

ON THE SPECTRAL FORM FACTOR FOR RANDOM MATRICES

GIORGIO CIPOLLONI, LÁSZLÓ ERDŐS[†], AND DOMINIK SCHRÖDER[‡]

ABSTRACT. In the physics literature the spectral form factor (SFF), the squared Fourier transform of the empirical eigenvalue density, is the most common tool to test universality for disordered quantum systems, yet previous mathematical results have been restricted only to two exactly solvable models [21, 22]. We rigorously prove the physics prediction on SFF up to an intermediate time scale for a large class of random matrices using a robust method, the multi-resolvent local laws. Beyond Wigner matrices we also consider the monoparametric ensemble and prove that universality of SFF can already be triggered by a single random parameter, supplementing the recently proven Wigner-Dyson universality [15] to larger spectral scales. Remarkably, extensive numerics indicates that our formulas correctly predict the SFF in the entire *slope-dip-ramp* regime, as customarily called in physics.

1. INTRODUCTION

Spectral statistics of disordered quantum systems tend to exhibit universal behavior and hence are widely used to study quantum chaos and to identify universality classes. In the chaotic regime, the celebrated Wigner-Dyson-Mehta eigenvalue gap statistics involving the well-known *sine-kernel* [42] tests this universality on the scale of individual eigenvalue spacing. On this small *microscopic* scale the universality phenomenon is the most robust and it depends only on the fundamental symmetry type of the model. On larger scales more details of the model influence the spectral statistics, nevertheless several qualitative and also quantitative universal patterns still prevail.

1.1. The spectral form factor and predictions from physics. In the physics literature the standard tool to investigate eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$ of a Hermitian $N \times N$ matrix (Hamiltonian) H on all scales at once is the *spectral form factor* (SFF) [40] defined as

$$\mathbf{SFF}(t) := \frac{1}{N^2} \sum_{i,j=1}^N e^{it(\lambda_i - \lambda_j)} = |\langle e^{itH} \rangle|^2 \quad (1.1)$$

with a real time parameter $t > 0$, i.e. it is the square of the Fourier transform of the empirical spectral density. Here we denoted the normalized trace of any $N \times N$ matrix A by $\langle A \rangle = \frac{1}{N} \text{Tr } A$. In case of random H , the expectation of $\mathbf{SFF}(t)$ is denoted by

$$K(t) := \mathbf{E}[\mathbf{SFF}(t)]. \quad (1.2)$$

Date: March 9, 2023.

2010 Mathematics Subject Classification. 60B20.

Key words and phrases. Wigner-Dyson universality, GUE ensemble, Slope-dip-ramp-plateau.

[†]Partially supported by ERC Advanced Grant "RMTBeyond" No. 101020331.

[‡]Supported by Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zürich Foundation.

For typical disordered Hamiltonians a key feature of $\mathbf{SFF}(t)$ is that for larger t (more precisely, in the *ramp* and *plateau* regimes, see later) it is strongly dependent on the sample, i.e. the standard deviation of $\mathbf{SFF}(t)$ is comparable with $K(t)$. In other words, $\mathbf{SFF}(t)$ is *not* self-averaging [45] despite the large summation in (1.1).

The spectral form factor and its expectation $K(t)$ have a very rich physics literature since they contain most physically relevant information about spectral statistics. Quantizations of integrable systems typically result in $K(t) \sim 1/N$ for all t where N is the dimension of the Hilbert space. Chaotic systems give rise to a linearly growing behavior of $K(t)$ for smaller t (so-called *ramp*) until it turns into a flat regime, the *plateau*. The turning point is around the Heisenberg time T_H , but the details of the transition depend on the symmetry class of H and on whether the eigenvalues are rescaled to take into account the non-constant density of states (in physics terminology: *unfolding the spectrum*). For example, in the time irreversible case (GUE symmetry class) the unfolded SFF has a sharp kink, while in the GOE symmetry class the kink is smoothed. The exact formulas can be computed from the Fourier transform of the two point eigenvalue correlation function of the corresponding Gaussian random matrix ensemble, see [42, Eqs. (6.2.17), (7.2.46)], the result is

$$K_{\text{GUE}}(\tau T_H) \approx \frac{1}{N} \times \begin{cases} \tau, & 0 < \tau \leq 1 \\ 1, & \tau \geq 1 \end{cases}, \quad K_{\text{GOE}}(\tau T_H) \approx \frac{1}{N} \times \begin{cases} 2\tau - \tau \log(1 + 2\tau), & 0 < \tau \leq 1 \\ 2 - \tau \log \frac{2\tau+1}{2\tau-1}, & \tau \geq 1 \end{cases}, \quad (1.3)$$

for any fixed $\tau > 0$ in the large N limit. Here we expressed the physical time t in units of the Heisenberg time, $\tau = t/T_H$, where T_H is given by $T_H = 2\pi\bar{\rho}$ with $\bar{\rho}$ being the average density. Choosing the standard normalisation for the independent (up to symmetry) matrix elements,

$$\mathbf{E}h_{ij} = 0, \quad \mathbf{E}|h_{ij}|^2 = \frac{1}{N}, \quad (1.4)$$

the limiting density of states is the semicircle law $\rho_{\text{sc}}(E) = \frac{1}{2\pi}\sqrt{(4-E^2)_+}$, so we have N eigenvalues in an interval of size 4, hence $\bar{\rho} = N/4$ and thus $T_H = \frac{\pi}{2}N$. In particular, in the original t variable

$$K_{\text{GUE}}(t) \approx \begin{cases} \frac{2t}{\pi N^2}, & \delta N \leq t \leq \frac{\pi}{2}N \\ \frac{1}{N}, & t \geq \frac{\pi}{2}N. \end{cases} \quad (1.5)$$

Note the lower bound on t : the formula holds in the large N limit in the regime where $t \geq \delta N$ for some fixed $\delta > 0$ that is independent of N . The corresponding formulas without unfolding the spectrum (i.e. for the quantity defined in (1.1)) are somewhat different, see e.g. [9, Eq. (4.8)] for the GUE case; they still have a ramp-plateau shape but the kink is smoothed.

The ramp-plateau picture and its sensitivity to the symmetry type has been established well beyond the standard mean field random matrix models. In fact, the Bohigas-Giannoni-Schmit conjecture [6] asserts that the formulas (1.3) are universal, i.e. they hold essentially for any chaotic quantum system, depending only on whether the system is without or with time reversal symmetry. The nonrigorous but remarkably effective semiclassical orbit theory [48, 43, 31, 4] based upon Gutzwiller's trace formula [27] and many follow-up works verified this conjecture for quantizations of a large family of classical chaotic systems, e.g. for certain billiards.

For smaller times, $t \ll T_H$, other details of H may become relevant. In particular the drop from $K(t=0) = 1$ to $K(t) \ll 1$ for $1 \ll t \ll T_H$ is first dominated by the typical non-analyticity of the density of states at the spectral edges giving rise to the *slope regime* up to an intermediate minimum

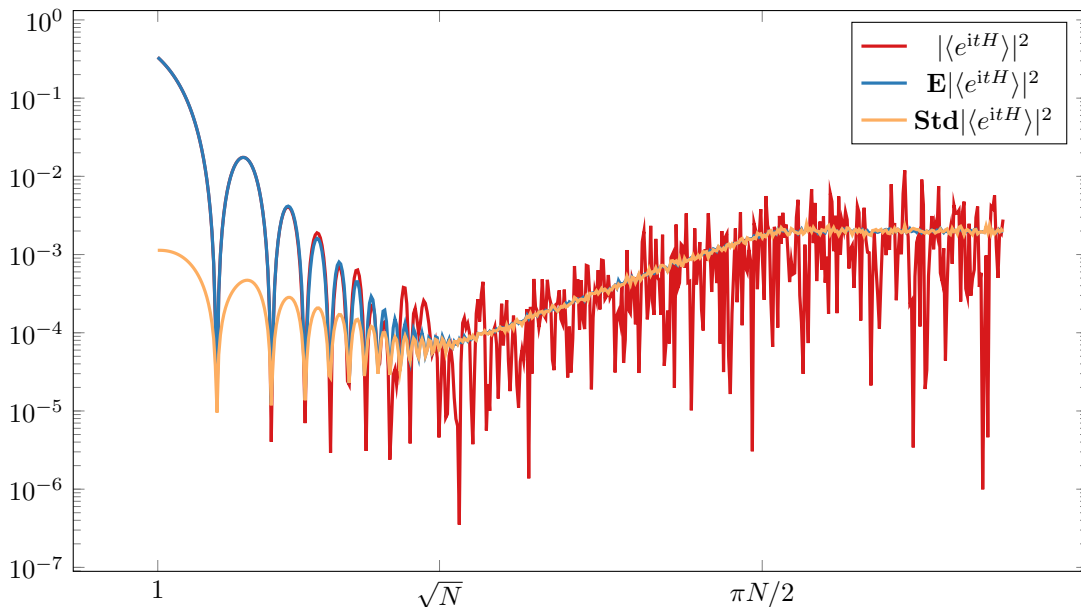


FIGURE 1. A typical *slope-dip-ramp-plateau* picture for the spectral form factor of a chaotic system. The figure on log-log scale shows the SFF of a single GUE realisation H of size 500×500 , as well as the empirical mean and standard deviation obtained from 500 independent realisations.

point of $K(t)$, called the *dip* (in the early literature the dip was called *correlation hole* [40], for a recent overview, see [17]).

Figure 1 shows the typical *slope-dip-ramp-plateau* picture for the GUE ensemble. Formula (1.5) is valid starting from scales $t \gg N^{1/2}$, while $K(t)$ is oscillatorily decreasing for $t \lesssim N^{1/2}$ with a dip-time $t_{\text{dip}} \sim N^{1/2}$. Thus $K(t)$ follows the universal behavior (1.5) only for $t \gg t_{\text{dip}}$. In this regime the fluctuation of the SFF is comparable with its expectation, $K(t)$, in fact $\langle e^{itH} \rangle$ is approximately Gaussian. In contrast, the dominant contribution to the slope regime, $t \ll t_{\text{dip}}$, is self-averaging with a relatively negligible fluctuation. However, if the edge effects are properly discounted (e.g. by considering the circular ensemble with uniform spectral density on the unit circle), i.e. the slope regime is entirely removed, then the Gaussian behavior holds for all $t \ll T_H$ with a universal variance given by (1.5).

In more recent works spectral form factors were studied for the celebrated Sachdev-Ye-Kitaev (SYK) model [46, 18, 32, 24, 23] which also exhibits a similar *slope-dip-ramp-plateau* pattern although the details are still debated in the physics literature and the numerics are much less reliable due to the exponentially large dimensionality of the model.

1.2. Our results. Quite surprisingly, despite its central role in the physics literature on quantum chaos, SFF has not been rigorously investigated in the mathematics literature up to very recently, when Forrester computed the large N limit of $K(t)$ rigorously for the GUE in [21] and

the Laguerre Unitary Ensemble (LUE) in [22] in the entire regime $t \ll N$. Both results rely on a remarkable identity from [9] (and its extension to the LUE case) and on previous stimulating work of Okuyama [44]. However, these methods use exact identities and thus are restricted to a few explicitly solvable invariant ensembles.

The main goal of the current paper is to investigate SFF beyond these special cases with a robust method, the multi-resolvent local laws. While our approach is valid for quite general ensembles, for definiteness we focus on two models: the standard Wigner ensemble (for both symmetry classes) and the novel *monoparametric ensemble* introduced recently [25] by Gharibyan, Pattison, Shenker and Wells. The latter consists of matrices of the form $H^s := s_1 H_1 + s_2 H_2$, where H_1 and H_2 are typical but *fixed* realisations of two independent Wigner matrices and $s = (s_1, s_2) \in S^1 \subset \mathbf{R}$ is a continuous random variable. The normalization $s_1^2 + s_2^2 = 1$ guarantees that the semicircle law for H^s is independent of s and it also shows that the model has effectively only one random parameter. One may also consider similar ensembles with finitely many parameters (see Remark 2.4) resulting in qualitatively the same behavior but with different power laws, see Table 1.

We study the statistics of H^s in the probability space of the single random variable s and probe how much universality still persists with such reduced randomness. We write \mathbf{E}_s for the expectation wrt. s and $\mathbf{E}_H, \mathbf{Std}_H$ for the expectation and standard deviation wrt. H_1 and H_2 .

Our main result is to prove a formula for the expectation and standard deviation of SFF for both ensembles up to an intermediate time. While this does not include the ramp regime, it already allows us to draw the following two main conclusions of the paper:

- (a) The expectation and standard deviation of $\mathbf{SFF}(t)$ for Wigner and monoparametric ensembles exhibit the same universal behavior to leading order for $1 \ll t \ll N^{1/4}$ if the trivial edge effects are removed. In the monoparametric case it is quite remarkable that already a single real random variable generates universality.
- (b) For the monoparametric ensemble $K(t) = \mathbf{E}_s[\mathbf{SFF}(t)]$ depends non-trivially on the fixed H_1, H_2 matrices, but for large t this dependence is a subleading effect whose relative size becomes increasingly negligible as a negative power of t . In particular, while the speed of convergence to universality is much slower for the monoparametric ensemble than for the Wigner case, it is improving for larger t .

The second item answers a question raised by the authors of [25] which strongly motivated the current work. In particular, sampling from s does not give a consistent estimator for $K(t)$, but the relative precision of such estimate improves for larger times.

We supplement these proofs with an extensive numerics demonstrating that both conclusions hold not only for $t \ll N^{1/4}$ but for the entire ramp regime, i.e. up to $t \ll T_H \sim N$. Note that recently we have proved [15] that the Wigner-Dyson-Mehta eigenvalue gap universality holds for the monoparametric ensemble, which strongly supports, albeit does not prove, that $K(t)$ in the plateau regime is also universal.

We remark that our method applies without difficulty for finite temperatures (expressed by a parameter $\beta > 0$) and for different-time autocorrelation functions, i.e. for

$$\langle e^{(-\beta+it)H} \rangle \langle e^{(-\beta-it')H} \rangle$$

as well, but for the simplicity of the presentation we focus on $\mathbf{SFF}(t)$ defined in (1.1), i.e. on $\beta = 0$ and $t = t'$.

1.3. Relations to previous mathematical results. Rigorous mathematics for the spectral form factor, even for Wigner matrices or even for GOE, significantly lags behind establishing the compelling physics picture about the slope-dip-ramp-plateau. Given the recently developed tools in random matrix theory, it may appear surprising that they do not directly answer the important questions on SFF. We now briefly explain why.

1.3.1. Limitations of the resolvent methods. For problems on macroscopic spectral scales (involving the cumulative effect of order N many eigenvalues), and to a large extent also on mesoscopic scales (involving many more than $O(1)$ eigenvalues), the *resolvent method* is suitable. This method considers the resolvent $G(z) = (H - z)^{-1}$ of H for a spectral parameter z away from (but typically still close to) the real axis and establishes that in a certain sense $G(z)$ becomes deterministic. This works for $\eta = \Im z \gg N^{-1}$ (in the bulk spectrum), i.e. on scales just above the eigenvalue spacing (note that the imaginary part of the spectral parameter sets a scale in the spectrum). Such results are called *local laws* and they can be extended to regular functions $f(H)$ by standard spectral calculus (Helffer-Sjöstrand formula, see (3.3) later).

However, the interesting questions about SFF concern a $1/N$ subleading fluctuation effect beyond the local laws. Indeed

$$\mathrm{Tr} e^{itH} = \sum_i e^{it\lambda_i}$$

is a special case of the well-studied *linear eigenvalue statistics*, $\mathrm{Tr} f(H) = \sum_i f(\lambda_i)$, with the regular test function $f(\lambda) = e^{it\lambda}$. To leading order it is deterministic and its fluctuation satisfies the central limit theorem (CLT) without the customary \sqrt{N} normalisation, i.e.

$$\sum_i f(\lambda_i) - \mathbf{E} \sum_i f(\lambda_i) \approx \mathcal{N}(0, V_f), \quad \text{with} \quad \mathbf{E} \sum_i f(\lambda_i) = N \int_{\mathbf{R}} f(x) \rho_{\mathrm{sc}}(x) dx + O_f(1). \quad (1.6)$$

is a normal random variable with variance¹

$$V_f = \frac{1}{4\pi^2} \iint_{-2}^2 \left| \frac{f(x) - f(y)}{x - y} \right|^2 \frac{4 - xy}{\sqrt{4 - x^2} \sqrt{4 - y^2}} dx dy. \quad (1.7)$$

The computation of higher moments of $\mathrm{Tr} f(H) - \mathbf{E} \mathrm{Tr} f(H)$ requires a generalization of the local laws to polynomial combinations of several G 's that are called *multi-resolvent local laws*.

