilon]\right) = 1 - \mathrm{o}(1).$$

The same result holds for Hermitian Wigner matrices after replacing the $\sqrt{2}$ factors with 1.

Remark 1.3. For the imaginary part of the logarithm, a similar estimate on the minimum holds, by considering the sup for the Wigner matrix $-H$:

$$\mathbb{P}\left(\inf_{|E|<2-\kappa} \frac{\operatorname{Im} L_N(E)}{\sqrt{2}\log N} \in [-1-\varepsilon, -1+\varepsilon]\right) = 1 - \mathrm{o}(1).$$

No such statement holds for the real part, as $\inf_{|E|<2-\varepsilon} \operatorname{Re} L_N(E) = -\infty$.

For Gaussian-divisible Wigner matrices, universality actually holds up to tightness.

Theorem 1.4. *Let H be a Gaussian-divisible symmetric Wigner matrix as in Definition 1.1. Then for any $\kappa > 0$, there exists a coupling between H and a GOE such that the following sequence of random variables is tight:*

$$\left(\sup_{|E|<2-\kappa} \operatorname{Re} L_N^H(E) - \sup_{|E|<2-\kappa} \operatorname{Re} L_N^{\rm GOE}(E)\right)_{N\geqslant 1}.$$

Corollary 1.5. *Conditional on the tightness of the following random variables for H in the integrable Gaussian orthogonal ensemble,*

$$\sup_{|E|<2-\kappa}\big(\operatorname{Re} L_N^H(E)-\sqrt{2}(\log N-\frac{3}{4}\log\log N)\big),$$

tightness also holds for H in the universal class of Gaussian-divisible symmetric Wigner matrices.

Natural analogues of Theorem 1.4 and Corollary 1.5 hold for the Hermitian symmetry class.

Remark 1.6. For the imaginary part of the logarithm, the same statements Theorem 1.4 and Corollary 1.5 are an immediate consequence of the homogenization of the Dyson Brownian motion from [20, Theorem 3.1], and an elementary bound on macroscopic linear statistics of Wigner matrices. The result is more subtle for the real part of the logarithm, as it involves a non-local observable of the spectrum.

Remark 1.7. We emphasize that tightness for the Gaussian ensembles is still elusive, despite the proof of this result for the circular ensembles [34]. Only the first order is established: for the GUE in [63] for the real part, in [37] for the imaginary part, and for the GOE in Theorem 1.2.

Our second result considers optimal rigidity of the particles. The first part establishes a high probability rigidity estimate with an optimal deviation including the multiplicative constant. The second establishes a rigidity estimate with much stronger control on the low probability exceptional set, which is still of optimal order in N.

For a given probability measure ν as in (1.5), the i-th quantiles of ν, denoted $\gamma_i=\gamma_i(N,\nu)$ for $1\leqslant i\leqslant N$, are defined through the relation

$$\int_{-\infty}^{\gamma_i}\mathrm{d}\nu=\frac{i-\frac{1}{2}}{N}. \tag{1.6}$$

Theorem 1.8. *Let H be a symmetric Wigner matrix as in Definition 1.1. The following holds.*

(i) For every $\kappa,\varepsilon>0$, we have

$$\mathbb{P}\left(\max_{\kappa N\leqslant k\leqslant(1-\kappa)N}\frac{\pi}{\sqrt{2}}\cdot\frac{\rho_{\rm sc}(\gamma_k)\,N\big(\lambda_k-\gamma_k\big)}{\log N}\in[1-\varepsilon,1+\varepsilon]\right)=1-\mathrm{o}(1),$$

(ii) For any $\kappa,\varepsilon,A>0$ there exists $C>0$ such that the following holds for all $N\in\mathbb{N}$. For all $k\in[\kappa N,(1-\kappa)N]$ and $u\in[0,A\sqrt{\log N}]$,

$$\mathbb{P}\left(|\lambda_k-\gamma_k|>u\cdot\frac{\sqrt{2}}{\pi\rho_{\rm sc}(\gamma_k)}\cdot\frac{\sqrt{\log N}}{N}\right)\leqslant Ce^{-(1-\varepsilon)u^2}. \tag{1.7}$$

For Hermitian Wigner matrices, (i) and (ii) also hold after replacing the $\sqrt{2}$ factor with 1.

A union bound in (ii) proves the optimal rigidity scale $(\log N)/N$ in the bulk of the spectrum: for every $D>0$ there exists $C>1$ such that for all $N\in\mathbb{N}$,

$$\mathbb{P}\left(\max_{\kappa N\leqslant k\leqslant(1-\kappa)N}|\lambda_k-\gamma_k|\geqslant\frac{C\log N}{N}\right)\leqslant CN^{-D}. \tag{1.8}$$

We next turn to our results on β-ensembles. We recall that the β-ensemble of dimension N, inverse temperature $\beta>0$, and potential $V\colon\mathbb{R}\to\mathbb{R}$ is the probability measure on the subset $\Delta_N=\{\boldsymbol{\lambda}=(\lambda_1,\dots\lambda_N)\in\mathbb{R}^N:\lambda_1\leqslant\lambda_2\leqslant\ldots\leqslant\lambda_N\}$ given by

$$\mathrm{d}\mu_N(\lambda_1,\dots,\lambda_N)=\frac{1}{Z_N}\prod_{1\leqslant k<l\leqslant N}|\lambda_k-\lambda_l|^\beta\exp\left(-\frac{\beta N}{2}\sum_{k=1}^N V(\lambda_k)\right)\mathrm{d}\lambda_1\dots\mathrm{d}\lambda_N, \tag{1.9}$$

where $Z_N=Z_N^{(\beta,V)}$ is a normalizing factor. In this paper, $\beta>0$ is fixed and our assumptions on V are the following.

(A1) V is analytic on $\mathbb{R}$.

(A2) At least **one** of the following growth conditions holds for V:

(i) Sub-quadratic:

$$\liminf_{x\to\pm\infty} \frac{V(x)}{2\ln|x|} > 1 \qquad \text{and} \qquad \limsup_{x\to\pm\infty} \frac{|V'(x)|}{|x|} < \infty. \tag{1.10}$$

(ii) Super-linear: There exist constants $M_0, C, c > 0$ such that

$$V'(x) \geqslant c \quad \text{and} \quad \sup_{y\in[M_0,x]} \frac{V'(y)}{y} \leqslant CV(x) \quad \text{for all} \quad x \geqslant M_0,$$

and similar estimates apply for $x \leqslant -M_0$, i.e. the above holds for $\widetilde{V}(x) := V(-x)$.

(A3) Under the previous assumptions, it is known that the expectation of the empirical spectral measure, given by $\mathbb{E}[N^{-1}\sum_{i=1}^N \delta_{\lambda_i}]$, converges weakly to an absolutely continuous probability measure μ_V with a continuous density, which we denote by ρ_V (see [26, Theorem 1] and [1, Proposition 1] for details). We assume that ρ_V is supported on a single interval $[A,B]$ and is positive on (A,B), with square root singularities at A and B. This means that there exists a $c>0$ and a function $r(E)\colon \mathbb{R}\to\mathbb{R}$ and

$$\rho_V(E) = \frac{1}{\pi}\sqrt{(E-A)(B-E)}\,r(E)\,\mathbb{1}_{[A,B]}. \tag{1.11}$$

Moreover, we assume that r does not vanish on $[A,B]$ and has an analytic extension to $\mathbb{C}$.[1]

(A4) Let

$$L_V(x) = \frac{V(x)}{2} - \int_{\mathbb{R}} \log|x-t|\,\mathrm{d}\mu_V(t).$$

There exists a constant ℓ_V such that $L_V(x) = \ell_V$ for $x\in[A,B]$, and $L_V(x) > \ell_V$ for $x \notin [A,B]$.

The following is an analogue of Theorem 1.2 for β-ensembles.

Theorem 1.9. *Let $(\lambda_1,\dots,\lambda_N)$ be distributed according to the density* (1.9) *with a potential V satisfying the above hypotheses. Take $\nu = \mu_V$ in the definition* (1.5)*. Then for any $\varepsilon,\kappa>0$ we have*

$$\mathbb{P}\left(\sup_{A+\kappa<E<B-\kappa} \sqrt{\frac{\beta}{2}}\frac{\operatorname{Re} L_N(E)}{\log N} \in [1-\varepsilon, 1+\varepsilon]\right) = 1-\mathrm{o}(1),$$

$$\mathbb{P}\left(\sup_{A+\kappa<E<B-\kappa} \sqrt{\frac{\beta}{2}}\frac{\operatorname{Im} L_N(E)}{\log N} \in [1-\varepsilon, 1+\varepsilon]\right) = 1-\mathrm{o}(1).$$

We next state the analogue of the first part of Theorem 1.8 for β-ensembles, which follows from the previous theorem.[2]

Corollary 1.10. *Under the same assumptions as Theorem 1.9 (in particular $\nu=\mu_V$ in (1.6)), we have for every $\kappa>0$ that*

$$\mathbb{P}\left(\max_{\kappa N\leqslant k\leqslant(1-\kappa)N} \pi\sqrt{\frac{\beta}{2}}\cdot\frac{\rho_V(\gamma_k)\,N\big(\lambda_k-\gamma_k\big)}{\log N} \in [1-\varepsilon,1+\varepsilon]\right) = 1-\mathrm{o}(1).$$

Remark 1.11. In Theorem 1.8 and Corollary 1.10 the quantiles γ_k can be replaced by $\mathbb{E}[\lambda_k]$, therefore answering the rigidity question as stated in Problem 2. Indeed, the bound $|\mathbb{E}[\lambda_k]-\gamma_k| = \mathrm{O}(N^{-1})$ holds thanks to [67, Theorems 1.4 and 1.5].

Remark 1.12. All results in this article have direct analogues for maxima on mesoscopic intervals which are supported in the bulk of the spectrum, and the proofs are the same up to notational changes. For example, in the case of symmetric Wigner matrices, for any deterministic interval $I = I(N) \subset [-2+\kappa, 2-\kappa]$ such that $\log|I|/\log N \to -1+\alpha$, $\alpha\in(0,1)$, and any $\varepsilon>0$ fixed, we have

$$\mathbb{P}\left(\sup_{E\in I} \frac{\operatorname{Re} L_N(E)}{\sqrt{2}\log N} \in [\alpha-\varepsilon,\alpha+\varepsilon]\right) = 1-\mathrm{o}(1). \tag{1.12}$$

[1] We remark that assumption (A3) is satisfied by a large class of potentials. For example, it suffices for V to be convex and twice differentiable [26, Example 1].

[2] An analogue of the rigidity scale with overwhelming probability, i.e. (1.8) was already shown in [25].

1.3 Related Works. The FHK conjectures were first stated in [47, 48] for Haar-distributed unitary random matrices and the Riemann zeta function. See [8] for a recent survey. See also [49] for the case of the GUE. While we do not discuss in this paper relations with ζ, we remark that the FHK conjecture for it, up to tightness of the analogue of the variable Z_N, has been established in [5, 6] after initial progress in [4, 54, 71]; see [55] for a survey.

On the random matrix side, the sharpest results available are for the circular β-ensembles. The leading and second order terms, for $\beta = 2$, were computed in [3] and [75]. For general β, the FHK conjecture up to tightness of the random variable Z_N, was obtained in [34], and the convergence was recently established in [76]. All these works rely on uncovering hierarchical structures in the spectra of random matrices, permitting the use of methods originally developed for branching processes [27, 28].

For other ensemble of random matrices, much less is known. As demonstrated in [37, 59] and also used by us, obtaining the leading order of the FHK conjectures (more precisely, a lower bound on the leading order) is closely related to proving convergence of powers of (a rescaled) version of the characteristic polynomial towards the Gaussian multiplicative chaos (GMC); we refer the reader to [23, 60, 61, 73, 80] for some works concerning the GMC for random matrices and further results in this direction. For $\beta = 2$, convergence toward the GMC of the characteristic polynomial (using Riemann–Hilbert techniques) was obtained in [17], in the so-called L^2 phase, which is not sufficient for obtaining leading order information on the maximum. In the context of more general Hermitian matrices, related results on the distribution of the characteristic polynomial of Gaussian β-ensembles (which again are not sufficient for controlling the maximum) were proved in [7, 33, 64, 65].

The fundamental reason one expects extreme values statistics such as (1.1) and a limiting Gaussian multiplicative chaos from random matrices is that they lie in the class of *logarithmically correlated fields*. For β-ensembles and Wigner matrices, this log-correlation has been proved in the sense of distributional convergence first, as follows e.g. from [57, 70], and more recently in the pointwise sense [24, 25]. For *Gaussian* log-correlated fields, a rich theory concerning the extremes is available, with the same universal scaling as in (1.1). In particular, the fluctuations of the analogue of Z_N are always of the form of two independent random variables, one being Gumbel and the second depending on the long-range behavior of the covariance. We refer the reader to [18, 19, 81] for an account of the theory in the canonical case of the Gaussian free field (from leading order computation to convergence of the maximum and details on the process of extrema), and to [42] for the universal description of the limit. Extending the theory beyond the Gaussian case (where extra tools, including comparison theorems, are available) toward its natural universality class has been a major challenge and has attracted a lot of recent activity. Beyond the models of random matrices and the Riemann zeta function already discussed, we mention here the sine-Gordon model [9] (where a renormalization procedures enables coupling to the Gaussian free field, yielding a full convergence result), the cover time for planar random walk [10, 11, 40] (where tightness has been proved), the maxima of Ginzburg–Landau fields [12], the maxima of characteristic polynomials of permutation matrices [38], where at this time only leading order information is available, and the model of two dimensional random polymers [32, 39], where not even the leading order convergence has been demonstrated.

We next turn to the topic of rigidity of eigenvalues, which has a long history, going back at least to [26]. The importance of obtaining some a-priori rigidity estimates for Wigner matrices was highlighted in [43], as part of their celebrated proof of the universality of spacing distribution for the Wigner ensemble. This work established the upper bound $|\lambda_k - \gamma_k| \leqslant N^{-1/2-\varepsilon}$ for some $\varepsilon > 0$.

Sharper estimates on rigidity for Wigner matrices were obtained in the seminal work [45], which bounded the fluctuations of the eigenvalues by $N^{-1+\varepsilon}$ for every $\varepsilon > 0$, with overwhelming probability. This result was then refined to show the bound $\mathrm{O}((\log N)^C/N)$ with overwhelming probability for some (potentially large) constant $C > 1$ [31, 50, 79]. To our knowledge, the sharpest result on rigidity prior to this work is contained in [51], who obtained the rigidity scale $(\log N)^2/N$. Our result on Gaussian decay far in the tail distribution, given in (1.7), is new even for the Gaussian ensembles.

A question related in spirit to the rigidity question is that of the maximal spacing between successive eigenvalues, going back to a question of Diaconis [41]. For the maximal spacing of GUE and CUE matrices matrices, the first order of the gap was computed in [13], and convergence of the rescaled maximal gap was established in [46]. Both of these works use determinantal methods. Universality and comparison results were obtained more recently in [20] and [66].

Finally, many other aspects of extreme value theory for random matrices have been very active recently. Notably we refer to important progress on the spectral radius of non-Hermitian random matrices, an example where universal fluctuations are known (which are not in the log-correlated universality class, see [35] and the references therein).

1.4 ***Proof Ideas.*** We now describe the main contours of the proofs in this paper. Even though our presentation of the main results starts with Wigner matrices, we describe the proofs first for β-ensembles (see Theorem 1.9), since the Wigner case is then based on a comparison which takes as input the results for the GOE and GUE. For concreteness, we focus on the description of the proofs for $\operatorname{Re} L_N(E)$.

The standard approach for estimating $\sup_{A+\kappa<E<B-\kappa} \operatorname{Re} L_N(E)$ from above has two components: first, one replaces the supremum over the interval $[A+\kappa, B-\kappa]$ by a maximum over a finite collection of E_i's, of spacing of order $1/N$. That this is enough has already been shown e.g. in [63, Corollary 5.4] (based on an idea from [34]).[3] After achieving this reduction, one uses a union bound together with a tail estimate on the law of $\operatorname{Re} L_N(E)$ with E fixed and deterministic. In particular, one needs to control exponential moments of the latter variable. Unlike the case treated in [63], we do not have at our disposal an integrable structure, and so explicit computations are not possible. Instead, we would like to use exponential estimates from [25] (in an improved form described in Theorem B.1).

Unfortunately, the estimates in Theorem B.1 do not apply directly for E_i, but rather only for $E_i + \mathrm{i}\eta_0$ where $\eta_0 = (\log N)^{1000}/N$ is as in (3.6). Because of that, we need to modify the above procedure and first move away from the real axis. Continuity arguments allow one to move to distance $1/N$ from the real line; to go beyond that, we need to use the very precise local law with Gaussian tail from [25, Remark 2.4] (which, after integration, give control on $L_N(E_i + \mathrm{i}\eta_0)$). This recent theorem provides essentially optimal bounds on the centered moments of the Stieltjes transform on all scales $\operatorname{Im} z > 0$, and is crucial for our work; weaker estimates, such as those available in the previous literature, would not have sufficed.

For the lower bound, due to the log-correlated structure of the field $L_N(\cdot)$, one could follow methods based on second moment analysis, including the insertion of appropriate barriers, as described e.g. in [81] for the Gaussian setup. There are several obstacles to that approach, including the need to obtain very precise decoupling inequalities for pairs of macroscopically separated energies E_i, based on Fourier transforms. Instead, we use the GMC approach (introduced in similar contexts in [37, 59]). Here again, the proof starts with the preliminary step of moving the problem off the real line and into the upper half plane (by distance η_0), in order to improve the regularity of L_N; this step is achieved using a Poisson integral representation of the harmonic function $\operatorname{rlog}_z(x) = \operatorname{Re}\log(z-x)$. Then, in the main technical step, we demonstrate that for every $\gamma \in (-\sqrt{2}, \sqrt{2})$, the random field with density

$$F(E) = \frac{e^{\sqrt{\beta}\gamma \operatorname{Re} \widetilde{L}_N(E+\mathrm{i}\eta_0)}}{\mathbb{E}\big[e^{\sqrt{\beta}\gamma \operatorname{Re} \widetilde{L}_N(E+\mathrm{i}\eta_0)}\big]} \tag{1.13}$$

with respect to Lebesgue measure, converges to a Gaussian multiplicative chaos as N grows. Here $\widetilde{L}_N$ is an appropriate centering of L_N. Following the general criteria in [37], this again follows from the controls provided by Theorem B.1. Once convergence to GMC has been achieved, the required lower bound follows (essentially because the GMC is supported on points E with $\operatorname{Re} L_N(E+\mathrm{i}\eta_0) > \sqrt{\beta}\gamma/2 - \delta$).

We now turn to the proof of our result on the log-characteristic polynomials of Wigner matrices, Theorem 1.2. This is fundamentally a universality result, stating that results established in Theorem 1.9 for the GOE/GUE does not depend on the distribution of the matrix entries. We adopt a dynamical approach to this question, in line with the general framework that has been developed to resolve the Wigner–Dyson–Mehta conjecture and other problems regarding the universality of local spectral statistics [44].

Our primary input is a method to couple characteristic polynomials. We consider the matrix-valued stochastic differential equation

$$\mathrm{d}H_t = \frac{1}{\sqrt{N}}\mathrm{d}B_t - \frac{1}{2}H_t\,\mathrm{d}t \tag{1.14}$$

with initial data H_0 given by a Wigner matrix, where B_t is a matrix of Brownian motions that are independent up to the symmetry $B_{ij} = B_{ji}$. The dynamics are chosen so that if H_0 is a GOE, then its distribution remains invariant for $t > 0$. It is well known that if the eigenvalues $(\lambda_i(t))_{i=1}^N$ of H_t evolve according to the Dyson Brownian motion, given by Equation (4.2):

$$\mathrm{d}\lambda_k = \frac{\mathrm{d}\beta_k}{\sqrt{N}} + \left(\frac{1}{N}\sum_{\ell\neq k}\frac{1}{\lambda_k - \lambda_\ell} - \frac{1}{2}\lambda_k\right)\mathrm{d}t, \tag{1.15}$$

[3] We will actually use a different method, that applies also to $\operatorname{Im} L_N$ and also allows one to work with mesoscopic intervals as in Remark 1.12. Our method builds on local laws up to microscopic scales from [25].

where the $\{\beta_k\}_{k=1}^N$ are independent, standard Brownian motions. To enforce a coupling, we let $(\mu_i(t))_{i=1}^N$ be a solution to (1.15) with the same choice of driving Brownian motions with the initial data $t = 0$ given by a GOE. Then the process $(\lambda_i(t) - \mu_i(t))_{i=1}^N$ satisfies a deterministic system of differential equations, which may be studied in detail using homogenization and the method of characteristics [20, 22]. Our main result on coupling, Proposition 4.3, is the estimate is that for any $z = E + \mathrm{i}\eta$ with $\eta \in (N^{-1}, 1)$, any time $t > \exp(-\bar{C}_0(\log\log N)^2)$ for appropriate $\bar{C}_0$, we have

$$\mathbb{P}\left(\max_{-2+\kappa<E<2+\kappa}\left|\sum\log\big(z-\mu_k(t)\big)-\sum\log\big(z-\lambda_k(t)\big)\right| > (\log N)^{1/2+\varepsilon}\right) = \mathrm{o}(1). \tag{1.16}$$

The crucial point here is that while t is relatively large, we are able to approach the real line up to the microscopic distance $1/N$, which is precisely the distance beyond which deterministic arguments do not yield control on the difference between $L_N(E)$ and $L_N(E + \mathrm{i}\eta)$. Note that the scale $1/N$ is below the scale of rigidity (which is of order $\log N/N$, as we prove in Theorem 1.8). The ability to nevertheless perform the coupling (using in a crucial way overcrowding estimates from [72] and a-priori suboptimal rigidity estimates from [45]) goes significantly beyond the earlier dynamics-based coupling of characteristic polynomials. See e.g. [20] for the existing sharpest result which requires $\operatorname{Im} z \geqslant N^{-1+\varepsilon}$.

Recalling that (1.15) is the eigenvalue evolution under (1.14), and that the desired result for the GOE follows from Theorem 1.9, we see that (1.16) implies matrices of the form $\sqrt{1-t}H + \sqrt{t}W$ satisfy the conclusions of Theorem 1.2, where H is Wigner matrix, W is a GOE, and t decays sufficiently slowly. It remains to extend the result from these *weakly Gaussian-divisible* matrices to the entire Wigner class. For this, we use a standard comparison argument based on four-moment matching [78]. It is well known that this technique shows that weakly Gaussian-divisible matrices are "dense" in the set of all Wigner matrices, in these sense that universality for sufficiently regular observables follows from establishing the expected behavior in the weakly Gaussian-divisible case. While the extremal statistic $\sup_{|E|<2-\kappa}\mathrm{Re}L_N(E)$ is non-local and not regular enough to directly apply results from the literature, ideas originally developed in [66] permit the comparison to proceed, and complete the proof of Theorem 1.2.

For Theorem 1.4, no density argument is needed but the relaxation step becomes particularly delicate as it needs to reach the tightness precision. Even worse, the maximum of the characteristic polynomials differences considered in (1.16) is probably not tight as $N \to \infty$, even for $t \asymp 1$. The main insight consists in proving that $\sum\log\big(E-\mu_k(t)\big)$ is very close to $\sum\log\big(E+X_N-\lambda_k(t)\big)$, up to error of order 1 where X_N is a *random shift*. This shift is small enough so that it only changes the size of the centering of the log-characteristic polynomial by an order 1. Choosing for E the location of the maximum for GOE completes the relaxation of the maximum, which actually requires many other ingredients as explained in the proof of Proposition 4.4.

We now turn to the proofs of rigidity. While the best previous rigidity for Wigner matrices was proved directly by resolvent methods [45], as a precursor to the local study of Dyson Brownian motion, we reverse this usual logic and derive optimal rigidity as a consequence of refined estimates on the local dynamics.

The estimates with optimal constant, Theorem 1.8 (i) and Corollary 1.10, are equivalent to the corresponding results for $\operatorname{Im} L_N(E)$, Theorem 1.2 and Theorem 1.9, see (3.25) for this classical equivalence. On the other hand, obtaining Theorem 1.8 (ii), which asserts rigidity for Wigner matrices with overwhelming probability, requires further novelties. The traditional four-moment comparison method is effective only for statements that hold with probability $1 - N^{-c}$, and therefore does not provide density of weakly Gaussian-divisible matrices for Theorem 1.8 (ii). However, iterative comparisons of moments of linear statistics have appeared in random matrix theory in [58, 79], which were recently strengthened in the context of eigenvector statistics towards comparison beyond the order of magnitude, up to optimal constants [15, 16]. We adapt this method to obtain the sharp Gaussian decay in (1.7) from the case of weakly Gaussian-divisible matrices. For this ensemble, we use our coupling (1.16), along with estimates specific to the GOE/GUE, to provide an optimal-order upper bound on the large moments (of order $\log N$) of the eigenvalues counting function (Lemma 4.5). With this estimate in hand for weakly Gaussian-divisible matrices, we use an inductive moment comparison (see Lemma 5.14), to obtain a similar estimate on the $\log N$ moment for arbitrary Wigner matrices. The desired rigidity result, and the precise tail bounds in Theorem 1.8, then follow by Markov's inequality.

To conclude this section on the developed methods, we mention that the upcoming work [36] obtains the analogue of Theorem 1.2 for non-Hermitian matrices with independent entries, with an approach relying on Fourier transforms of linear statistics instead of GMC and dynamics (the special case of Ginibre matrices was proved, also at leading order, in [59]). While this approach is particularly robust for the leading order of the characteristic polynomial and likely applies to Theorem 1.2, the dynamical method seems essential to our results on tightness (Theorem 1.4) and rigidity with overwhelming probability (Theorem 1.8, part (ii)).

1.5 Further Comments. Since this paper is already long and uses a multitude of tools, we have not discussed the edge of the spectrum, nor have we treated the case of β-ensembles with non-analytic potential. These extensions require work but seem within the reach of the tools developed here.

It is natural to expect that (1.2) holds. For Gaussian β-ensembles, the Jacobi representation in terms of independent variables, used in [7, 64, 65], could potentially be useful, as it was in the CβE case [34, 76]. In view of Theorem 1.4, any progress in that direction for $\beta = 1, 2$ would immediately translate to Gaussian divisible ensembles, at least at the level of tightness. More generally, we expect that the ideas developed here will be a useful basis for work on the higher order terms in Problem 1, or for studying FHK-type asymptotics and rigidity for other ensembles, including matrices of general Wigner type, adjacency matrices of random graphs, models arising in free probability, and non-Hermitian matrices.

In view of the first universal results on FHK asymptotics for Wigner matrices and β-ensembles, another natural question concerns universal limiting measures for random characteristic polynomials. We also expect that some methods from this paper would help towards the convergence of $|\det(E-H)|^\gamma \mathrm{d}E$ to a Gaussian multiplicative chaos.

1.6 Organization. Section 2 fixes our notation conventions and states essential results from previous works. In Section 3, we prove Theorem 1.9 and Corollary 1.10. Section 4 studies the short-time relaxation to equilibrium of Dyson Brownian motion for FHK-type observables, and Section 5 provides moment matching lemmas for these observables. The results in these sections are then used in Section 6 to prove Theorem 1.2 and Theorem 1.8. Appendix A establishes Proposition A.1, which controls diverging moments of the Stieltjes transform of Wigner matrices; these are used in Section 4. Appendix B proves Theorem B.1, on the Fourier–Laplace transform of the log-characteristic polynomial of β-ensembles near the real axis; this is used in Section 3.

Remark 1.13. Throughout, we suppress the dependence of the constants in our arguments on the constants in Definition 1.1 and the potential V from (1.9), since this dependence never affects our arguments. One could give explicit (suboptimal) error bounds in all our results in terms of these parameters, but for simplicity, we do not pursue this direction. Additionally, for brevity, we prove our results for real symmetric Wigner matrices, since the complex Hermitian case differs only in notation.

1.7 Acknowledgments. P. B. was supported by the NSF grant DMS-2054851. P. L. was supported by NSF postdoctoral fellowship DMS-2202891. O. Z. was supported by Israel Science Foundation grant number 421/20.

2 Preliminary Results

We begin by recalling some fundamental concepts and notation. For deterministic sequences $X = X_N$ and $Y = Y_N > 0$, we write $X = O(Y)$ if there exists a constant $C > 1$ such that $|X| \leqslant CY$ for all $N \geqslant 1$, and $X = o(Y)$ if $\lim_{N\to\infty} X/Y = 0$. We let $\mathbb{H} = \{z : \operatorname{Im} z > 0\}$ denote the complex upper half plane, and often use the notation $z = E + \mathrm{i}\eta$ for $z \in \mathbb{H}$, so that E and η stand for the real and imaginary parts of z, respectively. We often identify the complex plane $\mathbb{C}$ with $\mathbb{R}^2$, and use the notation $[E_1, E_2] \times [\eta_1, \eta_2]$ to denote the set $\{z \in \mathbb{C} : \operatorname{Re} z \in [E_1, E_2], \operatorname{Im} z \in [\eta_1, \eta_2]\}$. Our convention is that $\mathbb{N}$ denotes the set $\{1, 2, \dots\}$. The function log always denotes the natural logarithm. We write $\log_2(x) = \log\log x$.

We will frequently define constants that depend on some number of parameters. These will be introduced as $C(x_1, \dots, x_n)$, for parameters $x_1, \dots, x_n$, and subsequently referred to as C (suppressing the dependence on the parameters in the notation). These constants may change line to line without being renamed (while retaining the dependence on the same set of parameters). We usually write $C > 1$ for large constants, and $c > 0$ for small constants.

For $z = E + \mathrm{i}\eta \in \mathbb{H}$ we use the notations

$$\operatorname{Re}\log(z-\lambda) = \log|z-\lambda|, \quad \operatorname{Im}\log(z-\lambda) = \frac{\pi}{2} + \arctan\frac{\lambda - E}{\eta}, \tag{2.1}$$

which are coherent with our convention $\log(re^{\mathrm{i}\theta}) = \log(r) + \mathrm{i}\theta$ for any $r > 0$ and $\theta \in (-\pi, \pi]$.

We recall that Wigner matrices were defined in Definition 1.1. The Gaussian Orthogonal (resp. Unitary) Ensemble of dimension N, GOE_N (resp. GUE_N), is defined as the $N \times N$ Wigner matrix with independent entries H_{ij}, $i \leqslant j$, such that $\sqrt{N} H_{ij}$ is a real (resp. complex) Gaussian random variable with mean zero and variance $1 + \mathbb{1}_{i \neq j}$ (resp. $\operatorname{Re} H_{ij}$ and $\operatorname{Im} H_{ij}$ are independent, each with variance $(1 + \mathbb{1}_{i\neq j})/2$).

We say that an event $\mathcal{F} = \mathcal{F}(N)$ holds with overwhelming probability if for any $D > 0$, there exists a constant $C(D) > 1$ such that $\mathbb{P}(\mathcal{F}^c) \leqslant CN^{-D}$.

2.1 ***Semicircle law.*** Let Mat_N denote the set of $N \times N$ real symmetric matrices. Given $M \in \mathrm{Mat}_N$, we index the eigenvalues λ_i of M in increasing order: $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$. The resolvent of M is defined as by $G(z) = (M - z\mathrm{Id})^{-1}$ for $z \in \mathbb{H}$. The Stieltjes transform of M is defined for $z \in \mathbb{H}$ by

$$m_N(z) = \frac{1}{N} \operatorname{Tr} G = \frac{1}{N} \sum_i \frac{1}{\lambda_i - z}, \tag{2.2}$$

and Stieltjes transform of the semicircle law is given by

$$m_{\mathrm{sc}}(z) = \int_{\mathbb{R}} \frac{\rho_{\mathrm{sc}}(x)\,\mathrm{d}x}{x - z} = \frac{-z + \sqrt{z^2 - 4}}{2}. \tag{2.3}$$

Here $\sqrt{z^2-4} := \sqrt{z-2}\sqrt{z+2}$ is defined through the principal branch of the square root, extended to negative real numbers by $\sqrt{-x} = \mathrm{i}\sqrt{x}$ for $x > 0$.

We now recall some elementary bounds on $m_{\mathrm{sc}}(z)$.

Lemma 2.1 ([44, Lemma 6.2])**.** *There exists a constant $c > 0$ such that the following holds. For all $z = E + \mathrm{i}\eta$ such that $E \in [-10, 10]$ and $\eta \in (0, 10]$,*

$$c \leqslant |m_{\mathrm{sc}}(z)| \leqslant 1 - c\eta. \tag{2.4}$$

Set $\kappa = ||E| - 2|$. If $|E| \leqslant 2$, then

$$c\sqrt{\kappa + \eta} \leqslant \operatorname{Im} m_{\mathrm{sc}}(z) \leqslant c^{-1}\sqrt{\kappa + \eta}. \tag{2.5}$$

2.2 ***Wigner matrices.*** We recall the following fundamental estimates on Wigner matrices from [45, Theorem 2.1] and [45, Theorem 2.2]. In this theorem and throughout this paper, we will often use the control parameter

$$\varphi = \exp\left(C_0 (\log\log N)^2\right), \tag{2.6}$$

where $C_0 > 0$ is a constant depending only on the constant c from (1.3), whose value is fixed by the following lemma.

Theorem 2.2. *Let H be a Wigner matrix. Then there exists $C_0 > 0$ such that the following claims hold.*

(i) There exists $c > 0$ such that

$$\mathbb{P}\left(\bigcup_{z \in \mathbb{H}} \left\{ |m_N(z) - m_{\mathrm{sc}}(z)| \geqslant \frac{\varphi}{N\eta} \right\} \right) \leqslant c^{-1} \exp\left(-\varphi^c\right) \tag{2.7}$$

and

$$\mathbb{P}\left(\bigcup_{z \in \mathbb{H}} \left\{ \max_{i,j \in [\![1,N]\!]} |G_{ij}(z) - \delta_{ij} m_{\mathrm{sc}}(z)| \geqslant \varphi \sqrt{\frac{\operatorname{Im} m_{\mathrm{sc}}(z)}{N\eta}} + \frac{\varphi}{N\eta} \right\} \right) \leqslant c^{-1} \exp\left(-\varphi^c\right). \tag{2.8}$$

(ii) There exists $c > 0$ such that, defining $\hat{k} = \min(k, N + 1 - k)$ and γ_k as in (1.6) with $\nu = \rho_{\mathrm{sc}}$,

$$\mathbb{P}\left(\exists k \in [\![1, N]\!] : |\lambda_k - \gamma_k| \geqslant \varphi\, \hat{k}^{-\frac{1}{3}} N^{-\frac{2}{3}} \right) \leqslant c^{-1} \exp\left(-\varphi^c\right). \tag{2.9}$$

Remark 2.3. In [45], (2.7) and (2.8) were shown for z in a compact spectral domain. The extension to all $z \in \mathbb{H}$ follows from [14, Section 10].

2.3 ***Generic β-ensembles.*** We recall that β-ensembles were defined above in (1.9), and we retain the notation from the previous section. For $z \in \mathbb{H}$, let

$$s(z) = s_N(z) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\lambda_k - z}, \quad m_V(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_V(x)}{x-z}. \tag{2.10}$$

The following will be key to the proof of Theorem 1.9. It follows from [25, Remark 2.4]

Theorem 2.4. *Under the assumptions (A1), (A.2) (i), (A3) and (A4) there exist constants $C, c, \widetilde{\eta} > 0$ such that for any $q \geqslant 1$, $N \geqslant 1$ and $z = E + \mathrm{i}\eta$ with $0 < \eta \leqslant \widetilde{\eta}$ and $A - \eta \leqslant E \leqslant B + \eta$, we have*

$$\mathbb{E}\left[\, |s(z) - m_V(z)|^q \,\right] \leqslant \frac{(Cq)^{q/2}}{(N\eta)^q} + \frac{C^q e^{-cN}}{|z-A|^{q/2}|z-B|^{q/2}}.$$

We will also use the following rigidity estimate, which directly follows from [25, Lemma 3.8].

Lemma 2.5. *There exists $c(V) > 0$ such that for any $N \geqslant 1$ and $k \in [\![1, N]\!]$ we have*

$$\mu\left(|\lambda_k - \gamma_k| > N^{-\frac{2}{3}}(\widehat{k})^{-\frac{1}{3}}(\log N)^{23}\right) \leqslant c^{-1} e^{-c(\log N)^5}.$$

2.4 ***Resolvent identities.*** For $M \in \mathrm{Mat}_N$ and any differentiable $f : \mathbb{R} \to \mathbb{R}$, we set

$$\partial_{ij} f(H) = \left.\frac{d}{dt}\right|_{t=0} f\left(H + t\Delta^{(ij)}\right), \tag{2.11}$$

where $\Delta^{(ij)} \in \mathrm{Mat}_N$ is the matrix whose entries are zero except in the (i,j) and (j,i) positions, in which case they equal one: $\Delta^{(ij)}_{kl} = (\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il})(1+\delta_{ij})^{-1}$.

Given $M, \widetilde{M} \in \mathrm{Mat}_N$, with resolvents G and $\widetilde{G}$, respectively, it follows immediately from the definitions that

$$G - \widetilde{G} = G(\widetilde{M} - M)\widetilde{G}. \tag{2.12}$$

Additionally, the following resolvent identities are well known. The first can be found as [14, (3.6)]. The second is a straightforward consequence of (2.12) and the definition (2.11). The third is an immediate consequence of the spectral decomposition of $G(z)$ [14, (2.1)].

Lemma 2.6. *Given $M \in \mathrm{Mat}_N$, let $G(z) = (M - z\mathrm{Id})^{-1}$ denote its resolvent.*

(i) For any $i \in [\![1, N]\!]$,

$$\sum_{1 \leqslant j \leqslant N} |G_{ij}|^2 = \frac{\operatorname{Im} G_{ii}}{\operatorname{Im} z}. \tag{2.13}$$

(ii) For $i, j, k, l \in [\![1, N]\!]$,

$$\partial_{kl} G_{ij} = -(G_{ik}G_{lj} + G_{il}G_{kj})(1+\delta_{kl})^{-1}. \tag{2.14}$$

(iii) For $i, j \in [\![1, N]\!]$,

$$|G_{ij}(z)| \leqslant \eta^{-1}. \tag{2.15}$$

2.5 ***Eigenvalue overcrowding.*** We recall the following overcrowding estimate. It is contained in [72, Theorem 1.12], which is stated in greater generality for Wigner matrices with subgaussian entries, but we need only the special case of the Gaussian ensemble.

Theorem 2.7. *Let $\{\mu_i\}_{i=1}^N$ denote the eigenvalues of GOE_N. For any $0 < \gamma < 1$ there exist constants $C(\gamma), c(\gamma), \gamma_0(\gamma) > 0$ such that for any $k \in [\![\gamma_0^{-1}, \gamma_0 N]\!]$, $\varepsilon > 0$ and $E \in \mathbb{R}$, we have*

$$\mathbb{P}\left(\left|\left\{\mu_i \in \left[E - \frac{\varepsilon k}{N}, E + \frac{\varepsilon k}{N}\right]\right\}\right| \geqslant k\right) \leqslant (C\varepsilon)^{\frac{1}{2}(1-\gamma)k^2} + e^{-N^c}. \tag{2.16}$$

A similar bound holds for GUE_N, with $\frac{1}{2}(1-\gamma)k^2$ replaced by $(1-\gamma)k^2$.

3 Maximum for Log-Gases

This section contains the proofs of Theorem 1.9 and Corollary 1.10. The proof of Theorem 1.9 is contained in the first three subsections, and the proof of Corollary 1.10 is given in Section 3.4.

We give the proof of Theorem 1.9 here in the case where we suppose (A.2) (i) holds. Under the assumption (A.2) (ii), some extra care is needed such as working under conditional measures. We explain the necessary changes in Appendix B.6.

3.1 Upper bound for the real part. By monotonicity of $\eta \mapsto \log|E+\mathrm{i}\eta-\lambda|$, $\eta>0$, and the estimate $\int \log|E-\lambda|\,\mathrm{d}\rho_V(\lambda) = \int \log|E+\mathrm{i}\varepsilon-\lambda|\,\mathrm{d}\rho_V(\lambda)+\mathrm{O}(\varepsilon)$ uniformly in E, there exists a fixed $C>0$ such that

$$\sup_{E\in[A+\kappa,B-\kappa]} \operatorname{Re} L_N(E) \leqslant \sup_{E\in[A+\kappa,B-\kappa]} \operatorname{Re} L_N\left(E+\frac{\mathrm{i}}{N}\right)+C. \tag{3.1}$$

Let $J=[A+\kappa,B-\kappa]\cap N^{-1-c}\mathbb{Z}$, where $c>0$ is an arbitrary small constant. For any $E\in[A+\kappa,B-\kappa]$, let E' be the closest point in J, $z=E+\frac{\mathrm{i}}{N}$ and $z'=E'+\frac{\mathrm{i}}{N}$. Then from $\log(1+\varepsilon)=\varepsilon+\mathrm{O}(\varepsilon^2)$ and recalling the definition of $s(z)$ from (2.10), this implies (in this section we abbreviate $m=m_V$)

$$\begin{aligned}\operatorname{Re} L_N(z)-\operatorname{Re} L_N(z') &= \mathrm{O}((z-z')N(s(z')-m(z')))+\mathrm{O}\left((z-z')^2\sum\frac{1}{|z'-\lambda_i|^2}\right)+\mathrm{O}(1)\\ &= N^{-c}\,\mathrm{O}(|s(z')-m(z')|)+N^{-2c}\,\mathrm{O}(\operatorname{Im} s(z'))+\mathrm{O}(1).\end{aligned} \tag{3.2}$$

Next, Theorem 2.4 (with $q=\log N$) together with Markov's inequality gives

$$\max_{E\in[A+\kappa,B-\kappa]} \mathbb{P}\left(|s(z)-m(z)|>(\log N)^{7/10}\right)\leqslant N^{-200}. \tag{3.3}$$

for large enough N. Together with the boundedness of m_V on compact sets of $\mathbb{C}$ (see (B.3)), this gives

$$\mathbb{P}\left(\exists E'\in J: |s(z')|\geqslant(\log N)^{7/10}\right)\leqslant N^{-100}. \tag{3.4}$$

We conclude that

$$\mathbb{P}\left(\sup_{E\in[A+\kappa,B-\kappa]}\operatorname{Re} L_N(E)\leqslant\sup_{E\in J}\operatorname{Re} L_N(E+\mathrm{i}N^{-1})+(\log N)^{9/10}\right)\geqslant 1-\mathrm{O}(N^{-100}). \tag{3.5}$$

We now control the increments of L_N along the line segment $\{\operatorname{Re} z=E, N^{-1}<\operatorname{Im} z<\eta_0\}$ using Markov's inequality, where we set

$$\eta_0=\frac{(\log N)^{1000}}{N} \tag{3.6}$$

throughout this section. For $E\in J$, we denote $z=z(E)=E+\mathrm{i}/N$ and $\widetilde{z}=E+\mathrm{i}\eta_0$. Then for any fixed $\varepsilon>0$ and $p\in\mathbb{N}$, we have by a union bound that

$$\begin{aligned}&\mathbb{P}\left(\exists E\in J: \operatorname{Re} L_N(z)>\operatorname{Re} L_N(\widetilde{z})+\varepsilon\log N\right)\\ &\leqslant CN^{1+c}(\varepsilon\log N)^{-2p}\max_{E\in J}\mathbb{E}\left(\int_{[N^{-1},\eta_0]^{2p}}\prod_{i=1}^{p}(N(s-m_V)(E+\mathrm{i}\eta_i))\prod_{i=p+1}^{2p}(N\overline{(s-m_V)}(E+\mathrm{i}\eta_i))\right)\mathrm{d}\eta_1\dots\mathrm{d}\eta_{2p}.\end{aligned} \tag{3.7}$$

We now suppose that $p=O(N(\log\log N)^{-1})$. Theorem 2.4 gives, for $E\in[A+\kappa,B-\kappa]$,

$$\mathbb{E}[|(s-m_V)(E+\mathrm{i}\eta)|^p]\leqslant\frac{(Cp)^{p/2}}{(N\eta)^p}+C^pe^{-\widetilde{c}N}\leqslant\frac{(Cp)^{p/2}}{(N\eta)^p}, \tag{3.8}$$

for some $C=C(V,\kappa)$, where the latter inequality holds because we assume $\eta<\eta_0$. Equation (3.8) and Hölder's inequality give

$$\mathbb{E}\left[\prod_{i=1}^{p}|N(s-m_V)(E+\mathrm{i}\eta_i)|\prod_{i=p+1}^{2p}|N\overline{(s-m_V)}(E+\mathrm{i}\eta_i)|\right]\leqslant(Cp)^p\prod_{i=1}^{2p}\frac{1}{\eta_i}. \tag{3.9}$$

Inserting the previous display in (3.7), we obtain

$$\mathbb{P}\left(\exists E\in J:\operatorname{Re}L_N(z)>\operatorname{Re}L_N(\widetilde{z})+\varepsilon\log N\right)\leqslant\frac{N^{1+c}(Cp)^p(\log\log N)^{2p}}{(\varepsilon\log N)^{2p}}\leqslant N^{1+c}\frac{(Ap)^p(\log\log N)^{2p}}{(\log N)^{2p}}\leqslant N^{-100},\tag{3.10}$$

where A is a new constant depending on C and ε, and the latter inequality is obtained by setting $p=B\frac{\log N}{\log\log N}$ for sufficiently large B. We note that for the above reasoning, the Gaussian-like moment growth $(Cq)^{q/2}$ in Theorem 2.4 is crucial (as opposed to an exponential-like growth of $(Cq)^q$) .

Moreover, from Markov's inequality and Theorem B.1, for any fixed $\lambda>0$ we have

$$\begin{aligned}\mathbb{P}\left(\exists E\in J:\operatorname{Re}L_N(\widetilde{z})>(1+\varepsilon)\sqrt{\frac{2}{\beta}}\log N\right)&\leqslant N^{1+c}\,e^{-\lambda(1+\varepsilon)\sqrt{\frac{2}{\beta}}\log N}\max_{E\in J}\mathbb{E}\big[e^{\lambda\operatorname{Re}L_N(\widetilde{z})}\big]\\&\leqslant C\,N^{1+c}\max_{E\in J}e^{\frac{\sigma(\lambda,0,\widetilde{z})}{2}+\mu(\lambda,0,\widetilde{z})-\lambda(1+\varepsilon)\sqrt{\frac{2}{\beta}}\log N},\end{aligned}$$

where we refer to (B.6) and (B.7) for the definitions of σ and μ. From Lemma B.2, we have $\mu(\lambda,0,\widetilde{z})=\mathrm{O}(1)$ and $\sigma(\lambda,0,\widetilde{z})=(1+\mathrm{o}(1))\lambda^2\frac{\log N}{\beta}$ uniformly in N, $E\in J$, and λ in any compact subset of $\mathbb{R}_+$. Choosing $\lambda=\sqrt{2\beta}$ this implies that

$$\mathbb{P}\left(\exists E\in J:\operatorname{Re}L_N(\widetilde{z})>(1+\varepsilon)\sqrt{\frac{2}{\beta}}\log N\right)\leqslant e^{-(2\varepsilon-c-\mathrm{o}(1))\log N}\to 0.\tag{3.11}$$

With the choice $0<c<2\varepsilon$, equations (3.5), (3.10), (3.11) conclude the proof that

$$\mathbb{P}\left(\exists E\in[A+\kappa,B-\kappa]:\operatorname{Re}L_N(E)>(1+\varepsilon)\sqrt{\frac{2}{\beta}}\log N+2\varepsilon\log N\right)\to 0.$$

3.2 Upper bound for the imaginary part. The proof of the upper bound for $\operatorname{Im}L_N$ is the same as the one for the real part up to the following complication: There is no analogue of (3.1), as $\eta\mapsto\sum_i\operatorname{Im}\log(E+\mathrm{i}\eta-\lambda_i)$ is not monotone. To circumvent this problem, we observe that the error made by shifting $\operatorname{Im}\log$ from the real axis to a scale η can be bounded in terms of a linear combination of the real and imaginary parts of the Stieltjes transform at scale η.

More precisely, note that, from $\arctan(x)-\arctan(+\infty)=-\int_x^\infty\frac{du}{1+u^2}=-\frac{1}{x}+\mathrm{O}(\frac{1}{x^2})$ as $x\to\infty$, we have

$$\arctan((\lambda-E)/\eta)-\frac{\pi}{2}\cdot\operatorname{sgn}(\lambda-E)=-\frac{\eta(\lambda-E)}{(\lambda-E)^2+\eta^2}+\mathrm{O}\left(\frac{\eta^2}{(\lambda-E)^2+\eta^2}\right).$$

As a consequence, for $z=E+\frac{\mathrm{i}}{N}$,

$$\Big|\operatorname{Im}L_N(E)-\operatorname{Im}L_N(z)\Big|\leqslant 10|s(z)|.\tag{3.12}$$

As in the previous paragraph, let $J=[A+\kappa,B-\kappa]\cap N^{-1-c}\mathbb{Z}$, where $c>0$ is an arbitrary small constant. For any $E\in[A+\kappa,B-\kappa]$, let E' be the closest point in J and $z'=E'+\frac{\mathrm{i}}{N}$. Then the mean value theorem yields

$$s(z)-s(z')=\mathrm{O}\left(N^{-1-c}\max_{x\in[A+\kappa,B-\kappa]}\big|s'(x+\mathrm{i}N^{-1})\big|\right),\tag{3.13}$$

with an implicit constant uniform in the choice of E.

Note that $|s'(z)|\leqslant\eta^{-1}\operatorname{Im}s(z)\leqslant\eta^{-2}$ (where the last inequality follows from (2.15)), so $s(z)$ is η^{-2}-Lipschitz continuous. Taking a union bound over a mesh with spacing size $O(N^{-10})$, with (3.3) we obtain that

$$\mathbb{P}\left(\max_{x\in[A+\kappa,B-\kappa]}\operatorname{Im}s(x+\mathrm{i}N^{-1})>(\log N)^{7/10}\right)\leqslant N^{-100}.$$

Together with (3.12) and (3.13), and again using $|s'(z)|\leqslant\eta^{-1}\operatorname{Im}s(z)$, this gives

$$\mathbb{P}\left(\forall E\in[A+\kappa,B-\kappa],\Big|\operatorname{Im}L_N(E)-\operatorname{Im}L_N(z)\Big|\leqslant 10(|s(z')|+N^{-c/2})\right)\geqslant 1-N^{-100}.\tag{3.14}$$

Moreover, the same estimate as (3.2) holds for $\operatorname{Im} L_N$, so that

$$\mathbb{P}\left(\forall E\in[A+\kappa,B-\kappa],\left|\operatorname{Im} L_N(E)-\operatorname{Im} L_N(z')\right|\leqslant 100(|s(z')|+N^{-c/2})\right)\geqslant 1-N^{-100}. \tag{3.15}$$

Together with (3.4), we conclude that

$$\mathbb{P}\left(\sup_{E\in[A+\kappa,B-\kappa]}\operatorname{Im} L_N(E)\leqslant \sup_{E\in J}\operatorname{Im} L_N(E+\mathrm{i}N^{-1})+(\log N)^{9/10}\right)\geqslant 1-\mathrm{O}(N^{-100}). \tag{3.16}$$

Further, the analogues of (3.10) and (3.11) hold, with the same proofs up to notational changes, and together with (3.16) they conclude the proof of the upper bound for the imaginary part in Theorem 1.9.

3.3 Lower bound for the real and imaginary parts. We start with the proof for the real part. The proof for the imaginary part is essentially the same, and is described at the end of this section.

First step: shift to the upper half plane. For any $z=E+\mathrm{i}\eta$ with $\eta>0$, by harmonicity of $z\in\mathbb{H}\mapsto\log|z-\lambda|$ we have

$$\log|z-\lambda|=\int_{\mathbb{R}}\log|x-\lambda|\cdot\frac{\eta}{\eta^2+(E-x)^2}\frac{\mathrm{d}x}{\pi}.$$

This implies

$$\operatorname{Re} L_N(z)=\int_{\mathbb{R}}\operatorname{Re} L_N(x)\cdot\frac{\eta_0}{\eta_0^2+(E-x)^2}\frac{\mathrm{d}x}{\pi} \tag{3.17}$$

for $z\in\mathbb{R}+\mathrm{i}\eta_0$, with η_0 defined in (3.6). On the other hand,

$$\int_{[A+\kappa,B-\kappa]^c}|\operatorname{Re} L_N(x)|\cdot\frac{\eta_0}{\eta_0^2+(E-x)^2}\frac{\mathrm{d}x}{\pi}\leqslant\sum_i\int_{[A+\kappa,B-\kappa]^c}|\log|x-\lambda_i|-\log|x-\gamma_i||\cdot\frac{\eta_0}{\eta_0^2+(x-E)^2}\frac{\mathrm{d}x}{\pi}$$
$$+\int_{[A+\kappa,B-\kappa]^c}|N\int\log|x-\lambda|\mathrm{d}\rho_V(\lambda)-\sum_i\log|x-\gamma_i||\cdot\frac{\eta_0}{\eta_0^2+(x-E)^2}\frac{\mathrm{d}x}{\pi}. \tag{3.18}$$

For $x\in[A+\kappa,B-\kappa]^{\mathrm{c}}$ and $E\in[A+2\kappa,B-2\kappa]$, we have $|x-E|>\kappa$, so for such x and E we have $\frac{1}{\eta_0^2+(x-E)^2}\leqslant\frac{C}{1+x^2}$. Therefore the first sum on the right-hand side of (3.18) is smaller than

$$C\eta_0\sum_i\int_{\mathbb{R}}\frac{|\log|x-\lambda_i|-\log|x-\gamma_i||}{1+x^2}\mathrm{d}x\leqslant C\eta_0\sum_i|\lambda_i-\gamma_i|\cdot|\log|\lambda_i-\gamma_i||. \tag{3.19}$$

This implies that on the rigidity event from Lemma 2.5, the first term on the right-hand side of (3.18) is $\mathrm{O}((\log N)^{50}\eta_0)=\mathrm{o}(1)$; an error $\mathrm{o}(\log N)$ would be enough for the proof of the leading order of the maximum, so there is substantial margin here. From rigidity of β-ensembles, Lemma 2.5, we conclude that with probability $1-\mathrm{O}(N^{-10})$, for any $z\in[A+2\kappa,B-2\kappa]+\mathrm{i}\eta_0$ we have

$$\left|\operatorname{Re} L_N(z)-\int_{[A+\kappa,B-\kappa]}\operatorname{Re} L_N(x)\cdot\frac{\eta_0}{\eta_0^2+(E-x)^2}\frac{\mathrm{d}x}{\pi}\right|\leqslant 1. \tag{3.20}$$

Note that for $E\in[A+2\kappa,B-2\kappa]$ we have $\int_{[A+\kappa,B-\kappa]}\frac{\eta_0}{\eta_0^2+(E-x)^2}\frac{\mathrm{d}x}{\pi}=1+\mathrm{O}(\eta_0)$, so from the above equation we conclude that

$$\mathbb{P}\left(\sup_{z\in[A+2\kappa,B-2\kappa]+\mathrm{i}\eta_0}\operatorname{Re} L_N(z)\leqslant\sup_{E\in[A+\kappa,B-\kappa]}\operatorname{Re} L_N(E)+1\right)=1-\mathrm{o}(1). \tag{3.21}$$

Second step: lower bound for the smoothed field. Similarly to [37] and [61], the proof of the lower bound for $\max_{z\in[A+2\kappa,B-2\kappa]+\mathrm{i}\eta_0}\operatorname{Re} L_N(z)$ will be an straightforward corollary of the convergence of the corresponding field to a Gaussian multiplicative chaos measure, in the full subcritical phase.

More precisely, consider a centered Gaussian field G_η defined on $[A+2\kappa,B-2\kappa]$ with covariance $\mathbb{E}[G_\eta(E_1)G_\eta(E_2)]=\sigma(1,0,(z_1,z_2))$, where we denote $z_i=E_i+\mathrm{i}\eta$ and refer to (B.6) for the definition of σ. The existence of this field for $\eta > \eta_0$ follows from positivity of the covariance, which is a byproduct of Theorem B.1 (the limit of positive matrices is positive). From Lemma B.2, and noting that the covariance σ is defined in terms of the kernel c studied in that lemma, we have

$$\mathbb{E}[G_\eta(E_1)G_\eta(E_2)] = -\frac{1}{\beta}\log|z_1 - \bar{z}_2| + O_\kappa(1) \tag{3.22}$$

uniformly for $E_1, E_2 \in [A+2\kappa, B-2\kappa]$.

It is well known that for any $|\gamma| < \sqrt{2}$, there exists a random measure μ_γ, called the Gaussian multiplicative chaos with parameter γ, such that the following holds. For any continuous $f\colon \mathbb{R}\to\mathbb{R}$ with compact support in $(A+2\kappa, B-2\kappa)$, we have the distributional convergence

$$\lim_{\eta\to 0^+}\int f(E)\frac{e^{\sqrt{\beta}\gamma G_\eta(E)}}{\mathbb{E}[e^{\sqrt{\beta}\gamma G_\eta(E)}]}\mathrm{d}E = \int f\,\mathrm{d}\mu_\gamma.$$

The limiting random variable can be written $\int f\mathrm{d}\mu_\gamma$, for a certain random measure μ_γ, called the Gaussian multiplicative chaos with parameter γ. We refer for example to [37, Section 2.1] for a modern treatment of the existence and non-triviality of this limit.

In the following result, we denote $\widetilde{L}_N(z) = L_N(z) - \mu(1,0,z)$, with μ defined in (B.7)

Proposition 3.1. *For any $|\gamma| < \sqrt{2}$ and any continuous f with compact support in $(A+2\kappa, B-2\kappa)$, the following convergence in distribution holds:*

$$\lim_{N\to\infty}\int_{A+2\kappa}^{B-2\kappa} f(E)\frac{e^{\sqrt{\beta}\gamma\,\mathrm{Re}\,\widetilde{L}_N(E+\mathrm{i}\eta_0)}}{\mathbb{E}[e^{\sqrt{\beta}\gamma\,\mathrm{Re}\,\widetilde{L}_N(E+\mathrm{i}\eta_0)}]}\mathrm{d}E = \int f\mathrm{d}\mu_\gamma.$$

Proof. The proof is a direct application of the general criterion for convergence of a non-Gaussian random field to a Gaussian multiplicative chaos; see [37, Theorem 2.4] (which is a restatement from [62, Theorem 1.7]), and the key technical input, the Laplace transform of the log-characteristic polynomial from Theorem B.1. More precisely, [37, Theorem 2.4] states that a sufficient condition for the conclusion of Proposition 3.1 is that there is a constant $c > 0$ such that for any fixed p and $\zeta_1,\dots,\zeta_p \in \mathbb{R}$, uniformly in $\mathbf{z} \in ([A+2\kappa, B-2\kappa]\times[\eta_0, c])^p$, we have

$$\mathbb{E}_\mu\left[e^{\sum_{k=1}^p \zeta_k \mathrm{Re}\widetilde{L}_N(z_k)}\right] = \mathbb{E}\left[e^{\sum_{k=1}^p \zeta_k L(z_k)}\right](1+\mathrm{o}(1)) \tag{3.23}$$

as $N\to\infty$, where L is a centered Gaussian field defined on $[A+2\kappa, B-2\kappa]\times[\eta_0, c]$ characterized by $\mathrm{Var}[\sum_{k=1}^p \zeta_k L(z_k)] = \sigma(\boldsymbol{\zeta}, 0, \mathbf{z})$. Equation (3.23) is a direct consequence of Theorem B.1. Note that we consider $\widetilde{L}_N$ here instead of L_N because [37, Theorem 2.4] does not explicitly cover the possibility of a limiting shift. ☐

We now apply [37, Theorem 3.4]. The assumptions of this theorem are easily verified: [37, Assumption 3.1] follows from 3.23 and [37, Assumption 3.3] follows from Proposition 3.1 and a standard approximation argument to allow indicators for f. Then [37, Theorem 3.4] gives, for any fixed $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{z\in[A+2\kappa,B-2\kappa]+\mathrm{i}\eta_0}\mathrm{Re}\,\widetilde{L}_N(z) \geqslant \left(\sqrt{\frac{2}{\beta}} - \varepsilon\right)\log N\right) = 1 - \mathrm{o}(1).$$

From Lemma B.2, $\mu(1,0,z)$ is uniformly bounded, so from the previous equation there exists $C > 0$ such that

$$\mathbb{P}\left(\max_{z\in[A+2\kappa,B-2\kappa]+\mathrm{i}\eta_0}\mathrm{Re}\,L_N(z) \geqslant \left(\sqrt{\frac{2}{\beta}} - \varepsilon\right)\log N - C\right) = 1 - \mathrm{o}(1). \tag{3.24}$$

Equations (3.21) and (3.24) conclude the proof of the lower bound in Theorem 1.9, for $\mathrm{Re}\,L_N$.

Lower bound for the imaginary part. The proof for $\mathrm{Im}\,L_N$ is identical because Theorem B.1 also gives its joint Laplace transform, with limiting covariance $\sigma(0,\boldsymbol{\zeta},\mathbf{z}) = \sigma(\boldsymbol{\zeta},0,\mathbf{z}) + \mathrm{O}(1)$, a fact easily checked with (B.6) and Lemma B.2. The only small difference is about the analogue of (3.18), i.e. bounding $\int_{[A+\kappa,B-\kappa]^c}|\mathrm{Im}\,L_N(x)|\cdot\frac{\eta}{\eta^2+(E-x)^2}\frac{\mathrm{d}x}{\pi}$. By rigidity of the eigenvalues we have $|\mathrm{Im}\,L_N(E)| \leqslant N^\varepsilon$ for all E (this is a consequence of the implications (3.28) below), with overwhelming probability, and this is enough to conclude.

3.4 **Proof of Corollary 1.10**.

From the definition of Im log before (1.5), we have

$$\operatorname{Im} L_N(E) = \pi\left(\sum_{i=1}^{N} \mathbb{1}_{\{\lambda_i > E\}} - N\int_E^{\infty} \rho_V(s)\,\mathrm{d}s\right).$$

Let $k \in [\![0, N-1]\!]$, $n \in \mathbb{N}$, and $M \in [n, n+1]$ be parameters such that $n+2 < k < N-(n+2)$. For $E \in [\gamma_k, \gamma_{k+1})$, the previous display, together with the definition (1.6) of γ_k, yields the following implications:

$$\begin{aligned}
\operatorname{Im} L_N(E) > \pi M \;&\Rightarrow\; \lambda_{k-n+2} > \gamma_k \quad\;\; \Rightarrow\; \lambda_{k-n+2} - \gamma_{k-n+2} > \gamma_k - \gamma_{k-n+2},\\
\operatorname{Im} L_N(E) < \pi M \;&\Rightarrow\; \lambda_{k-n-2} < \gamma_{k+1} \;\Rightarrow\; \lambda_{k-n-2} - \gamma_{k-n-2} < \gamma_{k+1} - \gamma_{k-n-2}.
\end{aligned} \tag{3.25}$$

For the upper bound on the eigenvalue deviations, we take $M = \pi^{-1}\sqrt{\frac{2}{\beta}}(1+\varepsilon)\log N$. By Theorem 1.9, we have for any $\delta > 0$ that $\sup_{A+\delta<E<B-\delta} \operatorname{Im} L_N(E) \leqslant \pi M$ with high probability. Then taking $n = \lfloor M \rfloor$ and δ sufficiently small (in a way that depends only on κ), (3.25) implies that for every $j \in [\![\kappa N, (1-\kappa)N]\!]$, we have

$$\lambda_j - \gamma_j < \gamma_{j+n+2} - \gamma_j. \tag{3.26}$$

Further, if we define the quantity $\varepsilon_{j,m}$ implicitly by

$$\gamma_{j+m} - \gamma_j = \frac{m}{N\rho_V(\gamma_k)}(1+\varepsilon_{j,m}),$$

then (1.6) implies that

$$\sup_{j\in[\![\kappa N,(1-\kappa)N]\!]}\ \sup_{m\leqslant(\kappa/2)\sqrt{N}} \varepsilon_{j,m} = \mathrm{o}(1). \tag{3.27}$$

Combining (3.26) with (3.27) completes the proof of the upper bound. The proof for the lower bound is similar and hence omitted.

Remark 3.2. The reasoning in this proof of Corollary 1.10 could be reversed to show that Corollary 1.10 implies the bound on the imaginary part in Theorem 1.9, which demonstrates that these statements are logically equivalent. The relevant implications are now (using the same notation)

$$\begin{aligned}
\lambda_{k-n-2} - \gamma_{k-n-2} > \gamma_{k+1} - \gamma_{k-n-2} \;&\Rightarrow\; \operatorname{Im} L_N(E) > \pi M,\\
\lambda_{k+n+2} - \gamma_{k+n+2} < \gamma_k - \gamma_{k+n+2} \;&\Rightarrow\; \operatorname{Im} L_N(E) < \pi M.
\end{aligned} \tag{3.28}$$

4 Relaxation

This section proves convergence to equilibrium of $\sup \operatorname{Im} L_N$ (subsection 4.1) and $\sup \operatorname{Re} L_N$ (subsection 4.2) for the matrix Ornstein–Uhlenbeck dynamics (4.1) defined below. Indeed we will prove that theorems 1.2, 1.4 and 1.8 (i) hold for matrices of type H_t, for large enough t. We will finally prove relaxation of large moments of $\operatorname{Im} L_N$, which corresponds to a proof of Theorem 1.8 (i) for such weakly Gaussian-divisible random matrices H_t (subsection 4.3).

As explained in the introduction, this is an essential step in the proof for Wigner matrices, which then proceeds by density of the weakly Gaussian-divisible ensemble in Wigner matrices (the moment matching from Section 5).

For local statistics in the bulk of the spectrum, relaxation was first proved in [43] by a method based on entropy dissipation, up to an averaging on the energy level which prevents from considering observables such as $\sup \operatorname{Re} L_N, \sup \operatorname{Im} L_N$. Another method for relaxation was introduced in [22], through a coupling of the spectrum of H_t with a GOE. In this approach relaxation follows from homogenization of the Dyson Brownian motion: The difference between both spectra satisfies a deterministic, non-local parabolic equation at leading order, locally and with probability $1 - \mathrm{o}(1)$.

While ergodicity of $\operatorname{Im} L_N$ is closely related to relaxation of local spectral statistics, ergodicity of $\operatorname{Re} L_N$ requires convergence to equilibrium along the full spectrum. Moreover part (ii) of Theorem 1.8 requires probability bounds stronger than $1 - \mathrm{o}(1)$. Fortunately, the homogenization theory from [22] was greatly strengthened in [68] and in [20], as it holds with the probability bound $1 - N^{-D}$ for arbitrary D. For our paper, the homogenization from [20] is most pertinent as it holds for very large times, a key fact for our observable L_N, which reaches equilibrium only for $t = N^{-\mathrm{o}(1)}$. The methods from [20] also directly cover relaxation of $\operatorname{Re} L_N$, another decisive fact as the sum of the errors from the local homogenization in Proposition 4.1 below exceeds the required $\mathrm{o}(\log N)$ accuracy to catch the maximum of $\operatorname{Re} L_N$.

4.1 ***The eigenvalues.*** We first provide a quantitative relaxation of the eigenvalues (Proposition 4.1), which is a variant of [20, Theorem 3.1] and relies on this work.

Let H be a Wigner matrix. We first recall the definition of Dyson Brownian motion with initial data $H_0 = H$. As noted above in Section 1.4, for concreteness we consider just the real symmetric case, as the complex Hermitian case is analogous.

Let $B \in \mathrm{Mat}_N$ be such that the entries $\{B_{ij}\}_{i<j}$ and $B_{ii}/\sqrt{2}$ are independent standard Brownian motions, and $B_{ij} = B_{ji}$. Consider the matrix Ornstein–Uhlenbeck process

$$\mathrm{d}H_t = \frac{1}{\sqrt{N}}\mathrm{d}B_t - \frac{1}{2}H_t\,\mathrm{d}t. \tag{4.1}$$

If the eigenvalues of H_0 are distinct, it is well known that the eigenvalues $\boldsymbol{\lambda}(t) = (\lambda_1(t), \lambda_2(t), \dots, \lambda_N(t))$ of H_t are given by the strong solution of the system of stochastic differential equations

$$\mathrm{d}\lambda_k = \frac{\mathrm{d}\beta_k}{\sqrt{N}} + \left(\frac{1}{N}\sum_{\ell\neq k}\frac{1}{\lambda_k - \lambda_\ell} - \frac{1}{2}\lambda_k\right)\mathrm{d}t, \tag{4.2}$$

where the $\{\beta_k\}_{k=1}^N$ are independent, standard Brownian motions. (See, for example, [2, Lemma 4.3.3].)

We now let $\boldsymbol{\mu}(t) = (\mu_1(t), \mu_2(t), \dots, \mu_N(t))$ be a strong solution of the same SDE (4.2) with initial condition $\boldsymbol{\mu}(0) = (\mu_1, \mu_2, \dots, \mu_N)$, where $\{\mu_k\}_{k=1}^N$ are the eigenvalues of a GOE_N, denoted GOE:

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

For any $z \in \mathbb{H}$, we define

$$z_t = \frac{e^{t/2}(z + \sqrt{z^2-4}) + e^{-t/2}(z - \sqrt{z^2-4})}{2}, \tag{4.3}$$

where $\sqrt{z^2-4}$ is defined using a branch cut in the segment $[-2,2]$, as with m_{sc} in (2.3). For $z \in \mathbb{R}$, we define $z_t = \lim_{\eta\to 0^+}(z + \mathrm{i}\eta)_t$. The following key estimate on the difference between $\boldsymbol{\lambda}(t)$ and $\boldsymbol{\mu}(t)$ follows from the main result in [20]. We recall the notations φ from (2.6) and γ_k from (1.6), and let L^H and L^{GOE} denote the observable (1.5) defined using the eigenvalues of H and GOE, respectively.

Proposition 4.1. *Fix $\kappa, \varepsilon > 0$. Then for any $D > 0$ there exist $C(\varepsilon, \kappa, D) > 0$ such that for all $t \in (\varphi^C/N, 1)$, $E \in [-2+\kappa, 2-\kappa]$, and $k \in [\![1, N]\!]$ such that $\gamma_k \in [-2+\kappa, 2-\kappa]$, we have*

$$\mathbb{P}\Big(\Big|\lambda_k(t) - \mu_k(t) - \frac{\operatorname{Im} L_N^H(E_t) - \operatorname{Im} L_N^{\mathrm{GOE}}(E_t)}{N \operatorname{Im} m_{\mathrm{sc}}(E_t)}\Big| > \frac{N^{1+\varepsilon}\max(|E-\gamma_k|, N^{-1})}{N^2 t}\Big) \leqslant CN^{-D}. \tag{4.4}$$

Proof. The key to the proof is [20, Theorem 3.1], which states that there exists $C(D) > 0$ such that

$$\mathbb{P}\left(\left|\big(\lambda_k(t) - \mu_k(t)\big) - \bar{\mathrm{u}}_k(t)\right| > \frac{N^\varepsilon}{N^2 t}\right) \leqslant CN^{-D} \tag{4.5}$$

for $t \in (\varphi^C/N, 1)$, where we define

$$\bar{\mathrm{u}}_k(t) = \frac{1}{N \operatorname{Im} m_{\mathrm{sc}}(\gamma_k^t)}\sum_{j=1}^N \operatorname{Im}\left(\frac{1}{\gamma_j - \gamma_k^t}\right)\big(\lambda_j(0) - \mu_j(0)\big), \qquad \gamma_k^t = (\gamma_k)_t. \tag{4.6}$$

Moreover, from [20, Lemma 3.4], for all $\gamma_k, \gamma_\ell \in [-2+\kappa, 2-\kappa]$ we have

$$\mathbb{P}\left(|\bar{\mathrm{u}}_k(t) - \bar{\mathrm{u}}_\ell(t)| \geqslant C\varphi\frac{|k-\ell|}{N^2 t}\right) \leqslant CN^{-D}. \tag{4.7}$$

Let $E \in [-2+\kappa, 2-\kappa]$ be given, and fix some $\ell = \ell(E, N)$ such that $|E - \gamma_\ell| = \min_{j\in[\![1,N]\!]}|E - \gamma_j|$. The definition of γ_k in (1.6) (with $\nu = \rho_{\mathrm{sc}}$) gives

$$|k - \ell| < CN|\gamma_k - \gamma_\ell| \leqslant CN\left(|\gamma_k - E| + |E - \gamma_\ell|\right) \leqslant 2CN|\gamma_k - E|, \tag{4.8}$$

for some constant $C>0$. Then equations (4.5) and (4.7) together with the previous line imply that

$$\mathbb{P}\left(\ \left|\big(\lambda_k(t)-\mu_k(t)\big)-\bar{\mathrm{u}}_\ell(t)\right|>\frac{CN^{1+\varepsilon}\max(|E-\gamma_k|,N^{-1})}{N^2t}\right)\leqslant CN^{-D}, \tag{4.9}$$

where we increased C if necessary and used $N^\varepsilon\geqslant\varphi$ for sufficiently large N (depending on ε). We therefore just need to bound $\left|\frac{\operatorname{Im}L_N^H(E_t)-\operatorname{Im}L_N^{\mathrm{GOE}}(E_t)}{N\operatorname{Im}m_{\mathrm{sc}}(E_t)}-\bar{\mathrm{u}}_\ell(t)\right|$. We write

$$\frac{\operatorname{Im}L_N^H(E_t)-\operatorname{Im}L_N^{\mathrm{GOE}}(E_t)}{N\operatorname{Im}m_{\mathrm{sc}}(E_t)}=\frac{1}{N\operatorname{Im}m_{\mathrm{sc}}(E_t)}\operatorname{Im}\sum_{j=1}^N\log\left(1+\frac{\lambda_j(0)-\mu_j(0)}{\mu_j(0)-E_t}\right). \tag{4.10}$$

On the rigidity event from (2.9), a Taylor expansion of the logarithm gives, with overwhelming probability,

$$\operatorname{Im}\sum_{j=1}^N\log\left(1+\frac{\lambda_j(0)-\mu_j(0)}{\mu_j(0)-E_t}\right)=\operatorname{Im}\sum_{j=1}^N\frac{\lambda_j(0)-\mu_j(0)}{\mu_j(0)-E_t}+\mathrm{O}\left(\sum_{j=1}^N\left|\frac{\lambda_j(0)-\mu_j(0)}{\mu_j(0)-E_t}\right|^2\right). \tag{4.11}$$

For the error term, on the rigidity event from (2.9) we can write

$$\sum_{j=1}^N\left|\frac{\lambda_j(0)-\mu_j(0)}{\mu_j(0)-E_t}\right|^2\leqslant\frac{C\varphi^2}{Nt}\cdot\frac{1}{N}\sum_{j=1}^N\frac{\operatorname{Im}E_t}{|\mu_j(0)-E_t|^2}=\frac{C\varphi^2}{Nt}\ \operatorname{Im}m_N(E_t)\leqslant\frac{C\varphi^2}{Nt}, \tag{4.12}$$

where we used $ct\leqslant\operatorname{Im}E_t\leqslant Ct$ to bound $\operatorname{Im}m_N(E_t)\leqslant C$ using (2.7). This estimate on $\operatorname{Im}m_N(E_t)$ also shows that the second term in (4.11) is negligible when inserted in (4.10).

Finally, we need to bound

$$\begin{aligned}\frac{1}{N\operatorname{Im}m_{\mathrm{sc}}(E_t)}\operatorname{Im}\sum_{j=1}^N\frac{\lambda_j(0)-\mu_j(0)}{\mu_j(0)-E_t}-\bar{\mathrm{u}}_\ell(t)=\frac1N\sum_{j=1}^N(\lambda_j(0)-\mu_j(0))\left(\frac{1}{\operatorname{Im}m_{\mathrm{sc}}(E_t)}-\frac{1}{\operatorname{Im}m_{\mathrm{sc}}(\gamma_\ell^t)}\right)\operatorname{Im}\frac{1}{\mu_j(0)-E_t}\\+\frac1N\sum_{j=1}^N\frac{\lambda_j(0)-\mu_j(0)}{\operatorname{Im}m_{\mathrm{sc}}(\gamma_\ell^t)}\ \operatorname{Im}\left(\frac{1}{\mu_j(0)-E_t}-\frac{1}{\gamma_j-\gamma_\ell^t}\right).\end{aligned}$$

For the first sum, from $|\operatorname{Im}m_{\mathrm{sc}}(E_t)-\operatorname{Im}m_{\mathrm{sc}}(\gamma_\ell^t)|\leqslant C|E_t-\gamma_\ell^t|\leqslant CN^{-1}$, $\operatorname{Im}m_{\mathrm{sc}}(E_t)\geqslant c$, and $\operatorname{Im}m_{\mathrm{sc}}(\gamma_\ell^t)\geqslant c$, on the rigidity event from (2.9) we obtain

$$\frac1N\sum_{j=1}^N|\lambda_j(0)-\mu_j(0)|\left(\frac{1}{\operatorname{Im}m_{\mathrm{sc}}(E_t)}-\frac{1}{\operatorname{Im}m_{\mathrm{sc}}(\gamma_\ell^t)}\right)\operatorname{Im}\frac{1}{\mu_j(0)-E_t}\leqslant\frac{C\varphi}{N^3}\cdot\sum\frac{t}{|\gamma_j-E_t|^2}\leqslant\frac{C\varphi}{N^2}.$$

On the same rigidity event, the second sum is bounded by

$$\frac1N\sum_{j=1}^N|\mu_j(0)-\lambda_j(0)|\left|\operatorname{Im}\frac{E_t-\gamma_\ell^t}{(\mu_j(0)-E_t)(\gamma_j-\gamma_\ell^t)}\right|\leqslant\frac{C\varphi}{N^3}\sum_j\left(\frac{1}{|\mu_j(0)-E_t|^2}+\frac{1}{|\gamma_j-\gamma_\ell^t|^2}\right)\leqslant\frac{C\varphi}{N^2t}.$$

We have thus obtained

$$\left|\frac{\operatorname{Im}L_N^H(E_t)-\operatorname{Im}L_N^{\mathrm{GOE}}(E_t)}{N\operatorname{Im}m_{\mathrm{sc}}(E_t)}-\bar{\mathrm{u}}_\ell(t)\right|\leqslant C\frac{\varphi^2}{N^2t}, \tag{4.13}$$

which concludes the proof. □

Remark 4.2. A stronger result than (4.4) actually holds, in terms of the probability bound: for any $\kappa,\varepsilon>0$ there exists $C,\delta,N_0>0$ (depending on κ and ε) such that for any $t\in(\varphi^C/N,1)$, $E\in[-2+\kappa,2-\kappa]$, $k\in[\![1,N]\!]$ satisfying $\gamma_k\in[-2+\kappa,2-\kappa]$ and $N\geqslant N_0$, we have

$$\mathbb{P}\left(\left|\lambda_k(t)-\mu_k(t)-\frac{\operatorname{Im}L_N^H(E_t)-\operatorname{Im}L_N^{\mathrm{GOE}}(E_t)}{N\operatorname{Im}m_{\mathrm{sc}}(E_t)}\right|>\frac{N^{1+\varepsilon}\max(|E-\gamma_k|,N^{-1})}{N^2t}\right)\leqslant e^{-\delta\varphi^\delta}.$$

Indeed the proof of (4.5) from [20] relies on the rigidity estimate (2.9), which holds with probability $e^{-c\varphi^c}$, so that the probability bound N^{-D} in [20, Theorem 3.1] can be strengthened to $e^{-\delta\varphi^\delta}$ for a fixed, small enough δ, by elementary changes in the proof.

This improved probability bound is not necessary for the proofs of Theorem 1.2 and Theorem 1.8 (i). It will be used in the proof of Theorem 1.8 (ii).