Applying (1.6)–(1.7) to $f(x) = e^{itx}$ we obtain, roughly,

$$\mathbf{SFF}(t) = \frac{1}{N^2} \left| \mathrm{Tr} e^{itH} \right|^2 \approx \left[\frac{J_1(2t)}{t} + O\left(\sqrt{\frac{V_f}{N^2}}\right) \right]^2, \quad t \gg 1, \quad (1.8)$$

using that

$$\int_{\mathbf{R}} f(x) \rho_{\mathrm{sc}}(x) dx = \int_{\mathbf{R}} e^{itx} \rho_{\mathrm{sc}}(x) dx = \frac{J_1(2t)}{t},$$

where J_1 is the first Bessel function of the first kind. Note that V_f in (1.7) scales essentially as the $H^{1/2}$ Sobolev norm of f hence $V_f \sim t$ for our $f(x) = e^{itx}$ in the regime $t \gg 1$. Therefore the size of the fluctuation term in (1.8) is $V_f/N^2 \sim t/N^2$ and it competes with the deterministic term $(J_1/t)^2 \sim t^{-3}$. The dip time $t_{\mathrm{dip}} \sim \sqrt{N}$ is obtained as the threshold where the fluctuation (the

¹Eq. (1.7) is for matrices whose second and fourth moments coincide with the ones of GUE, otherwise there are additional terms, see e.g. [13, Theorem 2.4].

linear ramp function) becomes bigger than the slope function $(J_1/t)^2$. This argument, however, is heuristic as it neglects the error terms in (1.6) that also depend on t via f .

CLT for linear statistics (1.6) for Wigner matrices H has been proven [33, 34, 1, 47, 49, 26, 29, 30, 13, 36, 41, 38, 3, 28] for test functions of the form $f(x) = g(N^a(x - E))$ with some fixed reference point $|E| < 2$, scaling exponent $a \in [0, 1)$ and smooth function g with compact support, i.e. for macroscopic ($a = 0$) and mesoscopic ($0 < a < 1$) test functions living on a *single* scale N^{-a} . These proofs give optimal error terms for such functions but they were not optimized for dealing with functions that oscillate on a mesoscopic scale *and* have macroscopic support, like $f(x) = e^{itx}$ for some $t \sim N^\alpha$, $\alpha > 0$. The only CLT-type result for a special two-scale observable is in [2] where the eigenvalue counting function smoothed on an intermediate scale $N^{-1/3}$ was considered.

Quite remarkably, extensive numerics shows that the formulas (1.6)–(1.7) for $f(x) = e^{itx}$ are in perfect agreement with the expected behavior of $K(t)$ in the entire slope-dip-ramp regime all the way up to $t \ll N$, i.e. the CLT for linear statistics correctly predicts SFF well beyond its proven regime of validity. In the current paper we optimise the error terms specifically for e^{itx} and thus we could cover the regime $t \ll N^{5/11}$ for the variance in (1.6) (corresponding to $\mathbf{E}[\mathbf{SFF}(t)]$).

1.3.2. Limitations of Dyson Brownian motion techniques. For the microscopic scale (i.e. comparable with the eigenvalue spacing, $1/N$ in the bulk) the resolvent is heavily fluctuating as it strongly depends on single eigenvalues. Local laws cannot access them, but in this regime another approach, the careful analysis of the *Dyson Brownian Motion (DBM)* becomes applicable. While these two approaches are complementary and apparently cover all scales, the actual methods require additional conditions that seriously restrict their use for SFF.

The formulas (1.3) are obtained by computing the Fourier transform of the two point correlation function of the rescaled (*unfolded*) eigenvalues. Indeed, in the GUE case $K_{\text{GUE}}(t)$ in (1.3) is just the Fourier transform of $p_2(x, y) - 1 + \delta(x - y)$ in the difference variable $x - y$, where

$$p_2(x, y) := 1 - \left(\frac{\sin(\pi(x - y))}{\pi(x - y)} \right)^2,$$

is the two point function, given by the celebrated Wigner-Dyson sine kernel, and $K_{\text{GOE}}(t)$ has a similar origin. Wigner-Dyson theory is designed for microscopic scales, i.e. to describe eigenvalue correlations on scales comparable with the local level spacing Δ , this is encoded in the fact that (1.3) holds for any fixed $\tau > 0$ in the $N \rightarrow \infty$ limit (equivalently that (1.5) holds only for $t \geq \delta N$ since $\Delta \sim 1/N$ in the bulk). While this is a very elegant argument supporting (1.3), mathematically it is quite far from a rigorous proof.

The mathematical proofs of the sine-kernel universality use test functions that are rapidly decaying beyond scale Δ . The typical statements (so called *fixed energy universality* [7, 39]) show that for any fixed energy E in the bulk

$$\sum_{i < j} g(N\rho_{\text{sc}}(E)(\lambda_i - E), N\rho_{\text{sc}}(E)(\lambda_j - E)) \rightarrow \iint_{\mathbf{R}} g(x, y)p_2(x, y) dx dy$$

in the large N limit, for any smooth, compactly supported functions $g: \mathbf{R}^2 \rightarrow \mathbf{R}$. The current methods for proving the Wigner-Dyson universality cannot deal with functions that are macroscopically supported, like $g(x, y) = e^{it(x-y)}$ with a fast oscillation $t \sim N$.

1.4. Summary. Using multi-resolvent local laws we prove a CLT for linear statistics of monoparametric ensembles (Theorem 2.5) with covariance

$$\mathbf{Cov}(\mathrm{Tr} f(H^s), \mathrm{Tr} g(H^r)) \approx \frac{1}{\pi^2} \iint f'(x)g'(y) \log \left| \frac{1 - \langle s, r \rangle m_{\mathrm{sc}}(x) \overline{m_{\mathrm{sc}}(y)}}{1 - \langle s, r \rangle m_{\mathrm{sc}}(x) m_{\mathrm{sc}}(y)} \right| dx dy$$

with an additional term depending on the fourth cumulant. Due to a careful analysis of the error terms this allows us to prove the expected behavior on the expectation and standard deviation of the SFF for Wigner matrices for $t \ll N^{5/17}$ (Theorem 2.7) and for the monoparametric ensemble for $t \ll N^{1/4}$ (Theorem 2.8). Beyond these regime the spectral form factor is not understood mathematically apart from the special GUE and LUE cases. However, we can still use our predictions from the CLT for linear statistics (1.6) to derive an Ansatz for the behavior of $\mathbf{SFF}(t)$ in the entire $t \ll N$ regime. In particular, we show that the SFF is universal for the monoparametric ensemble. We find numerically that our theory correctly reproduces $\mathbf{SFF}(t)$ for any $t \ll N$ and it also coincides with the physics predictions for the GUE case.

Notations and conventions. For positive quantities f and g we will frequently use the notation $f \approx g$ meaning that $f/g \rightarrow 1$ in a limit that is always clear from the context. Similarly, $f \ll g$ means that $f/g \rightarrow 0$. Finally, the relation $f \sim g$ means that there exist two positive constants c, C such that $c \leq f/g \leq C$.

We say that an event holds "with very high probability" if for any fixed $D > 0$ the probability of the event is bigger than $1 - N^{-D}$ if $N \geq N_0(D)$, for some $N_0(D) > 0$.

Acknowledgement. We are grateful to the authors of [25] for sharing with us their insights and preliminary numerical results. We are especially thankful to Stephen Shenker for very valuable advice over several email communications. Helpful comments on the manuscript from Peter Forrester and from the anonymous referees are also acknowledged.

2. STATEMENT OF THE MAIN RESULTS

Our new results mainly concern the monoparametric ensemble but for comparison reasons we also prove the analogous results for the Wigner ensemble. We start with the two corresponding definitions.

Definition 2.1. *The Wigner ensemble consists of Hermitian $N \times N$ random matrices H with the following properties. The off-diagonal matrix elements below the diagonal are independent, identically distributed (i.i.d) real ($\beta = 1$) or complex ($\beta = 2$) random variables; in the latter case we assume that $\mathbf{E}h_{ij}^2 = 0$. The diagonal elements are i.i.d. real random variables with $\mathbf{E}h_{ii}^2 = 2/(N\beta)$. Besides the standard normalisation (1.4), we also make the customary moment assumption: for every $q \in \mathbf{N}$ there is a constant C_q such that*

$$\mathbf{E}|\sqrt{N}h_{ij}|^q \leq C_q. \quad (2.1)$$

In the case of Gaussian distributions, it is called the Gaussian Orthogonal or Unitary Ensemble (GOE/GUE), for the real and complex cases, respectively.

Remark 2.2. *The assumptions $\mathbf{E}h_{ij}^2 = 0$ in the complex case, and $\mathbf{E}h_{ii}^2 = 2/(\beta N)$ are made purely for convenience. All results can easily be generalised beyond this case but we refrain from doing so for notational simplicity.*

Definition 2.3. *The monoparametric ensemble consists of Hermitian $N \times N$ random matrices of the form*

$$H = H^s := s_1 H_1 + s_2 H_2, \quad (2.2)$$

where H_1, H_2 are independent Wigner matrices satisfying² $\mathbf{E}|h_{ij}^{(1)}|^4 = \mathbf{E}|h_{ij}^{(2)}|^4$ and $s = (s_1, s_2) \in S^1$ is a random vector, independent of H_1, H_2 . On the distribution of s we assume that it has an square integrable density $\rho(s)$ independent of N . We write \mathbf{E}_s for the expectation wrt. s and $\mathbf{E}_H, \mathbf{Std}_H$ for the expectation and standard deviation wrt. the Wigner matrices H_1 and H_2 .

The parameter space $S^1 \subset \mathbf{R}^2$ inherits the usual scalar product and norms from \mathbf{R}^2 , so for $s, r \in S^1$ we have

$$\langle s, r \rangle := s_1 r_1 + s_2 r_2, \quad \|s\|_p := (|s_1|^p + |s_2|^p)^{1/p}.$$

We also introduce the entrywise product of two vectors:

$$s \odot r := (s_1 r_1, s_2 r_2).$$

For a fixed s , H^s is just the weighted sum of two Wigner matrices, and, due to the normalisation, itself is just a Wigner matrix. However, the concept of monoparametric ensemble views H^s as a random matrix in the probability space of the single random variable s for a typical but fixed (quenched) realization of H_1 and H_2 . While Wigner matrices have a large ($\sim N^2$) number of independent random degrees of freedom, the monoparametric ensemble is generated by one single random variable hence, naively, much less universality properties are expected. Nevertheless, the standard Wigner-Dyson local eigenvalue universality holds [15].

Remark 2.4. *In [15] we considered the un-normalized monoparametric model $H^s := H_1 + sH_2$, for some real valued random variable s , whose density of states is a rescaled semicircular distribution. In this paper we prefer to work with more homogeneous models since the formulas are somewhat nicer, but our main results also apply to this inhomogeneous model with some slightly different exponents in the error terms. One may also consider a different un-normalized ensemble, $s_1 H_1 + s_2 H_2$ with $s \in \mathbf{R}^2$ having an absolutely continuous distribution, which is effectively a two parameter model. Similar results also hold for the multi-parametric analogue of (2.2), i.e. $s_1 H_1 + \dots + s_k H_k$ for $s \in S^{k-1}$, see Remark 2.6 and Section 2.4 later. Despite all these options, for definiteness, the main body of this paper concerns the homogenous monoparametric model from Definition 2.3.*

2.1. Central limit theorem for sum of Wigner matrices. To understand the effect of the random s , we study the joint statistics of H^s and H^r for two different fixed realisations r, s in the probability space of H_1, H_2 , i.e. we aim at the correlation effects between H^s and H^r . We introduce the short-hand notations

$$\langle f \rangle_{\text{sc}} := \int_{-2}^2 f(x) \frac{\sqrt{4-x^2}}{2\pi} dx, \quad \langle f \rangle_{1/\text{sc}} := \int_{-2}^2 f(x) \frac{1}{\pi\sqrt{4-x^2}} dx, \quad \kappa_4 := N^2 \mathbf{E}|h_{12}|^4 - 1 - \frac{2}{\beta}. \quad (2.3)$$

To estimate the error term in the following theorem we introduce a parameter $1 \leq \tau \ll N$ and the weighted norm

$$\|f\|_\tau := \tau^2 \|f\|_\infty + \tau \|f\|_{H^1} + \|f\|_{H^2}, \quad (2.4)$$

where $\|f\|_{H^k}^2 := \sum_{j \leq k} \int_{\mathbf{R}} |f^{(j)}|^2$ is the usual Sobolev norm. For the applications later, the parameter τ will be optimized.

²We assume equal fourth cumulants merely for notational convenience. Our proof verbatim covers also the more general case.

Theorem 2.5. For $s \in S^1$ and test functions $f \in H^2(\mathbf{R})$ the family of random variables $\text{Tr } f(H^s)$ is approximately Gaussian of mean

$$\mathbf{E} \text{Tr } f(H^s) = N \langle f \rangle_{\text{sc}} + \kappa_4 \|s\|_4^4 \left\langle \frac{x^4 - 4x^2 + 2}{2} f \right\rangle_{1/\text{sc}} + \mathbf{1}(\beta = 1) \left[\frac{f(2) + f(-2)}{4} - \frac{\langle f \rangle_{1/\text{sc}}}{2} \right] + \mathcal{O}(\mathcal{E}_1), \quad (2.5)$$

and fluctuation

$$\mathbf{E} \prod_{i=1}^p \left(\text{Tr } f_i(H^{s^i}) - \mathbf{E} \text{Tr } f_i(H^{s^i}) \right) = \sum_{P \in \text{Pair}([p])} \prod_{(i,j) \in P} v^{s^i s^j}(f_i, f_j) + \mathcal{O}_p(\mathcal{E}_p), \quad (2.6)$$

for any fixed $p \in \mathbf{N}$, functions $f_1, \dots, f_p \in H^2(\mathbf{R})$, and parameters $s^1, \dots, s^p \in S^1$, where³

$$v^{sr}(f, g) := \frac{1}{\beta \pi^2} \iint_{-2}^2 f'(x) g'(y) V^{sr}(x, y) dx dy + \frac{\kappa_4}{2} \langle s \odot s, r \odot r \rangle \langle (2 - x^2) f \rangle_{1/\text{sc}}^2 \quad (2.7)$$

$$V^{sr}(x, y) := \log \left| 1 - \langle s, r \rangle m_{\text{sc}}(x) \overline{m_{\text{sc}}(y)} \right| - \log |1 - \langle s, r \rangle m_{\text{sc}}(x) m_{\text{sc}}(y)|.$$

Here \mathcal{E}_p are error terms which for any $1 \leq \tau \ll N$ and any $\xi, \epsilon > 0$ may be estimated by⁴

$$\mathcal{E}_1 := \frac{N^\xi \|f\|_\tau}{N^{1/2} \tau^{1/2}}, \quad \mathcal{E}_p := N^\xi \left(\frac{1}{N^{1/2} \tau^{3/2}} + \frac{N^\epsilon}{N} + \frac{N^{-\epsilon}}{\tau^{2p-1}} \right) \left(1 + \frac{\tau^2}{N^{1-2\epsilon}} \right) \prod_{i \in [p]} \|f_i\|_\tau, \quad (2.8)$$

for $p \geq 2$. Additionally, if $s^1 = \dots = s^p$, i.e. in the Wigner case, we have the improved bound

$$\mathcal{E}_p := \frac{1}{N^{1/2} \tau^{3/2}} \prod_{i \in [p]} \|f_i\|_\tau \quad (2.9)$$

and the first term of (2.7) for $\beta = 2$ coincides with (1.7).

We note that (2.7) generalizes the standard variance calculation yielding (1.7) to $s \neq r$, see Section 3.2.4.