4.2 ***The characteristic polynomial.*** After the relaxation of individual eigenvalues in the previous subsection, we study the relaxation of L_N, an a priori more intricate problem as $\operatorname{Re} L_N$ depends on the full spectrum. Our results are of two types: relaxation of the full characteristic polynomial in the rectangle $[-2+\kappa, 2-\kappa]\times[N^{-1},1]$ and $t\in[\varphi^{-K},1]$, up to an error $(\log N)^{1/2}$ (Proposition 4.3), and relaxation of its maximum on $[-2+\kappa,2-\kappa]$ for $t=\Omega(1)$, up to tightness (Proposition 4.4) .

Proposition 4.3. *For all $K>10$, $\varepsilon,\kappa>0$, the following holds. For $z=E+\mathrm{i}\eta$, uniformly in $\eta\in(N^{-1},1)$ and $t\in[\varphi^{-K},1]$, we have*

$$\mathbb{P}\left(\max_{-2+\kappa<E<2+\kappa}\left|\sum\log\left(z-\mu_k(t)\right)-\sum\log\left(z-\lambda_k(t)\right)\right|>(\log N)^{1/2+\varepsilon}\right)=\mathrm{o}(1). \tag{4.14}$$

Proof. If $\varphi^{100}/N<\eta<1$, this follows from [20] and Corollary A.4. Indeed, by integrating over the parameter $\nu\in[0,1]$ in [20, Proposition 2.11], we have

$$\max_{-2+\kappa<E<2+\kappa}\left|\sum_{k=1}^N\log\frac{z-\mu_k(t)}{z-\lambda_k(t)}-\sum_{k=1}^N\log\frac{z_t-\mu_k(0)}{z_t-\lambda_k(0)}\right|<1 \tag{4.15}$$

with overwhelming probability. Together with Corollary A.4 (noting $z_t\geqslant C^{-1}t$ for some $C(\kappa)>0$ by an explicit computation using (4.3)), it implies that

$$\max_{-2+\kappa<E<2+\kappa}\left|\sum\log\left(z-\mu_k(t)\right)-\sum\log\left(z-\lambda_k(t)\right)\right|<(\log N)^{\varepsilon} \tag{4.16}$$

with probability $1-\mathrm{o}(1)$.

We now consider the case where $N^{-1}<\operatorname{Im}\eta<\varphi^{100}/N$, and denote $\widetilde{z}=E+\mathrm{i}\varphi^{100}/N$. Given that (4.16) holds for $\widetilde{z}$, we only need to prove that

$$\max_{-2+\kappa<E<2+\kappa}\left|\sum_{k=1}^N\log\frac{z-\mu_k(t)}{z-\lambda_k(t)}-\sum_{k=1}^N\log\frac{\widetilde{z}-\mu_k(t)}{\widetilde{z}-\lambda_k(t)}\right|<(\log N)^{1/2+\varepsilon} \tag{4.17}$$

with probability $1-\mathrm{o}(1)$ to complete the proof.

We divide the sum in (4.17) into two parts, depend on whether $|\gamma_k-E|>N^{-1/2}$ or not. For the terms such that $|\gamma_k-E|>N^{-1/2}$, by the rigidity bound (2.9) and the Taylor expansion for $x\mapsto\log(1+x)$ we have

$$\begin{aligned}\sum_{|\gamma_k-E|>N^{-1/2}}\left|\sum\log(\widetilde{z}-\mu_k(t))-\sum\log(z-\mu_k(t))\right|&=\sum_{|\gamma_k-E|>N^{-1/2}}\left|\log\left(1+\frac{\widetilde{z}-z+\mu_k(t)-\lambda_k(t)}{z-\lambda_k(t)}\right)\right|\\&\leqslant C\sum_{|\gamma_k-E|>N^{-1/2}}\left|\frac{\widetilde{z}-z}{z-\gamma_k}\right|+\left|\frac{\mu_k(t)-\lambda_k(t)}{z-\gamma_k}\right|\leqslant\frac{C\varphi^{200}}{N}\frac{1}{|z-\gamma_k|}\leqslant\frac{\varphi^{300}}{N^{1/2}}=\mathrm{o}(1).\end{aligned}$$

The same holds when replacing $\boldsymbol{\mu}$ with $\boldsymbol{\lambda}$. Hence to prove (4.17) we only need to obtain the following bounds for the terms such that $|\gamma_k-E|<N^{-1/2}$:

$$\max_{-2+\kappa<E<2+\kappa}\left|\sum_{|\gamma_k-E|<N^{-1/2}}\log(z-\mu_k(t))-\sum_{|\gamma_k-E|<N^{-1/2}}\log(z-\lambda_k(t))\right|<(\log N)^{1/2+\varepsilon}, \tag{4.18}$$

and the same bound with z replaced by $\widetilde{z}$. In the following we prove only (4.18), as the proof for $\widetilde{z}$ is the same. For this, we now bound

$$\left|\sum_{|\gamma_k-E|<N^{-1/2}}\log(z-\mu_k(t))-\sum_{|\gamma_k-E|<N^{-1/2}}\log(z-\lambda_k(t))\right| \tag{4.19}$$

$$\leqslant\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\|\mu_k(t)-E|\leqslant(\log N)^{\varepsilon}N^{-1}}}\left|\log\frac{z-\mu_k(t)}{z-\lambda_k(t)}\right|+\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\(\log N)^{\varepsilon}N^{-1}\leqslant|\mu_k(t)-E|\leqslant\varphi^3N^{-1}}}\left|\log\frac{z-\mu_k(t)}{z-\lambda_k(t)}\right| \tag{4.20}$$

$$+\left|\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\\varphi^{100}N^{-1}\leqslant|\mu_k(t)-E|}}\log\frac{z-\mu_k(t)}{z-\lambda_k(t)}\right|. \tag{4.21}$$

We begin by considering the contribution from the first sum in (4.20). First, suppose that $|z-\mu_k| \geqslant |z-\lambda_k|$. Using $|\log|1+z|| < \log(1+|z|)$ for $|1+z|>1$ and Proposition 4.1, we obtain that with high probability

$$\left|\log \frac{|z-\mu_k(t)|}{|z-\lambda_k(t)|}\right| \leqslant \log\left(1+\left|\frac{\mu_k(t)-\lambda_k(t)}{z-\mu_k(t)}\right|\right)$$
$$\leqslant \log\left(1+\left|\frac{\operatorname{Im} L_N^H(E_t)-\operatorname{Im} L_N^{\mathrm{GOE}}(E_t)}{N \operatorname{Im} m_{\mathrm{sc}}(E_t)\,(z-\mu_k(t))}+\mathrm{O}\left(\frac{N^{1+\varepsilon}\max(|E-\gamma_k|,N^{-1})}{N^2 t|z-\mu_k(t)|}\right)\right|\right). \quad (4.22)$$

By Corollary A.4, we have for all $\eta' \in (0,1)$ that $\mathbb{P}(\mathcal{G}_{\eta'}) = 1-\mathrm{o}(1)$, where

$$\mathcal{G}_{\eta'} = \left\{\max_{-2+\kappa<E<2+\kappa}\left|\sum \log\left(z_t'-\mu_k(0)\right)-\sum\log\left(z_t'-\lambda_k(0)\right)\right| \leqslant (\log N)^{\varepsilon}\right\}, \quad z'=E+\mathrm{i}\eta' \quad (4.23)$$

and the implicit constant in the o(1) term depends only on the choices of K, ε, and κ. On $\mathcal{G}_\eta$, using $\operatorname{Im} z \geqslant N^{-1}$ and $|N(z-\mu_k(t))| \geqslant 1$ in (4.22), we obtain

$$\left|\log\frac{|z-\mu_k(t)|}{|z-\lambda_k(t)|}\right| \leqslant \log\left(1+2(\log N)^{\varepsilon}+\frac{N^{\varepsilon}\max(|E-\gamma_k|,N^{-1})}{t}\right) < C\varepsilon \log\log N. \quad (4.24)$$

In (4.24), we used that $|\mu_k(t)-E| \leqslant (\log N)^{\varepsilon}/N$ implies $|E-\gamma_k| < C\varphi/N$ when rigidity (2.9) holds. The same bound as (4.24) naturally holds for $|z-\mu_k| > |z-\lambda_k|$. Thus for the first sum in (4.20) we conclude that

$$\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ |\mu_k(t)-E|\leqslant(\log N)^{\varepsilon}N^{-1}}} \left|\log\frac{|z-\mu_k(t)|}{|z-\lambda_k(t)|}\right| \leqslant C\varepsilon\log\log N\left|\left\{|\mu_k-E| \leqslant \frac{(\log N)^{\varepsilon}}{N}\right\}\right|. \quad (4.25)$$

Theorem 2.7 applied with $k=(\log N)^{1/2}$ gives

$$\mathbb{P}\left(\left|\left\{|\mu_k-E|\leqslant\frac{(\log N)^{\varepsilon}}{N}\right\}\right| > (\log N)^{1/2}\right) = \mathrm{o}(1). \quad (4.26)$$

Then (4.25) and (4.26) together imply that, with probability $1-\mathrm{o}(1)$,

$$\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ |\mu_k(t)-E|\leqslant(\log N)^{\varepsilon}N^{-1}}} \left|\log\frac{z-\mu_k(t)}{z-\lambda_k(t)}\right| \leqslant C(\log N)^{1/2+\varepsilon}. \quad (4.27)$$

We next consider the second sum on the left side of (4.20), and work on the event $\mathcal{G}_\eta$. By Taylor expansion and Proposition 4.1,

$$\left|\log\frac{z-\mu_k(t)}{z-\lambda_k(t)}\right| \leqslant \left|\log\left(1+\frac{L^H(E,t)-L^{\mathrm{GOE}}(E,t)}{N(z-\mu_k(t))}+\mathrm{O}\left(\frac{N^{1+\varepsilon}\max(|E-\gamma_k|,N^{-1})}{N^2 t|z-\mu_k(t)|}\right)\right)\right| \quad (4.28)$$
$$= \mathrm{O}\left(\frac{L^H(E,t)-L^{\mathrm{GOE}}(E,t)}{N(z-\mu_k(t))}\right)+\mathrm{O}\left(\frac{N^{1+\varepsilon}\max(|E-\gamma_k|,N^{-1})}{N^2 t|z-\mu_k(t)|}\right) \quad (4.29)$$
$$= \frac{C(\log N)^{\varepsilon/2}}{N|z-\mu_k(t)|}+\frac{CN^{-1+\varepsilon}t^{-1}\varphi}{N|z-\mu_k(t)|} \leqslant \frac{C(\log N)^{\varepsilon/2}}{N|z-\mu_k(t)|}. \quad (4.30)$$

In the second term in (4.29), we used rigidity and $|\mu_k(t)-E|<\varphi N^{-1}$ to show that $|E-\gamma_k| \leqslant 2\varphi N^{-1}$.

We now bound

$$\sum_{(\log N)^{\varepsilon}/N<|\mu_k-E|<\varphi^{100}/N} \frac{1}{|z-\mu_k(t)|} \leqslant \sum_{1\leqslant j\leqslant A(\log\log N)^2} \frac{1}{(2^j/N)}\left|\left\{|\mu_k-E| \in [2^j/N, 2^{j+1}/N]\right\}\right|, \quad (4.31)$$

where $A>0$ is a constant depending only on the constant C_0 used to define φ in (2.6). Set $\tau=(\log N)^{-1/2}$ and define the event

$$\mathcal{A}_j^E = \left\{\left|\{|\mu_k-E|\in[2^j/N,2^{j+1}/N]\}\right| \leqslant \frac{2^j}{\tau}\right\}. \quad (4.32)$$

From Theorem 2.7 applied with $\gamma = 1/2$ and $k = \tau^{-1}2^j$, we have

$$\mathbb{P}\left(\left(\mathcal{A}_j^E\right)^{\mathrm{c}}\right) \leqslant \exp\left(\frac{1}{4}\log(C\tau)(2^j/\tau)^2\right) + \exp\left(-N^c\right) \tag{4.33}$$

for all $j \leqslant A(\log\log N)^2$. Define

$$\mathcal{A} = \bigcap_{\substack{1\leqslant k\leqslant N\\ 0\leqslant j\leqslant A(\log\log N)^2}} \mathcal{A}_j^{\gamma_k}. \tag{4.34}$$

From (4.33), we obtain using a union bound that

$$\mathbb{P}(\mathcal{A}) = 1 - \mathrm{o}(1). \tag{4.35}$$

On the event where both $\mathcal{A}$ and the rigidity estimate (2.9) hold, by (4.31) we have

$$\frac{1}{N}\sum_{(\log N)^\varepsilon/N<|\mu_k-E|<\varphi^{100}/N} \frac{1}{|z-\mu_k(t)|} \leqslant A(\log N)^{1/2}(\log\log N)^2 \leqslant (\log N)^{1/2+\varepsilon/4}. \tag{4.36}$$

Combining (4.30), and (4.36), we find

$$\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ (\log N)^\varepsilon N^{-1}\leqslant|\mu_k(t)-E|\leqslant\varphi^{100}N^{-1}}} \left|\log\frac{z-\mu_k(t)}{z-\lambda_k(t)}\right| \leqslant (\log N)^{1/2+\varepsilon}. \tag{4.37}$$

Finally, we consider the sum in (4.21). By Taylor expansion and rigidity (2.9), we have

$$\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ \varphi^{100}N^{-1}\leqslant|\mu_k(t)-E|}} \log\left(\frac{z-\mu_k(t)}{z-\lambda_k(t)}\right)$$

$$= \sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ \varphi^{100}N^{-1}\leqslant|\mu_k(t)-E|}} (\lambda_k(t)-\mu_k(t))\frac{1}{z-\lambda_k} + \mathrm{O}\left(\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ \varphi^{100}N^{-1}\leqslant|\mu_k(t)-E|}} \frac{|\lambda_k(t)-\mu_k(t)|^2}{|z-\lambda_k(t)|^2}\right). \tag{4.38}$$

The above error term is again bounded by rigidity, as it is of order $\frac{\varphi^2}{N^2}\sum_{k\geqslant\varphi^{100}}\frac{N^2}{k^2} = \mathrm{o}(1)$. For the first term on the left side of (4.38), we use Proposition 4.1 to write it as

$$\frac{L^H(E,t)-L^{\mathrm{GOE}}(E,t)}{N}\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ \varphi^{100}N^{-1}\leqslant|\mu_k(t)-E|}} \frac{1}{z-\lambda_k} + \mathrm{O}\left(\frac{N^\varepsilon}{N^{3/2}t}\right)\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ \varphi^{100}N^{-1}\leqslant|\mu_k(t)-E|}} \frac{1}{|z-\gamma_k|}, \tag{4.39}$$

with overwhelming probability. The second error term above is again o(1) by rigidity. For the first term in (4.39), we work on $\mathcal{G}_\eta$ to obtain the high-probability bound

$$\frac{(\log N)^\varepsilon}{N}\left|\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ \varphi^{100}N^{-1}\leqslant|\mu_k(t)-E|}} \frac{1}{z-\lambda_k}\right| \leqslant \frac{(\log N)^\varepsilon}{N}\left|\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ \varphi^{100}N^{-1}\leqslant|\mu_k(t)-E|}} \frac{1}{z-\gamma_k}\right| \tag{4.40}$$

$$+ \frac{(\log N)^\varepsilon}{N}\left|\sum_{\substack{|\gamma_k-E|<N^{-1/2}\\ \varphi^{100}N^{-1}\leqslant|\mu_k(t)-E|}} \frac{|\lambda_k-\gamma_k|}{|z-\gamma_k|^2}\right|. \tag{4.41}$$

The second term above is again negligible by rigidity. Let k_0 be the index that minimizes $|E-\gamma_k|$ for $k \leqslant N$. In the right side of (4.40), the contribution from $|k-k_0| > N^{1/4}$ is negligible by rigidity. In this term, we may then replace the summation over indices such that $|\gamma_k - E| < N^{-1/2}$ and $\varphi^{100}N^{-1} \leqslant |\mu_k(t) - E|$ with one over indices k such that $N\varphi^2 < |k-k_0| < N^{1/4}$. In this replacement, any indices with $|\mu_k(t)-E| \leqslant \varphi^{100}N^{-1}$ may be restored as necessary using the arguments leading to (4.37).

Note that $\operatorname{Im}\frac{1}{z-\gamma_k} = \frac{\eta}{|z-\gamma_k|^2}$ and

$$\frac{c|k-k_0|}{N} \leqslant |E-\gamma_k| \leqslant \frac{c^{-1}|k-k_0|}{N} \tag{4.42}$$

for $k \geqslant \varphi^2 N$ by rigidity and $|\gamma_{k_0} - E| \leqslant CN^{-1}$. Then

$$\sum_{\substack{|\gamma_k - E| < N^{-1/2} \\ \varphi^{100}N^{-1} \leqslant |\mu_k(t)-E|}} \operatorname{Im}\frac{1}{z-\gamma_k} \leqslant \sum_{N\varphi^2 < k-k_0 < N^{1/4}} \frac{CN^2\eta}{|k-k_0|^2} \leqslant \frac{CN\eta}{\varphi^2}. \tag{4.43}$$

Further,

$$\operatorname{Re} \sum_{N\varphi^2 < k-k_0 < N^{1/4}} \left(\frac{1}{z-\gamma_k} + \frac{1}{z-\gamma_{2k_0-k}} \right) = \\ \sum_{\varphi^2 N < k-k_0 < N^{1/4}} \left(\frac{2E-\gamma_k-\gamma_{2k_0-k}}{|z-\gamma_k|^2} + (E-\gamma_{2k_0-k}) \left(\frac{1}{|z-\gamma_k|^2} - \frac{1}{|z-\gamma_{2k_0-k}|^2} \right) \right). \tag{4.44}$$

By a direct computation, we have

$$2E - \gamma_k - \gamma_{2k_0-k} = \mathrm{O}\big((k-k_0)^2/N^2\big). \tag{4.45}$$

Together with (4.42) this gives

$$\sum_{\varphi^2 N < k-k_0 < N^{1/4}} \frac{2E-\gamma_k-\gamma_{2k_0-k}}{|z-\gamma_k|^2} = \mathrm{O}(N^{1/4}). \tag{4.46}$$

Also, using (4.45) and (4.42), we compute

$$\frac{1}{|z-\gamma_k|^2} - \frac{1}{|z-\gamma_{2k_0-k}|^2} = \frac{(|z-\gamma_{2k_0-k}| - |z-\gamma_k|)(|z-\gamma_{2k_0-k}| + |z-\gamma_k|)}{|z-\gamma_k|^2|z-\gamma_{2k_0-k}|^2} \tag{4.47}$$

$$= \mathrm{O}\left(\frac{CN^{-3}|k-k_0|^3}{N^{-4}(k-k_0)^4} \right) = \mathrm{O}\left(\frac{N}{|k-k_0|} \right). \tag{4.48}$$

Putting (4.46) and (4.47) into (4.44), and combining this with (4.43) and our previous bounds, we find

$$\left| \sum_{\substack{|\gamma_k - E| < N^{-1/2} \\ \varphi^{100}N^{-1} \leqslant |\mu_k(t)-E|}} \log\left(\frac{z-\mu_k(t)}{z-\lambda_k(t)} \right) \right| = \mathrm{o}(1). \tag{4.49}$$

We finish the proof by combining (4.27), (4.37), and (4.49). □

The following Proposition directly implies Theorem 1.4. In the statement and its proof, we abbreviate

$$L_N^{\mathbf{x}}(z) = \sum_{k=1}^N \log(z-x_k) - N\int \log(z-\lambda)\rho_{\mathrm{sc}}(\lambda)\mathrm{d}\lambda. \tag{4.50}$$

Note that the proposition below uses Theorem 1.2 in its proof, but there is no circularity in the sense that the proposition is not used in proof of Theorem 1.2.

Proposition 4.4. *Let $\kappa, t > 0$ be fixed. Then tightness holds for the sequence of random variables*

$$\Big(\sup_{|E|<2-\kappa} \operatorname{Re} L_N^{\boldsymbol{\mu}_t}(E) - \sup_{|E|<2-\kappa} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E) \Big)_{N\geqslant 1}.$$

Proof. For the proof we define

$$m_N = \sup_{|E|<2-\kappa} \operatorname{Re} L_N^{\boldsymbol{\mu}_t}(E).$$

We will prove that for any $\varepsilon > 0$ there exists $C > 0$ such that for all N we have

$$\mathbb{P}\left(\sup_{|E|<2-\kappa} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E) \geqslant m_N - C\right) \geqslant 1-\varepsilon,$$

and the same result holds when permuting $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$. The claimed tightness then follows directly.

In the following, we denote

$$\ell_E = \frac{\operatorname{Im} L_N^{\boldsymbol{\lambda}_0}(E_t) - \operatorname{Im} L_N^{\boldsymbol{\mu}_0}(E_t)}{N \operatorname{Im} m_{\rm sc}(E_t)}, \quad \eta = \frac{1}{N}, \quad \widetilde{\eta} = \frac{\varphi^{100}}{N}.$$

Consider the following events depending on N and sometimes additional parameters $M, \delta > 0$:

$$A = A(M) = \bigcap_{|E|<2-\kappa} \left\{ \left| \operatorname{Re} L_N^{\boldsymbol{\mu}_t}(E+\mathrm{i}\widetilde{\eta}) - \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\widetilde{\eta}) \right| \leqslant M \right\}, \tag{4.51}$$

$$B = B(M) \tag{4.52}$$

$$= \bigcap_{|E|<2-\kappa} \left\{ \left| (\operatorname{Re} L_N^{\boldsymbol{\mu}_t}(E+\mathrm{i}\eta) - \operatorname{Re} L_N^{\boldsymbol{\mu}_t}(E+\mathrm{i}\widetilde{\eta})) - (\operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\eta+\ell_E) - \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\widetilde{\eta}+\ell_E)) \right| \leqslant M \right\}, \tag{4.53}$$

$$C = C(M) = \bigcap_{|E|<2-\kappa} \left\{ \left| \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\widetilde{\eta}+\ell_E) - \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\widetilde{\eta}) \right| \leqslant M \right\}, \tag{4.54}$$

$$D = D(M) = \left\{ \sup_{E\in[-2+\kappa, 2-\kappa]} (|L_N^{\boldsymbol{\lambda}_0}(E_t)| + |L_N^{\boldsymbol{\mu}_0}(E_t)|) \leqslant M \right\}, \tag{4.55}$$

$$E = E(\delta) = \left\{ \sup_{E\in[-2+\kappa-\delta, 2-\kappa+\delta]} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\eta) \leqslant (1+\frac{10}{N\delta})| \sup_{E\in[-2+\kappa-2\delta, 2-\kappa+2\delta]} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E)| + 3 \right\}, \tag{4.56}$$

$$F = F(\delta) = \left\{ \sup_{E\in[-2+\kappa-2\delta, -2+\kappa]\cup[2-\kappa, 2-\kappa+2\delta]} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t} \leqslant \frac{\log N}{10} \right\}, \tag{4.57}$$

$$G = \left\{ \sup_{E\in[2+\kappa, 2-\kappa]} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t} \geqslant \frac{\log N}{2} \right\} \cap \left\{ \sup_{E\in[2+\frac{\kappa}{2}, 2-\frac{\kappa}{2}]} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t} \leqslant 10 \log N \right\}. \tag{4.58}$$

From (3.1), there is some fixed $C_1 > 0$ such that, for any N, there exists $|E_0| < 2-\kappa$ such that

$$\operatorname{Re} L_N^{\boldsymbol{\mu}_t}(E_0+\mathrm{i}\eta) \geqslant m_N - C_1.$$

Therefore, on A we have

$$\operatorname{Re} L_N^{\boldsymbol{\mu}_t}(E_0+\mathrm{i}\eta) - \operatorname{Re} L_N^{\boldsymbol{\mu}_t}(E_0+\mathrm{i}\widetilde{\eta}) + \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E_0+\mathrm{i}\widetilde{\eta}) \geqslant m_N - C_1 - M,$$

and on $A \cap B$ we can write

$$\operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E_0+\mathrm{i}\eta+\ell_{E_0}) - \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E_0+\mathrm{i}\widetilde{\eta}+\ell_{E_0}) + \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E_0+\mathrm{i}\widetilde{\eta}) \geqslant m_N - C_1 - 2M.$$

On $A \cap B \cap C$ we therefore have

$$\operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E_0+\mathrm{i}\eta+\ell_{E_0}) \geqslant m_N - C_1 - 3M.$$

Assuming our parameters satisfy $M/N = \mathrm{o}(\delta)$ and $\frac{\log N}{N\delta} = \mathrm{o}(1)$, on $A \cap B \cap C \cap D \cap E \cap G$ this yields

$$\sup_{|E|<2-\kappa+2\delta} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E) \geqslant m_N - C_1 - 4M - 10.$$

Finally on $A \cap B \cap C \cap D \cap E \cap F \cap G$ we obtain

$$\sup_{|E|<2-\kappa} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E) \geqslant m_N - C_1 - 4M - 10.$$

The proof will therefore be complete if we obtain that, for fixed $\varepsilon > 0$, each one of the ensembles A, B, C, D, E, F, G has probability larger than $1-\varepsilon$ for large enough N. For $A(M), C(M), D(M), E(\delta)$ and $F(\delta)$ this will be true for the choice $\delta = N^{-1+\theta}$, $\theta \in (0, \frac{1}{10})$ arbitrary, and M fixed, large enough.

First, Corollary A.4 gives $\mathbb{P}(D(M)) > 1-\varepsilon$ for some fixed $M = M(\varepsilon, \kappa)$ and large enough N. Then from (4.15) we have $\mathbb{P}(A(M+1)) \geqslant \mathbb{P}(A(M+1) \cap D(M)) \geqslant 1-\varepsilon$ for large enough N.

To bound $\mathbb{P}(C)$, note that $\inf_{|E|<2-\kappa} \operatorname{Im} m_{\rm sc}(E_t) > c(\kappa) > 0$ so that on $D(M)$ we have $\ell_E = \mathrm{O}(1/N)$. Therefore on the rigidity event from (2.9) and on $D(M)$, a Taylor expansion gives

$$\left|\operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\widetilde\eta+\ell_E) - \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\widetilde\eta)\right| \leqslant \ell_E \left|\sum_i \frac{1}{E+\mathrm{i}\widetilde\eta - \lambda_i(t)}\right| + \mathrm{O}\left(\ell_E^2 \sum_i \frac{1}{|E+\mathrm{i}\widetilde\eta-\lambda_i(t)|^2}\right) \leqslant C_2 N \ell_E. \quad (4.59)$$

for some fixed $C_2 = C_2(\kappa)$. Hence there exists $\widetilde M = \widetilde M(\kappa, \varepsilon)$ such that $\mathbb{P}(C(\widetilde M)) > 1-\varepsilon$ for large enough N.

Moreover, Theorem 1.2 implies $\mathbb{P}(G) \geqslant 1-\varepsilon$ for large enough N, and similarly (1.12) implies $\mathbb{P}(F) \geqslant 1-\varepsilon$ for $\theta < 1/10$.

We now prove that for $\theta > 0$ we have $\mathbb{P}(E) > 1-\varepsilon$ for large enough N. First, similarly to (3.20), with probability $1-\mathrm{O}(N^{-20})$ we have, for any $|E| < 2-\kappa+\delta$, for our choice $\delta = N^{-1+\theta}$,

$$\left|\operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\eta) - \int_{[-2+\frac{\kappa}{2}, 2-\frac{\kappa}{2}]} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(u) \cdot \frac{\eta}{\eta^2+(E-u)^2}\frac{\mathrm{d}u}{\pi}\right| \leqslant 1.$$

Moreover, from Theorem 1.2, with probability $1-\mathrm{o}(1)$ we have

$$\int_{[-2+\frac{\kappa}{2}, -2+\kappa-2\delta]\cup[2-\kappa+2\delta, 2-\frac{\kappa}{2}]} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(u) \cdot \frac{\eta}{\eta^2+(E-u)^2}\frac{\mathrm{d}u}{\pi} \leqslant 10\log N \cdot \frac{\eta}{\delta} \leqslant 1$$

for any $|E| < 2-\kappa+\delta$. With above two equations we obtain, with probability $1-\mathrm{o}(1)$, for any $|E| < 2-\kappa+\delta$,

$$\operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(E+\mathrm{i}\eta) \leqslant \int_{[-2+\kappa-2\delta, 2-\kappa+2\delta]} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(u) \cdot \frac{\eta}{\eta^2+(E-u)^2}\frac{\mathrm{d}u}{\pi} + 2$$
$$\leqslant \sup_{|u|<2-\kappa+2\delta} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(u) \cdot \int_{[-2+\kappa-\delta, 2-\kappa+\delta]} \frac{\eta}{\eta^2+(E-u)^2}\frac{\mathrm{d}u}{\pi} + 2 \leqslant (1+10\frac{\eta}{\delta})|\sup_{|u|<2-\kappa+2\delta} \operatorname{Re} L_N^{\boldsymbol{\lambda}_t}(u)| + 2,$$

which concludes the proof that $\mathbb{P}(E) > 1-\varepsilon$ for large enough N.

Finally, we prove that $\mathbb{P}(B) > 1-\varepsilon$ for large enough N. First, we can easily ignore the contribution from eigenvalues close to the edge, because for any $|E| < 2-\kappa$, we have

$$\sum_{j:||\gamma_j|-2|<\frac{\kappa}{10}} |(\log(E+\mathrm{i}\eta-\mu_j(t)) - \log(E+\mathrm{i}\widetilde\eta-\mu_j(t))) - (\log(E+\mathrm{i}\eta+\ell_E-\lambda_j(t)) - \log(E+\mathrm{i}\widetilde\eta+\ell_E-\lambda_j(t)))|$$
$$\leqslant \sum_{j:||\gamma_j|-2|<\frac{\kappa}{10}} \widetilde\eta \left|\frac{1}{E+\mathrm{i}\eta-\mu_j(t)} - \frac{1}{E+\mathrm{i}\eta+\ell_E-\lambda_j(t)}\right|$$
$$+ C \sum_{j:||\gamma_j|-2|<\frac{\kappa}{10}} \widetilde\eta^2 \left|\frac{1}{|E+\mathrm{i}\eta-\mu_j(t)|^2} + \frac{1}{|E+\mathrm{i}\eta+\ell_E-\lambda_j(t)|^2}\right|$$
$$\leqslant N^{-1+\widetilde\varepsilon},$$

where the last inequality holds on D and the rigidity event, for any fixed $\widetilde\varepsilon > 0$ and N large enough.

Now fix α such that $\gamma_\alpha - 2 < \kappa/10$. From Proposition 4.1, with high probability we have, for any $k \in [\![\alpha N, (1-\alpha)N]\!]$,

$$\left|\lambda_k(t) - \mu_k(t) - \ell_{\gamma_k}\right| < N^{-2+\widetilde\varepsilon}.$$

Thus choosing k_0 such that $|E-\gamma_{k_0}| \leqslant \frac{C_3}{N}$, $C_3 = C_3(\kappa)$, and $\ell_{\gamma_k} - \ell_E = \mathrm{O}(\varphi^2(|k-k_0|+1)/N^2)$ (from (4.7) and (4.13)) we have

$$\left|\lambda_k(t) - \mu_k(t) - \ell_E\right| < C_4 \frac{N^{\widetilde\varepsilon} + |k-k_0|}{N^2}$$

for some $C_4 = C_4(\kappa)$. Note that for k close to k_0, the above error term is much smaller than the regularization scale η, so that by Taylor expansion we obtain

$$\sum_{\hat{j}\geqslant\alpha N} |\log(E+\mathrm{i}\eta-\mu_j(t)) - \log(E+\mathrm{i}\eta+\ell_E-\lambda_j(t))| \leqslant \frac{C_5}{N^2}\sum_{\hat{j}\geqslant\alpha N}\left|\frac{N^{\widetilde{\varepsilon}}+|j-k_0|}{\frac{|j-k_0|+1}{N}}\right| \leqslant C_6$$

for some constants $C_5(\kappa,\varepsilon)$, $C_6(\kappa,\varepsilon)$. This concludes the proof with the choice $M = C_6+1$. □

4.3 Large moments. We now prove quantitative relaxation for large moments of $\operatorname{Im} L_N$. We denote $\overline{\mathrm{Tr}}h(M) = \mathrm{Tr}h(M) - N\int h(x)\rho_{\mathrm{sc}}(x)\mathrm{d}x$

Lemma 4.5. *Let $H = H_0$ be a symmetric Wigner matrix and $\kappa,\varepsilon,A>0$. Then there exists $N_0 = N_0(\kappa,\varepsilon,A)$ such that for any $E\in[-2+\kappa,2-\kappa]$, $1\leqslant p\leqslant A\log N$ and $t\in[\varphi^A,1]$ we have*

$$\mathbb{E}\left[(\overline{\mathrm{Tr}}\mathbb{1}_{[E,\infty)}(H_t))^{2p}\right] \leqslant \left(\frac{2}{e\pi^2}+\varepsilon\right)^p p^p(\log N)^p.$$

The same bound holds for Hermitian Wigner matrices, with the prefactor $(\frac{1}{e\pi^2}+\varepsilon)^p$.

Proof. We first prove an equivalent concentration result for the Gaussian ensembles, and then extend it to the Gaussian divisible ensemble thanks to our results on relaxation of the Dyson Brownian motion.

First step: Concentration for GUE and GOE. The proof first requires some concentration estimates for the GUE and GOE. From [56, Theorem 1.1], there is a C_1 such that uniformly in γ in any compact set and $x\in[-2+\kappa,2-\kappa]$ we have

$$\mathbb{E}_{\mathrm{GUE}}[e^{\gamma\pi\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(G)}] \leqslant C_1 e^{\frac{\gamma^2}{4}\log N}. \tag{4.60}$$

With Markov's inequality, optimization in γ gives

$$\mathbb{P}_{\mathrm{GUE}}(|\overline{\mathrm{Tr}}\mathbb{1}_{[E,\infty)}(G)|>x) \leqslant C_1 e^{\frac{\gamma^2}{4}\log N-\gamma\pi x} \leqslant C_1 e^{-\frac{\pi^2x^2}{\log N}}\mathbb{1}_{x\leqslant\frac{B\log N}{2}} + C_1 e^{\frac{B^2\pi^2}{4}\log N - B\pi^2 x}\mathbb{1}_{x\geqslant\frac{B\log N}{2}} \tag{4.61}$$

for any fixed, arbitrarily large B, uniformly in $-2+\kappa\leqslant E\leqslant 2-\kappa$, with $C_1 = C_1(B,\kappa)$.

For the GOE, we follow [74, Proof of Lemma 23] and write the equality in law

$$\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(G_2) = \frac{1}{2}[\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(G_1)+\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(G_1')] + X \tag{4.62}$$

where G_2 is a GUE, G_1 and G_1' are independent GOE, and X is a random variable satisfying $|X|\leqslant 2$ almost surely. Together with (4.60) this implies

$$\left(\mathbb{E}_{\mathrm{GOE}}[e^{\gamma\pi\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(G)}]\right)^2 \leqslant e^{4\gamma\pi}\mathbb{E}_{\mathrm{GUE}}[e^{2\gamma\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(G)}] \leqslant C_2 e^{\gamma^2\log N}.$$

Similarly to (4.61), we conclude that

$$\mathbb{P}_{\mathrm{GOE}}(|\overline{\mathrm{Tr}}\mathbb{1}_{[E,\infty)}(G)|>x) \leqslant C_2 e^{-\frac{\pi^2x^2}{2\log N}}\mathbb{1}_{x\leqslant B\log N} + C_2 e^{\frac{B^2\pi^2}{2}\log N - B\pi^2 x}\mathbb{1}_{x\geqslant B\log N} \tag{4.63}$$

for any fixed, arbitrarily large B, uniformly in $-2+\kappa\leqslant E\leqslant 2-\kappa$, with $C_2 = C_2(B,\kappa)$.

Second step: The weakly Gaussian-divisible ensemble. We only consider the symmetric universality class (the proof for the Hermitian one is identical from now), a weakly Gaussian divisible matrix H_t with spectrum $\boldsymbol{\lambda}(t)$ coupled with the spectrum $\boldsymbol{\nu}(t)$ of G, a GOE matrix. Define the good set

$$\mathscr{G} = \bigcap_{\kappa N\leqslant k\leqslant(1-\kappa)N}\left\{\left|\lambda_k(t)-\mu_k(t)-\frac{\operatorname{Im}L_N^H(\gamma_k^t)-\operatorname{Im}L_N^G(\gamma_k^t)}{N\operatorname{Im}m_{\mathrm{sc}}(\gamma_k^t)}\right|\leqslant\frac{N^{\varepsilon}}{N^2t}\right\}$$
$$\bigcap_{1\leqslant j\leqslant N,s\in\{0,t\}}\left\{|\lambda_j(s)-\gamma_j|+|\mu_j(s)-\gamma_j|\leqslant 2\varphi\,\hat{j}^{-\frac13}N^{-\frac23}\right\}.$$

From Remark 4.2 and (2.9) we have $\mathbb{P}(\mathscr{G}) \geqslant 1 - e^{-\delta\varphi^\delta}$ for a fixed $\delta > 0$. For $p = \mathrm{O}(\log N)$ this implies

$$\mathbb{E}\left[|\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(H_t)|^{2p}\mathbb{1}_{\mathscr{G}^c}\right] \leqslant N^{2p} \cdot e^{-\delta\varphi^\delta/2} \ll 1.$$

We now bound

$$\mathbb{E}[|\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(H_t)|^{2p}\mathbb{1}_{\mathscr{G}}] = 2p\int_0^{100\varphi} u^{2p-1}\left(\mathbb{P}\left(\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(H_t) > u, \mathscr{G}\right) + \mathbb{P}\left(\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(H_t) < -u, \mathscr{G}\right)\right)\mathrm{d}u.$$

We only consider $\mathbb{P}\left(\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(H_t) > u, \mathscr{G}\right)$, as the same proof applies to the other term. We define γ_k the quantile closest to x, $n = \lfloor u/\pi \rfloor$ and $j = k - n + 2$. With (3.25) and the definition of $\mathscr{G}$, there is some $c = c(\kappa) > 0$ such that for any $\theta \in [0, 1/10]$ (eventually we will choose $\theta \to 0$) we have,

$$\begin{aligned}\mathbb{P}\left(\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(H_t) > u, \mathscr{G}\right) &\leqslant \mathbb{P}\left(\lambda_j(t) - \gamma_j > \gamma_k - \gamma_j, \mathscr{G}\right)\\ &\leqslant \mathbb{P}_{\mathrm{GOE}}\left(\mu_j(t) - \gamma_j > (1-\theta)(\gamma_k - \gamma_j)\right) + \mathbb{P}(|\operatorname{Im} L_N^H(\gamma_j^t)| > c\theta u) + \mathbb{P}(|\operatorname{Im} L_N^G(\gamma_j^t)| > c\theta u).\end{aligned}$$

We first bound the above GOE probability for different ranges of u (remember $j = j(u)$). From (3.28) and (4.63) we have the following: For any $\varepsilon, \kappa, B > 0$ there exists $C_3(\varepsilon, \kappa, B) > 0$ such that for any $x \in [-2+\kappa, 2-\kappa]$ and $u \in [0, 100\varphi]$ we have

$$\mathbb{P}_{\mathrm{GOE}}\left(\mu_j(t) - \gamma_j > (1-\theta)(\gamma_k - \gamma_j)\right) \leqslant C_3 e^{-(1-\varepsilon)(1-\theta)^2\frac{u^2\pi^2}{2\log N}}\mathbb{1}_{u<B\log N} + C_3 e^{\frac{B^2\pi^2}{2}\log N - B(1-\theta)(1-\varepsilon)\pi^2 u}\mathbb{1}_{u\geqslant B\log N}.$$

This implies

$$\begin{aligned}&2p\int_0^{100\varphi} u^{2p-1}\mathbb{P}_{\mathrm{GOE}}\left(\mu_j(t) - \gamma_j > (1-\theta)(\gamma_k - \gamma_j)\right)\mathrm{d}u\\ &\quad\leqslant C_3 2p\int_0^\infty u^{2p-1}e^{-(1-\varepsilon)(1-\theta)^2\frac{u^2\pi^2}{2\log N}}\mathrm{d}u + C_3 2p\int_{B\log N}^\infty u^{2p-1}e^{\frac{B^2\pi^2}{2}\log N - B(1-\theta)(1-\varepsilon)\pi^2 u}\mathbb{1}_{x\geqslant B\log N}\mathrm{d}u.\end{aligned} \tag{4.64}$$

The first term is bounded with $C_4(B,\kappa,\varepsilon)\left(\frac{2}{e\pi^2} + \alpha(\varepsilon,\theta)\right)^p p^p(\log N)^p$ by induction on p, where $\alpha(\varepsilon,\theta) \to 0$ as $\varepsilon, \theta \to 0$.

For the second term, if $p < B(\log N)/10$ it is bounded with

$$C_3 2pe^{-\frac{B^2\pi^2(1-\theta)(1-\varepsilon)\log N}{2}}(B\log N)^{2p} \leqslant 2C_3\Big(\frac{2}{e\pi^2} + \nu(\theta,\varepsilon)\Big)^p p^{p+1}(\log N)^p,$$

where $\nu(\theta,\varepsilon) \to 0$ as $\theta, \varepsilon \to 0$, and we have used $\sup_{x>0} x^p e^{-\frac{x\pi^2}{2}} = p^p(2/(e\pi^2))^p$. We have therefore proved that for any $\alpha > 0$, for $\varepsilon \leqslant \varepsilon_0(\kappa,\alpha)$ and $\theta \leqslant \theta_0(\kappa,\alpha)$, $p < B(\log N)/10$ and $N \geqslant N_1(\alpha,\kappa,B)$ we have

$$2p\int_0^{100\varphi} u^{2p-1}\mathbb{P}_{\mathrm{GOE}}\left(\mu_j(t) - \gamma_j > (1-\theta)(\gamma_k - \gamma_j)\right)\mathrm{d}u \leqslant \big(\frac{2}{e\pi^2} + \alpha\big)^p p^p(\log N)^p.$$

We now consider $2p\int_0^{100\varphi} u^{2p-1}\mathbb{P}(|\operatorname{Im} L_N^{\mathrm{GOE}}(\gamma_j^t)| > c\theta u)\mathrm{d}u$. Note that this could not be directly interpreted as a moment because $j = j(u)$. A direct calculation based on (A.3) gives, for any $p < D(\log N)/10$,

$$2p\int_0^{100\varphi} u^{2p-1}\mathbb{P}(|\operatorname{Im} L_N^{\mathrm{GOE}}(\gamma_j^t)| > c\theta u)\mathrm{d}u \leqslant \frac{C_7^p}{\theta^{2p}}(\log\log N)^{2p}p^{\frac{3p}{2}},$$

where $C_7 = C_7(\kappa, D)$. The same estimate holds for $2p\int_0^{100\varphi} u^{2p-1}\mathbb{P}(|\operatorname{Im} L_N^H(\gamma_j^t)| > c\theta u)\mathrm{d}u$. We choose $\theta \to 0$ satisfying $\theta \geqslant (\log N)^{-1/100}$, so that for any $p < D(\log N)/10$ and $N \geqslant N_2(\kappa,\alpha,D)$ we have

$$\frac{C_7^p}{\theta^{2p}}(\log\log N)^{2p}p^{\frac{3p}{2}} \leqslant \big(\frac{2}{e\pi^2} + \alpha\big)^p p^p(\log N)^p.$$

This concludes the proof. □

5 Moment Matching

This section contains moment matching lemmas that are used in the next section to establish our main results for Wigner matrices. Section 5.1 provides a comparison result for the real part of the log-characteristic polynomial. Section 5.2 establishes results for the deviations of the eigenvalues from their classical locations.

5.1 Real part of log-characteristic polynomial. Given parameters $r>0$ and $x\in[N^{-1},1]$, we define the line segment

$$\mathcal{L}_{r,x}=\{z=E+\mathrm{i}\eta\in\mathbb{H}:|E|<2-r,\eta=x\}. \tag{5.1}$$

Given $M\in \mathrm{Mat}_N$ with eigenvalues $\{\lambda_i\}_{i=1}^N$, we will study the observable

$$\max_{i\in J}\left(\sum_j \log|z_i-\lambda_j|-N\int_{\mathbb{R}}\log|z_i-\lambda|\,\mathrm{d}\rho_{\mathrm{sc}}(\lambda)\right), \tag{5.2}$$

where J is an index set satisfying $|J|\leqslant CN$ for some constant $C>0$ and the points $\{z_i\}_{i\in J}$ satisfy $z_i\in\mathcal{L}_{r,x}$. We set

$$\alpha_i=\sum_j\log|z_i-\lambda_j|-N\int_{\mathbb{R}}\log|z_i-\lambda|\,d\rho_{\mathrm{sc}}(\lambda),\qquad \boldsymbol{\alpha}=(\alpha_i)_{i\in J}. \tag{5.3}$$

Using the fundamental theorem of calculus, we write

$$\log|z_i-\lambda_j|=\log|\lambda_j-E-\mathrm{i}N^{100}|-\mathrm{Im}\int_x^{N^{100}}\frac{d\eta}{\lambda_j-E-\mathrm{i}\eta}. \tag{5.4}$$

Then

$$\alpha_i=\sum_j\log|\lambda_j-E-\mathrm{i}N^{100}|-\mathrm{Im}\int_x^{N^{100}}Nm_N(E+\mathrm{i}\eta)\,d\eta-N\int_{\mathbb{R}}\log|z_i-\lambda|\,d\rho_{\mathrm{sc}}(\lambda). \tag{5.5}$$

We also define a regularized version of α_i:

$$\widetilde{\alpha}_i=\sum_j\log|-E-\mathrm{i}N^{100}|-\mathrm{Im}\int_x^{N^{100}}Nm_N(E+\mathrm{i}\eta)\,d\eta-N\int_{\mathbb{R}}\log|z_i-\lambda|\,d\rho_{\mathrm{sc}}(\lambda).$$

(Of course, the first sum is simply $N\log|-E-\mathrm{i}N^{100}|$, but for comparison with (5.5), we write it in this form.) As before, we write $\widetilde{\boldsymbol{\alpha}}=(\widetilde{\alpha}_i)_{i\in J}$, and suppress the dependence of $\widetilde{\boldsymbol{\alpha}}$ on x and r in the notation. The following lemma shows that $\max_{i\in J}\widetilde{\alpha}_i$ is a good substitute for $\max_{i\in J}\alpha_i$ (that is, (5.2)).

Lemma 5.1. *Let H be a Wigner matrix and fix $r>0$. Then there exists $C(r)>0$ such that for all $x\in[N^{-1},1]$,*

$$\sup_{z\in\mathcal{L}_{r,x}}\mathbb{P}\left(\|\boldsymbol{\alpha}-\widetilde{\boldsymbol{\alpha}}\|_\infty>CN^{-10}\right)\leqslant CN^{-D}. \tag{5.6}$$

Proof. This follows from differentiating $y\mapsto\log|y-E-\mathrm{i}N^{100}|$ in y, then using the eigenvalue rigidity estimate (2.9) and the fundamental theorem of calculus. □

Given a vector $\boldsymbol{w}\in\mathbb{R}^{|J|}$ and parameters $\delta,\nu>0$, we introduce the regularized maximum observable denoted

$$F(\boldsymbol{w})=F_{\delta,\nu}(\boldsymbol{w})=\frac{1}{\delta}\log\left(\sum_{i\in J}\exp\left(\delta\nu w_i\right)\right). \tag{5.7}$$

The unusual notation δ for an inverse temperature aims at avoiding confusion with the β-ensembles. The following lemma is elementary and its proof is omitted.

Lemma 5.2. *For any $\boldsymbol{w}\in\mathbb{R}^{|J|}$, we have*

$$\left|\sup_{i\in J}\nu w_i-F_\delta(\boldsymbol{w})\right|<\frac{2\log N}{\delta}. \tag{5.8}$$

For the rest of this section, we fix $\delta = (\log N)^2$ and $\nu = 1/\log N$, so that

$$F(\widetilde{\boldsymbol{\alpha}}) = \frac{1}{\delta}\left(\sum_{i\in J}\exp(\delta\nu\widetilde{\alpha}_i)\right) \tag{5.9}$$

approximates (5.2) with $O\left((\log N)^{-1}\right)$ error, with high probability, by (5.8) and (5.6).

Definition 5.3. For any $w \in [0,1]$, $M = (m_{ij})_{1\leqslant i,j\leqslant N} \in \mathrm{Mat}_N$, and indices $a, b \in [\![1,N]\!]$, we define $\Theta_w^{(a,b)}M \in \mathrm{Mat}_N$ as follows. If $(i,j) \notin \{(a,b),(b,a)\}$, let the (i,j) entry of $\Theta_w^{(a,b)}M$ be equal to m_{ij}. If $(i,j) \in \{(a,b),(b,a)\}$, then let the (i,j) entry equal $wm_{a,b} = wm_{b,a}$. We also set $\Theta_w^{(a,b)}G(z) = (\Theta_w^{(a,b)}M - z)^{-1}$.

We recall that φ was defined in (2.6).

Lemma 5.4. *Let H be a Wigner matrix and fix $D, r > 0$. There exists $C(D,r) > 0$ such that*

$$\sup_{x\in[N^{-1},1]}\sup_{z\in\mathcal{L}_{r,x}}\mathbb{P}\left(\sup_{w\in[0,1]}\sup_{a,b,i\in[\![1,N]\!]}|\Theta_w^{(a,b)}G_{ii}(z)| > C\varphi^{10}\right) \leqslant CN^{-D}. \tag{5.10}$$

Proof. For the unperturbed matrices, $w = 1$, this is an immediate consequence of (2.8). The statement for a general rank-one perturbations can be deduced from the unperturbed case using a resolvent expansion; see [66, (4.54)] and the following material for details. □

Lemma 5.5. *Let H be a Wigner matrix and fix $D, r > 0$. There exists $C(D,r) > 0$ such that for all $x \in [N^{-1},1]$,*

$$\mathbb{P}\left(\sup_{w\in[0,1]}\sup_{k\in[\![1,5]\!]}\sup_{a,b,c,d\in[\![1,N]\!]}\left|\partial_{ab}^k\widetilde{\alpha}_i(\Theta_w^{(c,d)}H)\right| > C\varphi^{11k}\right) \leqslant CN^{-D}, \tag{5.11}$$

and

$$\sup_{w\in[0,1]}\sup_{k\in[\![1,5]\!]}\sup_{a,b,c,d\in[\![1,N]\!]}\left|\partial_{ab}^k\widetilde{\alpha}_i(\Theta_w^{(c,d)}H)\right| \leqslant CN^C \tag{5.12}$$

almost surely for all N.

Proof. The first and third terms in the definition of $\widetilde{\alpha}_i$ are constants and have derivative zero. For the second, we see using (2.14) that

$$\partial_{ab}Nm_N = \partial_{ab}\sum_i G_{ii} = -\sum_i G_{ia}G_{bi}. \tag{5.13}$$

Therefore

$$N\left|\partial_{ab}m_N\right| \leqslant \sum_i |G_{ia}G_{bi}| \leqslant C\sum_i\left(|G_{ia}|^2 + |G_{bi}|^2\right) \leqslant \frac{C}{\eta}\left(|G_{aa}| + |G_{bb}|\right), \tag{5.14}$$

where we used (2.13) in the last inequality. Similarly, for the higher derivatives we have

$$N\left|\partial_{ab}^k m_N\right| \leqslant \frac{C}{\eta}\left(|G_{aa}| + |G_{bb}| + |G_{ab}|\right)^k \leqslant \frac{C\varphi^{k10}}{\eta} \tag{5.15}$$

by (5.10). Then

$$\left|\partial_{ab}^k\widetilde{\alpha}_i\right| = \left|\partial_{ab}^k\operatorname{Im}\int_{N^{-1}}^{N^{100}} Nm_N(E+\mathrm{i}\eta)\,d\eta\right| \leqslant C\varphi^{k10}\int_{N^{-1}}^{N^{100}}\frac{1}{\eta}\,d\eta \leqslant C\varphi^{1+k10}, \tag{5.16}$$

where we increased the value of C. The remaining claim is similar and uses the trivial bound $|G_{ij}| \leqslant \eta^{-1}$ from (2.15). This completes the proof. □

Lemma 5.6. *Let H be a Wigner matrix and fix $D, r > 0$. Then there exist $C(D,r) > 0$ such that for all $x \in [N^{-1}, 1]$*

$$\mathbb{P}\left(\sup_{k\in[\![1,5]\!]}\sup_{a,b,c,d\in[\![1,N]\!]}\sup_{w\in[0,1]}\left|\partial_{ab}^k F\left(\widetilde{\boldsymbol{\alpha}}(\Theta_w^{(c,d)}H)\right)\right| > C\varphi^{12k}\right) \leqslant CN^{-D}. \tag{5.17}$$

Also, we have almost surely that for all $x \in [N^{-1}, 1]$

$$\sup_{k\in[\![1,5]\!]}\sup_{a,b,c,d\in[\![1,N]\!]}\sup_{w\in[0,1]}\left|\partial_{ab}^k F\left(\widetilde{\boldsymbol{\alpha}}(\Theta_w^{(c,d)}H)\right)\right| \leqslant CN^{C}. \tag{5.18}$$

Proof. First, we claim that the partial derivatives of $F_\delta(w)$ with respect to the entries of the vector $w \in \mathbb{R}^N$ satisfy

$$\sum_{\underline{j}}\left|\frac{\partial^d F_\delta(w)}{\partial_{j_1}\dots\partial_{j_d}}\right| \leqslant C\delta^{d-1}, \tag{5.19}$$

for any $d \in \mathbb{N}$. Here the sum runs over all multi-indices $\underline{j} = (j_1, \cdots, j_d)$ with values in $[1,N]^d$, $\partial_j = \partial_{w_j}$, and $C = C(d) > 0$ is a constant. This inequality follows by straightforward differentiation, and complete details are given in [66, Lemma 3.4].

Using the chain rule, (5.19) and (5.11) imply (5.17). Similarly, (5.19) and (5.12) imply (5.18). □

Theorem 5.7. *Fix $r > 0$. Let H and M be Wigner matrices such that $\mathbb{E}[H_{11}^k] = \mathbb{E}[m_{11}^k]$ for $1 \leqslant k \leqslant 3$ and $\left|\mathbb{E}[H_{11}^4] - \mathbb{E}[m_{11}^4]\right| \leqslant K_1 N^{-2}\varphi^{-K_2}$ for some $K_1, K_2 \geqslant 0$. Let $S\colon \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying $\|S^{(k)}\|_\infty \leqslant K_1$ for $k \in [\![0,5]\!]$. Then there exists $C(K_1) > 0$ such that if $K_2 > C$, then for all $x \in [N^{-1}, 1]$ we have*

$$\left|\mathbb{E}_H S\left(F_\delta(\widetilde{\boldsymbol{\alpha}})\right) - \mathbb{E}_M S\left(F_\delta(\widetilde{\boldsymbol{\alpha}})\right)\right| \leqslant C\varphi^{-K_2/2}. \tag{5.20}$$

Proof. We fix $z \in \mathcal{L}_{r,x}$ and omit it from the notation. Fix any bijection

$$\psi\colon \{(i,j) : 1 \leqslant i \leqslant j \leqslant N\} \to [\![1, \gamma_N]\!], \tag{5.21}$$

where $\gamma_N = N(N+1)/2$, and define the matrices $H^1, H^2, \dots, H^{\gamma_N}$ by

$$H_{ij}^\gamma = \begin{cases} H_{ij} & \text{if } \psi(i,j) \leqslant \gamma \\ m_{ij} & \text{if } \psi(i,j) > \gamma \end{cases} \tag{5.22}$$

for $i \leqslant j$.

Fix some $\gamma \in [\![1, \gamma_N]\!]$ and consider the indices (i,j) such that $\psi(i,j) = \gamma$. Define $T\colon \mathrm{Mat}_N \to \mathbb{R}$ by $T(X) = S\left(F_\delta(\widetilde{\boldsymbol{\alpha}}(X))\right)$. We Taylor expand $T\left(H^\gamma\right)$ in the (i,j) entry and write $\partial = \partial_{ij}$ to find

$$\begin{aligned} T\left(H^\gamma\right) - T\left(\Theta_0^{(i,j)}H^\gamma\right) &= \partial T\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij} + \frac{1}{2!}\partial^2 T\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^2 + \frac{1}{3!}\partial^3 T\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^3 \\ &\quad + \frac{1}{4!}\partial^4 T\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^4 + \frac{1}{5!}\partial^5 T\left(\Theta_{w_1(\gamma)}^{(i,j)}H^\gamma\right)H_{ij}^5, \end{aligned}$$

where $w_1(\gamma) \in [0,1]$ is a random variable depending on H_{ij}. Similarly, we expand $T\left(H^{\gamma-1}\right)$ in the (i,j) entry to obtain

$$T\left(H^{\gamma-1}\right) - T\left(\Theta_0^{(i,j)}H^\gamma\right) = \partial T\left(\Theta_0^{(i,j)}H^\gamma\right)m_{ij} + \frac{1}{2!}\partial^2 T\left(\Theta_0^{(i,j)}H^\gamma\right)m_{ij}^2 + \frac{1}{3!}\partial^3 T\left(\Theta_0^{(i,j)}H^\gamma\right)m_{ij}^3 \tag{5.23}$$

$$+ \frac{1}{4!}\partial^4 T\left(\Theta_0^{(i,j)}H^\gamma\right)m_{ij}^4 + \frac{1}{5!}\partial^5 T\left(\Theta_{w_2(\gamma)}^{(i,j)}H^\gamma\right)m_{ij}^5, \tag{5.24}$$

where $w_2(\gamma) \in [0,1]$ is a random variable depending on m_{ij}. Subtracting the previous two equations and taking expectation, we obtain

$$\mathbb{E}\left[T\left(H^\gamma\right)\right] - \mathbb{E}\left[T\left(H^{\gamma-1}\right)\right] = \frac{1}{4!}\mathbb{E}\left[\partial^4 T\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^4\right] - \frac{1}{4!}\mathbb{E}\left[\partial^4 T\left(\Theta_0^{(i,j)}H^\gamma\right)m_{ij}^4\right] \tag{5.25}$$

$$+ \frac{1}{5!}\mathbb{E}\left[\partial^5 T\left(\Theta_{w_1(\gamma)}^{(i,j)}H^\gamma\right)H_{ij}^5\right] - \frac{1}{5!}\mathbb{E}\left[\partial^5 T\left(\Theta_{w_2(\gamma)}^{(i,j)}H^\gamma\right)m_{ij}^5\right], \tag{5.26}$$

where we used that $\Theta_0^{(i,j)} H^\gamma$ is independent from H_{ij} and m_{ij}, and that $\mathbb{E}[h_{ij}^k] = \mathbb{E}[m_{ij}^k]$ for $k \in [\![1,3]\!]$.

Because H_{ij} and m_{ij} are independent from $\Theta_0^{(i,j)} H^\gamma$, we have

$$\mathbb{E}\left[\partial^4 T\left(\Theta_0^{(i,j)} H^\gamma\right) H_{ij}^4\right] - \mathbb{E}\left[\partial^4 T\left(\Theta_0^{(i,j)} H^\gamma\right) m_{ij}^4\right] = \mathbb{E}\left[\partial^4 T\left(\Theta_0^{(i,j)} H^\gamma\right)\right] \mathbb{E}\left[H_{ij}^4 - m_{ij}^4\right]. \tag{5.27}$$

By (5.17), (5.18), and the assumptions on S, there exists $C(K_1) > 0$ such that

$$\left|\mathbb{E}\left[\partial^4 T\left(\Theta_0^{(i,j)} H^\gamma\right)\right]\right| \leqslant C\varphi^{50}. \tag{5.28}$$

We conclude using (5.25) and the assumption on the fourth moments of H_{ij} and m_{ij} that

$$\left|\mathbb{E}\left[\partial^4 T\left(\Theta_0^{(i,j)} H^\gamma\right) H_{ij}^4\right] - \mathbb{E}\left[\partial^4 T\left(\Theta_0^{(i,j)} H^\gamma\right) m_{ij}^4\right]\right| \leqslant CK_1 N^{-2} \varphi^{50-K_2}. \tag{5.29}$$

The fifth order terms may be bounded similarly, and in fact are lower order since the fifth powers h_{ij}^5 and m_{ij}^5 contribute an additional factor of $N^{-1/2}$. Summing the Taylor expansions over all $O(N^2)$ indices (i,j), we conclude that

$$\big|\mathbb{E}\big[T(H)\big] - \mathbb{E}\big[T(M)\big]\big| \leqslant CK_1 \varphi^{50-K_2}. \tag{5.30}$$

This completes the proof. □

5.2 *Maximal deviation from classical location.* Using the rigidity and local law from Theorem 2.2, the proof of the following lemma is nearly identical to [66, Lemma 3.2].

Lemma 5.8. *Fix $\kappa > 0$. For all $i \in [\![\kappa N, (1-\kappa)N]\!]$, there exist smooth functions $\widetilde{\lambda}_i \colon \mathrm{Mat}_N \to \mathbb{R}$ such that the following holds. Suppose that H is a real symmetric Wigner matrix. There exist constants $C_1, C_2 > 0$ such that, uniformly in i and $k \in [\![1,5]\!]$,*

$$\big|\widetilde{\lambda}_i(H) - \lambda_i(H)\big| \leqslant \frac{1}{N\varphi^{C_2}}, \qquad \sup_{w\in[0,1]} \sup_{a,b,c,d\in[\![1,N]\!]} \big|\partial_{ab}^k \widetilde{\lambda}_i(\Theta_w^{(c,d)} H)\big| \leqslant \frac{\varphi^{C_1}}{N} \tag{5.31}$$

with probability at least $1 - c^{-1}\exp(-c\varphi)$. Here $\partial_{ab} = \partial_{X_{ab}}$ denotes the partial derivative with respect to the (a,b)-th matrix element.

Further, uniformly in i and $k \in [\![1,5]\!]$, we have the deterministic bound

$$\sup_{w\in[0,1]} \sup_{a,b,c,d\in[\![1,N]\!]} \big|\partial_{ab}^k \widetilde{\lambda}_i(\Theta_w^{(c,d)} H)\big| \leqslant C_1 N^{C_1}. \tag{5.32}$$

We write $\boldsymbol{\lambda} = (\lambda_i)_{i\in[\![1,N]\!]}$ and $\widetilde{\boldsymbol{\lambda}} = (\lambda_i)_{i\in[\![\kappa N,(1-\kappa)N]\!]}$, using the notation of the previous lemma. Set $J \subset [\![\kappa N, (1-\kappa)N]\!]$ and define the smoothed maximal deviation of a vector $\boldsymbol{v} \in \mathbb{R}^{|J|}$ from the classical eigenvalue locations γ_i by

$$\widehat{F}_\delta(\boldsymbol{v}) = \frac{1}{\delta} \log\left(\sum_{i\in J} \exp\left(\delta\nu_i(v_i - \gamma_i)\right) + \exp\left(\delta\nu_i(\gamma_i - v_i)\right)\right), \tag{5.33}$$

where we set

$$\nu_k = \sqrt{\frac{\pi}{2}} \cdot \frac{k\rho_{\mathrm{sc}}(\gamma_k)}{\log N}, \qquad \delta = (\log N)^2. \tag{5.34}$$

We omit the proof of the following derivative bounds, since it is similar to the proof of Lemma 5.6.

Lemma 5.9. *Let H be a Wigner matrix and fix $D, r > 0$. Then there exist $C(D,r) > 0$ such that*

$$\mathbb{P}\left(\sup_{k\in[\![1,5]\!]} \sup_{a,b,c,d\in[\![1,N]\!]} \sup_{w\in[0,1]} \left|\partial_{ab}^k \widehat{F}\left(\widetilde{\boldsymbol{\lambda}}(\Theta_w^{(c,d)} H)\right)\right| > C\varphi^{Cj}\right) \leqslant CN^{-D}. \tag{5.35}$$

Also, we have almost surely that

$$\sup_{k\in[\![1,5]\!]} \sup_{a,b,c,d\in[\![1,N]\!]} \sup_{w\in[0,1]} \left|\partial_{ab}^k \widehat{F}\left(\widetilde{\boldsymbol{\lambda}}(\Theta_w^{(c,d)} H)\right)\right| \leqslant CN^{Cj}. \tag{5.36}$$

Using the observable $F_\delta(\widetilde{\boldsymbol{\lambda}})$ and Lemma 5.9, the proof of the following comparison result is similar to Lemma 5.7.

Theorem 5.10. *Fix $\kappa > 0$. Let H and M be Wigner matrices such that $\mathbb{E}[H_{11}^k] = \mathbb{E}[m_{11}^k]$ for $1 \leqslant k \leqslant 3$ and $\left|\mathbb{E}[H_{11}^4] - \mathbb{E}[m_{11}^4]\right| \leqslant K_1 N^{-2}\varphi^{-K_2}$ for some $K_1, K_2 \geqslant 0$. Let $S\colon \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying $\|S^{(k)}\|_\infty \leqslant K_1$ for $k \in [\![0,5]\!]$. Then there exists $C(K_1)$ such that, if $K_2 > C$, then*

$$\left|\mathbb{E}_H S\left(\widehat{F}_\delta(\widetilde{\boldsymbol{\lambda}})\right) - \mathbb{E}_M S\left(\widehat{F}_\delta(\widetilde{\boldsymbol{\lambda}})\right)\right| \leqslant C\varphi^{-K_2/2}, \tag{5.37}$$

and for any $i \in [\![\kappa N, (1-\kappa)N]\!]$.

5.3 Large moments. For some parameter $\eta_1 > 0$ we define the function $f = f_E$ by

$$f = 0 \text{ on } (-\infty, E] \cup [3,\infty), \qquad f = 1 \text{ on } [E+\eta_1, 2.5] \tag{5.38}$$

$$\|f^{(k)}\|_{L^\infty(-\infty,2)} \leqslant 100 \cdot \eta_1^{-k}, \quad \|f^{(k)}\|_{L^\infty(2,\infty)} \leqslant 100, \tag{5.39}$$

for $k = 1, 2$. All results in this section hold for $\eta_1 \in [1/N, c]$, but we now fix

$$\eta_1 = \frac{\sqrt{\log N}}{N},$$

which will be enough to prove part (ii) of Theorem 1.8, in Subsection 6.4.

Moreover, given $M \in \mathrm{Mat}_N$ we define

$$X(M) = \sum_{i=1}^N f_E(\widetilde{\lambda}_i) - N\int_{\mathbb{R}} f_E(x)\,\mathrm{d}\rho_{\mathrm{sc}}(x), \quad X_p(M) = X(M)^{2p}. \tag{5.40}$$

Lemma 5.11. *Let H be a Wigner matrix, and fix $A, \kappa > 0$. Then there exists $C(A,\kappa) > 0$ such that*

$$\mathbb{P}\left(\sup_{k\in[\![1,5]\!]}\ \sup_{a,b,c,d\in[\![1,N]\!]}\ \sup_{w\in[0,1]} \left|\partial_{ab}^k X\left(\Theta_w^{(c,d)}H\right)\right| > C\varphi^{Ck}\right) \leqslant C\exp(-C^{-1}\varphi). \tag{5.41}$$

Also, we have

$$\sup_{k\in[\![1,5]\!]}\ \sup_{a,b,c,d\in[\![1,N]\!]}\ \sup_{w\in[0,1]} \left|\partial_{ab}^k X\left(\Theta_w^{(c,d)}H\right)\right| \leqslant CN^C. \tag{5.42}$$

Proof. Since the proof is similar to previous estimates, we give details only in the $j = 2$ case to illustrate the general principles involved. We have

$$\partial_{ab}^2 f(\widetilde{\lambda}) = f''(\widetilde{\lambda})(\partial_{ab}\widetilde{\lambda})^2 + f'(\widetilde{\lambda})\partial_{ab}^2\widetilde{\lambda}. \tag{5.43}$$

The f'' term is the most dangerous. As in the proof of the previous lemma, there are at most $N\eta_1(\log N)^C$ eigenvalues in the interval I_A where f' is nonzero. By Lemma 5.8 we have

$$f_0''(\widetilde{\lambda})(\partial_{ab}\widetilde{\lambda})^2 \leqslant \eta_1^{-2}N^{-2}\varphi^{2C} \leqslant C\varphi^{4C}. \tag{5.44}$$

where we used $\log N \leqslant \varphi$ and increased the constant C if necessary. The claim then follows after redefining C. This shows (5.41) for $j = 2$; the other j are similar. The bound (5.42) follows from the second inequality in (5.39). ☐

Lemma 5.12. *Let H be a Wigner matrix, and fix $A, \kappa > 0$. There exists $c(A,\kappa) > 0$ such that for all $E \in [-2+\kappa, 2-\kappa]$,*

$$\sup_{c,d\in[\![1,N]\!]}\ \sup_{w\in[0,1]} \mathbb{P}\left(\left|X\left(\Theta_w^{(c,d)}H\right)\right| \geqslant c^{-1}\varphi\right) \leqslant c^{-1}\exp(-c(\log N)^{c\log\log N}) \tag{5.45}$$

Proof. We give the details only for $w = 0$, since the case of general w follows by a straightforward perturbation argument. By Equation 5.46, it suffices to bound $\sum_{i=1}^N f(\lambda_i) - N\int_{\mathbb{R}} f\,\mathrm{d}\rho_{\mathrm{sc}}$, and this is immediate from (2.9). ☐

Lemma 5.13. *Let H be a Wigner matrix, and fix $A,\kappa > 0$. There exists $c(\kappa) > 0$ such that, for all $E \in [-2+\kappa, 2-\kappa]$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} f_E(\widetilde{\lambda}_i) - \sum_{i=1}^{N} f_E(\lambda_i)\right| > (\log N)^{1/2}\right) \leqslant c^{-1}\exp(-c\varphi^c). \tag{5.46}$$

Proof. Note that f' is supported on the interval $I_A = [E, E+\eta_1]$. For λ_i outside the support of f', it is straightforward to replace λ_i with $\widetilde{\lambda}_i$. There are at most N such eigenvalues, and $f_0(\lambda_i) - f_0(\widetilde{\lambda}_i) = O(N^{-1}\varphi^{-c})$ by Lemma 5.8, so the overall error from all such λ_i is $O(\varphi^{-c})$.

For the other eigenvalues, we know that f' can be as large as $\eta_1^{-1} = N(A\log N)^{-1}$. In the interval I_A there are at most $N\eta_1(\log N)^C$ eigenvalues with probability at least $1 - c^{-1}\exp(-c\varphi^c)$ by the rigidity estimate (2.9). We have $|\widetilde{\lambda} - \lambda| \leqslant N^{-1}\varphi^{-c}$. Then the accumulated error is

$$N^{-1}\varphi^{-c}\eta_1^{-1}N\eta_1(\log N)^C = \varphi^{-c}(\log N)^C = o(1). \tag{5.47}$$

which is acceptable. □

The next lemma considers a moment-matching argument for diverging moments. Such estimates for large moments appeared first in random matrix theory in [58, 79]. For the application to the accurate Gaussian decay exponent in Theorem 1.8 (ii), optimal upper bounds sharper than in [58, 79] are required, which correspond to the best possible θ below. A similar result in this direction was obtained in the context of eigenvectors in [15].

Lemma 5.14. *There exists $M_0 > 0$ such that the following holds. Let H and R be two Wigner matrices satisfying $\mathbb{E}[H_{ij}^k] = \mathbb{E}[R_{ij}^k]$ for $k \in [\![1,3]\!]$ and $\left|\mathbb{E}[H_{ij}^4] - \mathbb{E}[R_{ij}^4]\right| \leqslant N^{-2}s$ where $s < \varphi^{-M_0}$. Assume that there is $N_0(A,\kappa)$ and $\theta(A,\kappa)$ such that*

$$\mathbb{E}\big[X_p(R)\big] \leqslant \theta^p(\log N)^p p^p$$

for all $E \in [-2+\kappa, 2-\kappa]$, $p \leqslant A\log N$ and $N \geqslant N_0$. Then there is $N_1(A,\kappa)$ such that for all $N \geqslant N_1$ we have

$$\mathbb{E}\big[X_p(H)\big] \leqslant (1+\varphi^{-5})^p\theta^p(\log N)^p p^p \leqslant 3\theta^p(\log N)^p p^p. \tag{5.48}$$

Proof. Fix any bijection

$$\varphi\colon \{(i,j) : 1 \leqslant i \leqslant j \leqslant N\} \to [\![1,\gamma_N]\!], \tag{5.49}$$

where $\gamma_N = N(N+1)/2$, and define the matrices $H^1, H^2, \dots, H^{\gamma_N}$ by

$$H_{ij}^{\gamma} = \begin{cases} H_{ij} & \text{if } \varphi(i,j) \leqslant \gamma \\ R_{ij} & \text{if } \varphi(i,j) > \gamma \end{cases} \tag{5.50}$$

for $i \leqslant j$. We also fix z throughout the argument.

Fix some $\gamma \in [\![1,\gamma_N]\!]$ and consider the indices (i,j) such that $\varphi(i,j) = \gamma$. For any $m \geqslant 1$, we may Taylor expand $X_m(H^\gamma)$ in the (i,j) entry, write $\partial = \partial_{ij}$, and find

$$X_m(H^\gamma) - X_m\left(\Theta_0^{(i,j)}H^\gamma\right) = \partial X_m\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij} + \frac{1}{2!}\partial^2 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^2 + \frac{1}{3!}\partial^3 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^3 \tag{5.51}$$

$$+ \frac{1}{4!}\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^4 + \frac{1}{5!}\partial^5 X_m\left(\Theta_{w_1(\gamma)}^{(i,j)}H^\gamma\right)H_{ij}^5, \tag{5.52}$$

where $w_1(\gamma) \in [0,1]$ is a random variable depending on H_{ij}. Similarly, we may expand $X_m\left(H^{\gamma-1}\right)$ in the (i,j) entry to obtain a similar expansion with $w_2(\gamma) \in [0,1]$, a random variable depending on R_{ij}. Subtracting the expansion of $X_m(H^\gamma)$ from (5.51) and (5.52), and taking expectation, we find

$$\mathbb{E}\left[X_m(H^\gamma)\right] - \mathbb{E}\left[X_m\left(H^{\gamma-1}\right)\right] = \frac{1}{4!}\mathbb{E}\left[\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^4\right] - \frac{1}{4!}\mathbb{E}\left[\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)R_{ij}^4\right] \tag{5.53}$$

$$+ \frac{1}{5!}\mathbb{E}\left[\partial^5 X_m\left(\Theta_{w_1(\gamma)}^{(i,j)}H^\gamma\right)H_{ij}^5\right] - \frac{1}{5!}\mathbb{E}\left[\partial^5 X_m\left(\Theta_{w_2(\gamma)}^{(i,j)}H^\gamma\right)R_{ij}^5\right], \tag{5.54}$$

where we used that $\Theta_0^{(i,j)}H^\gamma$ is independent from H_{ij} and R_{ij}, and that $\mathbb{E}[h_{ij}^k] = \mathbb{E}[r_{ij}^k]$ for $k \in [\![1,3]\!]$.

We now proceed by induction, with the induction hypothesis at step $m \in \mathbb{N}$ being that

$$\mathbb{E}X_n\left(\Theta_w^{(a,b)}H^\gamma\right) \leqslant (1+\varphi^{-5})^n\theta^n(\log N)^n n^n \tag{5.55}$$

holds for all $0 \leqslant n \leqslant m$ and choices of $w \in [0,1]$ and $(a,b) \in [\![1,N]\!]^2$.

The base case $m = 0$ is trivial. Assuming the induction hypothesis holds for $m-1$, we will show it holds for m. Using the independence of H_{ij} and R_{ij} from $\Theta_0^{(i,j)}H^\gamma$, we may rewrite the first two terms terms on the right side of (5.53) as

$$\mathbb{E}\left[\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^4\right] - \mathbb{E}\left[\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)R_{ij}^4\right] = \mathbb{E}\left[\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)\right]\mathbb{E}\left[H_{ij}^4 - R_{ij}^4\right]. \tag{5.56}$$

For the second factor, we recall that $\left|\mathbb{E}\left[H_{ij}^4\right] - \mathbb{E}\left[R_{ij}^4\right]\right| \leqslant N^{-2}s = N^{-2}\varphi^{-M}$. For the first, we abbreviate $X_m = X_m\left(\Theta_0^{(i,j)}H^\gamma\right)$, write $X_m^{(\ell)}$ for the ℓth derivative of X_m with respect to the (i,j) coordinate, and compute

$$\begin{aligned}\partial^4 X_m = \partial^4\left(X^{2m}\right) &= 2mX^{2m-1}X^{(4)} + 3(2m)(2m-1)X^{2m-2}(X^{(2)})^2 \\ &\quad + 2m(2m-1)(2m-2)(2m-3)X^{2m-4}(X^{(1)})^4 \\ &\quad + 4(2m)(2m-1)X^{2m-2}X^{(1)}X^{(3)} + 6(2m)(2m-1)(2m-2)X^{2m-3}(X^{(1)})^2X^{(2)}.\end{aligned} \tag{5.57}$$

The terms with even powers of X may be bounded using the induction hypothesis (5.55) for $n \leqslant m-1$ and Lemma 5.11. The bound the odd powers, we additionally use (5.45) to show

$$\mathbb{E}|X^{2p-1}| \leqslant \varphi\mathbb{E}X^{2p-2} + (N)^{2p-1}c^{-1}\exp(-c(\log N)^{c\log\log N}), \tag{5.58}$$

where we observe that the second term is o(1) for $p \leqslant A\log N$. Let $\mathcal{B}$ be the set where (5.41) holds. We find[4]

$$\mathbb{E}\left|\mathbb{1}_{\mathcal{B}}\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)\right| \leqslant C\varphi^C\theta^m(\log N)^m m^m. \tag{5.59}$$

Here $C > 0$ is a constant that is independent of m. We also have by Lemma 5.11 that

$$\left|\mathbb{E}\left[\mathbb{1}_{\mathcal{B}^c}\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)\right]\right| \leqslant \widetilde{C}N^{-100}. \tag{5.60}$$

for some $\widetilde{C}$ which does not depend on $m \leqslant A\log N$, i,j,N.

It follows from (5.59), (5.60), and $m \leqslant \varphi$ that, if M_0 is chosen large enough relative to C, then

$$\left|\frac{1}{4!}\mathbb{E}\left[\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)H_{ij}^4\right] - \frac{1}{4!}\mathbb{E}\left[\partial^4 X_m\left(\Theta_0^{(i,j)}H^\gamma\right)R_{ij}^4\right]\right| \leqslant \varphi^{-10}N^{-2}\theta^m(\log N)^m m^m \tag{5.61}$$

holds uniformly in $N \geqslant N_0$ and $m \leqslant A\log N$, where N_0 does not depend on m.

Let $\mathcal{D}$ be the event where $\sup_{i,j}|R_{ij}| + |H_{ij}| \leqslant N^{-1/2+\delta_1}$ holds. Since the variables R_{ij} and H_{ij} are subexponential, we have

$$\mathbb{P}\left(\mathcal{D}^c\right) \leqslant D_1\exp\left(-d_1(\log N)^{d_1\log\log N}\right), \tag{5.62}$$

for some constants $D_1(\delta_1), d_1(\delta_1) > 0$.