Remark 2.6. Theorem 2.5 verbatim holds true also for the multi-parametric model

$$s_1 H_1 + \dots + s_k H_k$$

upon interpreting $\langle s, r \rangle$ and $\|s\|_p$ as the Euclidean inner product and p -norm in \mathbf{R}^k . Similarly, Theorem 2.5 also applies to the un-normalised case $s \in \mathbf{R}^2$ for which on the rhs. of (2.5) the function f has to be replaced by $f(\|s\|\cdot)$ with $\|\cdot\| := \|\cdot\|_2$ and v^{sr} from (2.7) has to be replaced by

$$\begin{aligned} \tilde{v}^{sr}(f, g) &:= \frac{\|s\| \|r\|}{\beta \pi^2} \iint_{-2}^2 f'(\|s\|x) g'(\|r\|y) V_{\frac{s}{\|s\|}, \frac{r}{\|r\|}}(x, y) dx dy \\ &+ \frac{\kappa_4}{2} \langle s \odot s, r \odot r \rangle \langle (2 - x^2) f(\|s\|x) \rangle_{1/\text{sc}} \langle (2 - x^2) f(\|r\|x) \rangle_{1/\text{sc}}. \end{aligned} \quad (2.10)$$

³For the applications in this paper, SFF in the regime $t \gg 1$, the first term in (2.7) is the only relevant one.

⁴The exponent in (2.8) can be optimized depending on τ and f .

2.2. SFF for Wigner and monoparametric ensemble. In this section we specialise Theorem 2.5 to the SFF case. We define the approximate expectation (rescaled by $1/N$)

$$\begin{aligned} e_N^s(t) &:= e(t) + \frac{1}{N} \left[\kappa_4 \|s\|_4^4 \left(1 - \frac{6}{t^2}\right) J_0(2t) + \kappa_4 \|s\|_4^4 \left(\frac{6}{t^3} - \frac{4}{t}\right) J_1(2t) - \mathbf{1}(\beta = 1) \frac{J_0(2t) - \cos(2t)}{2} \right] \\ e(t) &:= \frac{J_1(2t)}{t} \end{aligned} \tag{2.11}$$

in terms of the Bessel functions J_k of the first kind. We also define the approximate variance

$$\begin{aligned} v_{\pm, \kappa}^{sr}(t) &:= v^{sr}(e^{it\cdot}, e^{\pm it\cdot}) = v_{\pm}^{sr}(t) + \kappa_4 \langle s \odot s, r \odot r \rangle J_2(2t)^2, \\ v_{\pm}^{sr}(t) &:= \frac{t^2}{\beta \pi^2} \iint_{-2}^2 \cos(t(x \pm y)) V^{sr}(x, y) dx dy, \end{aligned} \tag{2.12}$$

From Theorem 2.5, choosing $f_i(x) = e^{\pm itx}$ and $\tau = t$, and recalling that $\langle e^{\pm itH^s} \rangle = N^{-1} \text{Tr } e^{\pm itH^s}$, we readily conclude the following asymptotics for SFF of the Wigner and monoparametric ensemble.

Theorem 2.7 (SFF for the Wigner ensemble). *For deterministic $t > 0$ (possibly N -dependent) we have*

$$\begin{aligned} \mathbf{E}_H |\langle e^{itH} \rangle|^2 &= E_{\text{wig}}(t)(1 + o(1)) \quad \text{for } t \ll N^{5/11}, \\ \mathbf{Var}_H |\langle e^{itH} \rangle|^2 &= S_{\text{wig}}(t)^2(1 + o(1)) \quad \text{for } t \ll N^{5/17}, \end{aligned} \tag{2.13}$$

and we have the asymptotics

$$\begin{aligned} E_{\text{wig}}(t) &:= e(t)^2 + \frac{v_{-, \kappa}^{ee}(t)}{N^2} \approx \begin{cases} \frac{J_1(2t)^2}{t^2}, & 1 \ll t \ll \sqrt{N} \\ \frac{2}{\pi} \frac{t}{N^2}, & \sqrt{N} \ll t \ll N, \end{cases} \\ S_{\text{wig}}(t) &:= \left(\frac{v_{+, \kappa}^{ee}(t)^2 + v_{-, \kappa}^{ee}(t)^2}{N^4} + 2e(t)^2 \frac{v_{+, \kappa}^{ee}(t) + v_{-, \kappa}^{ee}(t)}{N^2} \right)^{1/2} \approx \begin{cases} \frac{2J_1(2t)}{\sqrt{\pi}tN}, & 1 \ll t \ll \sqrt{N} \\ \frac{2}{\pi} \frac{t}{N^2}, & \sqrt{N} \ll t \ll N, \end{cases} \end{aligned} \tag{2.14}$$

where we set $e := (1, 0) \in S^1$.

This result shows that $S_{\text{wig}}(t) \ll E_{\text{wig}}(t)$ in the slope regime, $t \ll \sqrt{N}$, and $S_{\text{wig}}(t) \approx E_{\text{wig}}(t)$ in the ramp regime, $\sqrt{N} \ll t \ll N$ (see the first plot in Figure 2). In particular, in the ramp regime the SFF is a non-negative random variable whose fluctuations are of the same size as its expectation. Thus the SFF is not self-averaging in the ramp regime, while it is self-averaging in the slope regime but only owing to the dominance of the function $e(t)$ representing the edge effect. If one discounts the edge effect, i.e. artificially removes $e(t)$, then $S_{\text{wig}}(t) \approx E_{\text{wig}}(t)$ would hold for all $1 \ll t \ll N$, demonstrating the universal behavior of SFF in the entire slope-dip-ramp regime.

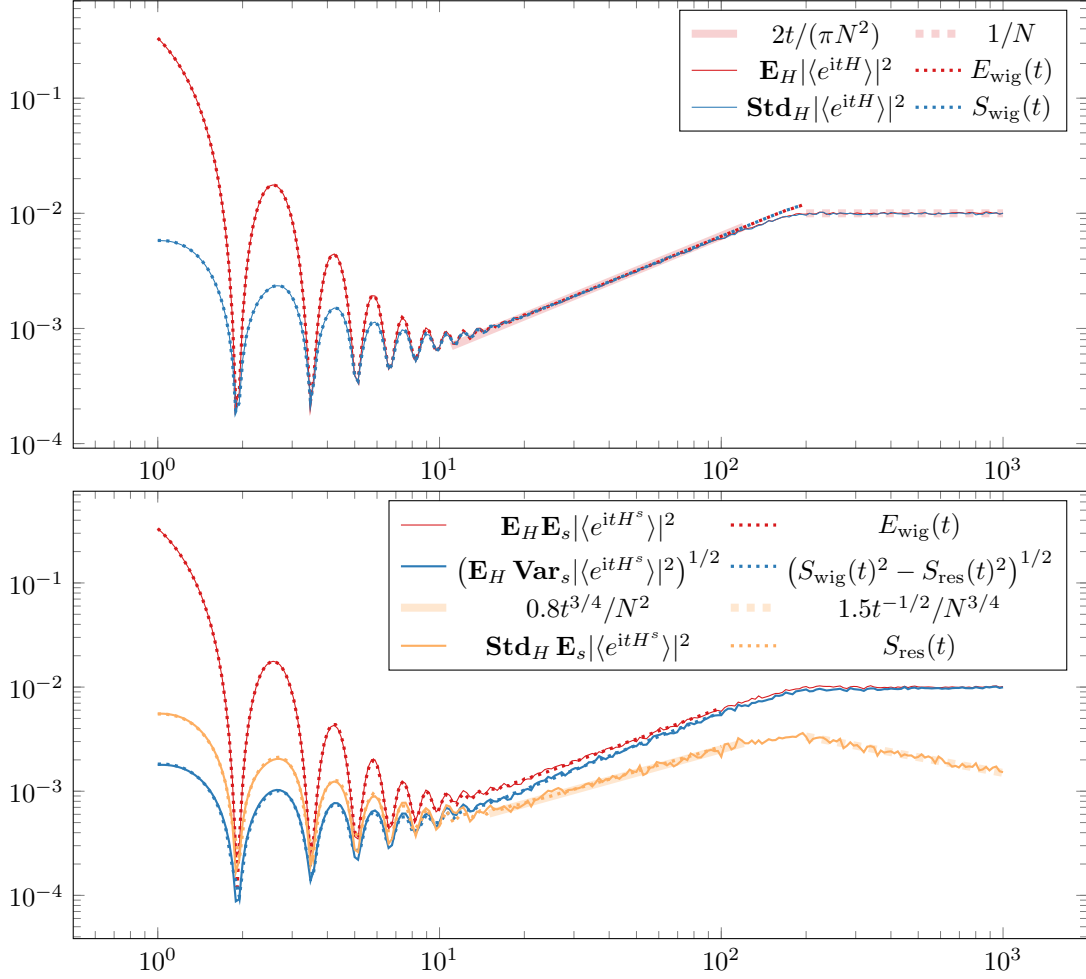


FIGURE 2. In the first plot we compare the empirical mean (red) and standard deviation (blue) of $|\langle e^{itH} \rangle|^2$ obtained from sampling 10,000 independent 100×100 GUE matrices H with our approximation (2.13). In the second plot we similarly compare the empirical mean (red) and variance (blue), with respect to s , obtained from sampling 500 independent scalar random variables s (from the uniform distribution on S^1) and 500 independent 100×100 GUE matrix pairs H_1, H_2 , with the prediction (2.15). We also test the precision of the latter GUE-pair sampling by finding the empirical standard deviation (with respect to H_1, H_2) of the empirical mean of the monoparametric SFF (orange). We observe that for both ensembles our resolvent approximation seems valid for all $t < N$.

Theorem 2.8 (SFF for the monoparametric ensemble (2.2)). *For $t > 0$ (possibly N -dependent) we have*

$$\begin{aligned} \mathbf{E}_H \mathbf{E}_s |\langle e^{itH^s} \rangle|^2 &= E_{\text{wig}}(t)(1 + o(1)) \quad \text{for } t \ll N^{3/7} \\ \mathbf{E}_H \mathbf{Var}_s |\langle e^{itH^s} \rangle|^2 &= \left(S_{\text{wig}}(t)^2 - S_{\text{res}}(t)^2 \right) (1 + o(1)) \quad \text{for } t \ll N^{5/17} \\ \mathbf{Var}_H \mathbf{E}_s |\langle e^{itH^s} \rangle|^2 &= S_{\text{res}}(t)^2 (1 + o(1)) \quad \text{for } t \ll N^{1/4} \end{aligned} \quad (2.15)$$

where the function

$$S_{\text{res}}(t) := \sqrt{\mathbf{E}_s \mathbf{E}_r \left(\frac{v_{+, \kappa}^{sr}(t)^2 + v_{-, \kappa}^{sr}(t)^2}{N^4} + 2e(t)^2 \frac{v_{+, \kappa}^{sr}(t) + v_{-, \kappa}^{sr}(t)}{N^2} \right)} \quad (2.16)$$

satisfies

$$S_{\text{res}}(t) \sim \begin{cases} \frac{\psi(t)}{N^{5/4}}, & 1 \ll t \ll \sqrt{N} \\ \frac{t^{3/4}}{N^2}, & \sqrt{N} \ll t \ll N, \end{cases} \quad (2.17)$$

where $\psi(t) \sim 1$ is a positive function with some oscillation.

In particular, this result immediately shows the following concentration effect:

Corollary 2.9. *For $1 \ll t \ll N^{1/4}$ it holds that*

$$\mathbf{Var}_H \mathbf{E}_s |\langle e^{itH^s} \rangle|^2 \lesssim \frac{1}{\sqrt{t}} \mathbf{Var}_H |\langle e^{itH} \rangle|^2, \quad (2.18)$$

i.e. averaging in s reduces the size of the fluctuation of the SFF by a factor of $t^{-1/4}$.

Note that

$$S_{\text{res}}(t) \lesssim t^{-1/4} S_{\text{wig}}(t) \quad (2.19)$$

both in the slope and ramp regimes showing that not only the expectation but also the variance of the SFF for the monoparametric ensemble coincide with those for the Wigner ensemble to leading order, hence they follow the universal pattern (red and blue curves in the second plot in Figure 2). However, the dependence of $\mathbf{E}_s[\mathbf{SFF}(t)]$ on the fixed Wigner matrix pair (H_1, H_2) is still present, albeit to a lower order, expressed by the *residual standard deviation* $S_{\text{res}}(t)$ whose relative size decreases as $t^{-1/4}$ as t increases (orange curves in Figure 2). It is quite remarkable that a single random mode s generates almost the entire randomness in the ensemble that is responsible for the universality of SFF. A similar phenomenon was manifested in the Wigner-Dyson universality proven in [15].

Remark 2.10. *Based upon extensive numerics (see Figure 2) we believe that (2.13), (2.15) and (2.18) hold up to any $t \ll N$, i.e. in the entire slope-dip-ramp regime and not only up to some fractional power of N as stated and proved rigorously. The proof for the entire regime $t \ll N$ is out of reach with the current technology based upon the multi-resolvent local law Lemma 3.3 whose error term does not trace the expected improvement due to different spectral parameters $z_1 \neq z_2$. We expect that the entire ramp regime $t \ll N$ should be accessible by resolvent techniques if a sharp version of Lemma 3.3, tracing the gain from $z_1 \neq z_2$, was available.*

Remark 2.11. We stated Theorems 2.7 and 2.8 only for the first two moments but the CLT from Theorem 2.5 allows us to compute arbitrary moments $\mathbf{E}|\langle e^{itH} \rangle|^{2m}$ for the Wigner case and $\mathbf{E}_s|\langle e^{itH^s} \rangle|^{2m}$ for the monoparametric case (together with their concentration in the (H_1, H_2) -space), albeit with worsening error estimates. This would lead to rigorous results of the type (2.13) and (2.15) but for a shorter time scale $t \ll N^{c(m)}$ with some $c(m) > 0$. However, in the spirit of Remark 2.10, we believe that $\langle e^{itH^s} \rangle$ can be approximated for any $t \ll N$, to leading order, by a family of complex Gaussians $\xi(t, s)$ of mean and variance

$$\mathbf{E}\xi(t, s) = e(t), \quad \mathbf{E}(\xi(t, s) - e(t))(\xi(t', s') - e(t')) = \frac{1}{N^2} v^{ss'}(e^{it'}, e^{it'}) \quad (2.20)$$

with v^{sr} from (2.7). Note that (2.20) also specifies the covariance of $\xi(t, s)$ and $\overline{\xi(t', s')} = \xi(-t', s')$ for different times.

The next lemma, to be proven in Section 3.2.4, provides explicit asymptotic formulas for $v_{\pm}^{ss}(t)$, in particular they imply the asymptotics in (2.14) together with $e(t) \sim t^{-3/2}$ (up to some oscillation due to the Bessel function) in the large t regime.

Lemma 2.12. For $s = r$ the functions $v_{\pm}^{ss}(t)$ appearing in (2.12) can be expressed as

$$\begin{aligned} v_{-}^{ss}(t) &= t^2 \left[J_0(2t)^2 + 2J_1(2t)^2 - J_0(2t)J_2(2t) \right] = \frac{2t}{\pi} - \frac{1 + 2\sin(4t)}{16\pi t} + \mathcal{O}(t^{-2}) \\ v_{+}^{ss}(t) &= -tJ_0(2t)J_1(2t) = \frac{\cos(4t)}{2\pi} - \frac{2 + \sin(4t)}{16\pi t} + \mathcal{O}(t^{-2}). \end{aligned} \quad (2.21)$$

The relation in (2.17) requires a stationary phase calculation that will be done separately in Section 5.

2.3. Implications for sampling. Determining the standard deviation of $|\langle e^{itH} \rangle|^2$ is important for numerical testing of (2.13). By taking the empirical average \mathbf{E}_H^n of n independent Wigner matrices we may approximate the true expectation $\mathbf{E}_H|\langle e^{itH} \rangle|^2$ at a speed

$$\mathbf{E}_H^n|\langle e^{itH} \rangle|^2 = \mathbf{E}_H|\langle e^{itH} \rangle|^2 + \Omega\left(n^{-1/2} \mathbf{Std}_H|\langle e^{itH} \rangle|^2\right) = E_{\text{wig}}(t) + \Omega(n^{-1/2} S_{\text{wig}}(t)), \quad (2.22)$$

c.f. the top of Figure 3. Here $\Omega(\dots)$ indicates an oscillatory error term of the given size. In the *ramp regime* the fluctuation of $\mathbf{E}_H^n|\langle e^{itH} \rangle|^2$ thus scales like $t/(\sqrt{n}N^2)$ using (2.14). In particular, this fluctuation vanishes as the sample size n goes to infinite, hence the statistics via sampling to test (2.13) is *consistent*.