For the terms in (5.54), we compute

$$\left|\mathbb{E}\left[\partial^5 X_m\left(\Theta_{w_1(\gamma)}^{(i,j)}H^\gamma\right)H_{ij}^5\right]\right| \leqslant \left|\mathbb{E}\left[\mathbb{1}_{\mathcal{D}}\partial^5 X_m\left(\Theta_{w_1(\gamma)}^{(i,j)}H^\gamma\right)H_{ij}^5\right]\right| + \left|\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^c}\partial^5 X_m\left(\Theta_{w_1(\gamma)}^{(i,j)}H^\gamma\right)H_{ij}^5\right]\right| \tag{5.63}$$

$$\leqslant CN^{-5/2+5\delta_1}\left(\mathbb{E}\left[\left|\partial^5 X_m\left(\Theta_{w_1(\gamma)}^{(i,j)}H^\gamma\right)\right|\right] + 1\right), \tag{5.64}$$

[4] We note that the constants in the probability bound given by Lemma 5.11 do not depend on γ, since the matrices H^γ verify Definition 1.1 simultaneously for the appropriate choice of constants. Therefore, the C in (5.60) is uniform in γ.

where in the last line we used (5.62) and the constant C comes from Lemma 5.11. Then repeating the previous argument for the fourth order term given in (5.59) and (5.60), we find that there exists $N_1(A)$ such that, for $\delta_1 < 1/100$, $m \leqslant A \log N$ and $N \geqslant N_1$ we have

$$\left|\frac{1}{5!}\mathbb{E}\left[\partial^5 X_m\left(\Theta^{(i,j)}_{w_1(\gamma)} H^\gamma\right) H_{ij}^5\right]\right| \leqslant Cm^5 N^{-5/2+5\delta_1}\varphi^C\theta^m(\log N)^m m^m$$
$$\leqslant CN^{-2-1/4}\theta^m(\log N)^m m^m \leqslant N^{-2-1/8}\theta^m(\log N)^m m^m. \quad (5.65)$$

Combining (5.61) and (5.65) yields

$$\left|\mathbb{E}\left[X_m\left(H^\gamma\right)\right] - \mathbb{E}\left[X_m\left(H^{\gamma-1}\right)\right]\right| \leqslant \varphi^{-5}N^{-2}\theta^m(\log N)^m m^m, \quad (5.66)$$

and summing (5.66) over all γ_N pairs (i,j), we find

$$\left|\mathbb{E}\left[X_m\left(R\right)\right] - \mathbb{E}\left[X_m\left(H^\gamma\right)\right]\right| \leqslant \varphi^{-5}\theta^m(\log N)^m m^m \quad (5.67)$$

for any γ. Together with our hypothesis on $\mathbb{E}\big[X_p(R)\big]$ this gives

$$\mathbb{E}\left[X_m\left(H^\gamma\right)\right] \leqslant (1+\varphi^{-5})^m\theta^m(\log N)^m m^m. \quad (5.68)$$

This verifies the induction hypothesis (5.55) when $w = 1$.

To address other values of w, we consider the following expansion:

$$X_m\left(H^\gamma\right) - X_m\left(\Theta^{(a,b)}_w H^\gamma\right) = \sum_{\ell=1}^{4}\frac{1}{\ell!}\partial^\ell X_m\left(\Theta^{(a,b)}_0 H^\gamma\right) H_{ij}^\ell(1-w^\ell) \quad (5.69)$$
$$+ \frac{1}{5!}\partial^5 X_m\left(\Theta^{(a,b)}_{\tau_1} H^\gamma\right) H_{ij}^5 - \frac{1}{5!}\partial^5 X_m\left(\Theta^{(a,b)}_{\tau_w} H^\gamma\right) H_{ij}^5 w^5. \quad (5.70)$$

Here $\tau_1, \tau_w \in [0,1]$ are random variables. The same argument that gave the bound (5.61) shows that the expectation of the right side of (5.69) is bounded with $\frac{1}{2}\theta^m(\log N)^m m^m$. Note this bound holds because of the additional factors of $N^{-1/2}$ coming from moments of H_{ij}, which are enough even for $\ell = 1,2,3$ as we don't sum over N^2 terms. The expectation of (5.70) is also bounded by $(1+\varphi^{-5})^m\theta^m(\log N)^m m^m$ by the reasoning leading to (5.65).

This proves

$$\sup_{w\in[0,1]}\ \sup_{a,b\in[\![1,n]\!]} \mathbb{E}\left[X_m\left(\Theta^{(a,b)}_w H^\gamma\right)\right] \leqslant \left(1+\varphi^{-5}\right)^m\theta^m(\log N)^m m^m. \quad (5.71)$$

and completes the induction. The second inequality in (5.48) follows because $p \leqslant A\log N$. □

6 Maximum for Wigner Matrices

This section proves Theorem 1.2 and Theorem 1.8 by combining the dynamics from Section 4 and the moment matching results from Section 5. It also relies heavily on Section 3, both its results (as the GOE serves as the base point of our comparison), and for methods used there to smooth the corresponding fields.

As we proceed by comparison, we will need to specify the matrix ensembles related to the characteristic polynomials: We will write L_N^H for the quantity (1.5), when considering Wigner matrices as in Definition 1.1, and we will write $L_N^{\rm GOE}$ for the same quantity when the eigenvalues of H are replaced by those of ${\rm GOE}_N$.

We will first prove Theorem 1.2 for the real part, and then part (i) of Theorem 1.8 on the deviations of $\lambda_i - \gamma_i$, which is equivalent to Theorem 1.2 for the imaginary part (see Section 3.4). Indeed, while the proof for $\operatorname{Re} L_N$ will go through a regularization similarly to the proof of Theorem 1.9 in Section 3, we cannot directly follow the same path for $\operatorname{Im} L_N^H$: for the upper bound, a priori smoothing $\operatorname{Im} L_N(E)$ into $\operatorname{Im} L_N(E+\frac{\mathrm{i}}{N})$ as in (3.5) is not possible because a local law allowing (3.4) is not known in the case of Wigner matrices.

In all the following proofs, we will need an intermediate weakly Gaussian-divisible random matrix ensemble as in the following result, which is an immediate consequence of [44, Lemma 16.2].

Lemma 6.1. *Let H be a Wigner matrix. Then there exist constants $C, c > 0$ such that the following holds for any $t \in (0, c)$. There exists a Wigner matrix $\widetilde{H}$ such that*

$$\widetilde{H}(t) = \sqrt{1-t}\widetilde{H} + \sqrt{t}W \tag{6.1}$$

is a Wigner matrix satisfying

$$\mathbb{E}\left[\widetilde{H}_{ij}(t)^k\right] = \mathbb{E}\left[H_{ij}^k\right], \qquad \left|\mathbb{E}\left[\widetilde{H}_{ij}(t)^4\right] - \mathbb{E}\left[H_{ij}^4\right]\right| \leqslant CtN^{-2} \tag{6.2}$$

for $k \in [\![1,3]\!]$. Here $W \in \mathrm{Mat}_N$ is a Wigner matrix and each W_{ij} is a mean zero Gaussian random variable (independent from $\widetilde{H}$).

6.1 Upper bound for the real part in Theorem 1.2. We start with the deterministic bound (3.1) in the particular case $\nu = \rho_{\mathrm{sc}}(x)\mathrm{d}x$, so that

$$\sup_{E \in [A+\kappa, B-\kappa]} \operatorname{Re} L_N^H(E) \leqslant \sup_{E \in J} \operatorname{Re} L_N^H\left(E + \frac{\mathrm{i}}{N}\right) + C_1 \tag{6.3}$$

where $J \subset [A+\kappa, B-\kappa]$ has cardinality at most $C_2 N$, and C_1, C_2 are absolute constants.

Set $t = \varphi^{-K}$, where $K > 0$ is a parameter. Let $\widetilde{H}(t)$ be the matrix (6.1) given by Lemma 6.1. For any $\varepsilon > 0$, let f_ε be a smooth function such that $0 \leqslant f_\varepsilon(x) \leqslant 1$ for $x \in \mathbb{R}$, $f_\varepsilon(x) = 0$ for $x \leqslant \sqrt{2} + \varepsilon/2$, and $f_\varepsilon(x) = 1$ for $x \geqslant \sqrt{2} + \varepsilon$. By Theorem 5.7, Lemma 5.1 and Lemma 5.2, if K is chosen large enough, we have

$$\mathbb{E}\left[f_\varepsilon\left((\log N)^{-1} \sup_{z \in J + \frac{\mathrm{i}}{N}} \operatorname{Re} L_N^{\widetilde{H}(t)}(z)\right)\right] - \mathbb{E}\left[f_\varepsilon\left((\log N)^{-1} \sup_{z \in J + \frac{\mathrm{i}}{N}} \operatorname{Re} L_N^{H}(z)\right)\right] = \mathrm{O}\left(\frac{1}{\log N}\right). \tag{6.4}$$

By Proposition 4.3, there exists a coupling of $\widetilde{H}(t)$ and GOE_N such that

$$\mathbb{P}\left(\sup_{z \in J + \frac{\mathrm{i}}{N}} \left|L_N^{\widetilde{H}(t)}(z) - L_N^{\mathrm{GOE}}(z)\right| > (\log N)^{\frac{1}{2}+\varepsilon}\right) = \mathrm{o}(1). \tag{6.5}$$

Finally, from (3.10) and (3.11), and recalling that the GOE is a β-ensemble with $\beta = 1$ and a quadratic potential (see, e.g., [44, (4.4)]), we have

$$\mathbb{P}\left((\log N)^{-1} \sup_{z \in J + \frac{\mathrm{i}}{N}} \operatorname{Re} L_N^{\mathrm{GOE}}(z) > \sqrt{2} + \frac{\varepsilon}{2}\right) = \mathrm{o}(1). \tag{6.6}$$

To conclude, we observe that from (6.3), (6.4), (6.5), (6.6) we have

$$\mathbb{P}\left((\log N)^{-1} \sup_{E \in [A+\kappa, B-\kappa]} \operatorname{Re} L_N^H(E) > \sqrt{2} + 2\varepsilon\right) = \mathrm{o}(1).$$

6.2 Lower bound for the real part in Theorem 1.2. Let $I = [-2+2\kappa, 2-2\kappa] \cap N^{-1}\mathbb{Z}$. We start with a direct analogue of (3.21), with identical proof:

$$\mathbb{P}\Big(\sup_{z \in I + \mathrm{i}\eta_0} \operatorname{Re} L_N^H(z) \leqslant \sup_{E \in [-2+\kappa, 2-2\kappa]} \operatorname{Re} L_N^H(E) + 1\Big) = 1 - \mathrm{o}(1), \tag{6.7}$$

where η_0 is defined in (3.6). We take $t = \varphi^{-K}$, where $K > 0$ is a parameter, and $\widetilde{H}(t)$ as in Lemma 6.1. Let f_ε be a smooth function such that $0 \leqslant f_\varepsilon(x) \leqslant 1$ for $x \in \mathbb{R}$, $f_\varepsilon(x) = 0$ for $x \geqslant 1 - \varepsilon/2$, and $f_\varepsilon(x) = 1$ for $x \leqslant 1 - \varepsilon$. By Theorem 5.7 and Lemma 5.2, if K is chosen large enough, we have

$$\mathbb{E}\left[f_\varepsilon\left((\log N)^{-1} \sup_{z \in I + \mathrm{i}\eta_0} \operatorname{Re} L_N^{\widetilde{H}(t)}(z)\right)\right] - \mathbb{E}\left[f_\varepsilon\left((\log N)^{-1} \sup_{z \in I + \mathrm{i}\eta_0} \operatorname{Re} L_N^{H}(z)\right)\right] = \mathrm{O}\left(\frac{1}{\log N}\right). \tag{6.8}$$

By Proposition 4.3, there exists a coupling of $\widetilde{H}(t)$ and GOE_N such that

$$\mathbb{P}\left(\sup_{z \in I + \mathrm{i}\eta_0} \left|L_N^{\widetilde{H}(t)}(z) - L_N^{\mathrm{GOE}}(z)\right| > (\log N)^{\frac{1}{2}+\varepsilon}\right) = \mathrm{o}(1). \tag{6.9}$$

Assuming that

$$\mathbb{P}\left((\log N)^{-1}\sup_{z\in I+\mathrm{i}\eta_0} L_N^{\mathrm{GOE}}(z)<\sqrt{2}-\frac{\varepsilon}{2}\right)=\mathrm{o}(1), \tag{6.10}$$

the desired lower bound follows from (6.7), (6.8), (6.9), (6.10):

$$\mathbb{P}\left((\log N)^{-1}\sup_{E\in[2+\kappa,2-\kappa]}\operatorname{Re} L_N^{H}(E)<\sqrt{2}-2\varepsilon\right)=\mathrm{o}(1).$$

We now prove (6.10). Equation (3.24) yields

$$\mathbb{P}\left((\log N)^{-1}\max_{z\in[-2+2\kappa,2-2\kappa]+\mathrm{i}\eta_0}\operatorname{Re} L_N^{\mathrm{GOE}}(z)\leqslant\sqrt{2}-\frac{\varepsilon}{4}\right)=\mathrm{o}(1). \tag{6.11}$$

Moreover,

$$\mathbb{P}\left(\exists z,w\in[-2+2\kappa,2-2\kappa]+\mathrm{i}\eta_0:|z-w|<\frac{1}{N},|L_N^{\mathrm{GOE}}(z)-L_N^{\mathrm{GOE}}(w)|>(\log N)^{9/10}\right)$$
$$\leqslant\mathbb{P}\left(\exists z\in[-2+2\kappa,2-2\kappa]+\mathrm{i}\eta_0:|s(z)|>(\log N)^{7/10}\right)=\mathrm{o}(1), \tag{6.12}$$

where the last inequality follows from a union bound, Theorem 2.4, Markov's inequality, and a straightforward mesh argument (similar to the one before (3.14)). Equations (6.11) and (6.12) give (6.10) and conclude the proof.

6.3 Extremal deviation with optimal constant. We now prove part (i) of Theorem 1.8.

As before, set $t=\varphi^{-K}$, where $K>0$ is a large parameter and $\widetilde{H}(t)$ be the matrix (6.1) given by Lemma 6.1. Let f_ε be a smooth function such that $0\leqslant f_\varepsilon(x)\leqslant 1$ for $x\in\mathbb{R}$, $f_\varepsilon(x)=0$ for $|x|\in[\sqrt{2}-\frac{\varepsilon}{2},\sqrt{2}+\varepsilon/2]$, and $f_\varepsilon(x)=1$ for $x\in[0,\sqrt{2}-\varepsilon]\cup[\sqrt{2}+\varepsilon,\infty)$. By Theorem 5.10, Lemma 5.8 and Lemma 5.2, we have

$$\left(\mathbb{E}^{\widetilde{H}(t)}-\mathbb{E}^{H}\right)\left[f_\varepsilon\left(\frac{\pi N}{\log N}\max_{k\in[\![\kappa N,(1-\kappa)N]\!]}\rho_{\mathrm{sc}}(\gamma_k)|\lambda_k-\gamma_k|\right)\right]=\mathrm{O}\left(\frac{1}{\log N}\right). \tag{6.13}$$

By Proposition 4.1 and Corollary A.4, there exists a coupling of $\widetilde{H}(t)$ (with eigenvalues $\boldsymbol{\lambda}$) and GOE_N (with eigenvalues $\boldsymbol{\mu}$) such that

$$\mathbb{P}\left(\max_{k\in[\![\kappa N,(1-\kappa)N]\!]}|\lambda_k-\mu_k|>\frac{(\log N)^{\frac{1}{2}+\varepsilon}}{N}\right)=\mathrm{o}(1). \tag{6.14}$$

Finally, from Corollary 1.10 we have

$$\mathbb{P}^{\mathrm{GOE}}\left(\frac{\pi N}{\log N}\max_{k\in[\![\kappa N,(1-\kappa)N]\!]}\rho_{\mathrm{sc}}(\gamma_k)|\lambda_k-\gamma_k|\notin\left[\sqrt{2}-\frac{\varepsilon}{2},\sqrt{2}+\frac{\varepsilon}{2}\right]\right)=\mathrm{o}(1). \tag{6.15}$$

From (6.13), (6.14), and (6.15) we have

$$\mathbb{P}^{H}\left(\frac{\pi N}{\log N}\max_{k\in[\![\kappa N,(1-\kappa)N]\!]}\rho_{\mathrm{sc}}(\gamma_k)|\lambda_k-\gamma_k|\notin[\sqrt{2}-\varepsilon,\sqrt{2}+\varepsilon]\right)=\mathrm{o}(1).$$

6.4 Rigidity with optimal order. We finally prove part (ii) of Theorem 1.8, building on the key relaxation and moment matching results, namely lemmas 4.5 and 5.14.

First step: smoothed indicator for Gaussian divisible ensemble. We specify $f=f_E$ from (5.38) to be of type $f=\int\eta_1^{-1}h((x-E)/\eta_1)\mathbb{1}_{[x,\infty)}\mathrm{d}x$ where h is positive, smooth, compactly supported on $[0,1]$ and $\int h=1$. Then f satisfies the bounds (5.39). Moreover, for t and H as in Lemma 4.5, we have

$$\mathbb{E}[|\overline{\mathrm{Tr}}f(H_t)|^{2p}]\leqslant\int\eta_1^{-1}h((x-E)/\eta_1)\mathbb{E}[|\overline{\mathrm{Tr}}\mathbb{1}_{[x,\infty)}(H_t)|^{2p}]\mathrm{d}x\leqslant\left(\frac{2}{e\pi^2}+\varepsilon\right)^{p}p^{p}(\log N)^{p}, \tag{6.16}$$

where we have used convexity of $x \mapsto x^{2p}$ in the first inequality and the result of Lemma 4.5 in the second inequality.

Second step: smoothed spectrum for Gaussian divisible ensemble. We now prove that the actual spectrum $\boldsymbol{\lambda}$ can be replaced by the smoothed one $\widetilde{\boldsymbol{\lambda}}$ in (6.16).

Let $A = \{\left|\sum_{i=1}^N f(\widetilde{\lambda}_i) - \sum_{i=1}^N f(\lambda_i)\right| > \sqrt{\log N}\}$, and remember the notation (5.40). From Cauchy Schwarz and Lemma 5.46 we have

$$\mathbb{E}\left[X_p(H_t)\mathbb{1}_A\right] \ll N^{2\log N}\exp(-c\varphi/2) \to 0.$$

Moreover, by definition of A, for any $\varepsilon > 0$ and $\lambda > 0$ to be chosen we bound

$$\begin{aligned}\mathbb{E}\left[X_p(H_t)\mathbb{1}_{A^c}\right] &\leqslant \mathbb{E}\left[(|\overline{\mathrm{Tr}}f(H_t)| + \sqrt{\log N})^{2p}\right] \\ &\leqslant \mathbb{E}\left[(|\overline{\mathrm{Tr}}f(H_t)| + \frac{|\overline{\mathrm{Tr}}f(H_t)|}{\lambda})^{2p}\mathbb{1}_{|\overline{\mathrm{Tr}}f(H_t)|>\lambda\sqrt{\log N}}\right] + \mathbb{E}\left[(\lambda\sqrt{\log N} + \sqrt{\log N})^{2p}\mathbb{1}_{|\overline{\mathrm{Tr}}f(H_t)|\leqslant\lambda\sqrt{\log N}]}\right] \\ &\leqslant (1+\lambda^{-1})^{2p}\left(\frac{2}{e\pi^2}+\varepsilon\right)^p p^p(\log N)^p + (\lambda+1)^{2p}(\log N)^p \leqslant \left(\frac{2}{e\pi^2}+2\varepsilon\right)^p p^p(\log N)^p\end{aligned}$$

for $N \geqslant N_0(\varepsilon,\kappa,A)$. We have used the definition of A in the first inequality, (6.16) in the third one and the choice $\lambda = p^{1/10}$ in the third one.

We have therefore proved that for any $\varepsilon,\kappa,A > 0$ there is a $N_1(\varepsilon,\kappa,A)$ such that

$$\mathbb{E}\left[X_p(H_t)\right] \leqslant \left(\frac{2}{e\pi^2}+\varepsilon\right)^p p^p(\log N)^p \tag{6.17}$$

for any $p \leqslant A\log N$, $E \in [-2+\kappa, 2-\kappa]$, and $t > \exp(-(\log N)^{1/10})$.

Third step: moment matching. Let H be the Wigner matrix of interest in part (ii) of Theorem 1.8. Considering the dynamics (4.1), the moment matching lemma [44, Lemma 16.2] gives existence of a Wigner matrix H_0 such that the matrix $R = H_t$ satisfies $\mathbb{E}[H_{ij}^k] = \mathbb{E}[R_{ij}^k]$ for $k \in [\![1,3]\!]$ and $\left|\mathbb{E}[H_{ij}^4] - \mathbb{E}[R_{ij}^4]\right| \leqslant CN^{-2}t$, where $C > 0$ depends only on the constants from Definition 1.1.

We choose $t = \varphi^{-M}$ for a fixed $M > M_0$, with M_0 the constant from Lemma 5.14. Note that $t > \exp(-(\log N)^{1/10})$ so that (6.17) holds. We can therefore apply Lemma 5.14 with $R = H_t$ and obtain that there is a $N_2(\varepsilon,\kappa,A)$ such that

$$\mathbb{E}\left[X_p(H)\right] \leqslant \left(\frac{2}{e\pi^2}+\varepsilon\right)^p p^p(\log N)^p \tag{6.18}$$

for any $p \leqslant A\log N$, $E \in [-2+\kappa, 2-\kappa]$ and $N \geqslant N_2$. As in the previous step but in the reverse direction now, with Lemma 5.46 we obtain that the same property holds for the actual eigenvalues: for any choice of the parameters, there is a $N_3(\varepsilon,\kappa,A)$ such that

$$\mathbb{E}\left[|\overline{\mathrm{Tr}}f(H)|^{2p}\right] \leqslant \left(\frac{2}{e\pi^2}+\varepsilon\right)^p p^p(\log N)^p \tag{6.19}$$

for any $p \leqslant A\log N$, $E \in [-2+\kappa, 2-\kappa]$ and $N \geqslant N_3$.

To conclude, note that for any fixed ε there is a $N_4(\varepsilon,\kappa)$ such that for any $u > 1$, the inequality $\lambda_k - \gamma_k > u\cdot\frac{\sqrt{2}}{\pi\rho_{\rm sc}(\gamma_k)}\cdot\frac{\sqrt{\log N}}{N}$ implies that for $E = \gamma_k + u\cdot(\frac{\sqrt{2}}{\pi\rho_{\rm sc}(\gamma_k)} - 1)\cdot\frac{\sqrt{\log N}}{N}$ we have $\overline{\mathrm{Tr}}f_E(H) > (1-\varepsilon)\frac{u\sqrt{2}}{\pi}\sqrt{\log N}$. With (6.19) this gives

$$\mathbb{P}\left(\lambda_k - \gamma_k > u\cdot\frac{\sqrt{2}}{\pi\rho_{\rm sc}(\gamma_k)}\cdot\frac{\sqrt{\log N}}{N}\right) \leqslant \left(\frac{2}{e\pi^2}+\varepsilon\right)^p p^p(\log N)^p\cdot((1-\varepsilon)\frac{u\sqrt{2}}{\pi}\sqrt{\log N})^{-2p} \leqslant (\frac{p}{e}+10\varepsilon)^p u^{-2p}.$$

Optimization in p then concludes the proof.

6.5 Gaussian divisible ensemble: universality up to tightness. Theorem 1.4 follows immediately from Proposition 4.4.

A High Moments of Linear Statistics for Wigner Matrices

The main goal of this appendix is to prove Proposition A.1, which provides estimates on large moments (growing in N) of the semicircle law. This proposition is used in the proof of Proposition 4.3, via its Corollary A.4, and Lemma 4.5. While a weaker result would suffice for the proof of Proposition 4.3, for example a bound on a fixed but large moment, it is indeed necessary to control growing moments for the application in Lemma 4.5.

Proposition A.1. *Let H be a real symmetric Wigner matrix. Fix $K, \kappa, A > 0$. For every $E \in [-2+\kappa, 2-\kappa]$, $\eta \in \left[\varphi^{-K}, \varphi^{K}\right]$, and $p \in \mathbb{N}$ with $p \leqslant A \log N$, there exists a constant $C(K,\kappa,A) > 0$ such that*

$$\mathbb{E}\left[\left|m_N(z) - m_{\rm sc}(z)\right|^p\right] \leqslant \left(\frac{C}{N\eta}\right)^p p^{3p/4}. \tag{A.1}$$

Remark A.2. A natural approach to bounds such as (A.1) relies on concentration for random matrices. However, even on close-to-macroscopic scales and for bounded or log-concave matrix entries, this method would not give accurate enough bounds for Proposition 4.3 and Lemma 4.5. Indeed the concentration from [52] yields $\mathbb{P}\left(N|m_N(z) - \mathbb{E}[m_N(z)]| > \lambda\right) \leqslant c^{-1} e^{-c\lambda^2/\eta^4}$ ($\eta = {\rm Im} z$), so that

$$\mathbb{E}\left[\left|m_N(z) - \mathbb{E}[m_N(z)]\right|^p\right] \leqslant \left(\frac{C}{N\eta^2}\right)^p p^{p/2}.$$

Integration of $|m_N - \mathbb{E}[m_N(z)]| \lesssim C/(N\eta^2)$ gives $|L_N| \lesssim \varphi^C$ for $\eta \asymp \varphi^{-C}$, an error bigger than the order of magnitude $\max_{|E|<2-\varepsilon} L_N \asymp \log N$ that we aim at.

Remark A.3. For our application to Lemma 4.5, it is critical that the exponent 3/4 in (A.1) is smaller than 1, i.e. one could not afford the exponential tail error $\left(\frac{C}{N\eta}\right)^p p^p$.

We defer the proof of Proposition A.1 to Appendix A.4, after establishing various preliminary results in the following subsections. Throughout, we use the notations defined in Section 4.1.

A.1 High-probability bound on the log-characteristic polynomial. We begin with an application of Proposition A.1 that provides estimates on the maximum of the log-characteristic polynomial smoothed at almost-macroscopic scale.

Corollary A.4. *Let $\boldsymbol{\lambda}$ be the spectrum of a $N \times N$ Wigner matrix and remember the notation (4.50). Let $\kappa > 0$ and denote $z = E + \mathrm{i}\eta$.*

(i) Fix $C_1 > 10$. There exists a constant $c(C_1, \kappa) > 0$ such that for every $N \geqslant 1$ and $u \in [1, c(\log N)^{3/4}]$, we have

$$\mathbb{P}\left(\max_{|E|<2-\kappa, \eta\in[\varphi^{-C_1},1]} \left|L_N^{\boldsymbol{\lambda}}(z)\right| > u\right) \leqslant c^{-1} e^{c^{-1}(\log\log N)^2 - c(\frac{u}{\log_2 N})^{4/3}}. \tag{A.2}$$

Moreover, for any $u \leqslant C_1\varphi$, denoting $q = C_1 \log n$ we have

$$\mathbb{P}\left(\left|L_N^{\boldsymbol{\lambda}}(z)\right| > u\right) \leqslant c^{-1} e^{-c(\frac{u}{\log_2 N})^{4/3}} \mathbb{1}_{u<c^{-1}(\log N)^{3/4}} + c^{-q} q^{\frac{3}{2}q} u^{-2q} \mathbb{1}_{u>c^{-1}(\log N)^{3/4}} \tag{A.3}$$

(ii) Let $0 < \eta_1 < \eta_2 < 0$ be fixed and $\mathscr{C}$ be a fixed smooth path in $[-2+\kappa, 2-\kappa] \times [\eta_1, \eta_2]$, of finite length. Then for every $\varepsilon > 0$ there exists $M > 0$ such that for any $N \geqslant 1$,

$$\mathbb{P}\left(\max_{\mathscr{C}} \left|L_N^{\boldsymbol{\lambda}}(z)\right| > M\right) \leqslant 1 - \varepsilon. \tag{A.4}$$

Proof. We start with (i). By the local semicircle law (2.7) we have

$$\partial_z\left(\sum_{k=1}^N \log(z - \lambda_k)\right) = \sum_{k=1}^N \frac{1}{z-\lambda_k} = -N m_N(z) = \mathrm{O}(\varphi^{C_1+4}), \tag{A.5}$$

with probability $1 - \mathrm{O}(\exp(-\varphi^c))$, uniformly in $|E| < 2-\kappa$ and $\eta \in [\varphi^{-C_1}, 1]$.

Let $\delta = \varphi^{-C_1-4}$, and let $\mathcal{M} = \mathcal{M}_\delta = \{m_i\}_{i=1}^{\lfloor 4\varphi^4\delta^{-1}\rfloor}$ be a collection of points in $[-2+\kappa, 2-\kappa]\times[\varphi^{-C_1}, 1]$ such that any $z \in [-2+\kappa, 2-\kappa]\times[\varphi^{-C_1}, 1]$ there exists m_i such that $|z-m_i| \leqslant 10\delta$. By (A.5) and the definition of $\mathcal{M}$, we have

$$\max_{E\in[-2+\kappa,2-\kappa],\eta\in[\varphi^{-C_1},1]} |L_N^{\boldsymbol{\lambda}}(z)| = \max_{z\in\mathcal{M}} |L_N^{\boldsymbol{\lambda}}(z)| + \mathrm{O}(\delta\varphi^{C_1+4}). \tag{A.6}$$

Since $\delta\varphi^{C_1+4} = \mathrm{O}(1)$, to establish (A.2), it therefore suffices to show that, for any $z \in \mathcal{M}$,

$$\mathbb{P}\left(\left|L_N^{\boldsymbol{\lambda}}(z)\right| > u\right) \leqslant c^{-1} e^{-c\left(\frac{u}{\log_2 N}\right)^{4/3}}, \tag{A.7}$$

and then use a union bound on $|\mathcal{M}| = \mathrm{O}(\varphi^{C_1+8})$ points (where we recall $\log_2(x) = \log\log x$ and the choice of φ in (2.6)). To prove (A.7), for some parameter $\eta_1 > \eta$ we first write

$$L_N^{\boldsymbol{\lambda}}(z) = N\int_{s=\eta}^{\eta_1} (m_N(E+\mathrm{i}s) - m_{\mathrm{sc}}(E+\mathrm{i}s))\mathrm{d}s + L_N^{\boldsymbol{\lambda}}(E+\mathrm{i}\eta_1).$$

To bound the above terms, we compute using the rigidity bound (2.9) and Taylor expansion of log that

$$\left|\sum_{i=1}^N \log(E+\mathrm{i}\eta_1-\lambda_i) - \sum_{i=1}^N \log(E+\mathrm{i}\eta_1-\gamma_i)\right| = \left|\sum_{i=1}^N \log\left(1+\frac{\lambda_i-\gamma_i}{E+\mathrm{i}\eta_1-\gamma_i}\right)\right| \leqslant \sum_{i=1}^N C|\lambda_i-\gamma_i|/\eta_1$$
$$\leqslant C\varphi^{1-C_2}\sum_{i=1}^N \min(i, N+1-i)^{-1/3}N^{-2/3} \leqslant C\varphi^{1-C_2}, \tag{A.8}$$

where we choose $\eta_1 = \varphi^{C_2}$ for some $C_2 > 0$. Similar reasoning shows that

$$\left|N\int \log(E+\mathrm{i}\eta_1-u)\rho_{\mathrm{sc}}(u)\mathrm{d}u - \sum_{i=1}^N \log(E+\mathrm{i}\eta_1-\gamma_i)\right| \leqslant C\varphi^{1-C_2},$$

where $C = C(C_1,\kappa) > 0$ may change from line to line below. The previous three equations give (for fixed, large enough C_2), for arbitrary $D > 0$ and $N \geqslant N_0(D)$,

$$\mathbb{P}\left(|L_N^{\boldsymbol{\lambda}}(z) - \int_\eta^{\eta_1} N\big(m_N(E+\mathrm{i}s) - m_{\mathrm{sc}}(E+\mathrm{i}s)\big)\,\mathrm{d}s| > 1\right) \leqslant N^{-D}.$$

We denote $\Delta_N(z) = N(m_N(z) - m_{\mathrm{sc}}(z))$. By Markov's inequality, for any $p \geqslant 1$,

$$\mathbb{P}\left(\left|\int_\eta^{\eta_1}\Delta_N(E+\mathrm{i}s)\mathrm{d}s\right| > u\right) \leqslant u^{-2p}\mathbb{E}\left[\left|\int_\eta^{\eta_1}\Delta_N(E+\mathrm{i}s)\mathrm{d}s\right|^{2p}\right]$$
$$\leqslant u^{-2p}\left|\int_\eta^{\eta_1}\left(\mathbb{E}\left|\Delta_N(E+\mathrm{i}s)\right|^{2p}\right)^{\frac{1}{2p}}\mathrm{d}s\right|^{2p} \leqslant u^{-2p}C^{2p}p^{\frac{3p}{2}}\left(\int_\eta^{\eta_1}\frac{\mathrm{d}\eta}{\eta}\right)^{2p}, \tag{A.9}$$

where the second inequality is obtained by expansion and Hölder's inequality, and the third inequality relies on Proposition A.1 for $p = \mathrm{O}(\log N)$. We now recall that $\log\varphi = \mathrm{O}\big((\log\log N)^2\big)$, so that the above probability is also bounded with

$$u^{-2p}C^{2p}p^{\frac{3p}{2}}\left(C\log\log N\right)^{2p}.$$

The choice $p = e^{-1}(u/\log_2 N)^{4/3}$ proves (A.7) and concludes the proof of (A.2). The proof of (A.3) is the same, with no need of discretization.

The proof of (ii) is simpler as it does not need any discretization and only requires finite moment estimates. Indeed it follows from the following the following two facts. First, $(L_N(z_0))_N$ is tight, where $z_0 \in [-2+\kappa, 2-\kappa]\times[\eta_1,\eta_2]$ is fixed. This follows from convergence in distribution of this linear statistic (see e.g. [70]). Then $\max_{\mathscr{C}}|L_N(z) - L_N(z_0)|$ is also tight because it is dominated by $\int_{\mathscr{C}}|\Delta_N(w)|\cdot|\mathrm{d}w|$, which is tight by Proposition A.1 with $p = 2$. □

A.2 Preliminaries. We first list some preliminary results necessary for the proof of Proposition A.1. We begin with a power counting lemma for resolvent entries. Given a parameter $A > 0$, we set

$$\mathcal{D}_A = \left\{ z = E + \mathrm{i}\eta \in \mathbb{H} : |E| < 2 - A^{-1}, \eta \geqslant \varphi^{-A} \right\}. \tag{A.10}$$

Throughout, we let H be a Wigner matrix and let G denote its resolvent.

Lemma A.5. *Fix $A > 0$. There exists $C(A) > 0$ and $N_0(A) > 0$ such that the following holds for all $z \in \mathcal{D}_A$ and $N \geqslant N_0$. For any $i, k \in [\![1, N]\!]$ and random variable $F(z)$ such that $|F| \leqslant (A\eta^{-1})^{A \log N}$ almost surely,*

$$\mathbb{E}\left[|F| \cdot \frac{1}{N} \sum_{j=1}^{N} |G_{ij}| \right] \leqslant \left(\frac{C}{N\eta} \right)^{1/2} \mathbb{E}\big[|F|\big] + N^{-A \log N}, \quad \mathbb{E}\left[|F| \cdot \frac{1}{N} \sum_{j=1}^{N} |G_{ij}||G_{jk}| \right] \leqslant \frac{C}{N\eta} \mathbb{E}\big[|F|\big] + N^{-A \log N}, \tag{A.11}$$

where we set $G = G(z)$, $F = F(z)$, and $z = E + \mathrm{i}\eta$. More generally, for any $n \leqslant A \log N$,

$$\mathbb{E}\left[|F| \cdot \frac{1}{N^n} \sum_{j_1, \dots, j_n = 1}^{N} |G_{ij_1} \cdots G_{j_n k}| \right] \leqslant \left(\frac{C}{N\eta} \right)^{n/2} \mathbb{E}\,|F| + N^{-A \log N}. \tag{A.12}$$

Proof. We give only the proof of (A.12), as the proofs of the remaining statements are similar. Suppose that n is odd. We apply the elementary inequality

$$\sum_{j=1}^{N} |G_{aj} G_{jb}| \leqslant \frac{1}{2} \sum_{j=1}^{N} |G_{aj}|^2 + |G_{jb}|^2 \tag{A.13}$$

for $j = j_1, j_3, j_5 \dots$ and the Ward identity (2.13) to show that the left side of (A.12) is bounded by

$$\mathbb{E}\left[|F| \cdot \frac{1}{N^n} \sum_{j_2, j_4, j_6, \dots} \frac{1}{2\eta} (\operatorname{Im} G_{ii} + \operatorname{Im} G_{j_2 j_2}) \cdots \frac{1}{2\eta} (\operatorname{Im} G_{j_{n-1} j_{n-1}} + \operatorname{Im} G_{kk}) \right], \tag{A.14}$$

where there are $(n+1)/2$ factors in the sum.

Let $\mathcal{A}$ be the high-probability set from (2.8). For $\eta > N^{-1/2}$, we have $\operatorname{Im} G_{jj} < C$ on $\mathcal{A}$, which implies

$$\mathbb{E}\left[\mathbb{1}_{\mathcal{A}} |F| \cdot \frac{1}{N^n} \sum_{j_2, j_4, j_6, \dots} \frac{1}{2\eta} (\operatorname{Im} G_{kk} + \operatorname{Im} G_{j_2 j_2}) \cdot \frac{1}{2\eta} (\operatorname{Im} G_{j_2 j_2} + \operatorname{Im} G_{j_4 j_4}) \dots \right] \tag{A.15}$$

$$\leqslant \frac{N^{(n-1)/2}}{N^n} \left(\frac{C}{\eta} \right)^{(n+1)/2} \mathbb{E}\big[|F|\big] = \left(\frac{C}{N\eta} \right)^{(n+1)/2} \mathbb{E}\big[|F|\big]. \tag{A.16}$$

using $C(N\eta)^{-1} \leqslant 1$. On $\mathcal{A}^c$ we use the trivial bound $\operatorname{Im} G_{ii} < \eta^{-1}$ and the strong probability estimate on $\mathbb{P}(\mathcal{A}^c)$ from (2.8). This gives

$$\mathbb{E}\left[|F| \cdot \mathbb{1}_{\mathcal{A}^c} \frac{1}{N^n} \sum_{j_2, j_4, j_6, \dots} \frac{1}{2\eta} (\operatorname{Im} G_{kk} + \operatorname{Im} G_{j_2 j_2}) \cdot \frac{1}{2\eta} (\operatorname{Im} G_{j_2 j_2} + \operatorname{Im} G_{j_4 j_4}) \dots \right] \tag{A.17}$$

$$\leqslant \frac{N^{(n+1)/2}}{N^n \eta^{n+1}} \cdot (A\eta^{-1})^{A \log N} \cdot c^{-1} \exp(-c(\log N)^{C_0 \log \log N}) \leqslant N^{-A \log N}, \tag{A.18}$$

by the assumptions on η and n (recall $z \in \mathcal{D}_A$), for sufficiently large N. The claim follows by combining (A.15) and (A.17).

The proof for even n is similar, using $|G_{j_n k}| \leqslant C$ on $\mathcal{A}$ to bound the left side of (A.12) by

$$\mathbb{E}\left[|F| \cdot \frac{1}{N^n} \sum_{j_2, j_4, j_6, \dots} \frac{1}{2\eta} (\operatorname{Im} G_{ii} + \operatorname{Im} G_{j_2 j_2}) \cdots \frac{1}{2\eta} (\operatorname{Im} G_{j_{n-2} j_{n-2}} + \operatorname{Im} G_{j_n j_n}) |G_{j_n k}| \right]. \tag{A.19}$$

□

Given a random variable X, we let $\kappa^{(j)}(X)$ denote the j-th cumulant of X. The following lemma is [69, Lemma 3.2]. Known as a cumulant expansion, it provides an extension of the well-known Gaussian integration by parts formula to non-Gaussian random variables. (Observe that (A.20) reduces to a single term when Y is Gaussian, as all higher cumulants of Y vanish in this case.)