In contrast, for the monoparametric ensemble, by taking the empirical average of n copies of s we naturally have

$$\mathbf{E}_s^n|\langle e^{itH^s} \rangle|^2 = \mathbf{E}_s|\langle e^{itH^s} \rangle|^2 + \Omega\left(k^{-1/2} S_{\text{wig}}(t)\right). \quad (2.23)$$

Replacing the first term by its expectation plus its fluctuation in the H -probability space, we also get

$$\mathbf{E}_s^n|\langle e^{itH^s} \rangle|^2 = \mathbf{E}_H \mathbf{E}_s|\langle e^{itH^s} \rangle|^2 + \Omega\left(\max(n^{-1/2} S_{\text{wig}}(t), S_{\text{res}}(t))\right), \quad (2.24)$$

where the error term contains both standard deviations and satisfies

$$\max(n^{-1/2} S_{\text{wig}}(t), S_{\text{res}}(t)) \sim \begin{cases} \frac{1}{Nt} \max\left\{\frac{1}{\sqrt{n}}, \frac{1}{t^{1/4}}\right\}, & 1 \ll t \ll \sqrt{N} \\ \frac{t}{N^2} \max\left\{\frac{1}{\sqrt{n}}, \frac{1}{t^{1/4}}\right\} & \sqrt{N} \ll t \ll N, \end{cases} \quad (2.25)$$

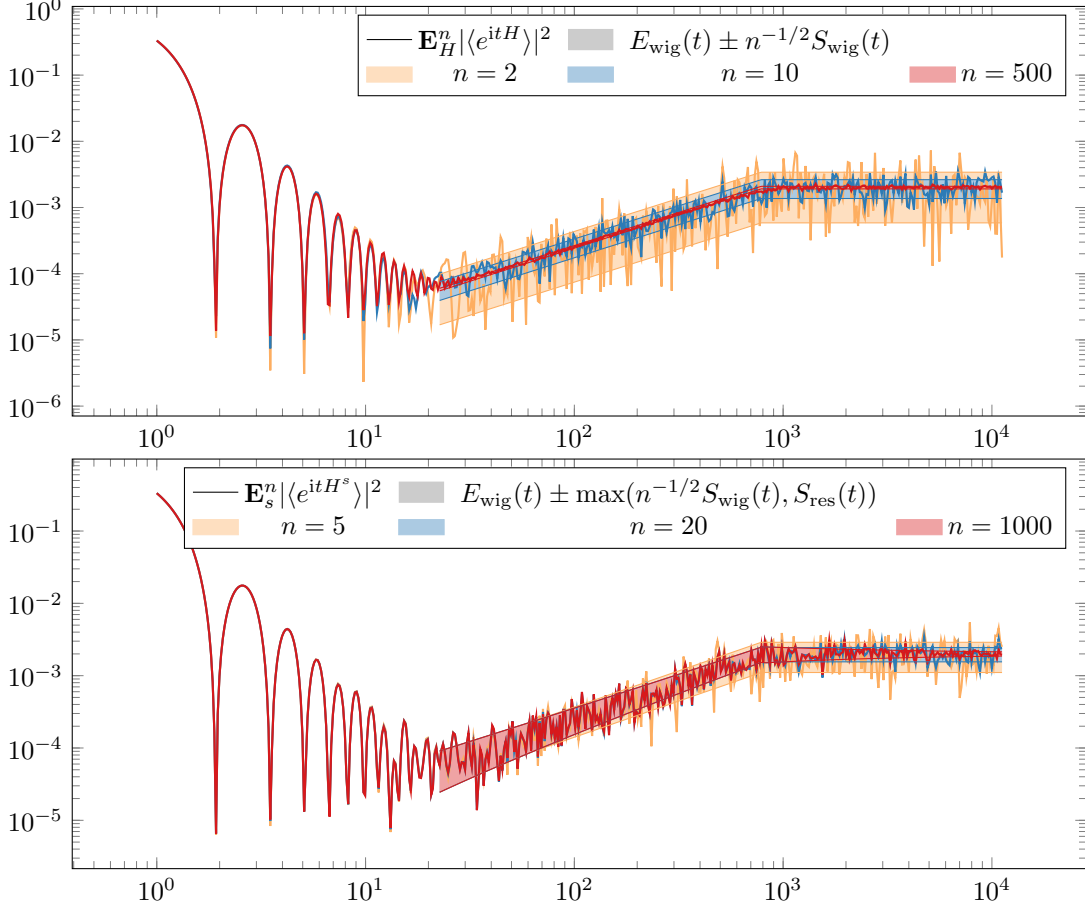


FIGURE 3. In the first plot we show the empirical mean of $|\langle e^{itH} \rangle|^2$ for k independent GUE matrices H . As expected the standard deviation of the sample average fluctuates within a strip of width $n^{-1/2} \mathbf{Std}_H |\langle e^{itH} \rangle|^2$, in particular the sample average exactly reproduces the mean if $n \rightarrow \infty$. In the second plot we show the empirical mean of $|\langle e^{itH^s} \rangle|^2$ for k independently sampled scalar random variables s for a fixed GUE matrix pair H_1, H_2 . We observe that while the sample mean approximates the true mean \mathbf{E}_s increasingly well as $n \rightarrow \infty$, the latter is still dependent on the chosen realisation of H_1, H_2 . Thus the empirical mean fluctuates in a strip of width $\max(n^{-1/2} S_{\text{wig}}(t), S_{\text{res}}(t))$ around the doubly averaged $\mathbf{E}_H \mathbf{E}_s |\langle e^{itH^s} \rangle|^2$.

due to (2.15) and (2.17). In particular, both in the slope and in the ramp regimes the size of the fluctuation of $\mathbf{E}_s^n |\langle e^{itH^s} \rangle|^2$ does not vanish even as the number of samples goes to infinity, $n \rightarrow \infty$, hence the statistics is *not consistent*, c.f. the bottom of Figure 3. However, this lack of consistency,

expressed by $S_{\text{res}}(t)$ is still negligible compared with the leading first term in (2.24) by a factor $t^{-1/4} \ll 1$ in the large t regime, see (2.19). We recall that mathematically rigorously we can prove all these facts only for $t \ll N^{1/4}$, i.e. well before the dip time, but the numerical tests leave no doubt on their validity in the entire regime $1 \ll t \ll N$.

2.4. Extensions. Beside the Wigner ensemble, we formulated our main results on SFF for the normalized monoparametric model in Theorem 2.8. We chose this model for definiteness, but our approach applies to the multi-parametric as well as to the un-normalised models introduced in Remark 2.4. Here we explain the modified results for these natural generalisations.

First, for the multi-parametric normalised model, $H^s = s_1 H_1 + \dots + s_k H_k$ with $k - 1$ effective parameters $s \in S^{k-1}$, Theorem 2.8 holds true verbatim modulo different sizes for the residual standard deviation $S_{\text{res}}(t)$. In fact, we have

$$S_{\text{res}}(t) \lesssim t^{-\frac{1}{2} + \frac{1}{4}(3-k)_+} S_{\text{wig}}(t), \quad (2.26)$$

see (5.4) later, hence $S_{\text{res}}(t)$ becomes less relevant compared with $S_{\text{wig}}(t)$ for larger $k > 2$, see (2.19). Consequently, the upper bounds on the times of proven validity in (2.15) slightly improve but they still remain below the dip time and we omit the precise formulas. We note that the t -power in (2.26) is not optimal for $k \geq 3$. A refined stationary phase estimate could be used to improve the estimate but we refrain from doing so since our primary interest is the mono-parametric model with few degrees of freedom.

Second, for the un-normalised model $H^s = s_1 H_1 + s_2 H_2$ with two effective parameters $s \in \mathbf{R}^2$, Theorem 2.8 also holds true modulo some minor changes. More precisely, (2.15) becomes

$$\begin{aligned} \mathbf{E}_H \mathbf{E}_s | \langle e^{itH^s} \rangle|^2 &= \mathbf{E}_s E_{\text{wig}}(\|s\|_2 t) (1 + o(1)) \quad \text{for } t \ll N^{3/7} \\ \mathbf{E}_H \mathbf{Var}_s | \langle e^{itH^s} \rangle|^2 &= \left(\mathbf{E}_s S_{\text{wig}}(\|s\|_2 t)^2 - \tilde{S}_{\text{res}}(t)^2 \right) (1 + o(1)) \quad \text{for } t \ll N^{5/17} \\ \mathbf{Var}_H \mathbf{E}_s | \langle e^{itH^s} \rangle|^2 &= \tilde{S}_{\text{res}}(t)^2 (1 + o(1)) \quad \text{for } t \ll N^{1/7}, \end{aligned} \quad (2.27)$$

with \tilde{S}_{res} obtained from replacing v^{sr} by \tilde{v}^{sr} from Remark 2.6 in (2.12). For $\tilde{S}_{\text{res}}(t)$ a stationary phase calculation gives the modified

$$\tilde{S}_{\text{res}}(t) \sim \begin{cases} \frac{\psi(t)}{N t^{7/4}}, & 1 \ll t \ll \sqrt{N} \\ \frac{t^{1/4}}{N^2}, & \sqrt{N} \ll t \ll N, \end{cases} \quad (2.28)$$

assuming that s has an absolutely continuous distribution with a differentiable, compactly supported density ρ on \mathbf{R}^2 with $\rho(0) = 0$. We will not prove the relation in these formulas in this paper, we only show how to obtain the necessary upper bound on them at the end of Section 5.

Note that now

$$\tilde{S}_{\text{res}}(t) \lesssim t^{-3/4} \mathbf{E}_s S_{\text{wig}}(\|s\|_2 t), \quad (2.29)$$

i.e. the fluctuation due to the residual randomness of (H_1, H_2) after taking the expectation in s remains negligible, in fact it is reduced compared with the normalised case (2.19). As a consequence $t^{1/4}$ in (2.25) is replaced by $t^{3/4}$.

Analogue results hold for the most general multi-parametric un-normalised model as well as to the mono-parametric inhomogeneous model $H^s = H_1 + s H_2$, $s \in \mathbf{R}$. We omit their precise formulation, the key point is that the analogue of (2.27) hold in all cases with a residual standard deviation $\tilde{S}_{\text{res}}(t)$ being smaller than the leading term $S_{\text{wig}}(t)$ by a polynomial factor in t (e.g. by

$t^{-1/2}$ for $H^s = H_1 + sH_2$). This guarantees that the universality of SFF holds for all these models. Table 1 summarizes the decay exponents of our main parametric models.

TABLE 1. For our three main parametric models the following table lists the size of the residual fluctuation compared to the fluctuation of the Wigner-SFF.

Quenched parametric model	Randomness
$s_1 H_1 + s_2 H_2$	$(s_1, s_2) \in S^1$ $S_{\text{res}}(t) \lesssim t^{-1/4} S_{\text{wig}}(t)$
$H_1 + s H_2$	$s \in \mathbf{R}$ $S_{\text{res}}(t) \lesssim t^{-1/2} S_{\text{wig}}(t)$
$s_1 H_1 + s_2 H_2$	$(s_1, s_2) \in \mathbf{R}^2$ $S_{\text{res}}(t) \lesssim t^{-3/4} S_{\text{wig}}(t)$

Outline. The rest of the paper is organised as follows. In Section 3 we outline the resolvent method and explain how via the Helffer-Sjöstrand representation a resolvent CLT implies the CLT for the linear statistics $\sum f(\lambda_i)$ of arbitrary test functions f from which our main results Theorems 2.5–2.8 follow. In Section 4 we present the proof of the resolvent CLT, while in Section 5 we conclude the proof of the asymptotics (2.17) via a stationary phase argument.

3. RESOLVENT METHOD

Let H be a Wigner matrix⁵ and $G(z) := (H - z)^{-1}$ its resolvent with a spectral parameter $z \in \mathbf{C} \setminus \mathbf{R}$. Define $m_{\text{sc}}(z)$, the Stieltjes transform of the semicircle law:

$$m(z) = m_{\text{sc}}(z) := \int_{\mathbf{R}} \frac{\rho_{\text{sc}}(x)}{x - z} dx, \quad \rho_{\text{sc}}(x) := \frac{\sqrt{(4 - x^2)_+}}{2\pi}. \quad (3.1)$$

The *local law* for a single resolvent states that the diagonal matrix $m(z) \cdot I$ well approximates the random resolvent $G(z)$ in the following sense (see e.g. [5, 20, 35]):

$$|\langle (G(z) - m(z))A \rangle| \lesssim N^\xi \frac{\|A\|}{N\eta}, \quad \langle \mathbf{x}, (G(z) - m(z))\mathbf{y} \rangle \lesssim N^\xi \frac{\|\mathbf{x}\| \|\mathbf{y}\|}{\sqrt{N\eta}} \quad (3.2)$$

with $\eta = |\Im z|$, for any fixed deterministic matrix A and deterministic vectors \mathbf{x}, \mathbf{y} . The first bound is called *averaged* local law, while the second one is the *isotropic* local law. The bounds (3.2) are understood in very high probability for any fixed $\xi > 0$.

The Helffer-Sjöstrand formula

$$\langle f(H) \rangle = \frac{2}{\pi} \int_{\mathbf{C}} \partial_{\bar{z}} f_{\mathbf{C}}(z) \langle G(z) \rangle d^2 z, \quad (3.3)$$

with $z = x + i\eta$ and $d^2 z := d\eta dx$, expresses the linear statistics of arbitrary functions as an integral of the resolvent $G(z)$ and the almost-analytic extension

$$f_{\mathbf{C}}(z) = f_{\mathbf{C}}(x + i\eta) := [f(x) + i\eta \partial_x f(x)] \chi(\tau\eta), \quad (3.4)$$

of f . Here the free parameter $\tau \in \mathbf{R}$ is chosen such that $N^{-1} \ll \tau^{-1} \lesssim 1$, and χ a smooth cut-off equal to 1 on $[-5, 5]$ and equal to 0 on $[-10, 10]^c$. The same τ was used to define the weighted

⁵The resolvent method extends to very general Hermitian matrices possibly with non-centered and correlated entries, see [19], but here we present only the Wigner case for simplicity.

H^2 -norm (2.4) and eventually we will optimize its value, a procedure that improves the standard error terms in the CLT. By (3.2) it follows that

$$\langle f(H) \rangle = \frac{2}{\pi} \int_{\mathbf{C}} \partial_{\bar{z}} f_{\mathbf{C}}(z) m(z) d^2 z + \mathcal{O}\left(N^\xi \frac{\|f\|_{H^2}}{N}\right) = \int_{-2}^2 \rho_{\text{sc}}(x) f(x) dx + \mathcal{O}\left(N^\xi \frac{\|f\|_{H^2}}{N}\right). \quad (3.5)$$

In order to compute the fluctuation in (3.5) via (3.3) we need to understand the correlation between $\langle G(z) \rangle, \langle G(z') \rangle$ for two different spectral parameters z, z' which turns out to be given by

$$\mathbf{Cov}(\langle G(z) \rangle, \langle G(z') \rangle) \approx \frac{1}{N^2} \frac{\langle G(z)^2 \rangle \langle G(z')^2 \rangle \langle G(z)G(z') \rangle (1 + \langle G(z)G(z') \rangle)}{\langle G(z) \rangle \langle G(z') \rangle}, \quad (3.6)$$

modulo some additional contribution from non-Gaussian fourth cumulant, see (3.8) for the final statement. While $G(z) \approx m(z)$, in general it is not true that $G(z)G(z') \approx m(z)m(z')$ since (3.2) allows only deterministic test matrices multiplying G . Nevertheless $G(z)G(z')$ is still approximable by a deterministic object:

$$G(z)G(z') \approx \frac{m(z)m(z')}{1 - m(z)m(z')}. \quad (3.7)$$

Statements of the form (3.7) with an appropriate error term are called *multi-resolvent local laws*.

We will apply this theory to the product of the resolvents G^s of $H^s = s_1 H_1 + s_2 H_2$ for two different parameters s , see the corresponding local law on $\langle G^s G^r \rangle$ in (3.11) later. Even though H_1 and H_2 as well as s and r are independent, the common (H_1, H_2) ingredients in H^s and H^r introduce a nontrivial correlation between these matrices. We therefore need to extend CLT for resolvents via multi-resolvent local laws to this parametric situation.