Lemma A.6. *Fix $\ell \in \mathbb{N}$, $Q > 0$, and $F \in C^{\ell+1}(\mathbb{R};\mathbb{C}^+)$. Let Y be a random variable such that $\mathbb{E}[Y] = 0$ with finite moments to order $\ell + 2$. Then*

$$\mathbb{E}\big[YF(Y)\big] = \sum_{r=1}^{\ell} \frac{\kappa^{(r+1)}(Y)}{r!}\mathbb{E}\left[F^{(r)}(Y)\right] + \mathbb{E}\Big[\Omega_\ell\big(YF(Y)\big)\Big], \tag{A.20}$$

where $\Omega_\ell(YF(Y))$ is an error term that satisfies

$$\Big|\mathbb{E}\Big[\Omega_\ell\big(YF(Y)\big)\Big]\Big| \leqslant C_\ell \mathbb{E}\left[|Y|^{\ell+2}\right] \sup_{|t|\leqslant Q}\left|F^{(\ell+1)}(t)\right| + C_\ell \mathbb{E}\left[|Y|^{\ell+2}\mathbb{1}(|Y|>Q)\right] \sup_{t\in\mathbb{R}}\left|F^{(\ell+1)}(t)\right|. \tag{A.21}$$

The constant C_ℓ satisfies $C_\ell \leqslant (C\ell)^\ell/\ell!$ for some $C > 0$ that does not depend on Q, F, or ℓ.

Lemma A.7. *Let H be a Wigner matrix in the sense of Definition 1.1. Then there exists a constant $C > 0$ such that*

$$\big|\kappa^{(k)}(H_{ij})\big| \leqslant \frac{(Ck)^{Ck}}{N^{k/2}}, \quad k \geqslant 3 \tag{A.22}$$

for all $i, j \in [\![1, N]\!]$.

Proof. The claim follows by expressing the cumulants in terms of moments, and using the moment bound

$$\mathbb{E}\big[|H_{ij}|^k\big] \leqslant \frac{(Ck)^{Ck}}{N^{k/2}}, \quad k \geqslant 3. \tag{A.23}$$

The bound (A.23) follows from the subexponential decay hypothesis (1.3). □

A.3 Main Calculation. Let $H = H(N)$ be a $N \times N$ Wigner matrix. We introduce the shorthand $m = m_N(z)$, where m_N denotes the Stieltjes transform of H. The proof of Proposition A.1 proceeds by bounding the moments

$$\mathbb{E}\big[|1 + zm + m^2|^{2D}\big], \qquad D \in \mathbb{N}. \tag{A.24}$$

To explain this strategy, observe that by the explicit form of m_{sc} in (2.3), we have $1 + zm_{\mathrm{sc}} + m_{\mathrm{sc}}^2 = 0$. Then, since $m \approx m_{\mathrm{sc}}$ for large N (by (2.8)), we have $1 + zm + m^2 \approx 0$. We will see later that the reverse implication also holds, so that sufficiently strong bounds on the moments in (A.24) imply the bounds on the moments of $|m_{\mathrm{sc}} - m|$ claimed in Proposition A.1. Therefore, we focus for now on (A.24).

Let $G = (H - z)^{-1}$ be the resolvent of H. To bound the moments (A.24), we use the definition of m in (2.2) to write

$$\mathbb{E}\big[|1 + zm + m^2|^{2D}\big] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}(1 + zG_{ii}) + m^2\right)(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)}^D\right], \tag{A.25}$$

which holds for any $D \in \mathbb{N}$. Set

$$P = P(m) = 1 + zm + m^2. \tag{A.26}$$

By the definition of the resolvent, we have $1 + zG_{ii} = (HG)_{ii}$, which implies

$$\mathbb{E}\big[|P|^{2D}\big] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{i,k=1}^{N} H_{ik}G_{ki}\right)P^{D-1}\overline{P}^D\right] + \mathbb{E}\left[m^2 P^{D-1}\overline{P}^D\right]. \tag{A.27}$$

Let $\ell \in N$ be a parameter, which will be fixed later. By a cumulant expansion using Lemma A.6 (setting $Y = H_{ik}$), we find that

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_{i,k} H_{ik}G_{ki}\right)P^{D-1}\overline{P}^D\right] = \frac{1}{N}\sum_{r=1}^{\ell}\frac{\kappa_{r+1}}{r!}\mathbb{E}\left[\sum_{i,k}(1 + \delta_{ik}\Delta_{r+1})\partial_{ik}^r\left(G_{ki}P^{D-1}\overline{P}^D\right)\right] + \mathbb{E}[\Omega_\ell]. \tag{A.28}$$

Here $\Omega_\ell = \Omega_\ell(z)$ is an error term that we will examine later, ∂_{ik} is the partial derivative in the matrix entry H_{ik}, κ_{r+1} is the $(r+1)$-th cumulant of H_{ij} for $i \neq j$, and Δ_r is equal to $(\kappa^{(r)}(H_{11}) - \kappa_r)/\kappa_r$.

We introduce the notation $I_{r,s}$ to denote the component of the r-th term of (A.28) where $r-s$ derivatives fall on G_{ik} and s derivatives fall on $P^{D-1}\overline{P}^D$:

$$I_{r,s} = \frac{\kappa_{r+1}}{N}\mathbb{E}\left[\sum_{i,k}(1+\delta_{ik}\Delta_{r+1})\left(\partial_{ik}^{r-s}G_{ki}\right)\partial_{ik}^{s}\left(P^{D-1}\overline{P}^D\right)\right]. \tag{A.29}$$

Then (A.28) becomes

$$\sum_{r=1}^{\ell}\sum_{s=0}^{r} w_{r,s}\mathbb{E}[I_{r,s}] + \mathbb{E}[\Omega_\ell], \qquad w_{r,s} = \frac{1}{r!}\binom{r}{s}. \tag{A.30}$$

We begin by bounding $\mathbb{E}[\Omega_\ell]$, then proceed to the $I_{r,s}$ terms.

A.3.1 Truncation.

Let $E^{[ik]}$ denote the $N \times N$ matrix with entries

$$(E^{[ik]})_{a,b} = \delta_{ia}\delta_{kb} + \delta_{ib}\delta_{ka} \text{ if } i \neq k, \quad (E^{[ik]})_{a,b} = \delta_{ia}\delta_{ib} \text{ if } i = k. \tag{A.31}$$

For $i,k \in [\![1,N]\!]$, we define $H^{(ik)} = H - H_{ik}E^{[ik]}$, which sets the (i,k) and (k,i) entries of H to zero.

Lemma A.8. *Suppose $i,k \in [\![1,N]\!]$, $D \leqslant \log N$, $A > 0$, and $z \in \mathcal{D}_A$. Define the function $F\colon \mathrm{Mat}_N \to \mathbb{C}$ by*

$$F_{ki}(M) = R_{ki}(M)P^{D-1}\overline{P}^D, \tag{A.32}$$

where here $R(M)$ denotes the resolvent of M. Choose $\ell \in \mathbb{N}$ such that $\ell < A\log N$. Then there exist constants $C, C_1(A) > 0$ such that

$$\mathbb{E}\left[\sup_{x\in\mathbb{R},|x|\leqslant N^{-1/4}}\left|\partial_{ik}^{\ell}F_{ki}\left(H^{(ik)} + xE^{[ik]}\right)\right|\right] \leqslant (4D+2\ell)^{\ell}C^{4D+\ell} + C_1 N^{-2\log N}. \tag{A.33}$$

Proof. Fix index pairs (a,b) and (i,k). By resolvent expansion (2.12), we have

$$\widetilde{G}_{ab} = G_{ab} + xH_{ik}\widetilde{G}_{ai}G_{kb} + xH_{ik}\widetilde{G}_{ak}G_{ib}, \tag{A.34}$$

where $\widetilde{G}$ is the resolvent of $H^{(ik)} + xE^{[ik]}$. By (1.3), we have $|H_{ik}| \leqslant 1$ with probability at least $1 - c^{-1}\exp(-N^{c/2})$. Combining this bound with (2.8) and (A.34), we obtain uniformly in $z \in \mathcal{D}_A$, with high probability, that

$$\sup_{|x|\leqslant N^{-1/4}}\max_{a,b}\left|\widetilde{G}_{ab}\right| \leqslant C. \tag{A.35}$$

Since P is a quadratic polynomial in m, F_{ki} is a polynomial of degree $4D-1$ in G_{ki} and m, with at most 2^{2D-1} terms. When ∂_{jk} acts on P or $\overline{P}$ it generates a new factor

$$2N^{-1}\sum_{i_l} G_{ji_l}G_{i_lk}, \tag{A.36}$$

with a new summation index i_l. Then $\partial_{ik}^{\ell}F_{ki}$ has degree $4D-1+\ell$ when considered as a polynomial in Green's function entries G_{ii}, G_{kk}, G_{ik}, m, and terms of the form (A.36). The number of such terms in $\partial_{ik}^{\ell}F_{ki}$ is bounded by $2^{2D-1}\times(4D-1+2\ell)^{\ell}$. Using (A.35), the contribution to the left side of (A.33) from the expectation on the set where (2.8) holds is bounded by the first term of (A.33) (after increasing C). On the low probability set where (A.35) does not hold, we use the trivial bound $|G_{ij}| \leqslant \eta^{-1}$ (from (2.15)) and the assumed lower bound $\eta > \varphi^{-A}$, which produces the second term of (A.33). □

Lemma A.9. *Let $A > 0$, $4\log N \leqslant \ell \leqslant A\log N$, $D \leqslant \log N$, and $z \in \mathcal{D}_A$. There exists a constant $C(A) > 0$ such that the term Ω_ℓ from* (A.28) *satisfies* $\sup_{z\in\mathcal{D}_A}\left|\mathbb{E}[\Omega_\ell]\right| \leqslant CN^{-\ell/4}$.

Proof. By Lemma A.6 with $Q = N^{-1/4}$,

$$\left|\mathbb{E}\big[\Omega_\ell(H_{ik}F_{ki})\big]\right| \leqslant C_\ell \mathbb{E}\left[|H_{ik}|^{\ell+2}\right] \mathbb{E}\left[\sup_{|x|\leqslant N^{-1/4}} \left|F^{(\ell+1)}(H^{(ik)} + xE^{[ik]})\right|\right] \tag{A.37}$$

$$+ C_\ell \mathbb{E}\left[|H_{ik}|^{\ell+2}\mathbb{1}\big(|H_{ik}| > N^{-1/4}\big)\right] \mathbb{E}\left[\sup_{x\in\mathbb{R}} \left|F_{ik}^{(\ell+1)}(H^{(ik)} + xE^{[ik]})\right|\right]. \tag{A.38}$$

The first term may be bounded using Lemma A.8, Lemma A.6, and the moment bound (A.23):

$$C_\ell \mathbb{E}\left[|H_{ik}|^{\ell+2}\right] \mathbb{E}\left[\sup_{|x|\leqslant N^{-1/4}} \left|F^{(\ell+1)}(H^{(ik)} + xE^{[ik]})\right|\right] \leqslant \frac{(C\ell)^{C\ell}}{\ell! N^{\ell/2}} \left((4D+2\ell)^\ell C^{4D+\ell} + CN^{-2\log N}\right). \tag{A.39}$$

For the second term, we use the trivial bound $|G_{ij}| \leqslant \eta^{-1}$ (from (2.15)) to obtain the deterministic bound

$$\sup_{x\in\mathbb{R}} \left|F_{ik}^{(\ell+1)}(H^{(ik)} + xE^{[ik]})\right| \leqslant 2^{2D}(4D+2\ell)^\ell \left(\frac{C}{\eta}\right)^{4D+\ell}, \tag{A.40}$$

and using Hölder's inequality and (1.3) we have

$$\mathbb{E}\left[|H_{ik}|^{\ell+2}\mathbb{1}(|H_{ik}| > N^{-1/4})\right] \leqslant \frac{(C\ell)^{C\ell}}{N^{\ell/2}} \exp(-cN^{c/4}). \tag{A.41}$$

Combining our estimates for the first and second terms of Ω_ℓ yields the conclusion after using $\ell \geqslant 4\log N$. □

A.3.2 Main Terms.

We will need to analyze the terms $I_{r,s}$ from (A.29) explicitly for $s \leqslant 1$. For the others, we can proceed on general combinatorial grounds. The following lemma collects our estimates on these terms. We set $P' = P'(m) = 2m + z$.

Lemma A.10. *Fix $A > 0$ and suppose $D \leqslant \log N$. For all $z \in \mathcal{D}_A$, we have*

$$\mathbb{E}[I_{1,0}] + \mathbb{E}[I_{1,1}] = -\mathbb{E}[m^2 P^{D-1}\overline{P}^D] + \Omega, \tag{A.42}$$

where $\Omega = \Omega(z) \geqslant 0$ is an error term satisfying

$$\Omega \leqslant \left(\frac{C}{N\eta}\right) \mathbb{E}\left[|P|^{2D-1}\right] + D\left(\frac{C}{N\eta}\right)^2 \mathbb{E}\left[|P|^{2D-2}\right] + CN^{-2\log N}. \tag{A.43}$$

Further, we have

$$\left|\mathbb{E}[I_{2,0}]\right| \leqslant C(\log N)N^{-1/2} \sum_{a=1}^{2D} \left(\frac{1}{N\eta}\right)^a \mathbb{E}\left[|P|^{2D-a}\right] + N^{-2\log N} \tag{A.44}$$

and

$$\left|\mathbb{E}[I_{2,1}]\right| \leqslant C(\log N)N^{-1/2} \sum_{a=1}^{2D} \left(\frac{1}{N\eta}\right)^a \mathbb{E}\left[|P|^{2D-a}\right] + N^{-2\log N}. \tag{A.45}$$

For $r \geqslant 3$ we have

$$\left|\mathbb{E}[I_{r,0}]\right| + \left|\mathbb{E}[I_{r,1}]\right| \leqslant \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \mathbb{E}\left[|P|^{2D-1}\right] + DC^r \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \frac{C}{N\eta} \mathbb{E}\left[|P|^{2D-2}\right] + CN^{-2\log N}. \tag{A.46}$$

Finally, for $r \geqslant 2$,

$$\left|\mathbb{E}[I_{r,r}]\right| \leqslant \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \sum_{s_0=2}^{r+1} \left(\frac{C}{N\eta}\right)^{1/2} \left(\frac{CD}{N\eta}\right)^{s_0-1} \mathbb{E}\left[|P|^{2D-s_0}\right] + CN^{-2\log N}, \tag{A.47}$$

and for pairs (r,s) such that $r > s \geqslant 2$,

$$\left|\mathbb{E}[I_{r,s}]\right| \leqslant \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \sum_{a=2}^{r+1} \left(\frac{CD}{N\eta}\right)^{a-1} \mathbb{E}\left[|P|^{2D-a}\right] + CN^{-2\log N}. \tag{A.48}$$

For clarity we prove the claims (A.42), (A.44), (A.45), (A.46), (A.47), and (A.48) separately.

Proof of (A.42). We write, using $\kappa_2 = N^{-1}$,

$$\mathbb{E}[I_{1,0}] = \frac{1}{N^2}\mathbb{E}\left[\sum_{i,k}(1+\delta_{ik}\Delta_{r+1})\,(\partial_{ik}G_{ki})\,P^{D-1}\overline{P}^D\right]. \tag{A.49}$$

We first bound the terms with $i = k$, which are sub-leading. Using Theorem 2.2, we have

$$\frac{1}{N^2}\left|\mathbb{E}\left[\sum_{i}(1+\Delta_{r+1})\,(\partial_{ii}G_{ii})\,P^{D-1}\overline{P}^D\right]\right| \leqslant CN^{-1}\mathbb{E}\left[|P|^{2D-1}\right] + CN^{-2\log N}. \tag{A.50}$$

For $i \neq k$, we have

$$(\partial_{ik}G_{ki})\,P^{D-1}\overline{P}^D = -G_{ii}G_{kk}P^{D-1}\overline{P}^D - G_{ki}G_{ki}P^{D-1}\overline{P}^D. \tag{A.51}$$

Considering the first term, we have

$$\frac{1}{N^2}\sum_{i,k} -G_{ii}G_{kk}P^{D-1}\overline{P}^D = -m^2P^{D-1}\overline{P}^D, \tag{A.52}$$

which matches the first term of (A.42). Considering the second term of (A.51) and using (A.11), we have

$$\frac{1}{N^2}\left|\sum_{i,k} G_{ki}G_{ki}P^{D-1}\overline{P}^D\right| \leqslant \frac{C}{N\eta}\mathbb{E}\left[|P|^{2D-1}\right] + N^{-2\log N}. \tag{A.53}$$

This completes the analysis of $I_{1,0}$. Next, we have

$$\mathbb{E}[I_{0,1}] = \frac{1}{N^2}\mathbb{E}\left[\sum_{i,k}(1+\delta_{ik}\Delta_{r+1})G_{ki}\cdot\partial_{ik}\left(P^{D-1}\overline{P}^D\right)\right], \tag{A.54}$$

and

$$\partial_{ik}\left(P^{D-1}\overline{P}^D\right) = -\frac{2(D-1)}{N}G_{ki}P'\sum_{j=1}^{N}G_{jk}G_{ij}P^{D-2}\overline{P}^D - \frac{2D}{N}G_{ki}\overline{P}'\sum_{j=1}^{N}\overline{G_{jk}G_{ij}}P^{D-1}\overline{P}^{D-1}. \tag{A.55}$$

As in (A.50), we can remove the $i = k$ terms with negligible error. For the first term of (A.55), we use (2.8) and (A.12) to get

$$\frac{1}{N^3}\left|\mathbb{E}\left[\sum_{i\neq k}G_{ki}P'\sum_{j=1}G_{jk}G_{ij}P^{D-2}\overline{P}^D\right]\right| \leqslant \left(\frac{C}{N\eta}\right)^2\mathbb{E}\left[|P'||P|^{2D-2}\right] + N^{-2\log N}. \tag{A.56}$$

The bound on the second term of (A.55) is similar. □

Proof of (A.44). We write $I_{2,0} = I_{2,0}^{(1)} + I_{2,0}^{(3)} + R_1$, where $I_{2,0}^{(1)}$ contains all terms with exactly one off-diagonal resolvent entry, $I_{2,0}^{(1)}$ contains all terms with three off-diagonal resolvent entries, and R_1 contains all other terms. Reasoning as in (A.50), we see that R_1 is negligible:

$$\left|\mathbb{E}[R_1]\right| \leqslant N^{-1}\mathbb{E}\left[|P|^{2D-1}\right] + CN^{-2\log N}. \tag{A.57}$$

By Theorem 2.2, we get

$$\left|\mathbb{E}[I_{2,0}^{(3)}]\right| = \frac{\kappa_3}{N}\left|\mathbb{E}\left[\sum_{i\neq k}(G_{ik})^3P^{D-1}\overline{P}^D\right]\right| \leqslant N^{-1/2}\left(\frac{C}{N\eta}\right)^{3/2}\mathbb{E}\left[|P|^{2D-1}\right] + CN^{-2\log N}. \tag{A.58}$$

To study $I_{2,0}^{(1)}$, we perform another cumulant expansion using (A.20):

$$z\mathbb{E}[I_{2,0}^{(1)}] = \sum_{r=1}^{\ell'}\sum_{s=0}^{r} w_{r,s}\mathbb{E}[\hat{I}_{r,s}] + \hat{\Omega}_{\ell'}, \tag{A.59}$$

where

$$\mathbb{E}[\hat{I}_{r,s}] = N\kappa_{r+1}N\kappa_3\mathbb{E}\left[\frac{1}{N^3}\sum_{i\neq k,j}\left(\partial_{kj}^{r-s}\left(G_{ji}G_{kk}G_{ii}\right)\right)\left(\partial_{kj}^{s}\left(P^{D-1}\overline{P}^{D}\right)\right)\right]. \tag{A.60}$$

We take $\ell' = 20\log N$ and see that

$$|\hat{\Omega}_{\ell'}| \leqslant CN^{-2\log N} \tag{A.61}$$

by a straightforward modification of the proof of Lemma A.9.

We begin with the terms $\hat{I}_{r,0}$ for $r \geqslant 1$. When $r = 1$, we define terms $\hat{I}_{1,0}^{(i)}$ for $i = 1,2,3$ by the decomposition

$$\begin{aligned}
\mathbb{E}[\hat{I}_{1,0}] = & - N\kappa_3\mathbb{E}\left[N^{-3}\sum_{i_1\neq i_2,i_3} G_{i_2i_1}G_{i_3i_3}G_{i_2i_2}G_{i_1i_1}P^{D-1}\overline{P}^{D}\right] && \text{(A.62)}\\
& - 3N\kappa_3\mathbb{E}\left[N^{-3}\sum_{i_1\neq i_2,i_3} G_{i_2i_3}G_{i_3i_1}G_{i_2i_2}G_{i_1i_1}P^{D-1}\overline{P}^{D}\right] && \text{(A.63)}\\
& - 2N\kappa_3\mathbb{E}\left[N^{-3}\sum_{i_1\neq i_2,i_3} G_{i_1i_2}G_{i_2i_3}G_{i_3i_1}G_{i_2i_2}P^{D-1}\overline{P}^{D}\right] && \text{(A.64)}\\
& = \mathbb{E}[\hat{I}_{1,0}^{(1)}] + 3\mathbb{E}[\hat{I}_{1,0}^{(2)}] + 2\mathbb{E}[\hat{I}_{1,0}^{(3)}]. && \text{(A.65)}
\end{aligned}$$

where we have decomposed the sum according to the number of off-diagonal resolvent entries in each product, Using Lemma A.5 on the off-diagonal resolvent entries, we find

$$\left|\mathbb{E}[\hat{I}_{1,0}^{(2)}]\right| + |\mathbb{E}[\hat{I}_{1,0}^{(3)}]| \leqslant N^{-1/2}\left(\frac{C}{N\eta}\right)\mathbb{E}\,|P|^{2D-1} + N^{-2\log N}. \tag{A.66}$$

Further, by using (2.8) and incurring a negligible error, we may replace $\mathbb{E}[\hat{I}_{1,0}^{(1)}]$ by

$$-m_{\mathrm{sc}}\mathbb{E}\left[N^{-2}\sum_{i_1\neq i_2} G_{i_2i_1}G_{i_2i_2}G_{i_1i_1}P^{D-1}\overline{P}^{D}\right] = -m_{\mathrm{sc}}\mathbb{E}[I_{2,0}^{(1)}]. \tag{A.67}$$

We now turn to terms $\hat{I}_{r,0}$ with $r > 1$. We observe that Lemma A.7 implies

$$(N\kappa_{r+1})\,(N\kappa_3) \leqslant (Cr)^{Cr}N^{-r/2} \tag{A.68}$$

and recall that $|w_{r,0}| \leqslant 1$. Since every product in $\hat{I}_{r,0}$ has at least one off-diagonal entry, by (2.8) we have the bound

$$w_{r,0}\left|\mathbb{E}[\hat{I}_{r,0}]\right| \leqslant (Cr)^{Cr}N^{-r/2}\left(\frac{1}{N\eta}\right)^{1/2}\mathbb{E}\,|P|^{2D-1} + N^{-2\log N}. \tag{A.69}$$

We next consider the terms $\hat{I}_{r,s}$ with $s = 1$. In this case, we first note that the order $r-1$ derivative of the product of resolvent entries in $\hat{I}_{r,1}$ contributes at least one off-diagonal resolvent entry. Next, we see that each $\partial_{ij}\left(P^{D-1}\overline{P}^{D}\right)$ contains either two factors of (A.36), or the derivative of (A.36). Using (A.68), this leads to the bound

$$\left|\mathbb{E}[\hat{I}_{r,1}]\right| \leqslant D(Cr)^{Cr}N^{-r/2}\left(\frac{1}{N\eta}\right)^{3/2}\mathbb{E}\left[|P'||P|^{2D-2}\right]. \tag{A.70}$$

We absorbed the combinatorial factor corresponding to the number of terms coming from the derivatives, which is bounded by C^r, into the prefactor.

Now we consider the case of $\hat{I}_{r,s}$ with $2 \leqslant s \leqslant r$. We begin by noting that (2.8) gives

$$\left|\mathbb{E}[\hat{I}_{r,s}]\right| \leqslant (Cr)^{Cr} N^{-r/2} \left|\mathbb{E}\left[\frac{1}{N^3}\sum_{i\neq k,j}\left(\partial_{jk}^{r-s}\left(G_{ji}G_{kk}G_{ii}\right)\right)\left(\partial_{jk}^{s}\left(P^{D-1}\overline{P}^{D}\right)\right)\right]\right| \tag{A.71}$$

$$\leqslant (Cr)^{Cr} N^{-r/2}\left(\frac{1}{N\eta}\right)^{1/2}\mathbb{E}\left[\frac{1}{N^2}\sum_{k,j}\left|\partial_{jk}^{s}\left(P^{D-1}\overline{P}^{D}\right)\right|\right]. \tag{A.72}$$

Here we used the fact that $\partial_{jk}^{r-s}\left(G_{ji}G_{kk}G_{ii}\right)$ always contains at least one off-diagonal resolvent entry, and again absorbed the combinatorial factor corresponding to the number of terms in this derivative into the prefactor.

For $s \geqslant 2$, the derivative $\partial_{jk}^{s}\left(P^{D-1}\overline{P}^{D}\right)$ is a sum of terms that may contain factors of P and P', and their conjugates. Any such term is a constant times a product of the form $P^{D-s_1}\overline{P}^{D-s_2}(P')^{s_3}\left(\overline{P'}\right)^{s_4}$, with $s_i \geqslant 0$ for $i \in [\![1,4]\!]$. A term with such a product came from ∂_{jk} acting s_1-1 times on P and s_2 times on $\overline{P}$, so we must have $s_1 - 1 \geqslant s_3$, $s_2 \geqslant s_4$, and $s_1 - 1 + s_2 \leqslant s$. We further see that ∂_{jk} acted s_1-1-s_3 times on P' and $s_2 - s_4$ times on $\overline{P'}$.

When ∂_{jk} acts on a power of P, $\overline{P}$, or their derivatives, it generates a new factor of $\partial_{jk} m = 2N^{-1}\sum_{i_l} G_{ji_l}G_{i_l k}$, where i_l is a new summation index (not appearing elsewhere), and a constant prefactor no greater than D (by the chain rule applied to P^k and analogous terms for $k \leqslant D$). The number of new summation indices is then

$$s_1 - 1 + s_2 + (s_1 - 1 - s_3) + (s_2 - s_4) = 2s_1 + 2s_2 - s_3 - s_4 - 2. \tag{A.73}$$

Further, this number does not decrease when ∂_{jk} acts on resolvent entries instead of P, $\overline{P}$, or their derivatives. We introduce $a = s_1 + s_2$ and $b = s_3 + s_4$. Then, using $a \leqslant s+1$, (A.73) yields

$$\left|\mathbb{E}[\hat{I}_{r,s}]\right| \leqslant (Cr)^{Cr} D^r N^{-r/2}\sum_{a=2}^{s+1}\sum_{b=0}^{a-2}\left(\frac{1}{N\eta}\right)^{1/2+2a-b-2}\mathbb{E}\left[|P'|^b|P|^{2D-a}\right] \tag{A.74}$$

$$+ (Cr)^{Cr} D^r N^{-r/2}\sum_{a=2}^{s+1}\left(\frac{1}{N\eta}\right)^{1/2+2a-1}\mathbb{E}\left[|P'|^{a-1}|P|^{2D-a}\right], \tag{A.75}$$

where the second term comes from terms such that $b = a-1$. After increasing C, we obtain

$$\left|\mathbb{E}[\hat{I}_{r,s}]\right| \leqslant (Cr)^{Cr} D^r N^{-r/2}\sum_{a=2}^{r+1}\left(\frac{1}{N\eta}\right)^{a}\mathbb{E}\left[|P|^{2D-a}\right] + CN^{-5\log N}. \tag{A.76}$$

Combining the definition of $I_{2,0}$, (A.58), (A.59), (A.61), and the estimates on the $\hat{I}_{r,s}$ terms, we obtain

$$\left|(z+m_{\mathrm{sc}})\mathbb{E}[I_{2,0}^{(1)}]\right| \leqslant \sum_{r=2}^{8\log N}(Cr)^{Cr}D^rN^{-r/2}\sum_{a=1}^{r}\left(\frac{1}{N\eta}\right)^{a}\mathbb{E}\left[|P|^{2D-a}\right] \tag{A.77}$$

$$+ \sum_{r=2}^{8\log N}(Cr)^{Cr}DN^{-r/2}\left(\frac{1}{N\eta}\right)^{3/2}\mathbb{E}\left[|P|^{2D-2}\right] \tag{A.78}$$

$$+ \sum_{r=2}^{8\log N}(Cr)^{Cr}N^{-r/2}\left(\frac{1}{N\eta}\right)^{1/2}\mathbb{E}\,|P|^{2D-1} \tag{A.79}$$

$$+ N^{-1/2}\left(\frac{C}{N\eta}\right)\mathbb{E}\,|P|^{2D-1} + N^{-2\log N}. \tag{A.80}$$

After some simplification and increasing the value of C, this implies

$$\left|(z+m_{\mathrm{sc}})\mathbb{E}[I_{2,0}^{(1)}]\right| \leqslant C(\log N)N^{-1/2}\sum_{a=1}^{2D}\left(\frac{1}{N\eta}\right)^{a}\mathbb{E}\left[|P|^{2D-a}\right] + N^{-2\log N}. \tag{A.81}$$

Using $|z + m_{\mathrm{sc}}(z)| > c(A) > 0$ on $\mathcal{D}_A$ (see (2.3)), we obtain the conclusion. □

Proof of (A.45). We have

$$\mathbb{E}[I_{2,1}] = N\kappa_3\mathbb{E}\left[N^{-2}\sum_{i,k}(1+\delta_{ik}\Delta_3)\,(\partial_{ik}G_{ki})\,\partial_{ik}\left(P^{D-1}\overline{P}^D\right)\right]. \tag{A.82}$$

Using the logic of the previous proof, it is straightforward to see that the leading-order contribution is given by

$$J = N\kappa_3\mathbb{E}\left[\frac{2(D-1)}{N^2}\sum_{i_1\neq i_2}G_{i_1i_1}G_{i_2i_2}\left(\frac{1}{N}\sum_{i_3=1}^{N}G_{i_2i_3}G_{i_3i_1}\right)P'P^{D-2}\overline{P}^D\right] \tag{A.83}$$

$$+ N\kappa_3\mathbb{E}\left[\frac{2D}{N^2}\sum_{i_1\neq i_2}G_{i_1i_1}G_{i_2i_2}\left(\frac{1}{N}\sum_{i_3=1}^{N}G_{i_2i_3}G_{i_3i_1}\right)P'P^{D-1}\overline{P}^{D-1}\right]. \tag{A.84}$$

For the first term, we use the resolvent expansion to write it as

$$2(D-1)N\kappa_3\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2,i_3,i_4}H_{i_2i_4}G_{i_4i_3}G_{i_1i_1}G_{i_2i_2}G_{i_3i_1}P'P^{D-2}\overline{P}^D\right]. \tag{A.85}$$

As in the previous proof, we now use the cumulant expansion (A.20) to calculate this term (expanding each term in the sum in the variable $H_{i_2i_4}$). The leading term in the expansion is

$$2(D-1)N\kappa_3\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2,i_3}mG_{i_2i_3}G_{i_2i_2}G_{i_3i_1}P'P^{D-2}\overline{P}^D\right], \tag{A.86}$$

and we obtain

$$|z+m_{\mathrm{sc}}(z)|\,|2(D-1)N\kappa_3|\,\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2,i_3}G_{i_2i_3}G_{i_2i_2}G_{i_3i_1}P'P^{D-2}\overline{P}^D\right]$$
$$\leqslant C(\log N)N^{-1/2}\sum_{a=1}^{2D}\left(\frac{1}{N\eta}\right)^a\mathbb{E}\left[|P|^{2D-a}\right]+N^{-2\log N}. \tag{A.87}$$

This controls the first term of J. By nearly identical reasoning, a similar bound holds for the second term of J. Using $|z+m_{\mathrm{sc}}(z)|\geqslant c$ for $z\in\mathcal{D}_A$ in (A.87), we obtain the result. □

Proof of (A.46). By Theorem 2.2 we have

$$|\mathbb{E}[I_{r,0}]| = \left|N\kappa_{r+1}\mathbb{E}\left[N^{-2}\sum_{i,k}(1+\delta_{ik}\Delta_{r+1})\,(\partial_{ik}^rG_{ki})\,P^{D-1}\overline{P}^D\right]\right| \tag{A.88}$$

$$\leqslant \frac{(Cr)^{Cr}}{N^{(r-1)/2}}\mathbb{E}\left[|P|^{2D-1}\right]+CN^{-2\log N}. \tag{A.89}$$

We absorbed the combinatorial factor 4^r representing the number of different terms coming from the derivatives of G_{ki} into the constant.

For $I_{r,1}$, we have

$$|\mathbb{E}[I_{r,1}]| \leqslant \frac{(Cr)^{Cr}}{N^{(r-1)/2}}\left|\mathbb{E}\left[N^{-2}\sum_{i,k}(1+\delta_{ik}\Delta_{r+1})\,(\partial_{ik}^{r-1}G_{ki})\,\partial_{ik}\left(P^{D-1}\overline{P}^D\right)\right]\right|. \tag{A.90}$$

From the derivative on $P^{D-1}\overline{P}^D$ we get $|P'||P|^{2D-2}$, a factor of $2N^{-1}\sum_a G_{ia}G_{ak}$, a factor of D, and some constant that is bounded uniformly in r. For the terms with at least 3 off-diagonal G_{ab}, we can use Theorem 2.2 and Lemma A.5 to get the bound

$$DC^r \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \mathbb{E}\left[N^{-3} \sum_{i,k,a} |G_{ki} G_{ia} G_{ak}||P'||P|^{2D-2} \right] \tag{A.91}$$

$$\leqslant DC^r \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \left(\frac{C}{N\eta} \right)^{3/2} \mathbb{E}\left[|P'||P|^{2D-2} \right] + CN^{-5\log N} \tag{A.92}$$

$$\leqslant DC^r \frac{(Cr)^{Cr}}{N^{(r-2)/2}} \left(\frac{C}{N\eta} \right)^{2} \mathbb{E}\left[|P|^{2D-2} \right] + CN^{-5\log N}. \tag{A.93}$$

For terms with only two off-diagonal entries, we have

$$DC^r \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \mathbb{E}\left[N^{-2} \sum_{i,k} G_{ii}^{r/2} G_{kk}^{r/2} \left(\frac{1}{N} \sum G_{ia} G_{ak} \right) P' P^{D-2} \overline{P}^D \right] \tag{A.94}$$

$$\leqslant DC^r \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \frac{C}{N\eta} \mathbb{E}\left[|P|^{2D-2} \right] + CN^{-2\log N} \tag{A.95}$$

and the same bound for the term where the derivative falls on $\overline{P}^D$. This completes the proof. □

Proof of (A.47) *and* (A.48). We treat all terms $I_{r,s}$ with $2 \leqslant s \leqslant r$ the same way. From the order s derivatives in $I_{r,s}$, we get a sum of monomials of the form

$$P^{D-s_1} \overline{P^{D-s_2}} (P')^{s_3} \overline{(P')}^{s_4} \prod_{d=1}^{n} Q_d, \tag{A.96}$$

with $1 \leqslant s_1 \leqslant D$ and $0 \leqslant s_2 \leqslant D$ depending on how the derivatives fall, and we have omitted the constant prefactor. Each Q_d represents a "fresh summation index" i_d in the following way. Any derivative on P, $\overline{P}$ generates a new summation index a with a factor $\partial_{ik} m = 2N^{-1} \sum_a G_{ia} G_{ak}$. Similarly a derivative on P' or $\overline{P}'$ gives $4N^{-1} \sum_a G_{ia} G_{ak}$. Each Q_d represents a sum corresponding to one of these new indices, potentially differentiated further. For example, Q_1 could be $2N^{-1} \sum_a G_{ia} G_{ak}$, or (applying a derivative ∂_{ik})

$$2N^{-1} \sum_a \left(G_{ii} G_{ak} G_{ak} + G_{ik} G_{ia} G_{ak} + G_{ia} G_{ia} G_{kk} + G_{ia} G_{ak} G_{ik} \right), \tag{A.97}$$

or any higher derivative. We consider any constant factors that are produced when a new index is generated, or when a Q_d term is differentiated, as part of the corresponding Q_d term. For example, in (A.97), we consider the factor 2 as part of Q_d.

We see that the monomial (A.96) came from $P^{D-1}\overline{P^D}$ because $s_1 - 1$ derivatives ∂_{ik} derivatives fell on P and s_2 on $\overline{P}$. This implies $s_3 \leqslant s_1 - 1$ and $s_4 \leqslant s_2$. The number of derivatives on P' was $s_1 - 1 - s_3$, and on $\overline{P'}$ it was $s_2 - s_4$. Then the total number of new indices is

$$n = 2s_1 + 2s_2 - s_3 - s_4 - 2. \tag{A.98}$$

Let the number of derivatives that fall on some on some Q-type term be s_5. Note that $n + s_5 = s$.

We next consider the constant factor associated to (A.96). There are two contributions to this: the number of times such a monomial appears through differentiation, and a factor from the derivatives of powers of P, P', and their conjugates. We bound the first contribution by the total number of monomials produced, which is crudely bounded by $(s+4)^s$, because a derivative of a term of the form (A.96) produces $n+4 \leqslant s+4$ new monomials of the same form, one for each choice of factor to differentiate. For the second, we see that the P-type terms appear with power at most D, and there are n total derivatives applied to them, so this contribution is bounded by D^n. We therefore see that the constant factor is no larger than $(Cr)^{Cr} D^n$.

Now consider bounding each monomial. We will bound the P and P' terms (and their conjugates) by their absolute values. For the Q terms, we will use Theorem 2.2 to bound the G_{ab} terms with no new index, and then invoke (A.12) in the $m = 2$ case. We must further track the constant pre-factors coming from derivatives of Q terms. Each Q term starts as $2N^{-1}\sum_a G_{ia}G_{ak}$ or $4N^{-1}\sum_a G_{ia}G_{ak}$, and each successive derivative multiplies the number of terms by 4. Recall there are s_5 such derivatives. Combining the bound on the number of new terms and the semicircle law bound, we get a C^{s_5} factor, which we bound by C^r.

We first consider the case $r = s$. Set $s_0 = s_1 + s_2$ and $s' = s_3 + s_4$. We recall there are $n = 2s_0 - s' - 2$ new indices. Using power counting (A.12) and noting the isolated off-diagonal G_{ik} term, which is not differentiated, we get

$$|\mathbb{E}I_{r,r}| \leqslant \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \sum_{s_0=2}^{r+1} \sum_{s'=0}^{s_0-1} \left(\frac{C}{N\eta}\right)^{1/2} \left(\frac{CD}{N\eta}\right)^{2s_0-s'-2} \mathbb{E}\left[|P'|^{s'}|P|^{2D-s_0}\right] \tag{A.99}$$

$$\leqslant \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \sum_{s_0=2}^{r+1} \left(\frac{C}{N\eta}\right)^{1/2} \left(\frac{CD}{N\eta}\right)^{s_0-1} \mathbb{E}\left[P|^{2D-s_0}\right] + CN^{-2\log N}, \tag{A.100}$$

where we increased C in the second line.

For $s \neq r$ we get

$$|\mathbb{E}I_{r,s}| \leqslant \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \sum_{s_0=2}^{r+1} \sum_{s'=1}^{s_0-2} \left(\frac{CD}{N\eta}\right)^{2s_0-s'-2} \mathbb{E}\left[|P'|^{s'}|P|^{2D-s_0}\right] \tag{A.101}$$

$$+ \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \sum_{s_0=2}^{r+1} \left(\frac{CD}{N\eta}\right)^{s_0-1} \mathbb{E}\left[|P'|^{s_0-1}|P|^{2D-s_0}\right] \tag{A.102}$$

The second term bounds the terms with $s_0 - 1 = s'$ that come from when ∂_{ik} lands $s_1 - 1$ times on P and s_2 times on $\overline{P}$, and their derivatives are not hit. By Theorem 2.2, we obtain the desired bound

$$|\mathbb{E}I_{r,s}| \leqslant \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \sum_{s_0=2}^{r+1} \left(\frac{CD}{N\eta}\right)^{s_0-1} \mathbb{E}\left[|P|^{2D-s_0}\right] + CN^{-5\log N}. \tag{A.103}$$

□

A.4 Proof of Moment Bound.