3.1. Resolvent CLT. The main technical result of the present paper is the following Central Limit Theorem for product of resolvents of the random matrix $H^s := s_1 H_1 + s_2 H_2$ with $s = (s_1, s_2) \in S^1$.

Proposition 3.1. *Fix $\epsilon > 0$, $p \in \mathbf{N}$, $s^1, \dots, s^p \in S^1$, $z_1, \dots, z_p \in \mathbf{C} \setminus \mathbf{R}$, and define $G_i := (H^{s^i} - z_i)^{-1}$. Then for any arbitrary small $\xi > 0$ and $\eta_* \geq N^{-1+\epsilon}$ it holds*

$$\mathbf{E}_H \prod_{i \in [p]} \langle G_i - \mathbf{E}_H G_i \rangle = \frac{1}{N^p} \sum_{P \in \text{Pair}([p])} \prod_{(i,j) \in P} V_{ij} + \mathcal{O}\left(N^\xi \Psi_p \left(\frac{1}{L^{1/2}} + \frac{1}{N\eta_*^2} + \frac{1}{N^2\eta_*^4}\right)\right), \quad \Psi_p := \prod_{i \in [p]} \frac{1}{N|\eta_i|}. \quad (3.8)$$

Here $\eta_i := \Im z_i$, $\eta_* := \min_i |\eta_i|$, $L := \min_i (N\eta_i \rho_i)$, and

$$V_{ij} := -\frac{2}{\beta} \partial_{z_i} \partial_{z_j} \log \left(1 - \langle s^i, s^j \rangle m_i m_j\right) - \langle s^i \odot s^i, s^j \odot s^j \rangle \kappa_4(m_i^2)'(m_j^2)', \quad (3.9)$$

where $m_i := m_{\text{sc}}(z_i)$, and $\kappa_4 := N^2 \mathbf{E} |h_{12}|^4 - 1 - 2/\beta$. Additionally, for the expectation we have

$$\mathbf{E}_H \langle G_i \rangle = m_i + \frac{\kappa_4}{N} \|s\|_4^4 m_i' m_i^3 + \mathbf{1}(\beta = 1) \frac{1}{N} \frac{m_i m_i'}{1 - m_i^2} + \mathcal{O}\left(N^\xi \frac{\sqrt{\rho_i}}{(N\eta_i)^{3/2}}\right), \quad (3.10)$$

with $\rho_i := \pi^{-1} |\Im m_i|$.

Remark 3.2. *For Wigner matrices, i.e. for $s^1 = \dots = s^p = (1, 0)$, the error term in (3.8) is given by $\Psi L^{-1/2}$, as a consequence of the fact that the error terms in the first and second line of (3.11) are replaced by $(N\eta_1 \eta_2)^{-1}$ and $(N\eta_1 \eta_2^2)^{-1}$, respectively (see e.g. [16, Remark 3.5]).*

We point out that similar resolvent CLT have often been used as a basic input to prove CLT for linear eigenvalue statistics of both Hermitian and non-Hermitian matrices down to optimal mesoscopic scales (see e.g. [11, 10, 14, 28, 29, 30, 37, 38, 36]). The main novelty here is to extend the resolvent CLT to the monoparametric ensemble.

Along the proof of Proposition 3.1 we establish the following multi-resolvent local laws.

Lemma 3.3. *For $G_i = G^{s^i}(z_i)$ we have the two- and three-resolvent local laws*

$$\begin{aligned} \left| \langle G_1 G_2 \rangle - \frac{m_1 m_2}{1 - \langle s^1, s^2 \rangle m_1 m_2} \right| &\lesssim \frac{N^\xi}{N |\eta_1 \eta_2|^{3/2}} \\ \left| \langle G_1 G_2^2 \rangle - \frac{m_1 m_2'}{(1 - \langle s^1, s^2 \rangle m_1 m_2)^2} \right| &\lesssim \frac{N^\xi}{N |\eta_1| |\eta_2| \eta_*^2} + \frac{1}{N^2 |\eta_1 \eta_2|^3}, \end{aligned} \quad (3.11)$$

where $m_i = m_{\text{sc}}(z_i)$, with very high probability for any fixed $\xi, \epsilon > 0$ and $|\Im z_i| \geq N^{-1+\epsilon}$.

The proofs of Proposition 3.1 and Lemma 3.3 will be presented in Section 4. In these proofs we will often use the standard *cumulant expansion* (see [8, 30, 34] in the random matrix context):

$$\mathbf{E}_H h_{ab} f(H) = \frac{1}{N} \mathbf{E}_H \partial_{ba} f(H) + \sum_{k=2}^R \sum_{q+q'=k} \frac{\kappa_{ab}^{q+1, q'}}{N^{(k+1)/2}} \mathbf{E}_H \partial_{ab}^q \partial_{ba}^{q'} f(H) + \Omega_R. \quad (3.12)$$

Here ∂_{ab} denotes the directional derivative $\partial_{h_{ab}}$, the first term in the rhs. represents the second order (Gaussian) contribution, while the sum in (3.12) represents the non-Gaussian contribution with $\kappa_{ab}^{p,q}$ denoting the joint cumulant of p copies of $N^{1/2} h_{ab}$ and q copies of $N^{1/2} \overline{h_{ab}}$. The cumulant expansion is typically truncated at a high (N -independent) level R with an error term Ω_R that is negligible. To see this, note that in our applications f will be a product of resolvents at spectral parameters z_i with $\eta_* = \min |\Im z_i| \gg 1/N$ hence derivatives of f remain bounded with very high probability by the isotropic local law (3.2) thus the tail of the series (3.12) decays as $N^{-(k+1)/2}$.

3.2. Proof of Theorem 2.5. The proof of Theorem 2.5 is divided into three steps: (i) computation of the expectation, (ii) computation of the variance, (iii) proof of Wick Theorem. The expectation is computed in Section 3.2.1, while the Wick Theorem and the explicit computation of the variance are proven in Section 3.2.2.

3.2.1. Expectation. Using the bound

$$|\partial_{\bar{z}} f_{\mathbf{C}}| \lesssim \eta |f''| + \tau |\chi'| [|f| + i\eta |f'|], \quad (3.13)$$

and $|\langle G^s - m \rangle| \lesssim N^\xi (N\eta)^{-1}$ by (3.2), with $m = m_{\text{sc}}$, we conclude that

$$\mathbf{E}_H \langle f(H^s) \rangle = \int_{\mathbf{R}} \int_{|\eta| \geq \eta_0} \partial_{\bar{z}} f_{\mathbf{C}}(z) \mathbf{E}_H \langle G^s(z) \rangle d\eta dx + \mathcal{O} \left(\frac{N^\xi \eta_0 \|f\|_{H^2}}{N} + N^\xi \eta_0^2 \|f\|_{H^2} \right), \quad (3.14)$$

for any $N^{-1} \ll \eta_0 \ll \tau^{-1}$. Note that we chose $\eta_0 \gg N^{-1}$ in order to use Proposition 3.1.

Plugging (3.10) into (3.14), and using (3.13) to estimate the error term, we get that

$$\begin{aligned}
\mathbf{E}_H \langle f(H^s) \rangle &= \int_{\mathbf{R}} \int_{|\eta| \geq \eta_0} \partial_{\bar{z}} f_{\mathbf{C}}(z) \left[m + \frac{\kappa_4}{N} \|s\|_4^4 m' m^3 + \mathbf{1}(\beta = 1) \frac{1}{N} \frac{mm'}{1 - m^2} \right] d\eta dx \\
&\quad + \mathcal{O} \left(\frac{N^\xi \eta_0 \|f\|_{H^2}}{N} + N^\xi \eta_0^2 \|f\|_{H^2} + \frac{N^\xi \|f\|_{H^2}}{N^{3/2} \tau^{1/2}} + \frac{N^\xi \tau^{3/2} \|f\|_\infty}{N^{3/2}} + \frac{N^\xi \tau^{1/2} \|f\|_{H^1}}{N^{3/2}} \right) \\
&= \int_{\mathbf{R}} \int_{|\eta| \geq \eta_0} \partial_{\bar{z}} f_{\mathbf{C}}(z) \left[m + \frac{\kappa_4}{N} \|s\|_4^4 m' m^3 + \mathbf{1}(\beta = 1) \frac{1}{N} \frac{mm'}{1 - m^2} \right] d\eta dx + \mathcal{O} \left(\frac{N^\xi \|f\|_\tau}{N^{3/2} \tau^{1/2}} \right),
\end{aligned} \tag{3.15}$$

where to go to the last line we chose $\eta_0 \sim N^{-1+\epsilon}$, for some very small $\epsilon > 0$, and we used the norm $\|f\|_\tau$ defined in (2.4).

Adding back the regime $|\eta| < \eta_0$ at the price of a negligible error smaller than the one in (3.15), by explicit computations (exactly as in [13, Section D.1]) in the leading term of (3.15), we conclude

$$\begin{aligned}
\mathbf{E}_H \langle f(H^s) \rangle &= \int_{-2}^2 \rho_{\text{sc}}(x) f(x) dx + \frac{\kappa_4}{2N} \|s\|_4^4 \int_{-2}^2 \frac{x^4 - 4x^2 + 2}{\pi \sqrt{4 - x^2}} f(x) dx \\
&\quad + \mathbf{1}(\beta = 1) \left[\frac{f(2) + f(-2)}{4N} - \frac{1}{2\pi N} \int_{-2}^2 \frac{f(x)}{\sqrt{4 - x^2}} dx \right] + \mathcal{O} \left(\frac{N^\xi \|f\|_\tau}{N^{3/2} \tau^{1/2}} \right).
\end{aligned} \tag{3.16}$$

3.2.2. *Second moment and Wick theorem.* Define

$$L_N(f, s) := N[\langle f(H^s) \rangle - \mathbf{E}_H \langle f(H^s) \rangle], \tag{3.17}$$

then in this section, using Proposition 3.1, we compute the leading order term of $\mathbf{E}_H L_N(f_1, s^1) L_N(f_2, s^2)$. More precisely, by (3.8) for $p = 2$, and using (3.13) to estimate the error term, it follows that

$$\begin{aligned}
&\mathbf{E}_H L_N(f_1, s^1) L_N(f_2, s^2) \\
&= \iiint_{\mathbf{R}} \iiint_{|\eta_1|, |\eta_2| \geq \eta_0} \partial_{\bar{z}_1} f_{\mathbf{C}}(z_1) \partial_{\bar{z}_2} f_{\mathbf{C}}(z_2) V_{12} \\
&\quad + \mathcal{O} \left(N^\xi \eta_0 (\|f_1\|_{H^2} \|f_2\|_\infty + \|f_2\|_{H^2} \|f_1\|_\infty) + \frac{N^\xi \|f_1\|_\tau \|f_2\|_\tau}{N^{1/2} \tau^{3/2}} + \frac{\|f_1\|_{H^2} \|f_2\|_{H^2}}{N^{1-\xi} \eta_0 \tau} \left(1 + \frac{1}{N \eta_0^2} \right) \right. \\
&\quad \quad + \frac{(\|f_1\|_{H^2} (\tau^2 \|f_2\|_\infty + \tau \|f_2\|_{H^1}) + \|f_2\|_{H^2} (\tau^2 \|f_1\|_\infty + \tau \|f_1\|_{H^1}))}{N^{1-\xi} \eta_0 \tau} \left(1 + \frac{1}{N \eta_0^2} \right) \\
&\quad \quad \left. + \frac{(\tau^2 \|f_1\|_\infty + \tau \|f_1\|_{H^1}) (\tau^2 \|f_2\|_\infty + \tau \|f_2\|_{H^1})}{N} \left(1 + \frac{\tau^2}{N^{1-2\epsilon}} \right) \right) \\
&= \iiint_{\mathbf{R}} \iiint_{|\eta_1|, |\eta_2| \geq N^{-\epsilon} \tau^{-1}} \partial_{\bar{z}_1} f_{\mathbf{C}}(z_1) \partial_{\bar{z}_2} f_{\mathbf{C}}(z_2) V_{12} \\
&\quad + \mathcal{O} \left(N^\xi \|f_1\|_\tau \|f_2\|_\tau \left(\frac{N^\epsilon}{N} + \frac{N^{-\epsilon}}{\tau^3} \right) \left(1 + \frac{\tau^2}{N^{1-2\epsilon}} \right) \right),
\end{aligned} \tag{3.18}$$

where to go to the last line we chose $\eta_0 \sim N^{-\epsilon} \tau^{-1}$, for any $\epsilon > 0$, and V_{12} is defined in (3.9). From (3.18), adding back the regimes $|\eta_i| < N^{-\epsilon} \tau^{-1}$ at the price of an error smaller than the one in the

last line of (3.18), we conclude (2.6) for $p = 2$ by explicit computation in deterministic term as in [13, Section D.2].

We conclude this section with the computation of higher moments:

$$\begin{aligned} \mathbf{E}_H \prod_{i \in [p]} L_N(f_i, s^i) &= \sum_{P \in \text{Pair}([p])} \prod_{(i,j) \in P} \iint_{\mathbf{R}} \iint_{|\eta_i|, |\eta_j| \geq N^{-\epsilon}} \partial_{\bar{z}_i} f_{\mathbf{C}}(z_i) \partial_{\bar{z}_j} f_{\mathbf{C}}(z_j) V_{ij} \\ &+ \mathcal{O} \left(\left(\frac{N^\xi}{N^{1/2} \tau^{3/2}} + \frac{N^\epsilon}{N} + \frac{N^{-\epsilon}}{\tau^{2p-1}} \right) \left(1 + \frac{\tau^2}{N^{1-2\epsilon}} \right) \prod_{i \in [p]} \|f_i\|_\tau \right), \end{aligned} \quad (3.19)$$

which concludes the proof of (2.6) for any $p \in \mathbf{N}$, after adding back the regimes $|\eta_i| < N^{-\epsilon} \tau^{-1}$ at the price of an error smaller than the one in the second line of (3.19).

3.2.3. Proof of Theorems 2.7 and 2.8. We just show how Theorem 2.7 follows by Theorem 2.5; the proof of Theorems 2.8 is completely analogous and so omitted. In particular, to make the presentation shorter we just show the details of the proof of the first equation in (2.13). Using Theorem 2.5 as an input, the proof of the second equation in (2.13) follows exactly in the same way.

First of all we write

$$\mathbf{E}_H |\langle e^{itH} \rangle|^2 = \mathbf{E}_H |\langle e^{itH} \rangle - \mathbf{E}_H \langle e^{itH} \rangle|^2 + |\mathbf{E}_H \langle e^{itH} \rangle|^2. \quad (3.20)$$

Then, using (2.6) with $p = 2$, $f_1(x) = e^{itx}$, $f_2(x) = e^{-itx}$, and $\tau = t$ to compute the leading order of the first term in (3.20), and (2.5) with $f(x) = e^{itx}$ to compute the leading order of the second term in (3.20), we conclude that

$$\mathbf{E}_H |\langle e^{itH} \rangle|^2 = E_{\text{wig}}(t) + \mathcal{O} \left(\frac{1}{N^{3/2}} + \frac{t^{5/2}}{N^{5/2}} \right), \quad (3.21)$$

with $E_{\text{wig}}(t)$ defined in (2.14). Finally, using the asymptotics of $E_{\text{wig}}(t)$ in (2.14) we readily conclude that the error term in (3.21) is much smaller than the leading term $E_{\text{wig}}(t)$ as long as $t \ll N^{5/11}$.

3.2.4. Variance calculations when $s = r$ and the proof of Lemma 2.12. We note that (2.7) generalises the standard variance calculation yielding (1.7) to $s \neq r$. For the case $s = r$ the two formulas can be seen to be equivalent using the identity

$$\begin{aligned} \frac{1}{2\pi^2} \iint_{-2}^2 f'(x) g'(y) \log \left| \frac{1 - m_{\text{sc}}(x) \overline{m_{\text{sc}}(y)}}{1 - m_{\text{sc}}(x) m_{\text{sc}}(y)} \right| dx dy \\ = \frac{1}{4\pi^2} \iint_{-2}^2 \frac{f(x) - f(y)}{x - y} \frac{g(x) - g(y)}{x - y} \frac{4 - xy}{\sqrt{4 - x^2} \sqrt{4 - y^2}} dx dy \end{aligned} \quad (3.22)$$

that can be proven by integration by parts and using $(m_{\text{sc}}(x) + x)m_{\text{sc}}(x) = -1$ from the explicit form of $m_{\text{sc}}(x)$ from (3.1).

Proof of Lemma 2.12. Using (3.22) the functions $v_{\pm}^{ss}(t)$ appearing in (2.12) can be expressed as

$$\begin{aligned} v_{-}^{ss}(t) &= \frac{1}{\pi^2} \int_{-1}^1 \int_{-1}^1 \frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}} \left(\frac{\sin(t(x-y))}{x-y} \right)^2 dx dy \\ &= \sum_{k=1}^{\infty} k J_k(2t)^2 = t^2 \left[J_0(2t)^2 + 2J_1(2t)^2 - J_0(2t)J_2(2t) \right] \\ &= \frac{2t}{\pi} - \frac{1+2\sin(4t)}{16\pi t} + \mathcal{O}(t^{-2}) \end{aligned} \quad (3.23)$$

and

$$\begin{aligned} v_{+}^{ss}(t) &= \frac{1}{4\pi^2} \int_{-1}^1 \int_{-1}^1 \frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}} \left(\frac{e^{2itx} - e^{2ity}}{x-y} \right)^2 dx dy \\ &= \sum_{k=1}^{\infty} (-1)^k k J_k(2t)^2 = -t J_0(2t) J_1(2t) \\ &= \frac{\cos(4t)}{2\pi} - \frac{2+\sin(4t)}{16\pi t} + \mathcal{O}(t^{-2}) \end{aligned} \quad (3.24)$$

where the series representations follow directly from [13, Remark 2.6] and the series evaluations follow from [50, V.§5.51(1)]. \square

4. CENTRAL LIMIT THEOREM FOR RESOLVENTS

The proof of Proposition 3.1 is divided into three parts: in Section 4.1 we compute the subleading order correction to $\mathbf{E}_H \langle G_i \rangle$, in Section 4.2 we explicitly compute the variance, and finally in Section 4.3 we prove a Wick Theorem. To keep our presentation simpler we only prove the CLT for resolvent in the complex case, the real case is completely analogous and so omitted (see e.g. [13, Section 4]).

4.1. Computation of the expectation. For $G = G^s(z)$ we have

$$I = s_1 \underline{H_1 G} + s_2 \underline{H_2 G} - \langle G \rangle G - zG, \quad \underline{H_i G} := H_i G + s_i \langle G \rangle G \quad (4.1)$$

so that $G \approx m$ for the solution m to the equation

$$-\frac{1}{m} = z + m, \quad m(z) = m_{\text{sc}}(z). \quad (4.2)$$

The fact that $G \approx m$ in averaged and isotropic sense follows from the single resolvent local law (3.2). This is a consequence of the fact that the term $\underline{H_i G}$ in (4.1) is designed in such a way $\mathbf{E} \underline{H_i G} \approx 0$ in averaged and isotropic sense. In fact, for Gaussian ensembles $\mathbf{E} \underline{H_i G} = 0$ and the deviation from zero for general ensembles is a lower order effect due to non-vanishing of higher order cumulants of the entry distribution. From (4.1) and (4.2) we obtain

$$(1 - m^2 \langle \cdot \rangle) [G - m] = -m (s_1 \underline{H_1 G} + s_2 \underline{H_2 G}) + m \langle G - m \rangle (G - m). \quad (4.3)$$

Additionally, we define $\rho(z) := \pi^{-1} |\Im m(z)|$. For simplicity of notation from now on we assume that $\Im z > 0$. We remark that by $1 - m^2 \langle \cdot \rangle$ in the lhs. of (4.3) we denote the operator acting on matrices $R \in \mathbf{C}^{N \times N}$ as $(1 - m^2 \langle \cdot \rangle) [R] = R - m^2 \langle R \rangle$.

We then start computing:

$$\mathbf{E}_H \langle G - m \rangle = -\frac{m'}{m} \mathbf{E}_H \langle s_1 \underline{H_1 G} + s_2 \underline{H_2 G} \rangle + \mathcal{O} \left(\frac{N^\xi}{N^2 \eta^2 \rho} \right), \quad (4.4)$$

for any small $\xi > 0$, where we used that $|1 - m^2| \gtrsim \rho$, that $m' = m^2/(1 - m^2)$, and that $|\langle G - m \rangle| \lesssim N^\xi (N\eta)^{-1}$ by (3.2). Then using cumulant expansion (see (3.12), ignoring the truncation error) we claim (and prove below) that

$$\begin{aligned} \mathbf{E}_H \langle s_1 \underline{H_1 G} + s_2 \underline{H_2 G} \rangle &= \mathbf{E}_H \frac{1}{N} \sum_{k \geq 2} \sum_{ab} \sum_{\alpha \in \{ab, ba\}^k} \left(\frac{\kappa^{(1)}(ab, \alpha)}{k!} s_1 \partial_\alpha^{(1)} + \frac{\kappa^{(2)}(ab, \alpha)}{k!} s_2 \partial_\alpha^{(2)} \right) G_{ba} \\ &= \frac{\kappa_4}{N} \|s\|_4^4 m^4 + \mathcal{O} \left(\frac{N^\xi \rho^{3/2}}{N^{3/2} \eta^{1/2}} + \frac{N^\xi \rho^{3/2}}{N^2 \eta^{3/2}} \right), \end{aligned} \quad (4.5)$$

where $\kappa^{(i)}(ab, \alpha)$ denotes the joint cumulant of the random variables $h_{ab}^i, h_{\alpha_1}^i, \dots, h_{\alpha_k}^i$, and $\partial_\alpha^{(i)} := \partial_{\alpha_1}^{(i)} \dots \partial_{\alpha_k}^{(i)}$, with $i = 1, 2$, where $\partial_{\alpha_j}^{(i)}$ denotes the directional derivative in the direction $h_{\alpha_j}^i$. Here $h_{\alpha_j}^i$ are the entries of H_i . Combining (4.5) with (4.4) we obtain exactly the expansion in (3.10) (recall that here we only present the proof in the complex case, the real case being completely analogous).

Proof of the second equality in (4.5). First of all we recall that by (2.1) it follows the bound $|\kappa^{(i)}(ab, \alpha)| \lesssim N^{-(k+1)/2}$, with $i = 1, 2$.

We start with $k = 2$. In this case we can neglect the summation when $a = b$ since it gives a contribution $N^{-3/2}$. Hence we can assume that $a \neq b$. In this case we have the bounds

$$N^{-5/2} \left| \sum_{a \neq b} G_{ab}^3 \right| \lesssim \frac{N^\xi \rho^{3/2}}{N^2 \eta^{3/2}}, \quad N^{-5/2} \left| \sum_{a \neq b} G_{aa} G_{bb} G_{ab} \right| \lesssim \frac{N^\xi}{N^{3/2}} + \frac{N^\xi \rho^{3/2}}{N^2 \eta^{3/2}}, \quad (4.6)$$

with very high probability. The first bound in (4.6) follows from the isotropic law in (3.2). The second bound in (4.6) follows by writing $G = m + (G - m)$ and using the isotropic resummation

$$\sum_{ab} (G - m)_{aa} G_{ab} = \sum_a \langle e_a, G \mathbf{1} \rangle, \quad (4.7)$$

with $e_a \in \mathbf{R}^N$ the unit vector in the a -direction and $\mathbf{1} := (1, \dots, 1) \in \mathbf{R}^N$.

For $k = 3$ whenever there are at least two off-diagonal G 's we get a bound $N^{-2} \eta^{-1} \rho$. The only way to get only diagonal G 's is that α is one of (ab, ba, ba) , (ba, ab, ba) , (ba, ba, ab) ; in this case $\kappa^{(i)}(ab, \alpha) = \kappa_4 / N^2$, with $\kappa_4 := \kappa^{(i)}(ab, ba, ab, ba)$. For these terms we have (see [13, Lemma 4.2] for the analogous proof for Wigner matrices)

$$\partial_\alpha^{(i)} G_{ba} = -2s_i^3 G_{aa}^2 G_{bb}^2 + \mathcal{O} \left(\frac{N^\xi \rho}{N^2 \eta} \right), \quad (4.8)$$

with very high probability, where the error comes from terms with at least two off-diagonal G 's. Hence we finally conclude that the terms $k = 3$ give a contribution:

$$-2\kappa_4 \frac{3}{3!} \|s\|_4^4 \frac{1}{N^3} \sum_{ab} G_{aa}^2 G_{bb}^2 = \frac{\kappa_4}{N} \|s\|_4^4 m^4 + \mathcal{O} \left(\frac{N^\xi \rho^{3/2}}{N^{3/2} \eta^{1/2}} + \frac{N^\xi \rho}{N^2 \eta} \right). \quad (4.9)$$

All the terms with $k \geq 4$ can be estimated trivially using that $|G_{ab}| \lesssim 1$ with very high probability by (3.2). \square

4.2. Computation of the variance. For the second moment, using (4.3), we compute

$$\mathbf{E}_H \langle G_1 - \mathbf{E}_H G_1 \rangle \langle G_2 - \mathbf{E}_H G_2 \rangle = -\mathbf{E}_H \left(\frac{m'_1}{m_1} \langle s_1^1 \underline{H}_1 G_1 + s_2^1 \underline{H}_2 G_1 \rangle + \frac{\kappa_4}{N} \|s^1\|_4^4 m'_1 m_1^3 \right) \langle G_2 - \mathbf{E}_H G_2 \rangle + \mathcal{O} \left(\frac{N^\xi \Psi_2}{L^{1/2}} \right) \quad (4.10)$$

where $s^i = (s_1^i, s_2^i) \in S^1$ and we used (3.10) to approximate $\langle G_i - \mathbf{E}_H G_i \rangle$ with $\langle G_i - m_i \rangle$. We made this replacement to use the equation for $G - m$ from (4.3).

Then performing cumulant expansion we compute:

$$\begin{aligned} & -\mathbf{E}_H \left(\frac{m'_1}{m_1} \langle s_1^1 \underline{H}_1 G_1 + s_2^1 \underline{H}_2 G_1 \rangle + \frac{\kappa_4}{N} \|s^1\|_4^4 m'_1 m_1^3 \right) \langle G_2 - \mathbf{E}_H G_2 \rangle \\ &= \frac{\langle s^1, s^2 \rangle m'_1 \mathbf{E}_H \langle G_1 G_2^2 \rangle}{m_1 N^2} - \frac{\kappa_4}{N} \|s^1\|_4^4 m'_1 m_1^3 \mathbf{E}_H \langle G_2 - \mathbf{E}_H G_2 \rangle \\ & \quad - \frac{m'_1}{m_1} \sum_{k \geq 2} \sum_{ab} \sum_{\alpha \in \{ab, ba\}^k} \left(\frac{\kappa^{(1)}(ab, \alpha)}{k! N} s_1^1 \partial_\alpha^{(1)} + s_2^1 \frac{\kappa^{(2)}(ab, \alpha)}{k! N} \partial_\alpha^{(2)} \right) \mathbf{E}_H [(G_1)_{ba} \langle G_2 - \mathbf{E}_H G_2 \rangle]. \end{aligned} \quad (4.11)$$

Using the local law (3.11) we conclude that

$$\begin{aligned} \frac{m'_1}{m_1} \langle s^1, s^2 \rangle \frac{\langle G_1 G_2^2 \rangle}{N^2} &= \langle s^1, s^2 \rangle \frac{m'_1 m'_2}{(1 - \langle s^1, s^2 \rangle m_1 m_2)^2 N^2} + \mathcal{O} \left(\frac{N^\xi}{N^3 \eta_1 \eta_2 \eta_*^2} + \frac{N^\xi}{N^4 |\eta_1 \eta_2|^3} \right) \\ &= -\frac{1}{N^2} \partial_{z_1} \partial_{z_2} \log(1 - \langle s^1, s^2 \rangle m_1 m_2) + \mathcal{O} \left(\frac{N^\xi}{N^3 \eta_1 \eta_2 \eta_*^2} + \frac{N^\xi}{N^4 |\eta_1 \eta_2|^3} \right), \end{aligned} \quad (4.12)$$

with very high probability.

We are now left with the third line of (4.11). The α -derivative in (4.11) may hit either $(G_1)_{ba}$ or $\langle G_2 - \mathbf{E}_H G_2 \rangle$. Define

$$\begin{aligned} \Phi_k &:= \frac{m'_1}{m_1} \sum_{ab} \sum_{\alpha \in \{ab, ba\}^k} \left(\frac{\kappa^{(1)}(ab, \alpha)}{k! N} s_1^1 \partial_\alpha^{(1)} + s_2^1 \frac{\kappa^{(2)}(ab, \alpha)}{k! N} \partial_\alpha^{(2)} \right) \mathbf{E}_H [(G_1)_{ba} \langle G_2 - \mathbf{E}_H G_2 \rangle] \\ &= \sum_{ab} \sum_{\alpha} \frac{s_1^1 \kappa^{(1)}(ab, \alpha)}{k! N} \mathbf{E}_H \left(\frac{m'_1}{m_1} \partial_{\alpha_1}^{(1)} \frac{(G_1)_{ba}}{k_1!} \right) \left(\partial_{\alpha_2}^{(1)} \frac{\langle G_2 - \mathbf{E}_H G_2 \rangle}{(k - k_1)!} \right) \\ & \quad + \sum_{ab} \sum_{\alpha} \frac{s_2^1 \kappa^{(2)}(ab, \alpha)}{k! N} \mathbf{E}_H \left(\frac{m'_1}{m_1} \partial_{\alpha_1}^{(2)} \frac{(G_1)_{ba}}{k_1!} \right) \left(\partial_{\alpha_2}^{(2)} \frac{\langle G_2 - \mathbf{E}_H G_2 \rangle}{(k - k_1)!} \right), \end{aligned} \quad (4.13)$$

where k_1 denotes the number of derivatives that hit $(G_1)_{ba}$. The summation \sum_{α} indicates the summation over tuples $\alpha_i^{k_i}$, with $i = 1, 2$ and $k_2 := k - k_1$. We now claim that

$$\Phi_k = -\mathbf{1}(k = 3) \left(\kappa_4 \frac{\langle s^1 \odot s^1, s^2 \odot s^2 \rangle}{2N^2} (m_2') (m_2')' + \frac{\kappa_4}{N} \|s^1\|_4^4 m'_1 m_1^3 \right) + \mathcal{O} \left(N^\xi \frac{\Psi_2}{L^{1/2}} \right). \quad (4.14)$$

Similarly to the proof of [13, Eq. (113)] we readily conclude that the terms in Φ_k in (4.13) with $k = 2$, or k_1 odd and $k \geq 4$, or $k \geq 3$ and k_1 even are bounded by $N^\xi \Psi_2 L^{-1/2}$. For $k = 3$ and $k_1 = 3$, analogously to (4.8)–(4.9) we obtain a contribution of

$$-\frac{\kappa_4}{N} \|s^1\|_4^4 m'_1 m_1^3 + \mathcal{O}\left(\frac{N^\xi}{N|\eta_1|L^{1/2}}\right) \quad (4.15)$$

to (4.14).

For $k = 3$ and $k_1 = 1$ we start computing the action of the α_1 -derivative on $(G_1)_{ba}$:

$$\sum_{\alpha_1} \partial_{\alpha_1}^{(i)} (G_1)_{ba} = -s_i^1 (G_1)_{ba}^2 - s_i^1 (G_1)_{aa} (G_1)_{bb} = -s_i^1 m_1^2 (1 + \delta_{ab}) + \mathcal{O}\left(N^\xi \sqrt{\frac{\rho_1}{N|\eta_1|}}\right), \quad (4.16)$$

with very high probability. Additionally, we have that (see [13, Lemma 4.2] for the analogous proof for Wigner matrices)

$$\partial_{ab,ba}^{(i)} \langle G_2 - \mathbf{E}_H G_2 \rangle = \frac{2m_2 m'_2}{N} (s_i^2)^2 + \mathcal{O}\left(\frac{N^\xi \rho_2^{1/2}}{(N|\eta_2|)^{3/2}}\right), \quad (4.17)$$

with very high probability. We thus conclude that the $(k, k_1) = (3, 1)$ contribution to (4.14) is

$$-\kappa_4 \frac{\langle s^1 \odot s^1, s^2 \odot s^2 \rangle}{2N^2} (m_1^2)' (m_2^2)' + \mathcal{O}\left(\frac{N^\xi \Psi_2}{L^{1/2}}\right), \quad (4.18)$$

where we used that only the terms with $\kappa_4 = \kappa^{(i)}(ab, ba, ab, ba)$ contribute. This concludes the proof of (3.8) for $p = 2$.