Proof of Proposition A.1. We proceed by induction to bound the powers $\mathbb{E}\big[|P|^p\big]$, with P as in (A.26). The base case $p = 0$ is trivial. For the induction step, suppose $D \leqslant (\log N)/2$, and that there exists $C_1(K,\kappa) > 0$ such that

$$\mathbb{E}\left[|P|^p\right] \leqslant \left(\frac{C_1}{N\eta}\right)^p p^{3p/4} \tag{A.104}$$

for all $p \leqslant 2D - 2$. We will show that if C_1 is chosen large enough, in a way that does not depend on D, then (A.104) also holds for $p = 2D$ and $p = 2D - 1$.

Set $\ell = 20\log N$. Combining (A.27), (A.28), Lemma A.9, and Lemma A.10, we have

$$\mathbb{E}\left[|P|^{2D}\right] \leqslant \left(\frac{C}{N\eta}\right) \mathbb{E}\left[|P|^{2D-1}\right] + D\left(\frac{C}{N\eta}\right)^2 \mathbb{E}\left[|P|^{2D-2}\right] \tag{A.105}$$

$$+ C(\log N)N^{-1/2} \sum_{a=1}^{2D} \left(\frac{1}{N\eta}\right)^a \mathbb{E}\left[|P|^{2D-a}\right] \tag{A.106}$$

$$+ \sum_{r=3}^{8\log N} \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \mathbb{E}\left[|P|^{2D-1}\right] + \sum_{r=3}^{8\log N} DC^r \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \frac{C}{N\eta} \mathbb{E}\left[|P|^{2D-2}\right] \tag{A.107}$$

$$+ \sum_{r=2}^{8\log N} \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \sum_{a=2}^{r+1} \left(\frac{C}{N\eta}\right)^{1/2} \left(\frac{CD}{N\eta}\right)^{a-1} \mathbb{E}\left[P|^{2D-a}\right] \tag{A.108}$$

$$+ \sum_{r=3}^{8\log N} \frac{(Cr)^{Cr}}{N^{(r-1)/2}} \sum_{a=2}^{r+1} \left(\frac{CD}{N\eta}\right)^{a-1} \mathbb{E}\left[|P|^{2D-a}\right] + CN^{-2\log N}. \tag{A.109}$$

We now use (A.104) in the above inequality. After increasing C, and choosing $C_1 > C$ in a way that only depends on C, K, and κ, we obtain

$$\mathbb{E}\left[|P|^{2D}\right] \leqslant \left(\frac{C_1}{N\eta}\right)^{2D} (2D)^{3D/2}. \tag{A.110}$$

Further, from Hölder's inequality, we obtain as desired that

$$\mathbb{E}\left[|P|^{2D-1}\right] \leqslant \mathbb{E}\left[|P|^{2D}\right]^{\frac{2D-1}{2D}} = \left(\frac{C_1}{N\eta}\right)^{2D-1} (2D)^{3(2D-1)/4}, \tag{A.111}$$

which completes the induction step.

We now recall the stability of the defining equation $u^2 = zu + 1 = 0$ for $m_{\rm sc}$ (see [14, Lemma 5.5]). Set $\mathcal{A}_z = \{|P(z)| \leqslant 1\}$. Then m satisfies $\mathbb{1}_{\mathcal{A}_z}\big|m(z) - m_{\rm sc}(z)\big| \leqslant C\big|P(z)\big|$ for some $C(\kappa) > 0$. The claimed result follows from (A.104) and the bound $|m(z) - m_{\rm sc}(z)| \leqslant 2\eta^{-1}$ on the set $\mathcal{A}_z^c$, which has negligible probability by Markov's inequality and (A.104). □

B Mesoscopic Fluctuations for β-ensembles

This appendix considers β-ensembles as defined in (1.9). We follow Johansson's loop equations method from [57] to establish Gaussian fluctuations of the characteristic polynomial at any mesoscopic scale larger than $(\log N)^C/N$. The main result is Theorem B.1 below. It is used in Section 3.

Compared to [57], our work presents two novelties. First, [57] considered macroscopic scales, while we prove a result for all mesoscopic scales. Second, [57] considers the Laplace transform, while we give asymptotics of the mixed Fourier–Laplace transform. We note that for the proof on the leading order of the maximum in Section 3, only the Laplace transform is needed, but the Fourier transform may be of independent interest (for example for future finer estimates).

We face the following difficulties in proving these generalizations.

(i) For the Laplace transform, to prove rigidity of the measures (1.9) perturbed on mesoscopic scales, we need precise a priori bounds. Our main tool is the local law with Gaussian tail from [25], as stated in Theorem 2.4.

(ii) For the Fourier transform, the loop equation method requires handling complex measures and the partition function may vanish. Despite this difficulty, asymptotics of characteristic functions were obtained in [22, Appendix A]. We follow the argument developed there, which proceeds through a Gronwall lemma.

B.1 Preliminary facts and notations. We consider the probability density (1.9), with V satisfying the assumptions of Section 1.2, i.e. (A1), (A2) (i), (A3) and (A4) (see Subsection B.6 regarding the Assumption (A2) (ii)). In this section we abbreviate the corresponding probability measure by $\mu = \mu_N$. We recall that the equilibrium density (1.11) is assumed to lie on a single interval $[A, B]$ and defines a function $r(E)$.We will also need the notations

$$\tau(s) = \sqrt{(s-A)(B-s)}, \quad b(z) = \sqrt{z-A}\sqrt{z-B}, \tag{B.1}$$

where we use the principal branch of the square root, extended to negative real numbers by $\sqrt{-x} = \mathrm{i}\sqrt{x}$ for $x > 0$. We will use the formula

$$\int_A^B \frac{\tau(s)}{s-z}\,\mathrm{d}s = \pi\left(\frac{A+B}{2} - z + b(z)\right), \tag{B.2}$$

which is just the usual formula for the Stieltjes transform of the semicircle law from (2.3), up to an affine change of variables. Then the Stieltjes transform m_V from (2.10) satisfies the equation

$$2m_V(z) + V'(z) = 2r(z)b(z) \tag{B.3}$$

for any $z \notin [A, B]$, where we recall that r from (1.11) is assumed to admit an analytic extension to $\mathbb{C}$.

Given a function $g\colon \mathbb{R}\to\mathbb{R}$, we consider the linear statistics

$$S_N(g) := \sum_{k=1}^{N} g(\lambda_k) - N\int g\,\mathrm{d}\mu_V.$$

The functions g considered in this appendix are

$$\mathrm{rlog}_z(\lambda) = \mathrm{Re}\log(z-\lambda),\quad \mathrm{ilog}_z(\lambda) = \mathrm{Im}\log(z-\lambda), \tag{B.4}$$

where Re log and Im log are defined in (2.1). The limiting covariance for these test functions will be written in terms of

$$v(z) = \frac{1}{2}\left(\frac{A+B}{2} - z + b(z)\right),\quad \gamma = \frac{(A-B)^2}{16},\quad c(z,w) = \log\left(1 - \frac{v(z)v(w)}{\gamma}\right),$$

where log is the usual complex logarithm.

The mesoscopic central limit theorem proved in this section will hold on any scale greater than a parameter

$$\eta_0 \in \Big[\frac{(\log N)^{1000}}{N}, \widetilde{\eta}\Big], \tag{B.5}$$

where $\widetilde{\eta}$ is given by Theorem 2.4.

B.2 Mixed Fourier-Laplace transform. Given points $\mathbf{z} = (z_i)_{i=1}^p$ in $\mathbb{C}\setminus\mathbb{R}$, we define the following quadratic form in complex vectors $\boldsymbol{\zeta} = (\zeta_i)_{i=1}^p$, $\boldsymbol{\xi} = (\xi_j)_{j=1}^p$, which will represent an asymptotic covariance:

$$\begin{aligned}\sigma(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z}) = -\frac{1}{2\beta}\sum_{i,j=1}^{p}\big[&(\zeta_j - \mathrm{i}\xi_j)(\zeta_i - \mathrm{i}\xi_i)c(z_j,z_i) + (\zeta_j - \mathrm{i}\xi_j)(\zeta_i + \mathrm{i}\xi_i)c(z_j,\bar z_i)\\ &+ (\zeta_j + \mathrm{i}\xi_j)(\zeta_i - \mathrm{i}\xi_i)c(\bar z_j,z_i) + (\zeta_j + \mathrm{i}\xi_j)(\zeta_i + \mathrm{i}\xi_i)c(\bar z_j,\bar z_i)\big].\end{aligned} \tag{B.6}$$

We also define the following function, which will represent an asymptotic shift:

$$\begin{aligned}\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z}) = &\sum_{j=1}^{p}(\zeta_j - \mathrm{i}\xi_j)\int_{z_j}^{z_j+\mathrm{i}\infty}\left(\frac14 - \frac{1}{2\beta}\right)\left((b'(z)-1) + \int_A^B \frac{r'(s)\tau(s)}{r(s)(s-z)}\frac{\mathrm{d}s}{\pi}\right)\frac{\mathrm{d}z}{b(z)}\\ &- \sum_{j=1}^{p}(\zeta_j + \mathrm{i}\xi_j)\int_{z_j}^{z_j+\mathrm{i}\infty}\left(\frac14 - \frac{1}{2\beta}\right)\left((b'(\bar z)-1) + \int_A^B \frac{r'(s)\tau(s)}{r(s)(s-\bar z)}\frac{\mathrm{d}s}{\pi}\right)\frac{\mathrm{d}z}{b(\bar z)}.\end{aligned} \tag{B.7}$$

Note that μ depends on the external potential V through r, while σ is independent of V.

Theorem B.1. *With the notation (B.4), let*

$$h = \sum_{i=1}^{p}(\zeta_i\,\mathrm{rlog}_{z_i} + \xi_i\,\mathrm{ilog}_{z_i}), \tag{B.8}$$

where $p\geqslant 1$ is fixed. Let $\kappa, M>0$. Then, uniformly in $\mathrm{Re}(\boldsymbol{\zeta},\boldsymbol{\xi})\in[-M,M]^{2p}$, $\mathrm{Im}(\boldsymbol{\zeta},\boldsymbol{\xi})\in\sqrt{\beta}\cdot[-\frac{1}{10p},\frac{1}{10p}]^{2p}$, and $\mathbf{z}\in([A+\kappa,B-\kappa]\times[\eta_0,\widetilde{\eta}])^p$, we have

$$\mathbb{E}_\mu\left[e^{S_N(h)}\right] = e^{\frac{\sigma(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})}{2} + \mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})}\cdot\left(1 + \mathrm{O}_{\kappa,M,p}\left(\frac{1}{\sqrt{N\eta_0}}\right)\right).$$

We now state an elementary lemma about the size of the variance and shift terms occurring in the above central limit theorem.

Lemma B.2. *Fix $M,\kappa>0$ and $p\in\mathbb{N}$. Then uniformly in $|\boldsymbol{\zeta}|,|\boldsymbol{\xi}|\in[-M,M]^p$, $\mathbf{z}\in([A+\kappa,B-\kappa]\times[0,1])^p$ we have $\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z}) = \mathrm{O}(1)$.*

Moreover, uniformly in $z,w\in[A+\kappa,B-\kappa]\times[0,1]$, we have $c(z,w) = \mathrm{O}_\kappa(1)$, while for $z\in[A+\kappa,B-\kappa]\times[0,1]$, and $w\in[A+\kappa,B-\kappa]\times[-1,0]$, we have

$$c(z,w) = \log(z-w) + f(z,w),$$

where f is a continuous function that satisfies $f(z,w) = \mathrm{O}_\kappa(1)$ uniformly for such z,w.

Proof. We start with the asymptotics for (B.7), which is a linear combination of terms of type

$$\int_w^{w+\mathrm{i}\infty} \frac{b'(z)-1}{b(z)}\mathrm{d}z, \tag{B.9}$$

$$\int_w^{w+\mathrm{i}\infty} \left(\int_A^B \frac{r'(s)}{r(s)(s-z)}\mathrm{d}s\right)\cdot \frac{1}{b(z)}\mathrm{d}z, \tag{B.10}$$

with bounded coefficients. Since $\operatorname{Re} w \in [A+\kappa, B-\kappa]$, we have $\int_w^{w+\mathrm{i}} \frac{b'(z)-1}{b(z)}\mathrm{d}z = \mathrm{O}(1)$. Moreover $b'(z)-1 = \mathrm{O}(1/|z|)$ and $b(z)\sim |z|$ as $\operatorname{Im} z \to \infty$, so $\int_{w+\mathrm{i}}^{w+\mathrm{i}\infty} \frac{b'(z)-1}{b(z)}\mathrm{d}z = \mathrm{O}(1)$. We have proved that (B.9) is $\mathrm{O}(1)$ as expected.

For the term (B.10), from our non-vanishing assumption on r it is of order at most

$$\int_w^{w+\mathrm{i}\infty} \left(\int_A^B \frac{\mathrm{d}s}{|s-z|}\right)\cdot \frac{1}{|b(z)|}\mathrm{d}|z| \leqslant C\int_w^{w+\mathrm{i}} \frac{\log \eta_z}{|b(z)|}\mathrm{d}|z| + C\int_{w_j+\mathrm{i}}^{w_j+\mathrm{i}\infty} \frac{1}{|z|\cdot|b(z)|}\mathrm{d}|z|.$$

The first term is $\mathrm{O}(1)$, since we have $|b(z)|\geqslant c$ for some $c>0$ by $\operatorname{Re} w_j \in [A+\kappa, B-\kappa]$. The second term is $\mathrm{O}(1)$ because of the quadratic decay at $\mathrm{i}\infty$.

For the asymptotics of $c(z,w)$, to simplify notations we can assume without loss of generality that $A=-2$ and $B=2$. Note that $v(z) = \frac{-z+\sqrt{z^2-4}}{2}$ conformally maps $\mathbb{C}\cup\{\infty\}\setminus[-2,2]$ into $\mathbb{D}$, so $c(z,w)$ is well-defined. Moreover, the image of $[2+\kappa, 2-\kappa]\times[0,1]$ by v is a subset of $\mathbb{H}\cap\mathbb{D}$ which has positive distance to -1 and 1. Then for any $z,w\in[2+\kappa,2-\kappa]\times[0,1]$ we have $|v(z)v(w)|<1-\varepsilon$ for some fixed $\varepsilon>0$, and hence $c(z,w)=\mathrm{O}(1)$.

For the case $\operatorname{Im} z>0, \operatorname{Im} w<0$, we denote $v(E^+) = \lim_{\eta\to0^+} v(E+\mathrm{i}\eta) = \frac{-E+\sqrt{4-E^2}\mathrm{i}}{2}$ and similarly $v(E^-) = \frac{-E-\sqrt{4-E^2}\mathrm{i}}{2}$. We have $v'(z) = -\frac{v(z)}{b(z)} = -\frac12 + \frac{z}{2\sqrt{z^2-4}}$. We therefore denote $v'(E^+) = -\frac12 - \frac{E}{2\sqrt{4-E^2}}\mathrm{i}$ and $v'(E^-) = -\frac12 + \frac{E}{2\sqrt{4-E^2}}\mathrm{i}$.

Note that $v(E^+)v(E^-)=1$ and $v(E^+)v'(E^-) = -v(E_-)v'(E^+) = \frac{1}{b(E_+)} = -\frac{\mathrm{i}}{\sqrt{4-E^2}}$. This implies, for $\operatorname{Im}\varepsilon_1>0$ and $\operatorname{Im}\varepsilon_2<0$, that $v(E+\varepsilon_1)v(E+\varepsilon_2)$ is equal to

$$\big(v(E^+)+v'(E^+)\varepsilon_1+\mathrm{O}(\varepsilon_1)\big)\cdot\big(v(E^-)+v'(E^-)\varepsilon_2+\mathrm{O}(\varepsilon_2)\big) = 1+\frac{\mathrm{i}}{\sqrt{4-E^2}}(\varepsilon_1-\varepsilon_2)+\mathrm{O}(|\varepsilon_1|^2+|\varepsilon_2|^2),$$

which implies $\log(1-v(z)v(w)) = \log(z-w)+\mathrm{O}(1)$ for $z\in[A+\kappa,B-\kappa]\times[0,1]$, and $w\in[A+\kappa,B-\kappa]\times[-1,0]$. ☐

B.3 Rigidity under biased measures. We start with an important preliminary bound on the Laplace transform, relying on [25].

Lemma B.3. *For any fixed $\kappa,\beta>0$, there exist $N_0(V,\kappa,\beta)\in\mathbb{N}$ and $C(V,\kappa,\beta)>0$ such that the following holds. Let $\widetilde{\eta}$ be given by Theorem 2.4, and fix any $M>0$. For any $z\in\mathbb{C}$ such that $\operatorname{Re} z\in[A+\kappa,B-\kappa]$ and $\operatorname{Im} z\in[N^{-1},\widetilde{\eta}]$, and any $N>N_0$ and $\zeta\in[-M,M]$,*

$$\log \mathbb{E}_\mu\left[e^{\zeta\left(\sum_{k=1}^N \mathrm{rlog}_z(\lambda_k) - N\int \mathrm{rlog}_z\,\mathrm{d}\rho_V\right)}\right] \in [-CM(\log N)^2, CM^2(\log N)^2]. \tag{B.11}$$

Moreover, the same estimate holds when considering ilog *instead of* rlog.

Proof. For the lower bound, by Jensen's inequality and the hypothesis $\zeta\in[-M,M]$ we have

$$\begin{aligned}\log \mathbb{E}_\mu\left[e^{\zeta\left(\sum_{k=1}^N \mathrm{rlog}_z(\lambda_k) - N\int \mathrm{rlog}_z\,\mathrm{d}\rho_V\right)}\right] &\geqslant \zeta\,\mathbb{E}_\mu\left[\sum_{k=1}^N \mathrm{rlog}_z(\lambda_k) - N\int \mathrm{rlog}_z\,\mathrm{d}\rho_V\right]\\ &\geqslant -M\sum_{k=1}^N \mathbb{E}_\mu\Big[\big|\mathrm{rlog}_z(\lambda_k)-\mathrm{rlog}_z(\gamma_k)\big|\Big] - M\left|N\int \mathrm{rlog}_z\,\mathrm{d}\rho_V - \sum_{k=1}^N \mathrm{rlog}_z(\gamma_k)\right|. \end{aligned}\tag{B.12}$$

The first summand above is easily bounded on the good set

$$\mathcal{G} = \bigcap_{\hat{k}>(\log N)^2}\{|\lambda_k-\gamma_k|\leqslant C(\log N)N^{-2/3}\hat{k}^{-1/3}\} \bigcap_{\hat{k}\leqslant(\log N)^2}\{|\lambda_k-\gamma_k|\leqslant C(\log N)^{10}N^{-2/3}\}$$

as follows. For $\hat{k} \leqslant (\log N)^2$, we use that λ_k, γ_k are near A or B, where rlog_z is C_1-Lipschitz continuous for some constant $C_1 > 0$ by the assumptions on $\mathrm{Re}\, z$ and $\mathrm{Im}\, z$. For the other values of $\hat{k}$, we use the mean value theorem. Then setting $E = \mathrm{Re}\, z$ and $\eta = \mathrm{Im}\, z$, and recalling that E is in the bulk of the spectrum, we have

$$\sum_{k=1}^{N} \mathbb{E}_\mu\left[\left|\mathrm{rlog}_z(\lambda_k) - \mathrm{rlog}_z(\gamma_k)\right|\mathbb{1}_{\mathcal{G}}\right]$$
$$\leqslant C_1 \sum_{\hat{k}\leqslant(\log N)^2} |\lambda_k - \gamma_k|\mathbb{1}_{\mathcal{G}} + \sum_{\hat{k}\geqslant(\log N)^2} |\lambda_k - \gamma_k| \max_{|\lambda-\gamma_k|\leqslant C(\log N)N^{-2/3}(\hat{k})^{-1/3}} |\mathrm{rlog}_z'(\lambda)|\mathbb{1}_{\mathcal{G}}$$
$$\leqslant CN^{-1/2} + \sum_{|E-\gamma_k|>(\log N)/N} \frac{C(\log N)}{|E-\gamma_k|} N^{-\frac{2}{3}}(\hat{k})^{-\frac{1}{3}} + \sum_{|E-\gamma_k|<(\log N)/N} \frac{C(\log N)}{\eta} N^{-1} \leqslant C(\log N)^2. \quad \text{(B.13)}$$

Moreover, for large enough N we have

$$\mathbb{P}(\mathcal{G}^{\mathrm{c}}) \leqslant N^{-100}$$

from [25, Corollary 1.5] for $\hat{k} > (\log N)^2$, and [21, Corollary 1.5 and Corollary 1.6] for $\hat{k} \leqslant (\log N)^2$. This implies

$$\sum_{k=1}^{N} \mathbb{E}_\mu\left[\left|\mathrm{rlog}_z(\lambda_k) - \mathrm{rlog}_z(\gamma_k)\right|\mathbb{1}_{\mathcal{G}^c}\right] \leqslant N^{-10}\mathbb{E}[|\mathrm{rlog}_z(\lambda_1)|]^{1/2} \leqslant N^{-2}$$

where we used that the latter expectation is finite, from the estimate $\mathbb{P}(|\lambda_1| > x) \leqslant (x-L)^N$ for some fixed L [26, Equation (3.3)].

For the second, deterministic, term in (B.12), we have by the intermediate value theorem that $N\int \mathrm{rlog}_z\, \mathrm{d}\mu_V - \sum \mathrm{rlog}_z(\gamma_k) = \sum(\mathrm{rlog}_z(\delta_k) - \mathrm{rlog}_z(\gamma_k))$ for some $\delta_k \in [\gamma_{k-1}, \gamma_{k+1}]$. Then the same reasoning as (B.13) applies, giving an analogous bound. This completes the proof of the lower bound in (B.11).

For the upper bound, let η' be a parameter to be fixed later. For all $x \in [\eta, \eta']$, we define

$$g(x) = \sum_{k=1}^{N} \mathrm{rlog}_{E+\mathrm{i}x}(\lambda_i) - N\int \mathrm{rlog}_{E+\mathrm{i}x}\, \mathrm{d}\rho_V, \qquad \delta(x) = N\left|s(E+\mathrm{i}x) - m_V(E+\mathrm{i}x)\right|.$$

From the local law, Theorem 2.4, we have for $\eta' \in [\eta, \widetilde{\eta}]$ that

$$\begin{aligned}
\mathbb{E}\left[e^{\xi\left(g(\eta)-g(\eta')\right)}\right] &\leqslant 2\sum_{p\geqslant 0} \frac{|\xi|^{2p}}{(2p)!}\mathbb{E}\left[\left(\int_\eta^{\eta'} \delta(x)\right)^{2p}\right] \\
&\leqslant 2\sum_{p\geqslant 0} \frac{|\xi|^{2p}}{(2p)!}\int_{[\eta,\eta']^{2p}} \mathbb{E}[\delta(\eta_1)\dots\delta(\eta_{2p})]\mathrm{d}\eta_1\dots\mathrm{d}\eta_{2p} \\
&\leqslant 2\sum_{p\geqslant 0} \frac{|\xi|^{2p}}{(2p)!}\int_{[\eta,\eta']^{2p}} (\mathbb{E}(\delta(\eta_1)^{2p})^{\frac{1}{2p}}\dots(\mathbb{E}(\delta(\eta_{2p})^{2p})^{\frac{1}{2p}})\mathrm{d}\eta_1\dots\mathrm{d}\eta_{2p} \\
&\leqslant 2\sum_{p\geqslant 0} \frac{|\xi|^{2p}}{(2p)!}\left(\int_{[\eta,\eta']}\left(\frac{(Cp)^p}{x^{2p}} + (CN)^{2p}e^{-cN}\right)^{\frac{1}{2p}}\mathrm{d}x\right)^{2p} \\
&\leqslant \sum_{p\geqslant 0}\frac{|2\xi|^{2p}}{(2p)!}\left(\int_{[\eta,\eta']}\frac{(Cp)^{1/2}}{x}\mathrm{d}x\right)^{2p} + \sum_{p\geqslant 0}\frac{|2\xi|^{2p}}{(2p)!}\left(\int_{[\eta,\eta']} CNe^{-c\frac{N}{2p}}\mathrm{d}x\right)^{2p},
\end{aligned} \quad \text{(B.14)}$$

where we successively used Hölder's inequality, the inequality $(a+b)^c \leqslant a^c + b^c$ for $a, b > 0$, $c \in [0,1]$ (with $c = 1/(2p)$ above), and $(x+y)^{2p} \leqslant 2^{2p-1}(x^{2p} + y^{2p})$. The first sum above is of order

$$\sum_{p\geqslant 0} \frac{|C\xi|^{2p}}{p^p}(\log N)^{2p} \leqslant \exp(C|\xi|^2(\log N)^2), \quad \text{(B.15)}$$

while the second one is bounded by

$$\sum_{p\geqslant 0} \frac{|C\xi|^{2p}}{(2p)!}(\eta' N)^{2p}e^{-cN} \leqslant e^{C|\xi|\eta' N - cN}, \quad \text{(B.16)}$$

where on the right-hand sides of (B.15) and (B.16), $C, c > 0$ are constants that depend on V but not on η'. We now fix η' small enough, as a function of M, so that these upper bound in (B.16) is $o(1)$ when $|\xi| < 2M$. In conclusion, we have shown (recalling (B.14)) that there exist constants $C, \eta' > 0$ such that for any $|\xi| < 2M$ we have

$$\mathbb{E}\left[e^{\xi(g(\eta)-g(\eta'))}\right] \leqslant \exp(CM^2(\log N)^2). \tag{B.17}$$

Moreover, as $\xi g(\eta')$ is a smooth linear statistic on a macroscopic scale, uniformly in $|\xi| < 2M$ we have (see e.g. [77, Theorem 1(i)])

$$\mathbb{E}\left[e^{\xi g(\eta')}\right] = \mathrm{O}(1), \tag{B.18}$$

where the implicit constant depends only on the choice of g and V. Equations (B.17) and (B.18) conclude the proof of the upper bound when $\eta < \eta'$, after combining them with the Cauchy–Schwarz inequality to estimate the left side of (B.11). For $\eta \in [\eta', \widetilde{\eta}]$, we can directly use $\mathbb{E}\left[e^{\xi g(\eta)}\right] = \mathrm{O}(1)$, as explained before the previous equation. $\square$

For fixed $\boldsymbol{\zeta}, \boldsymbol{\xi}$, and a function $h\colon \mathbb{R} \to \mathbb{R}$, we will need the following (complex) measures. They are modifications of the measure (1.9) and depend on the parameter $0 \leqslant t \leqslant 1$:

$$\mathrm{d}\mu_h^t(\boldsymbol{\lambda}) := \frac{e^{tS_N(h)}}{Z_h(t)}\mathrm{d}\mu(\boldsymbol{\lambda}),$$

where we assumed

$$Z_h(t) := \mathbb{E}_\mu[e^{tS_N(h)}] \neq 0. \tag{B.19}$$

In the next section we will use rigidity under biased measures under the following form.

Lemma B.4. *For any t such that $Z_h(t) \neq 0$, and any function $f\colon \mathbb{R} \to \mathbb{R}$ and measurable set $\mathscr{G}$,*

$$|\mathbb{E}_{\mu_h^t}[f\mathbb{1}_{\mathscr{G}}]| \leqslant \frac{Z_{\mathrm{Re}\, h}(t)}{|Z_h(t)|}\sup_{\mathscr{G}}|f|. \tag{B.20}$$

Moreover, any integer $1 \leqslant k \leqslant N$, we define $\mathscr{G} = \mathscr{G}_k = \{|\lambda_k - \gamma_k| < N^{-\frac{2}{3}}(\widehat{k})^{-\frac{1}{3}}(\log N)^{100}\}$. Then for all $M, p > 1$ there exists $c(M,p) > 0$ such that for any $(\boldsymbol{\zeta}, \boldsymbol{\xi}) \in \mathbb{C}^{2p}$ such that $\mathrm{Re}(\boldsymbol{\zeta}, \boldsymbol{\xi}) \in [-M, M]^{2p}$, $t \in [0,1]$ such that $Z_h(t) \neq 0$, and $N \geqslant 1$, we have

$$|\mathbb{E}_{\mu_h^t}[f\mathbb{1}_{\mathscr{G}_k^c}]| \leqslant \frac{\mathbb{E}_\mu[|f|^2]^{1/2}}{|Z_h(t)|}\cdot e^{-(\log N)^4}. \tag{B.21}$$

Proof. The first statement follows directly from

$$|\mathbb{E}_{\mu_h^t}[f\mathbb{1}_{\mathscr{G}}]| \leqslant \frac{\int e^{tS_N(\mathrm{Re}\, h)}|f|\mathbb{1}_{\mathscr{G}}\mathrm{d}\mu}{|Z_h(t)|}.$$

For the second statement, we use the Cauchy–Schwarz inequality twice, writing

$$|\mathbb{E}_{\mu_h^t}[f\mathbb{1}_{\mathscr{G}^c}]| \leqslant \frac{\mathbb{E}_\mu[|f|\mathbb{1}_{\mathscr{G}^c}e^{tS_N(\mathrm{Re}\, h)}]}{|Z_h(t)|} \leqslant \frac{\mathbb{E}_\mu[|f|^2]^{1/2}}{|Z_h(t)|}\mathbb{P}_\mu(\mathscr{G}^c)^{1/4}\mathbb{E}_\mu[e^{4tS_N(\mathrm{Re}\, h)}]^{1/4}.$$

One concludes with lemmas B.3 and 2.5. $\square$

B.4 Analysis of the loop equation. To prove Theorem B.1 we start with the identity $\frac{\mathrm{d}}{\mathrm{d}t}\log Z(t) = \mathbb{E}_{\mu_h^t}(S_N(h))$, and therefore want to estimate expectation of general linear statistics for the measure μ_h^t. The starting point for the proof of Theorem B.1 will be the first order loop equation (B.28) below.

Let $\rho_1^{(N,t)}(s)$ be the 1-point correlation function for the measure μ_h^t (see e.g. [44, Definition 2.4]), and let $m_{N,t}(z)$ be the Stieljes transform of $\rho_1^{(N,t)}(s)$. We introduce

$$\varphi(z) = \varphi_{N,t}(z) := m_{N,t}(z) - m(z). \tag{B.22}$$

For $z \in \mathbb{C} \setminus \mathbb{R}$, we define

$$\widetilde{\varphi}(z) := \frac{1}{2\pi b(z)}\left(\frac{2t}{\beta}\int_A^B \frac{h'(s)}{s-z}\tau(s)\,\mathrm{d}s - \left(\frac{2}{\beta}-1\right)\left(\pi(b'(z)-1) + \int_A^B \frac{r'(s)\tau(s)}{r(s)(s-z)}\,\mathrm{d}s\right)\right). \tag{B.23}$$

The main result of this subsection is the following one.

Lemma B.5. *For any $\kappa, M > 0$ and $p \geqslant 1$ the following holds. Uniformly in any $z = E + \mathrm{i}\eta$, with $\eta_0 \leqslant \eta \leqslant N^{10}$ (see (B.5)) and $A + \kappa \leqslant E \leqslant B - \kappa$, $(\boldsymbol{\zeta}, \boldsymbol{\xi}) \in \mathbb{C}^{2p}$ such that $\mathrm{Re}(\boldsymbol{\zeta}, \boldsymbol{\xi}) \in [-M, M]^{2p}$, and t such that $|Z_h(t)| \geqslant e^{-(\log N)^2}$, we have*

$$\varphi(z) = \frac{\widetilde{\varphi}(z)}{N} + \mathrm{O}_{\kappa,M,p}\left(\frac{(\|\boldsymbol{\xi}\|_\infty + \|\boldsymbol{\zeta}\|_\infty)(\log N)^{500}}{N^2\eta_0\eta} \cdot \frac{Z_{\mathrm{Re}\, h}(t)^2}{|Z_h(t)|^2}\right). \tag{B.24}$$

We now prove two lemmas which will be used as input to the proof of Lemma B.5.

Lemma B.6. *For any $\kappa, M > 0$, $p \geqslant 1$, we have uniformly in $\mathrm{Re}(\boldsymbol{\zeta}, \boldsymbol{\xi}) \in [-M, M]^{2p}$ and $z = E + \mathrm{i}\eta$ such that $\eta \neq 0$, $A + \kappa \leqslant E \leqslant B - \kappa$, that*

$$\varphi(z) = \mathrm{O}_{\kappa,M,p}\left(\frac{(\log N)^{200}}{N\eta} \cdot \frac{Z_{\mathrm{Re}\, h}(t)}{|Z_h(t)|}\right), \qquad \mathrm{Var}_{\mu_h^t}(s_N(z)) = \mathrm{O}_{\kappa,M,p}\left(\frac{(\log N)^{400}}{(N\eta)^2} \cdot \frac{Z_{\mathrm{Re}\, h}(t)^2}{|Z_h(t)|^2}\right).$$

Proof. These bounds are immediate consequences of Lemma B.4. Indeed, rewriting (B.22) gives

$$\varphi(z) = \mathbb{E}_{\mu_h^t}\left[\frac{1}{N}\sum_k\left(\frac{1}{\lambda_k - z} - \frac{1}{\gamma_k - z}\right)\right] + \frac{1}{N}\sum_k \frac{1}{\gamma_k - z} - \int \frac{\mathrm{d}\rho_V(\lambda)}{\lambda - z}.$$

The deterministic sum in the second term is easily seen to be $\mathrm{O}((N\eta)^{-1})$ by definition of the γ_k's. Denoting $\mathscr{G}_k = \{|\lambda_k - \gamma_k| < N^{-\frac{2}{3}}(\widehat{k})^{-\frac{1}{3}}(\log N)^{100}\}$, the expectation in the first term is bounded by

$$\mathbb{E}_{\mu_h^t}\left[\frac{1}{N}\sum_k\left|\frac{1}{\lambda_k - z} - \frac{1}{\gamma_k - z}\right|\mathbb{1}_{\cap_k \mathscr{G}_k}\right] + \sum_k \mathbb{E}_{\mu_h^t}\left[\frac{1}{N}\sum_k\left|\frac{1}{\lambda_k - z} - \frac{1}{\gamma_k - z}\right|\mathbb{1}_{\mathscr{G}_k^c}\right]$$
$$\leqslant \frac{(\log N)^{200}}{N\eta} \cdot \frac{Z_{\mathrm{Re}\, h}(t)}{|Z_h(t)|} + \frac{Ne^{-(\log N)^4}}{\eta|Z_h(t)|},$$

where we have used Lemma 2.5, (B.20) and (B.21). The second term is negligible because $Z_{\mathrm{Re}\, h}(t) \geqslant \exp(-CM(\log N)^2)$ from Lemma B.3. This concludes our estimate on φ, after using the elementary estimate $|Z_h(t)| \leqslant Z_{\mathrm{Re}\, h}(t)$.

The bound on $\mathrm{Var}_{\mu_h^t}(s_N(z))$ proceeds similarly, starting with

$$|\mathrm{Var}_{\mu_h^t}(s_N(z))| \leqslant 2\left|\mathbb{E}_{\mu_h^t}[|s_N(z) - m(z)|^2]\right| + 2|\varphi(z)|^2.$$

One applies the same reasoning as before to $\mathbb{E}_{\mu_h^t}[|s_N(z) - m(z)|^2]$ and obtains the bound $\frac{(\log N)^{300}}{(N\eta)^2} \cdot \frac{Z_{\mathrm{Re}\, h}(t)}{|Z_h(t)|}$, so that the bound on $\varphi(z)^2$ dominates. □

Lemma B.7. *For any $M > 0$, $p \geqslant 1$, uniformly in $z = E + \mathrm{i}\eta$ with $\eta \neq 0$, $A \leqslant E \leqslant B$, $A > 0$, $\mathrm{Re}(\boldsymbol{\zeta}, \boldsymbol{\xi}) \in [-M, M]^{2p}$, and uniformly in $h \in \mathscr{C}^2(\mathbb{R})$ we have*

$$\int_{\mathbb{R}} \frac{h'(s)}{s - z}\left(\rho_1^{(N,t)}(s) - \rho_V(s)\right)\mathrm{d}s =$$
$$\frac{(\log N)^{200}}{N} \cdot \frac{Z_{\mathrm{Re}\, h}(t)}{|Z_h(t)|}\,\mathrm{O}_{M,p}\left(\int \frac{|h''(s)|}{|z - s|}\mathrm{d}s + \int \frac{|h'(s)|}{|z - s|^2}\mathrm{d}s + \frac{e^{-(\log N)^2}}{\eta^2}(\|h'\|_\infty + \|h''\|_\infty)\right). \tag{B.25}$$

Proof. We apply Lemma B.4, and the proof is almost the same as Lemma 5.3 in [22], so we omit the details. The only differences are that the rigidity estimate is now known with multiplicative error $(\log N)^{100}$ instead of N^{ξ}, the probability error is now $e^{-(\log N)^2}$ instead of $e^{-N^{\xi}}$. □

Proof of Lemma B.5. We closely follow some steps in the proof of [25, Lemma 4.6], with the difference that we now work under complex measures. We first define

$$\psi(z) \coloneqq \frac{2t}{\beta N}\int_A^B \frac{h'(s)}{s - z}\rho_V(s)\,\mathrm{d}s - \frac{1}{N}\left(\frac{2}{\beta} - 1\right)m_V'(z) - \int_{\mathbb{R}} \frac{V'(s) - V'(z)}{s - z}\left(\rho_1^{(N,t)}(s) - \rho_V(s)\right)\mathrm{d}s \tag{B.26}$$

$$\mathrm{Err}(z) \coloneqq \varphi(z)^2 - \frac{2t}{\beta N}\int_{\mathbb{R}} \frac{h'(s)}{s - z}\left(\rho_1^{(N,t)}(s) - \rho_V(s)\right)\mathrm{d}s + \frac{1}{N}\left(\frac{2}{\beta} - 1\right)\varphi'(z) + \mathrm{Var}_{\mu_h^t}(s_N(z)). \tag{B.27}$$

Then, by the same proof as [25, Equation (4.7)] but with complex measures, we have

$$\big(2m_V(z)+V'(z)\big)\varphi(z)-\psi(z)+\mathrm{Err}(z)=0. \tag{B.28}$$

For the proof, we also need to work under the rigidity event $\mathscr{R}:=\bigcap_{1\leqslant k\leqslant N}\{|\lambda_k-\gamma_k|<(\log N)^{100}N^{-\frac{2}{3}}(\hat{k})^{-\frac{1}{3}}\}$, by introducing the new probability measure

$$\mathrm{d}\mu_h^{t,\mathscr{R}}(\lambda_1,\dots,\lambda_N)=\frac{\mathbb{1}_{\mathscr{R}}}{\mathbb{P}_{\mu_h^t}(\mathscr{R})}\,\mathrm{d}\mu_h^t(\lambda_1,\dots,\lambda_N).$$

Moreover, let $\rho_1^{(N,t,\mathscr{R})}(s)$ be the 1-point function under $\mu_h^{t,\mathscr{R}}$, $\varphi^{\mathscr{R}}(z):=\mathbb{E}_{\mu_h^{t,\mathscr{R}}}[s_N(z)]-m_V(z)$, and $\mathrm{Err}^{\mathscr{R}}(z)$ be defined as $\mathrm{Err}(z)$ but with $\mu_h^{t,\mathscr{R}}$, $\rho_1^{(N,t,\mathscr{R})}(s)$ and $\varphi^{\mathscr{R}}(z)$ instead of μ_h^t, $\rho_1^{(N,t)}(s)$ and $\varphi(z)$. Note that

$$\mathbb{P}_{\mu_h^t}(\mathscr{R})=1+\mathrm{O}\left(N|Z_h(t)|^{-1}e^{-(\log N)^4}\right)=1+\mathrm{O}(e^{-(\log N)^4/2}) \tag{B.29}$$

by (B.21) and our assumption on $Z_h(t)$. This easily implies that lemmas B.6 and B.7 still hold under $\mu_h^{t,\mathscr{R}}$, giving

$$\begin{aligned}
&\varphi^{\mathscr{R}}(z)=\mathrm{O}\left(\frac{(\log N)^{200}}{N\eta}\cdot\frac{Z_{\mathrm{Re}\,h}(t)}{|Z_h(t)|}\right),\qquad \mathrm{Var}_{\mu_h^{t,\mathscr{R}}}(s_N(z))=\mathrm{O}\left(\frac{(\log N)^{400}}{(N\eta)^2}\cdot\frac{Z_{\mathrm{Re}\,h}(t)^2}{|Z_h(t)|^2}\right),\\
&\int_{\mathbb{R}}\frac{h'(s)}{s-z}\left(\rho_1^{(N,t,\mathscr{R})}(s)-\rho_V(s)\right)\mathrm{d}s\\
&\quad=\frac{(\log N)^{200}}{N}\cdot\frac{Z_{\mathrm{Re}\,h}(t)}{|Z_h(t)|}\,\mathrm{O}\left(\int\frac{|h''(s)|}{|z-s|}\mathrm{d}s+\int\frac{|h'(s)|}{|z-s|^2}\mathrm{d}s+\frac{e^{-(\log N)^2}}{\eta^2}\left(\|h'\|_\infty+\|h''\|_\infty\right)\right).
\end{aligned} \tag{B.30}$$

Fix some $z=E+\mathrm{i}\eta$ with $\eta_0\leqslant\eta$ and $A-\kappa\leqslant E\leqslant B+\kappa$. We also assume $\eta<\kappa$ first. We consider the rectangle with vertices $A-\kappa\pm\mathrm{i}e^{-(\log N)^3}$, $B+\kappa\pm\mathrm{i}e^{-(\log N)^3}$, and denote by $\mathcal{C}$ the corresponding closed contour with positive orientation. We decompose this contour into $\mathcal{C}_{\mathrm{hor}}$, which consists only in the horizontal pieces, and $\mathcal{C}_{\mathrm{ver}}$, which consists only in the vertical pieces. By the loop equation (B.28) and (B.3), we have

$$\int_{\mathcal{C}_{\mathrm{hor}}}\frac{2r(w)b(w)\varphi(w)-\psi(w)+\mathrm{Err}(w)}{r(w)(z-w)}\,\mathrm{d}w=0.$$

Using (B.29), the hypothesis that $|Z_h(t)|>e^{-(\log N)^2}$, and (B.21), on $\mathcal{C}_{\mathrm{hor}}$ we have

$$\varphi^{\mathscr{R}}(w)=\varphi(w)+\mathrm{O}\left(e^{-(\log N)^4/5}\right), \tag{B.31}$$

and similarly $\mathrm{Err}^{\mathscr{R}}(w)=\mathrm{Err}(w)+\mathrm{O}\left(e^{-(\log N)^4/5}\right)$. Together with the facts that $|z-w|>\eta/2$ and $c\leqslant r,b\leqslant C$ on $\mathcal{C}_{\mathrm{hor}}$ (remember that r is continuous and has no zero on $[A,B]$) this implies

$$\int_{\mathcal{C}_{\mathrm{hor}}}\frac{2r(w)b(w)\varphi^{\mathscr{R}}(w)-\psi(w)+\mathrm{Err}^{\mathscr{R}}(w)}{r(w)(z-w)}\,\mathrm{d}w=\mathrm{O}\left(e^{-(\log N)^4/10}\right). \tag{B.32}$$

On the other hand, for w on $\mathcal{C}_{\mathrm{ver}}$, we have $2r(w)b(w)\varphi^{\mathscr{R}}(w)-\psi(w)+\mathrm{Err}^{\mathscr{R}}(w)=\mathrm{O}(e^{(\log N)^{5/2}})$, an easy estimate based on the following facts: (1) by the definition of, $\mu_h^{t,\mathscr{R}}$, all particles are at a distance larger than $\kappa/2$ from $\mathcal{C}_{\mathrm{ver}}$), (2) $c\leqslant r,b\leqslant C$ on $\mathcal{C}_{\mathrm{ver}}$, (3) $|Z_h(t)|>e^{-(\log N)^2}$ by assumption, (4) $Z_{\mathrm{Re}h}(t)\leqslant e^{CM(\log N)^2}$ by Lemma B.3. This implies

$$\int_{\mathcal{C}_{\mathrm{ver}}}\frac{2r(w)b(w)\varphi^{\mathscr{R}}(w)-\psi(w)+\mathrm{Err}^{\mathscr{R}}(w)}{r(w)(z-w)}\,\mathrm{d}w=\mathrm{O}\left(e^{-(\log N)^3/2}\right). \tag{B.33}$$

Combining (B.32) and (B.33), we get

$$\int_{\mathcal{C}}\frac{2r(w)b(w)\varphi^{\mathscr{R}}(w)-\psi(w)+\mathrm{Err}^{\mathscr{R}}(w)}{r(w)(z-w)}\,\mathrm{d}w=\mathrm{O}\left(e^{-(\log N)^3/2}\right). \tag{B.34}$$

We now estimate each term in the integral in (B.34) successively.