4.3. Asymptotic Wick Theorem. The proof of the Wick Theorem for resolvent is completely analogous to the one for Wigner matrices in [13, Section 4]. The only differences are that along the proof we have to carefully keep track of the s_i , as we did in Section 4.2, since in the Wigner case $s^1 = \dots = s^p = (1, 0)$, and that we have to use the three G 's local law in (3.11) with a weaker error term instead of the one in [13, Eq. (45)] to compute the leading order deterministic term (see (4.21)–(4.22) below).

Define

$$Y_S := \prod_{i \in S} \langle G_i - \mathbf{E}_H G_i \rangle, \quad (4.19)$$

with $S \subset \mathbf{N}$. Similarly to Section 4.2 we start computing

$$\begin{aligned} \mathbf{E}_H Y_{[p]} &= \sum_{i \in [2,p]} \frac{m'_1}{m_1} \frac{\langle s^1, s^i \rangle}{N^2} \mathbf{E}_H \langle G_1 G_i^2 \rangle Y_{[p] \setminus \{1,i\}} - \frac{\kappa_4}{N} \|s^1\|_4^4 m'_1 m_1^3 \mathbf{E}_H Y_{[2,p]} \\ &\quad - \sum_{k \geq 2} \sum_{ab} \sum_{\alpha \in \{ab, ba\}^k} \left(\frac{\kappa^{(1)}(ab, \alpha)}{k!N} s_1^1 \partial_\alpha^{(1)} + s_2^1 \frac{\kappa^{(2)}(ab, \alpha)}{k!N} \partial_\alpha^{(2)} \right) \mathbf{E}_H \left[\frac{m'_1}{m_1} (G_1)_{ba} Y_{[2,p]} \right] \\ &\quad + \mathcal{O}\left(N^\xi \frac{\Psi_p}{L^{1/2}}\right). \end{aligned} \quad (4.20)$$

Then proceeding analogously to (4.13)–(4.18) (see also [13, Eqs. (110)–(114)] for the Wigner case) we conclude that

$$\begin{aligned} \mathbf{E}_H Y_{[p]} &= \sum_{i \in [2, p]} \frac{m'_1 \langle s^1, s^i \rangle}{m_1 N^2} \mathbf{E}_H \langle G_1 G_i^2 \rangle Y_{[p] \setminus \{1, i\}} \\ &\quad - \sum_{i \in [2, p]} \kappa_4 \frac{\langle s^1 \odot s^1, s^i \odot s^i \rangle}{2N^2} (m_1^2)' (m_i^2)' \mathbf{E}_H Y_{[1, p] \setminus \{1, i\}} + \mathcal{O} \left(N^\xi \frac{\Psi_p}{L^{1/2}} \right). \end{aligned} \quad (4.21)$$

In order to compute the leading deterministic term of $\langle G_1 G_i^2 \rangle$ we use the local law (3.11) and get

$$\mathbf{E}_H Y_{[p]} = \frac{1}{N^2} \sum_{i \in [2, p]} V_{1, i} \mathbf{E}_H Y_{[p] \setminus \{1, i\}} + \mathcal{O} \left(N^\xi \Psi_p \left(\frac{1}{L^{1/2}} + \frac{1}{N\eta_*^2} + \frac{1}{N^2\eta_*^4} \right) \right). \quad (4.22)$$

Finally, proceeding iteratively we conclude (3.8).

4.4. Multi resolvents local laws. The goal of this section is to prove the local laws in (3.11). Starting from (4.3) we get

$$\begin{aligned} (1 - \langle s^1, s^2 \rangle m_1 m_2 \langle \cdot \rangle) G_1 G_2 &= m_1 m_2 + m_1 \langle G_2 - m_2 \rangle - m_1 (s_1^1 \underline{H}_1 G_1 G_2 + s_2^1 \underline{H}_2 G_1 G_2) \\ &\quad + m_1 \langle s^1, s^2 \rangle \langle G_1 G_2 \rangle (G_2 - m_2) + m_1 \langle G_1 - m_1 \rangle G_1 G_2. \end{aligned} \quad (4.23)$$

We estimate $|\langle G_1 G_2 \rangle| \lesssim N^\xi (\eta^*)^{-1}$ with very high probability, where $\eta^* := \eta_1 \vee \eta_2$, using $|\langle G_1 G_2 \rangle| \leq \langle |G_2| \rangle / \eta_1$ (in case $\eta^* = \eta_1$) and the rigidity of eigenvalues to estimate $\langle |G_2| \rangle \leq N^\xi$. Then by the single resolvent local law $|\langle G_i - m_i \rangle| \lesssim N^\xi (N\eta_i)^{-1}$ from (3.2) we obtain that

$$(1 - \langle s^1, s^2 \rangle m_1 m_2) \langle G_1 G_2 \rangle = m_1 m_2 - m_1 (s_1^1 \langle \underline{H}_1 G_1 G_2 \rangle + s_2^1 \langle \underline{H}_2 G_1 G_2 \rangle) + \mathcal{O} \left(\frac{N^\xi}{N|\eta_1||\eta_2|} \right), \quad (4.24)$$

with very high probability. Finally, using that

$$|\langle \underline{H}_i G_1 G_2 \rangle| \lesssim \frac{N^\xi}{\sqrt{N|\eta_1 \eta_2| \eta_*}}, \quad i \in [2] \quad (4.25)$$

with very high probability from an analogous proof to [15, Eq. (5.8)] (see also [12, Eq. (5.10c)]), and that

$$|1 - \langle s^1, s^2 \rangle m_1 m_2| \gtrsim \eta^*. \quad (4.26)$$

we conclude the first local law in (3.11).

For the second local law in (3.11) we start writing the equation for $G_1 G_2^2$:

$$\begin{aligned} G_1 G_2^2 &= m_1 m'_2 + m_1 (G_2^2 - m'_2) - m_1 (s_1^1 \underline{H}_1 G_1 G_2^2 + s_2^1 \underline{H}_2 G_1 G_2^2) \\ &\quad + m_1 \langle s^1, s^2 \rangle (\langle G_1 G_2 \rangle G_2^2 + \langle G_1 G_2^2 \rangle G_2) + m_1 \langle G_1 - m_1 \rangle G_1 G_2^2. \end{aligned} \quad (4.27)$$

Then, using the usual single G local law and the two G 's local law from (3.11), we conclude that

$$\begin{aligned} (1 - \langle s^1, s^2 \rangle m_1 m_2) \langle G_1 G_2^2 \rangle &= m_1 m'_2 + \langle s^1, s^2 \rangle \frac{m_1^2 m_2 m'_2}{1 - \langle s^1, s^2 \rangle m_1 m_2} \\ &\quad - m_1 (s_1^1 \underline{H}_1 G_1 G_2^2 + s_2^1 \underline{H}_2 G_1 G_2^2) + \mathcal{O} \left(\frac{N^\xi}{N|\eta_1||\eta_2| \eta_*} \right). \end{aligned} \quad (4.28)$$

Then, using that

$$|\langle H_i G_1 G_2^2 \rangle| \lesssim \frac{N^\xi}{N \sqrt{|\eta_1 \eta_2| \eta_*^2}}, \quad i \in [2], \quad (4.29)$$

with very high probability, and (4.26) we conclude (3.11). The proof of (4.29) follows analogously to the one of (4.25).

5. STATIONARY PHASE CALCULATIONS

The proof of (2.17) is a tedious stationary phase calculation since $v_\pm^{sr}(t)$, the leading part of $v_{\pm, \kappa}^{sr}(t)$ (see (2.12)), are given in terms of oscillatory integrals for $t \gg 1$ being the large parameter. Unlike in the $s = r$ case, no explicit formula similar to (2.21) is available. The main complication is that $V^{sr}(x, y)$ defined in (2.7) has logarithmic singularities, integrated against a fast oscillatory term from $f'g'$, so standard stationary phase formulas cannot directly be applied. Nevertheless, a certain number of integration by parts can still be performed before the derivative of the integrand stops being integrable and the leading term can be computed.

We will first give a proof of

$$\mathbf{E}_s \mathbf{E}_r v_-^{sr}(t) \sim \sqrt{t} \quad (5.1)$$

then we explain how to modify this argument to obtain

$$\mathbf{E}_s \mathbf{E}_r v_-^{sr}(t)^2 \sim t^{3/2}, \quad (5.2)$$

in both cases with a definite large t asymptotics with computable explicit constants. The proof reveals that the corresponding results for $\mathbf{E}_s \mathbf{E}_r v_+^{sr}(t)$ and $\mathbf{E}_s \mathbf{E}_r v_+^{sr}(t)^2$ guarantee only an upper bound with the same behavior

$$\mathbf{E}_s \mathbf{E}_r v_+^{sr}(t) \lesssim \sqrt{t}, \quad \mathbf{E}_s \mathbf{E}_r v_+^{sr}(t)^2 \lesssim t^{3/2} \quad (5.3)$$

depending on the distribution of s on S^1 , the matching lower bound may not necessarily hold. However, for our main conclusions like (2.19) only an upper bound on $S_{\text{res}}(t)$ is important.

All these exponents are valid for the $k = 2$ case, i.e. for $H^s = s_1 H_1 + s_2 H_2$. For the general multivariate model, $k \geq 3$, exactly the same proof gives the upper bounds

$$\mathbf{E}_s \mathbf{E}_r v_\pm^{sr}(t) \lesssim \min\{1, t^{\frac{3-k}{2}}\}, \quad \mathbf{E}_s \mathbf{E}_r v_\pm^{sr}(t)^2 \lesssim \min\{1, t^{\frac{5-k}{2}}\}. \quad (5.4)$$

The k -dependence of the exponent can directly be related to the tail behavior (5.5) and (5.8) below, so for simplicity we will carry out our main analysis only for $k = 2$. In fact, a more careful analysis yields somewhat better bounds than (5.4), but we will not pursue this improvement here.

We introduce a new random variable

$$U := \langle s, r \rangle$$

then clearly $|U| \leq 1$ and since $r, s \in S^k$ have a distribution with an L^2 density, it is easy to see that the density ρ^* of U is bounded by

$$\rho^*(U) \lesssim (1 - U^2)^{\frac{k-3}{2}}. \quad (5.5)$$

The fact that the main contribution to the lhs. of (5.4) comes from the regime $U \approx 1$ is a consequence of the singularity of the logarithm in (2.7) in this regime (see computations below).

Indeed, $U = \cos \alpha$ where α is the angle between r, s and near $U \approx \pm 1$ we have $1 \pm U \approx \frac{1}{2}\alpha^2(1 + O(\alpha^2))$. For example, for $k = 2$ we have

$$\mathbf{P}(1 - U = \epsilon + d\epsilon) = \frac{d\epsilon}{\sqrt{\epsilon}} \left(\int_{S^1} \rho^2(s) ds \right) (1 + O(\sqrt{\epsilon})) \quad (5.6)$$

$$\mathbf{P}(1 + U = \epsilon + d\epsilon) = \frac{d\epsilon}{\sqrt{\epsilon}} \left(\int_{S^1} \rho(s)\rho(s + \pi) ds \right) (1 + O(\sqrt{\epsilon})) \quad (5.7)$$

in the $\epsilon \ll 1$ regime. In particular, the bound in (5.5) is actually an asymptotics in the most critical $U \approx 1$ regime, while the regime $U \approx -1$ it may happen that the density ρ^* is much smaller than (5.5) predicts. For symmetric distribution, $\rho(s) = \rho(s + \pi)$, the two asymptotics are the same. Similar relations hold for $k \geq 3$, in which case we have

$$\mathbf{P}(1 \pm U = \epsilon + d\epsilon) \lesssim \epsilon^{\frac{k-3}{2}} d\epsilon \quad (5.8)$$

with an explicit asymptotics for $U \approx 1$.

So we will study

$$R_{\pm}(t) = t^2 \Re \int dU \rho^*(U) \iint_{-2}^2 dx dy e^{it(x \pm y)} \left[\log |1 - Um(x)\overline{m(y)}| - \log |1 - Um(x)m(y)| \right]. \quad (5.9)$$

Since $|m| \leq 1$, as long as $|U| \leq 1 - \delta$ for any small fixed $\delta > 0$, the arguments of the logarithms are separated away from zero and they allow to perform arbitrary number of integration by parts, each gaining a factor of $1/t$. There is a square root singularity of $m(x)$ and $m(y)$ at the spectral edges $2, -2$ which still allows one to perform one integration by parts in each variable since m' is still integrable. Therefore the contribution of the regime $|U| \leq 1 - \delta$ to (5.9) is of order $t^2(1/t)^2 = O(1)$, hence negligible compared with the target (5.1). In the sequel we thus focus on the important $U \approx \pm 1$ regimes, in particular every $\int dU$ integral is understood to be restricted to $|U| \geq 1 - \delta$.

Note that $\overline{m(y)} = -m(-y)$, so if U has a symmetric distribution (for example if $s \in S^1$ has a symmetric distribution), then by symmetry we have

$$R_-(t) = -R_+(t).$$

For definiteness, we focus on $R_-(t)$, the analysis of R_+ is analogous. From the explicit form $m(x) = \frac{1}{2}(-x + i\sqrt{4-x^2})$ a simple exercise shows that

$$|1 - Um(x)\overline{m(y)}|^2 \gtrsim (1 - U)^2 + (x - y)^2, \quad |1 - Um(x)m(y)|^2 \gtrsim (1 + U)^2 + (x + y)^2. \quad (5.10)$$

This shows that the critical regime is $U \approx 1$ and $x \approx y$ for the first integrand in (5.9) and $U \approx -1$, $x \approx -y$ for the second. Again, for definiteness, we focus on the first regime, i.e. on the first log-integrand in (5.9) and establish the following relations for large t and $k = 2$:

Lemma 5.1. *In the $k = 2$ case we have*

$$t^2 \int dU \rho^*(U) \Re \iint_{-2}^2 e^{it(x-y)} \log |1 - Um(x)\overline{m(y)}|^2 dx dy \sim \sqrt{t} \quad (5.11)$$

and

$$t^4 \int dU \rho^*(U) \left[\Re \iint_{-2}^2 e^{it(x-y)} \log |1 - Um(x)\overline{m(y)}|^2 dx dy \right]^2 \sim t^{3/2}, \quad (5.12)$$

for $t \geq 1$. For $t \gg 1$ an analogous asymptotic statement holds with explicitly computable positive constants that depend on the distribution of s .

Proof of Lemma 5.1. Introduce the variables

$$a := \frac{x+y}{2}, \quad b := \frac{x-y}{2}, \quad \text{i.e. } x = a+b, \quad y = a-b.$$

Since $|x|, |y| \leq 2$ we have

$$|a| \leq 2, \quad |b| \leq \min\{|2-a|, |2+a|\}. \quad (5.13)$$

In terms of these variables, we have

$$|1 - Um(x)\overline{m(y)}|^2 = \left(1 - U + 2U \frac{b^2}{b^2 + d^2}\right)^2 + \frac{4U^2 b^2 d^2}{(b^2 + d^2)^2}, \quad d := \frac{1}{2} [\sqrt{4 - (a+b)^2} + \sqrt{4 - (a-b)^2}]. \quad (5.14)$$

Here we also used the identity

$$1 - m(x)\overline{m(y)} = \frac{2b}{2b + m(x) - \overline{m(y)}} = \frac{2b}{b + id}$$

following from the equation $-m(x)^{-1} = x + m(x)$ and similarly for $m(y)$. In the regime (5.13) we have

$$|b| \leq \frac{1}{2}(4 - a^2), \quad |b| \leq \sqrt{4 - a^2}. \quad (5.15)$$

Note that by Taylor expansion around a and concavity of the function $x \rightarrow \sqrt{4 - x^2}$ in $x \in [-2, 2]$, we have

$$0 \leq \sqrt{4 - a^2} - d \lesssim \frac{b^2}{(4 - a^2)^{3/2}} \leq \frac{|b|}{\sqrt{4 - a^2}}, \quad \text{as well as } \frac{1}{2}\sqrt{4 - a^2} \leq d \leq \sqrt{4 - a^2}. \quad (5.16)$$

We define the function

$$F = F(U, a, b) := (1 - U)^2 + \frac{4U^2 b^2}{4 - a^2} \quad (5.17)$$

for $|U| \leq 1$, and a, b as in (5.13). We will use F to approximate

$$M = M(U, a, b) := |1 - Um(a+b)\overline{m(a-b)}|^2 \quad (5.18)$$

in the critical regime where $|U| \geq 1 - \delta$ and $|b| \leq \delta$ for some small fixed $\delta > 0$. We clearly have

$$M(U, a, b) \geq \frac{1}{4}F(U, a, b) \quad (5.19)$$

in the regime (5.13), where $|b| \leq \sqrt{4 - a^2} \leq 2d$, using (5.16).