We start with the part involving $\varphi^{\mathscr{R}}(w)$. The function $w \mapsto 2b(w)\varphi^{\mathscr{R}}(w)/(z-w)$ is analytic on and outside $\mathcal{C}$, except for the pole at z, and it behaves as $\mathrm{O}(w^{-2})$ as $|w| \to \infty$ because $b(w) = \mathrm{O}(w)$ and $\varphi^{\mathscr{R}}(w) = \mathrm{O}(w^{-2})$. Therefore, by the Cauchy integral formula with residue at infinity, we get

$$\int_{\mathcal{C}} \frac{2b(w)\varphi^{\mathscr{R}}(w)}{(z-w)}\,\mathrm{d}w = 4\mathrm{i}\pi b(z)\varphi^{\mathscr{R}}(z). \tag{B.35}$$

Now we evaluate the part involving $\psi(w)$. Recall the definition of $\psi(w)$ in (B.26) and note that the third term there is analytic in $w \in \mathbb{C}$. Moreover from (B.3) we have $m_V'(w) = -\frac{1}{2}V''(w) + (rb)'(w)$, where $V''(w)$ is also analytic in $w \in \mathbb{C}$. Since z is exterior to $\mathcal{C}$ and r has no zero inside $\mathcal{C}$ for κ chosen small enough, these analytic terms disappear in (B.34) and we get

$$\begin{aligned}\int_{\mathcal{C}} \frac{\psi(w)}{r(w)(z-w)}\,\mathrm{d}w &= \int_{\mathcal{C}} \left(\frac{2t}{\beta N}\int_A^B \frac{h'(s)}{s-w}\rho_V(s)\,\mathrm{d}s - \frac{1}{N}\left(\frac{2}{\beta}-1\right)(rb)'(w)\right)\frac{\mathrm{d}w}{r(w)(z-w)} \\ &= -\frac{4\pi t}{\beta N}\int_A^B \frac{h'(s)}{r(s)(z-s)}\rho_V(s)\,\mathrm{d}s - \frac{1}{N}\left(\frac{2}{\beta}-1\right)\int_A^B \frac{-2\mathrm{i}(r\tau)'(s)}{r(s)(z-s)}\,\mathrm{d}s,\end{aligned} \tag{B.36}$$

where, for the first term, we applied Cauchy's integral formula and, for the second term, we let the contour approach the segment $[A, B]$ and used $\lim_{y\to 0+}(rb)'(x \pm \mathrm{i}y) = \pm\mathrm{i}(r\tau)'(x)$ for all $x \in (A, B)$, recalling $\tau(x) = \sqrt{(x-A)(B-x)}$. Recalling the definition of $\widetilde{\varphi}(z)$ and that $\rho_V = \frac{1}{\pi}r\tau$ (recall (1.11) and (B.1)), we get

$$\int_{\mathcal{C}} \frac{\psi(w)\,\mathrm{d}w}{r(w)(z-w)} = -2\left(\frac{2t}{\beta N}\int_A^B \frac{h'(s)}{s-z}\tau(s)\,\mathrm{d}s + \frac{\mathrm{i}}{N}\left(\frac{2}{\beta}-1\right)\int_A^B \left(\frac{\tau'(s)}{s-z} + \frac{r'(s)\tau(s)}{r(s)(s-z)}\right)\mathrm{d}s\right) = \frac{4\mathrm{i}\pi b(z)}{N}\widetilde{\varphi}(z), \tag{B.37}$$

where we used that $\int_A^B \frac{\tau'(s)}{s-z}\,\mathrm{d}s = \int_A^B \frac{\tau(s)}{(s-z)^2}\,\mathrm{d}s = \pi(b'(z)-1)$, which follows from differentiating the identity (B.2).

Finally, we deal with the part involving $\mathrm{Err}^{\mathscr{R}}(w)$. We deform the contour $\mathcal{C}$ into $\mathcal{C}'$, the positively oriented rectangle with vertices $A-\kappa \pm \mathrm{i}\eta/2$, $B+\kappa \pm \mathrm{i}\eta/2$, which does not change the value of the integral because the deformation does not cross any poles. The function $w \mapsto \mathrm{Err}^{\mathscr{R}}(w)/(r(w)(z-w))$ is analytic on and between these contours (remember we assume $\eta < \kappa$ first, and κ is fixed and small enough), so

$$\begin{aligned}\int_{\mathcal{C}} \frac{\mathrm{Err}^{\mathscr{R}}(w)}{r(w)(z-w)}\,\mathrm{d}w = \int_{\mathcal{C}'} \frac{\mathrm{Err}^{\mathscr{R}}(w)}{r(w)(z-w)}\,\mathrm{d}w &= \mathrm{O}\left(\frac{(\log N)^{400}}{(N\eta)^2}\cdot\frac{Z_{\mathrm{Re}\,h}(t)^2}{|Z_h(t)|^2}\right) \\ &+ \mathrm{O}\left(\frac{(\log N)^{200}}{N^2}\sum_{\ell=1}^m \frac{1}{\eta_\ell(\eta_\ell \vee |z-z_\ell|)}\cdot\frac{Z_{\mathrm{Re}\,h}(t)}{|Z_h(t)|}\right),\end{aligned} \tag{B.38}$$

where we used that $|r(w)|$ is uniformly lower bounded and we applied (B.30) on the horizontal pieces of $\mathcal{C}'$, and on the vertical pieces we used that the same estimates hold substituting η with κ. Coming back to (B.34) and combining (B.35), (B.37) and (B.38), for $\eta \leqslant \kappa$ we have proved that

$$\varphi^{\mathscr{R}}(z) = \frac{\widetilde{\varphi}(z)}{N} + \mathrm{O}\left(\frac{(\log N)^{400}}{N^2\eta_0\eta}\cdot\frac{Z_{\mathrm{Re}\,h}(t)^2}{|Z_h(t)|^2}\right). \tag{B.39}$$

For $\eta \geqslant \kappa$, the result follows from the Cauchy formula $\frac{1}{s-z} = \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}''}\frac{\mathrm{d}w}{(w-z)(w-s)}$ where $\mathcal{C}''$ is the rectangle with vertices $A-\kappa \pm \mathrm{i}\kappa/10$ and $B+\kappa \pm \mathrm{i}\kappa/10$. More precisely we use (B.39) in the portion $|\mathrm{Im} z| > \eta_0$ of $\mathcal{C}''$ and the estimates from (B.30) with η replaced by 1 on the part $|\mathrm{Im} z| \leqslant \eta_0$.

Finally, the result follows from (B.39) and (B.31). □

B.5 Proof of Theorem B.1. Let $\hat{Z}_j = z_j + \mathrm{i}N^9$. With Z_h as in (B.19), we can write

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\log Z_h(t) = \mathbb{E}^{\mu_h^t}\left[\sum_{i=1}^p (\zeta_i \,\mathrm{rlog}_{z_i} + \xi_i \,\mathrm{ilog}_{z_i})\right] \\
&\quad = N\sum_{j=1}^p \frac{\zeta_j - \mathrm{i}\xi_j}{2}\int_{z_j}^{\hat{Z}_j}\varphi(z)\mathrm{d}z - N\sum_{j=1}^p \frac{\zeta_j + \mathrm{i}\xi_j}{2}\int_{z_j}^{\hat{Z}_j}\varphi(\bar{z})\mathrm{d}z + \mathbb{E}^{\mu_h^t}\left[\sum_{j=1}^p (\zeta_j \,\mathrm{rlog}_{\hat{Z}_j} + \xi_j \,\mathrm{ilog}_{\hat{Z}_j})\right], \quad \text{(B.40)}
\end{aligned}$$

where we have used $\mathrm{d}z = \mathrm{i}\,\mathrm{d}y$ and

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}y}\log\left|(x+\mathrm{i}y) - \lambda\right| &= -\frac{1}{2\mathrm{i}}\left(\frac{1}{\lambda - (x+\mathrm{i}y)} - \frac{1}{\lambda - (x-\mathrm{i}y)}\right), \\
\frac{\mathrm{d}}{\mathrm{d}y}\operatorname{Im}\log\left((x+\mathrm{i}y) - \lambda\right) &= -\frac{1}{2}\left(\frac{1}{\lambda - (x+\mathrm{i}y)} + \frac{1}{\lambda - (x-\mathrm{i}y)}\right).
\end{aligned}$$

To bound the last term in (B.40), we note that the inequality $\log(1+\varepsilon) = \mathrm{O}(\varepsilon)$ for $|\varepsilon| < 1/2$ gives

$$\begin{aligned}
&\mathbb{E}^{\mu_h^t}\left[\sum_i \log(\hat{Z}_j - \lambda_i) - N\int \log(\hat{Z}_j - \lambda)\mathrm{d}\rho_V(\lambda)\right] \\
&\qquad = \mathbb{E}^{\mu_h^t}\left[\sum_i \log(1 - \lambda_i/\hat{Z}_j)\right] - N\int \log(1 - \lambda/\hat{Z}_j)\mathrm{d}\rho_V(\lambda) = \mathrm{O}(N^{-7}).
\end{aligned}$$

We now insert the asymptotics from Lemma B.5 in (B.40). The error term in (B.24) contributes

$$\frac{Z_{\operatorname{Re} h}(t)^2}{|Z_h(t)|^2}\sum_{j=1}^p \frac{|\zeta_j| + |\xi_j|}{2}\int_{z_j}^{\hat{Z}_j}\frac{(\|\boldsymbol{\xi}\|_\infty + \|\boldsymbol{\zeta}\|_\infty)(\log N)^{500}}{N\eta_0\eta_z}|\mathrm{d}z| \leqslant C\frac{Z_{\operatorname{Re} h}(t)^2}{|Z_h(t)|^2}\frac{(\|\boldsymbol{\xi}\|_\infty^2 + \|\boldsymbol{\zeta}\|_\infty^2)(\log N)^{600}}{N\eta_0}.$$

Moreover, one easily checks that $\int_{\hat{Z}_j}^{\hat{Z}_j + \mathrm{i}\infty}\widetilde{\varphi}(z)\,\mathrm{d}z = \mathrm{O}(N^{-6})$, so that denoting

$$p(z) = \frac{1}{b(z)}\left(\frac{1}{4} - \frac{1}{2\beta}\right)\left((b'(z) - 1) + \int_A^B \frac{r'(s)\tau(s)}{r(s)(s-z)}\frac{\mathrm{d}s}{\pi}\right)$$

we have proved that, as long as $|Z_h(t)| \geqslant e^{-(\log N)^2}$ (necessary to apply Lemma B.5) we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\log Z_h(t) = &\sum_{j=1}^p \frac{\zeta_j - \mathrm{i}\xi_j}{2}\int_{z_j}^{z_j + \mathrm{i}\infty}\frac{t}{\pi\beta b(z)}\int_A^B \frac{h'(s)}{s-z}\tau(s)\,\mathrm{d}s\mathrm{d}z \\
&- \sum_{j=1}^p \frac{\zeta_j + \mathrm{i}\xi_j}{2}\int_{z_j}^{z_j + \mathrm{i}\infty}\frac{t}{\pi\beta b(\bar{z})}\int_A^B \frac{h'(s)}{s-\bar{z}}\tau(s)\,\mathrm{d}s\mathrm{d}z \\
&+ \sum_{j=1}^p (\zeta_j - \mathrm{i}\xi_j)\int_{z_j}^{z_j + \mathrm{i}\infty} p(z)\mathrm{d}z - \sum_{j=1}^p (\zeta_j + \mathrm{i}\xi_j)\int_{z_j}^{z_j + \mathrm{i}\infty} p(\bar{z})\mathrm{d}z \\
&+ \mathrm{O}\left(\frac{Z_{\operatorname{Re} h}(t)^2}{|Z_h(t)|^2}\frac{(\|\boldsymbol{\xi}\|_\infty^2 + \|\boldsymbol{\zeta}\|_\infty^2)(\log N)^{600}}{N\eta_0}\right). \qquad \text{(B.41)}
\end{aligned}$$

While the third line above cannot be simplified for general V, for our particular choice

$$h = \sum_{i=1}^p (\zeta_i \,\mathrm{rlog}_{z_i} + \xi_i \,\mathrm{ilog}_{z_i}),$$

the first and second lines can. Indeed,

$$h'(s) = \frac{1}{2}\sum_i \left(\frac{\zeta_i - \mathrm{i}\xi_i}{s - z_i} + \frac{\zeta_i + \mathrm{i}\xi_i}{s - \bar{z}_i}\right),$$

so that

$$\int_A^B \frac{h'(s)}{s-z}\tau(s)\mathrm{d}s = \frac{1}{2}\sum_i \int_A^B \left(\frac{\zeta_i - \mathrm{i}\xi_i}{z-z_i}\left(\frac{1}{s-z}-\frac{1}{s-z_i}\right)+\frac{\zeta_i+\mathrm{i}\xi_i}{z-\bar{z}_i}\left(\frac{1}{s-z}-\frac{1}{s-\bar{z}_i}\right)\right)\tau(s)\mathrm{d}s.$$

From (B.2) and the definition of v, we can write

$$\frac{1}{2\pi}\int_A^B \frac{h'(s)}{s-z}\tau(s)\mathrm{d}s = \frac{1}{2}\sum_i\left(\frac{(\zeta_i-\mathrm{i}\xi_i)(v(z)-v(z_i))}{z-z_i}+\frac{(\zeta_i+\mathrm{i}\xi_i)(v(z)-v(\bar{z}_i))}{z-\bar{z}_i}\right).$$

Note that $v'(z) = -v(z)/b(z)$ and $v^2 + (z - \frac{A+B}{2})v + \frac{1}{4}\left(\frac{A-B}{2}\right)^2 = 0$, so that, abbreviating $v_i = v(z_i)$ and $\gamma = ((A-B)/2)^2/4$ we obtain

$$\int_w^{w+\mathrm{i}\infty}\frac{v(z)-v(z_i)}{z-z_i}\frac{\mathrm{d}z}{b(z)} = \int_{v(w)}^0 \frac{v-v_i}{v+\frac{\gamma}{v}-(v_i+\frac{\gamma}{v_i})}\frac{\mathrm{d}v}{v} = \int_{v(w)}^0\frac{1}{v-\frac{\gamma}{v_i}}\mathrm{d}v = -\log(1-\frac{v(w)v(z_i)}{\gamma})$$

and similarly

$$\int_w^{w+\mathrm{i}\infty}\frac{v(\bar{z})-v(z_i)}{\bar{z}-z_i}\frac{\mathrm{d}z}{b(\bar{z})} = -\log\left(1-\frac{v(\bar{w})v(z_i)}{\gamma}\right).$$

We have therefore proved, using the notations (B.6) and (B.7),

$$\frac{\mathrm{d}}{\mathrm{d}t}\log Z_h(t) = t\,\sigma(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})+\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})+\mathrm{O}\left(\frac{Z_{\mathrm{Re}\,h}(t)^2}{|Z_h(t)|^2}\frac{(\log N)^{600}}{N\eta_0}\right)$$

where here and below, we abbreviate $\mathrm{O} = \mathrm{O}_{M,p,\kappa}$. From this equation we first conclude about the case of real-valued h. Then trivially $Z_{\mathrm{Re}\,h}(t) = Z_h(t)$ so integrating the above equation gives

$$Z_h(t) = \exp\left(\frac{t^2}{2}\sigma(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})+t\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})+\mathrm{O}\left(\frac{(\log N)^{600}}{N\eta_0}\right)\right).$$

From our assumption on η_0 the above error term is O(1). For the general complex case, we now have

$$\frac{\mathrm{d}}{\mathrm{d}t}\log Z_h(t) = t\,\sigma(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})+\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})+\mathrm{O}\left(\frac{(\log N)^{600}}{N\eta_0}\cdot\frac{e^{\frac{t^2}{2}\sigma(\mathrm{Re}\,\boldsymbol{\zeta},\mathrm{Re}\,\boldsymbol{\xi},\mathbf{z})}}{|Z_h(t)|}\right), \tag{B.42}$$

so that, taking the real part above, we have

$$\frac{\frac{\mathrm{d}}{\mathrm{d}t}|Z_h(t)|}{|Z_h(t)|} = t\,\sigma(\mathrm{Re}\,\boldsymbol{\zeta},\mathrm{Re}\,\boldsymbol{\xi},\mathbf{z})-t\,\sigma(\mathrm{Im}\,\boldsymbol{\zeta},\mathrm{Im}\,\boldsymbol{\xi},\mathbf{z})+\mathrm{Re}\,\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})+\mathrm{O}\left(\frac{(\log N)^{600}}{N\eta_0}\cdot\frac{e^{t^2\sigma(\mathrm{Re}\,\boldsymbol{\zeta},\mathrm{Re}\,\boldsymbol{\xi},\mathbf{z})}}{|Z_h(t)|^2}\right).$$

Defining $g(t) = |Z_h(t)|^2 e^{t^2\sigma(\mathrm{Im}\,\boldsymbol{\zeta},\mathrm{Im}\,\boldsymbol{\xi},\mathbf{z})-t^2\,\sigma(\mathrm{Re}\,\boldsymbol{\zeta},\mathrm{Re}\,\boldsymbol{\xi},\mathbf{z})-t\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})}$, the above equation implies

$$g'(t) = \mathrm{O}\left(\frac{(\log N)^{600}}{N\eta_0}\cdot e^{t^2\sigma(\mathrm{Im}\,\boldsymbol{\zeta},\mathrm{Im}\,\boldsymbol{\xi},\mathbf{z})}\right).$$

From the assumption $\mathrm{Im}(\boldsymbol{\zeta},\boldsymbol{\xi}) \in \sqrt{\beta}\cdot[-\frac{1}{10p},\frac{1}{10p}]^{2p}$, we have $\sigma(\mathrm{Im}\,\boldsymbol{\zeta},\mathrm{Im}\,\boldsymbol{\xi},\mathbf{z}) \leqslant \frac{1}{5}\log(N\eta_0)$, so that $g'(t) = \mathrm{O}\left(\frac{1}{\sqrt{N\eta_0}}\right)$ and we have proved that

$$|Z_h(t)| = e^{-\frac{t^2}{2}\sigma(\mathrm{Im}\,\boldsymbol{\zeta},\mathrm{Im}\,\boldsymbol{\xi},\mathbf{z})+\frac{t^2}{2}\,\sigma(\mathrm{Re}\,\boldsymbol{\zeta},\mathrm{Re}\,\boldsymbol{\xi},\mathbf{z})+t\,\mathrm{Re}\,\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})}\cdot\left(1+\frac{1}{\sqrt{N\eta_0}}\right).$$

Inserting this estimate in (B.42) finally gives

$$Z_h(t) = e^{\frac{t^2}{2}\,\sigma(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})+t\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})}\cdot\left(1+\frac{1}{\sqrt{N\eta_0}}\right),$$

We note that all equations since (B.41) hold only provided that $|Z_h(s)| \geqslant e^{-(\log N)^2}$ for $s \in [0,t]$, which is necessary to apply Lemma B.5. Therefore, denoting $t_0 = \max\{t \in [0,1] : Z_h(t) > e^{-(\log N)^2}\}$, for large enough N we have

$$Z_h(t_0) = e^{\frac{t_0^2}{2}\,\sigma(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})+t_0\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z})}\cdot(1+\frac{1}{\sqrt{N\eta_0}}) > e^{-(\log N)^2},$$

where we have used in the above inequality the easy estimates $\sigma(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z}) = \mathrm{O}(\log N)$ and $\mu(\boldsymbol{\zeta},\boldsymbol{\xi},\mathbf{z}) = \mathrm{O}(1)$. By continuity this implies $t_0 = 1$. The expected result therefore holds by taking $t = 1$.

B.6 Generalization to further potentials. The proof of the local law Theorem 2.4 is the only place requiring the sub-quadratic growth assumption from (1.10). Theorem 1.9 also holds for V growing at least linearly, as in Assumption (A2) (ii) through the following steps.

(i) Denoting $\mathbb{E}^{[A-\delta,B+\delta]}$ for the expectation conditional on all particles remaining in $[A-\delta,B+\delta]$, by [25, Equations (2.25), (2.14)], the following local law holds:

$$\mathbb{E}^{[A-\delta,B+\delta]}\left[|s_N(z)-m_V(z)|^{2q}\right]\leqslant\frac{(Cq)^q}{(N\eta)^{2q}}+\frac{C^qe^{-cN}}{|z-A|^q|z-B|^q}.$$

When compared to [25, Theorem 1.1], note the exponentially small second error term, possible thanks to working under the conditioned measure. This improvement is essential to the proof of Theorem 1.9.

(ii) Based on this local law for the conditioned measure, an analogue of the previous quantitative central limit theorem, Theorem B.1, can be proved under $\mathbb{E}^{[A-\delta,B+\delta]}$ for a function $\widetilde{L}_N$ coinciding with L_N on $[A-\delta/2,B+\delta/2]$ but compactly supported on $[A-\delta,B+\delta]$. This gives Theorem 1.9 for $\widetilde{L}_N$, and then for L_N as the probability of a particle outside $[A-\delta/2,B+\delta/2]$ is o(1) by rigidity.

References

[1] S. Albeverio, L. Pastur, and M. Shcherbina, *On the 1/n expansion for some unitary invariant ensembles of random matrices*, Communications in Mathematical Physics **224** (2001), no. 1, 271–305.

[2] G. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010.

[3] L.-P. Arguin, D. Belius, and P. Bourgade, *Maximum of the characteristic polynomial of random unitary matrices*, Comm. Math. Phys. **349** (2017), no. 2, 703–751.

[4] L.-P. Arguin, D. Belius, P. Bourgade, M. Radziwiłł, and K. Soundararajan, *Maximum of the Riemann zeta function on a short interval of the critical line*, Comm. Pure Appl. Math. **72** (2019), no. 3, 500–535.

[5] L.-P. Arguin, P. Bourgade, and M. Radziwiłł, *The Fyodorov-Hiary-Keating Conjecture. I*, arXiv preprint arXiv:2007.00988 (2020).

[6] ———, *The Fyodorov-Hiary-Keating Conjecture. II*, arXiv preprint arXiv:2307.00982 (2023).

[7] F. Augeri, R. Butez, and O. Zeitouni, *A CLT for the characteristic polynomial of random Jacobi matrices, and the GβE*, Prob. Theory Rel. Fields **186** (2023), 1–89.

[8] E. Bailey and J.P. Keating, *Maxima of log-correlated fields: some recent developments*, Journal of Physics A: Mathematical and Theoretical **55** (2022), no. 5, 053001.

[9] R. Bauerschmidt and M. Hofstetter, *Maximum and coupling of the sine-Gordon field*, Ann. Probab. **50** (2022), no. 2, 455–508.

[10] D. Belius and N. Kistler, *The subleading order of two dimensional cover times*, Probab. Theory Related Fields **167** (2017), no. 1-2, 461–552.

[11] D. Belius, J. Rosen, and O. Zeitouni, *Tightness for the cover time of the two dimensional sphere*, Probab. Theory Related Fields **176** (2020), no. 3-4, 1357–1437.

[12] D. Belius and W. Wu, *Maximum of the Ginzburg-Landau fields*, Ann. Probab. **48** (2020), no. 6, 2647–2679.

[13] G. Ben Arous and P. Bourgade, *Extreme gaps between eigenvalues of random matrices*, Ann. Probab. **41** (2013), no. 4, 2648–2681.

[14] F. Benaych-Georges and A. Knowles, *Local semicircle law for Wigner matrices*, Panor. Synthèses **53** (2017).

[15] L. Benigni and P. Lopatto, *Optimal delocalization for generalized Wigner matrices*, Adv. Math. **396** (2022).

[16] ———, *Fluctuations in local quantum unique ergodicity for generalized Wigner matrices*, Comm. Math. Phys. **391** (2022), no. 2, 401–454.

[17] N. Berestycki, C. Webb, and M. D. Wong, *Random Hermitian matrices and Gaussian multiplicative chaos*, Probab. Theory Related Fields **172** (2018), no. 1-2, 103–189.

[18] M. Biskup, *Extrema of the two-dimensional discrete Gaussian free field*, Springer Proc. Math. Stat., vol. 304, Springer, Cham, 2020.

[19] E. Bolthausen, J.-D. Deuschel, and G. Giacomin, *Entropic repulsion and the maximum of the two-dimensional harmonic crystal*, Ann. Probab. **29** (2001), no. 4, 1670–1692.

[20] P. Bourgade, *Extreme gaps between eigenvalues of Wigner matrices*, J. Eur. Math. Soc. **8** (2022), 2823–2873.

[21] P. Bourgade, L. Erdős, and H.-T. Yau, *Edge universality for beta ensembles*, Communications in Mathematical Physics **332** (2014), no. 1, 261–353.

[22] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin, *Fixed energy universality for generalized Wigner matrices*, Comm. Pure Appl. Math. **69** (2016), no. 10, 1815–1881.

[23] P. Bourgade and H. Falconet, *Liouville quantum gravity from random matrix dynamics*, arXiv preprint arXiv:2206.03029 (2022).

[24] P. Bourgade and K. Mody, *Gaussian fluctuations of the determinant of Wigner matrices*, Electronic Journal of Probability **24** (2019), 1–28.

[25] P. Bourgade, K. Mody, and M. Pain, *Optimal local law and central limit theorem for β-ensembles*, Communications in Mathematical Physics **390** (2022), no. 3, 1017–1079.

[26] A. Boutet De Monvel, L. Pastur, and M. Shcherbina, *On the statistical mechanics approach in the random matrix theory: integrated density of states*, Journal of Statistical Physics **79** (1995), no. 3, 585–611.

[27] M. Bramson, *Maximal displacement of branching Brownian motion*, Comm. Pure Appl. Math. **31** (1978), no. 5, 531–581.

[28] ______, *Convergence of solutions of the Kolmogorov equation to travelling waves*, Mem. Amer. Math. Soc. **44** (1983), no. 285, iv+190.

[29] M. Bramson, J. Ding, and O. Zeitouni, *Convergence in law of the maximum of the two-dimensional discrete Gaussian free field*, Comm. Pure Appl. Math. **69** (2016), no. 1, 62–123.

[30] ______, *Convergence in law of the maximum of nonlattice branching random walk*, Ann. Inst. Henri Poincaré Probab. Stat. **52** (2016), no. 4, 1897–1924.

[31] C. Cacciapuoti, A. Maltsev, and B. Schlein, *Bounds for the Stieltjes transform and the density of states of Wigner matrices*, Probability Theory and Related Fields **163** (2015), no. 1, 1–59.

[32] F. Caravenna, R. Sun, and N. Zygouras, *The two-dimensional KPZ equation in the entire subcritical regime*, Ann. Probab. **48** (2020), no. 3, 1086–1127.

[33] C. Charlier, B. Fahs, C. Webb, and M. D. Wong, *Asymptotics of Hankel determinants with a multi-cut regular potential and Fisher-Hartwig singularities*, to appear in Mem. Amer. Math. Soc. (2021).

[34] R. Chhaibi, T. Madaule, and J. Najnudel, *On the maximum of the* CβE *field*, Duke Math. J. **167** (2018), no. 12, 2243–2345.

[35] G. Cipolloni, L. Erdős, and Y. Xu, *Universality of extremal eigenvalues of large random matrices*, arXiv preprint arXiv:2312.08325.

[36] G. Cipolloni and B. Landon, *Maximum of the Characteristic Polynomial of IID Matrices*, arXiv preprint arXiv:2405.05045 (2024).

[37] T. Claeys, B. Fahs, G. Lambert, and C. Webb, *How much can the eigenvalues of a random Hermitian matrix fluctuate?*, Duke Math. J. **170** (2021), no. 9, 2085–2235.

[38] N. Cook and O. Zeitouni, *Maximum of the characteristic polynomial for a random permutation matrix*, Comm. Pure Appl. Math. **73** (2020), no. 8, 1660–1731.

[39] C. Cosco and O. Zeitouni, *Moments of partition functions of 2d Gaussian polymers in the weak disorder regime—I*, Comm. Math. Phys. **403** (2023), no. 1, 417–450.

[40] A. Dembo, Y. Peres, and O. Zeitouni, *Cover times for Brownian motion and random walks in two dimensions*, Ann. of Math. (2) **160** (2004), no. 2, 433–464.

[41] P. Diaconis, *Patterns in eigenvalues: the 70th Josiah Willard Gibbs lecture*, Bull. Amer. Math. Soc. (N.S.) **40** (2003), no. 2, 155–178.

[42] J. Ding, R. Roy, and O. Zeitouni, *Convergence of the centered maximum of log-correlated Gaussian fields*, The Annals of Probability **45** (2017), no. 6A, 3886–3928.

[43] L. Erdős, B. Schlein, and H.-T. Yau, *Universality of random matrices and local relaxation flow*, Invent. Math. **185** (2011), no. 1, 75–119.

[44] L. Erdős and H.-T. Yau, *A Dynamical Approach to Random Matrix Theory*, Courant Lecture Notes in Mathematics, vol. 28, 2017.

[45] L. Erdős, H.-T. Yau, and J. Yin, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math. **229** (2012), no. 3, 1435–1515.

[46] R. Feng and D. Wei, *Large gaps of CUE and GUE*, arXiv preprint arXiv:1807.02149 (2018).

[47] Y. Fyodorov, G. Hiary, and J. Keating, *Freezing Transition, Characteristic Polynomials of Random Matrices, and the Riemann Zeta Function*, Physical Review Letters **108** (2012).

[48] Y. Fyodorov and J. Keating, *Freezing transitions and extreme values: random matrix theory, and disordered landscapes*, Philosophical Transactions of the Royal Society A **372** (2014), no. 2007, 20120503.

[49] Y. Fyodorov and N. Simm, *On the distribution of the maximum value of the characteristic polynomial of GUE random matrices*, Nonlinearity **29** (2016), no. 9, 2837.

[50] F. Götze and A. Tikhomirov, *On the rate of convergence to the semi-circular law*, High dimensional probability VI, Progr. Probab., vol. 66, Birkhäuser/Springer, Basel, 2013, pp. 139–165.

[51] F. Götze, A. Naumov, A. Tikhomirov, and D. Timushev, *On the local semicircular law for Wigner ensembles*, Bernoulli **24** (2018), no. 3, 2358–2400.

[52] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electron. Comm. Probab. **5** (2000), 119–136.

[53] J. Gustavsson, *Gaussian fluctuations of eigenvalues in the GUE*, Ann. Inst. H. Poincaré Probab. Statist. **41** (2005), no. 2, 151–178 (English, with English and French summaries).

[54] A. J. Harper, *On the partition function of the Riemann zeta function, and the Fyodorov-Hiary-Keating conjecture*, arXiv preprint arXiv:1906.05783 (2019).

[55] ———, *The Riemann zeta function in short intervals [after Najnudel, and Arguin, Belius, Bourgade, Radziwiłłand Soundararajan]*, Astérisque, Séminaire Bourbaki. Vol. 2018/2019. **422** (2020), Exp. No. 1161, 391–414.

[56] A. Its and I. Krasovsky, *Hankel determinant and orthogonal polynomials for the Gaussian weight with a jump*, Integrable systems and random matrices, Contemp. Math., vol. 458, Amer. Math. Soc., Providence, RI, 2008, pp. 215–247.

[57] K. Johansson, *On fluctuations of eigenvalues of random Hermitian matrices*, Duke Math. J. **91** (1998), no. 1, 151–204.

[58] A. Knowles and J. Yin, *The isotropic semicircle law and deformation of Wigner matrices*, Comm. Pure Appl. Math. **66** (2013), no. 11, 1663–1750.

[59] G. Lambert, *Maximum of the characteristic polynomial of the Ginibre ensemble*, Communications in Mathematical Physics **378** (2020), no. 2, 943–985.

[60] ———, *Mesoscopic central limit theorem for the circular β-ensembles and applications*, Electronic Journal of Probability **26** (2021), 1–33.

[61] G. Lambert, T. Leblé, and O. Zeitouni, *Law of large numbers for the maximum of the two-dimensional Coulomb gas potential*, arXiv preprint arXiv:2303.09912 (2023).

[62] G. Lambert, D. Ostrovsky, and N. Simm, *Subcritical multiplicative chaos for regularized counting statistics from random matrix theory*, Comm. Math. Phys. **360** (2018), no. 1, 1–54.

[63] G. Lambert and E. Paquette, *The law of large numbers for the maximum of almost Gaussian log-correlated random fields coming from random matrices*, Probab. Theory Related Fields **173** (2019), 157–209. See arXiv 1611.08885 version 3.

[64] ———, *Strong approximation of Gaussian β-ensemble characteristic polynomials: the edge regime and the stochastic Airy function*, arXiv preprint arXiv:2009.05003 (2020).

[65] ———, *Strong approximation of Gaussian β-ensemble characteristic polynomials: the hyperbolic regime*, Ann. Appl. Probab. 33 **33** (2023), 549–612.

[66] B. Landon, P. Lopatto, and J. Marcinek, *Comparison theorem for some extremal eigenvalue statistics*, Annals Probab. **48** (2020), 2894–2919.

[67] B. Landon and P. Sosoe, *Applications of mesoscopic CLTs in random matrix theory*, Annals Appl. Probab. **30** (2020), 2769–2795.

[68] B. Landon, P. Sosoe, and H.-T. Yau, *Fixed energy universality for Dyson Brownian motion*, Advances in Mathematics **346** (2019), 1137–1332.

[69] J. O. Lee and K. Schnelli, *Local law and Tracy–Widom limit for sparse random matrices*, Probability Theory and Related Fields **171** (2018), no. 1, 543–616.

[70] A. Lytova and L. Pastur, *Central limit theorem for linear eigenvalue statistics of random matrices with independent entries*, Ann. Probab. **37** (2009), no. 5, 1778–1840.

[71] J. Najnudel, *On the extreme values of the Riemann zeta function on random intervals of the critical line*, Probab. Theory Related Fields **172** (2018), no. 1-2, 387–452.

[72] H. Nguyen, *Random matrices: overcrowding estimates for the spectrum*, J. Funct. Anal. **275** (2018), no. 8, 2197–2224.

[73] M. Nikula, E. Saksman, and C. Webb, *Multiplicative chaos and the characteristic polynomial of the CUE: The L^1-phase*, Transactions of the American Mathematical Society **373** (2020), no. 6, 3905–3965.

[74] S. O'Rourke, *Gaussian fluctuations of eigenvalues in Wigner random matrices*, J. Stat. Phys. **138** (2010), no. 6, 1045–1066.

[75] E. Paquette and O. Zeitouni, *The maximum of the CUE field*, Int. Math. Res. Not. IMRN **16** (2018), 5028–5119.

[76] ———, *The extremal landscape for the CβE ensemble*, arXiv preprint arXiv:2209.06743 (2022).

[77] M. Shcherbina, *Fluctuations of linear eigenvalue statistics of β matrix models in the multi-cut regime*, J. Stat. Phys. **151** (2013), no. 6, 1004–1034.

[78] T. Tao and V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math. **206** (2011), no. 1, 127–204.

[79] ———, *Random matrices: sharp concentration of eigenvalues*, Random Matrices Theory Appl. **2** (2013), no. 3.

[80] C. Webb, *The characteristic polynomial of a random unitary matrix and Gaussian multiplicative chaos—the L^2-phase*, Electron. J. Probab. **20** (2015), no. 104, 21.

[81] O. Zeitouni, *Branching random walks and Gaussian fields*, Probability and statistical physics in St. Petersburg, 2016, pp. 437–471.