For the difference function

$$\Delta(U, a, b) := M(U, a, b) - F(U, a, b) \quad (5.20)$$

an elementary calculation from (5.14)–(5.16) gives

$$|\Delta(U, a, b)| \lesssim \frac{b^2}{(4 - a^2)^{3/2}} \sqrt{F} \quad (5.21)$$

in the regime $|U| \geq 1 - \delta$ and $|b| \leq \delta$. Furthermore, similar estimates hold for the first derivative;

$$\left| \frac{d}{db} \Delta(U, a, b) \right| \lesssim \frac{|b| \sqrt{F}}{(4 - a^2)^{3/2}}, \quad \left| \frac{d}{da} \Delta(U, a, b) \right| \lesssim \frac{b^2 \sqrt{F}}{(4 - a^2)^{5/2}} \lesssim \frac{|b| \sqrt{F}}{(4 - a^2)^{3/2}}, \quad (5.22)$$

as well as for the second derivatives

$$\left| \frac{d^2}{db^2} \Delta(U, a, b) \right| \lesssim \frac{\sqrt{F}}{(4-a^2)^{3/2}}, \quad \left| \frac{d}{da} \frac{d}{db} \Delta(U, a, b) \right| \lesssim \frac{|b|\sqrt{F}}{(4-a^2)^{5/2}} \lesssim \frac{\sqrt{F}}{(4-a^2)^{3/2}}. \quad (5.23)$$

The proof of Lemma 5.1 consists of two parts. First we compute the integral with $\log F$, i.e. we show that

$$t^2 \int dU \rho^*(U) \Re \iint_{-2}^2 e^{it(x-y)} \log F\left(U, \frac{x+y}{2}, \frac{x-y}{2}\right) dx dy \sim \sqrt{t} \quad (5.24)$$

with an explicit positive constant factor in the asymptotic regime $t \gg 1$. Second, we show that the integrand in (5.11) can indeed be replaced with F up to a negligible error,

$$\left| t^2 \int dU \rho^*(U) \iint_{-2}^2 e^{it(x-y)} \left[\log \left| 1 - Um(x)\overline{m(y)} \right|^2 - \log F\left(U, \frac{x+y}{2}, \frac{x-y}{2}\right) \right] dx dy \right| \lesssim 1. \quad (5.25)$$

Part I. To prove (5.24), we use the a, b variables and the symmetry of F in a to restrict the a integration to $0 \leq a \leq 2$:

$$(5.24) = 4t^2 \Re \int dU \rho^*(U) \int_0^2 da \int_{-(2-a)}^{2-a} db e^{2itb} \log F(U, a, b). \quad (5.26)$$

Using integration by parts, we have

$$\begin{aligned} \int_{-(2-a)}^{2-a} db e^{2itb} \log \left[(1-U)^2 + \frac{4U^2 b^2}{4-a^2} \right] &= \frac{1}{2it} \left[e^{2it(2-a)} - e^{-2it(2-a)} \right] \log \left[(1-U)^2 + \frac{4U^2(2-a)}{2+a} \right] \\ &\quad - \frac{1}{2it} \frac{4U^2}{4-a^2} \int_{-(2-a)}^{2-a} db e^{2itb} \frac{2b}{(1-U)^2 + \frac{4U^2 b^2}{4-a^2}}. \end{aligned} \quad (5.27)$$

In the boundary terms we can perform one more integration by parts in the a variable when plugged into (5.26). Just focusing on the first boundary term in (5.27), using $|U| \leq 1$ we have

$$\left| \frac{1}{2it} e^{4it} \int_0^2 da e^{-2ita} \log \left[(1-U)^2 + \frac{4U^2(2-a)}{2+a} \right] \right| \lesssim \frac{1}{t^2} \int_0^2 \frac{da}{(1-U)^2 + U^2(2-a)} \lesssim \frac{|\log(1-U)|}{t^2}.$$

Since $\rho^*(U)$ is a density bounded by $(1-U^2)^{-1/2}$ in the $U \approx 1$ regime from (5.5), the logarithmic singularity is integrable showing that the two boundary terms in (5.27), when plugged into (5.26), give at most an $O(1)$ contribution, negligible compared with the target behavior of order \sqrt{t} in (5.1).

To compute the main (second) term in the rhs. of (5.27), we first extend the integration limits to infinity and claim that

$$\begin{aligned} t^2 \int dU \rho^*(U) \left| \frac{1}{2it} \int_0^2 da \frac{4U^2}{4-a^2} \int_{2-a}^\infty db e^{2itb} \frac{2b}{(1-U)^2 + \frac{4U^2 b^2}{4-a^2}} \right| \\ \lesssim t \int dU \rho^*(U) \int_0^2 \frac{da}{2-a} \left| \int_{2-a}^\infty db e^{2itb} \frac{2b}{(1-U)^2 + \frac{4U^2 b^2}{4-a^2}} \right| \end{aligned} \quad (5.28)$$

gives a negligible contribution to (5.26) (the lower limit is removed similarly). Indeed, we apply one more integration by parts inside the absolute value in (5.28):

$$\left| \int_{2-a}^{\infty} db e^{2itb} \frac{2b}{(1-U)^2 + \frac{4U^2b^2}{4-a^2}} \right| \lesssim t^{-1} \int_{2-a}^{\infty} \frac{db}{(1-U)^2 + \frac{U^2b^2}{4-a^2}} + t^{-1} \frac{2-a}{(1-U)^2 + (2-a)}.$$

Its contribution to the rhs of (5.28) is thus bounded by

$$\begin{aligned} & \int dU \rho^*(U) \int_0^2 \frac{da}{2-a} \left[\int_{2-a}^{\infty} \frac{db}{(1-U)^2 + \frac{U^2b^2}{4-a^2}} + \frac{2-a}{(1-U)^2 + (2-a)} \right] \\ & \lesssim \int \frac{dU}{\sqrt{1-U^2}} \left[\int_0^2 \frac{da}{\sqrt{2-a}} \frac{1}{1-U + \sqrt{2-a}} + |\log(1-U)| \right] \lesssim 1. \end{aligned}$$

Summarizing, we just proved that

$$\begin{aligned} (5.24) &= -2t\Im \int dU \rho^*(U) \int_0^2 da \frac{4U^2}{4-a^2} \int_{-\infty}^{\infty} db e^{2itb} \frac{2b}{(1-U)^2 + \frac{4U^2b^2}{4-a^2}} + O(1) \\ &= \frac{t}{\pi} \int dU \rho^*(U) \int_0^2 da e^{-t\sqrt{4-a^2}(1-U)/U} + O(1) \\ &= \frac{c_0 t}{\pi} \int dU \frac{1}{\sqrt{1-U}} \int_0^2 da e^{-t\sqrt{4-a^2}(1-U)/U} + O(1) \\ &= \frac{c_0 \sqrt{t}}{\pi} \int_0^{\infty} \frac{e^{-v}}{\sqrt{v}} dv \int_0^2 \frac{da}{(4-a^2)^{1/4}} + O(1) \\ &= \frac{\Gamma(3/4)}{\sqrt{2}\Gamma(5/4)} c_0 \sqrt{t} + O(1), \end{aligned} \tag{5.29}$$

where in the second line we used residue calculation, in the third line we used that

$$\rho^*(U) = \frac{c_0}{\sqrt{1-U}} + O(1)$$

in the regime $U \approx 1$ with some positive constant $c_0 > 0$ depending on the distribution of s (see (5.6)), and finally in the fourth line we used that for large t the main contribution to the integral comes from $U \approx 1$ in order to simplify the integrand. This completes the proof of (5.24).

Part II. We now prove (5.25). After changing to the a, b variables and considering only the $0 \leq a \leq 2$ regime for definiteness, we perform an integration by parts in b that gives

$$\begin{aligned} (5.25) &\lesssim t \int dU \rho^*(U) \left| \int_0^2 da e^{2ita} \left[\log M(U, a, b) - \log F(U, a, b) \right] db \right| \\ &\quad + t \int dU \rho^*(U) \int_0^2 da \left| \int_{-(2-a)}^{2-a} e^{2itb} \partial_b \left[\log M(U, a, b) - \log F(U, a, b) \right] db \right| \end{aligned} \tag{5.30}$$

recalling the definition of M from (5.18). The first term in (5.30) is the boundary term, which is negligible after one more integration by parts using the ∂_a derivative estimate from (5.22).

In the second term we perform one more integration by parts to obtain

$$(5.25) \lesssim t \int dU \rho^*(U) \left| \int_0^2 da e^{2ita} \partial_b \left[\log M(U, a, b) - \log F(U, a, b) \right] db \right| \\ + \int dU \rho^*(U) \int_0^2 da \int_{-(2-a)}^{2-a} \left| \partial_b^2 \left[\log M(U, a, b) - \log F(U, a, b) \right] \right| db, \quad (5.31)$$

where the first term comes from the boundary. In this term we can perform one more integration by parts in a . The corresponding boundary terms are easily seen to be order one and the main term is analogous to the first term in the rhs of (5.31) just we have the mixed $\partial_a \partial_b$ derivative. Recalling $\Delta = M - F$ from (5.20), we use the estimate

$$|\partial_b^2 [\log M - \log F]| \lesssim \frac{|\partial_b^2 \Delta|}{F} + \frac{|\partial_b^2 F|}{F^2} |\Delta| + \frac{|\partial_b \Delta| |\partial_b M + \partial_b F|}{F^2} + (\partial_b F)^2 \frac{|\Delta|}{F^3}$$

in the situation where $M \gtrsim F > 0$ are positive functions (see (5.19)). Similar bound holds for the mixed derivative.

Therefore, we can estimate both integrals in (5.31) as follows:

$$(5.25) \lesssim \int dU \rho^*(U) \int_0^2 da \int_{-(2-a)}^{2-a} \frac{1}{(4-a^2)^{3/2}} \frac{1}{\left[(1-U)^2 + \frac{b^2}{4-a^2} \right]^{1/2}} db \\ \lesssim \int \frac{dU}{\sqrt{1-U^2}} \int_0^2 \frac{da}{4-a^2} \int_0^{\sqrt{2-a}} \frac{du}{\left[(1-U)^2 + u^2 \right]^{1/2}} \\ \lesssim \int \frac{dU}{\sqrt{1-U^2}} \int_0^{\sqrt{2}} \frac{|\log u| + 1}{\left[(1-U)^2 + u^2 \right]^{1/2}} du \\ \lesssim \int \frac{dU |\log(1-U)|^2}{(1-U)^{1/2}} \lesssim 1. \quad (5.32)$$

Here we used the bounds (5.21), (5.22) and (5.23) and that $|b| \leq 2-a \lesssim 4-a^2$ to simplify some estimates. For computing the derivatives of F we used its explicit form (5.17). This completes the proof of (5.25) and thus also the proof of (5.11) in Lemma 5.1.

The proof of (5.12) is very similar. We again approximate $M = |1 - Um(x)\overline{m(y)}|^2$ by F at the expense of negligible errors. We omit these calculations as they are very similar those for (5.11) and focus only on the main term which is (see the analogous (5.26))

$$16t^4 \int dU \rho^*(U) \left[\Re \int_0^2 da \int_{-(2-a)}^{2-a} db e^{2itb} \log F(U, a, b) \right]^2. \quad (5.33)$$

After one integration by parts and neglecting the lower order boundary terms, we have the following analogue of (5.29):

$$\begin{aligned}
4t^2 \int dU \rho^*(U) & \left[\Re \int_0^2 da \frac{U^2}{4-a^2} \int_{-\infty}^{\infty} db e^{2itb} \frac{2b}{(1-U)^2 + \frac{4U^2 b^2}{4-a^2}} \right]^2 \\
& = \frac{t^2}{\pi^2} \int dU \rho^*(U) \left[\int_0^2 da e^{-t\sqrt{4-a^2}(1-U)/U} \right]^2 \\
& \approx \frac{c_0 t^{3/2}}{\pi^2} \int_0^{\infty} \frac{dv}{\sqrt{v}} \left(\int_0^2 da e^{-\sqrt{4-a^2}v} \right)^2 \\
& = \frac{c_0 t^{3/2}}{\pi^2} \iint_0^2 \frac{da_1 da_2}{(\sqrt{4-a_1^2} + \sqrt{4-a_2^2})^{1/2}} \int_0^{\infty} \frac{e^{-v}}{\sqrt{v}} dv \sim t^{3/2}
\end{aligned} \tag{5.34}$$

as the leading term. This proves (5.12) and completes the proof of Lemma 5.1. \square

We close this section by commenting on the proof of the upper bound in (2.28). Recall from (2.16) that the essential part of $\tilde{S}_{\text{res}}(t)$ in the slope regime is given by $\mathbf{E}_s \mathbf{E}_r \tilde{v}^{sr}(t)$ expressed by the oscillatory integrals

$$R_{\pm}(t) := t^2 \iint_{\mathbf{R}^2} \rho(s) \rho(r) ds dr \iint_{-2}^2 dx dy e^{it(\|s\|x \pm \|r\|y)} A(U, x, y) \tag{5.35}$$

with

$$A(U, x, y) := \log \left| 1 - Um(x) \overline{m(y)} \right| - \log |1 - Um(x)m(y)|,$$

where $U = \frac{\langle s, r \rangle}{\|s\| \|r\|}$ is the cosine of the angle between the vectors $s, r \in \mathbf{R}^2$. Assuming for the moment that ρ , the density of s , is rotationally symmetric, $\rho(s) = \rho(\|s\|)$ with a slight abuse of notations, we have

$$\begin{aligned}
R_{\pm}(t) & \sim t^2 \int_{-1}^1 \frac{dU}{\sqrt{1-U^2}} \iint_{-2}^2 dx dy A(U, x, y) \int_0^{\infty} e^{itx\sigma} \rho(\sigma) \sigma d\sigma \int_0^{\infty} e^{\pm ity\sigma'} \rho(\sigma') \sigma' d\sigma' \\
& \sim t \int_{-1}^1 \frac{dU}{\sqrt{1-U^2}} \iint_{-2}^2 dx dy \widehat{\rho}(tx) \widehat{\rho(\sigma)\sigma}(\pm ty) \frac{d}{dx} A(U, x, y)
\end{aligned} \tag{5.36}$$

performing an integration by parts in x and ignoring lower order boundary term. In the last step we also computed the Fourier transform (we used that $\rho(0) = 0$ to extend ρ to \mathbf{R}). The main contribution comes from the regime where A is nearly singular, and considering (5.10), we just focus on the regime $U \sim 1$ and $x \sim y$, the singularity from the other logarithmic term is treated analogously. Similarly to the proof of (5.25) we may ignore the edge regime, and effectively we have

$$\left| \frac{d}{dx} \log \left| 1 - Um(x) \overline{m(y)} \right| \right| \lesssim \frac{1}{(1-U) + |x-y|}. \tag{5.37}$$

Thus we can continue estimating the last line of (5.36)

$$|(5.36)| \lesssim t \int_{-1}^1 \frac{dU}{\sqrt{1-U^2}} \iint_{-2}^2 dx dy \frac{|\widehat{\rho}(tx) \widehat{\rho(\sigma)\sigma}(ty)|}{(1-U) + |x-y|} \lesssim t^{-1/2}.$$

Here we used the regularity of ρ , so that the last two factors essentially restrict the integration to the regime $|x|, |y| \lesssim 1/t$. The final inequality is obtained just by scaling.

To understand $\tilde{S}_{\text{res}}(t)$ in the ramp regime, we need to compute $\mathbf{E}_s \mathbf{E}_r \tilde{v}_{\pm}^{sr}(t)^2$, i.e. integrals of the following type:

$$\begin{aligned}
& t^4 \left| \iint_{\mathbf{R}^2} \rho(s) \rho(r) ds dr \right| \left| \iint_{-2}^2 dx dy e^{it(\|s\|x \pm \|r\|y)} A(U, x, y) \right|^2 \\
&= t^4 \int_{-1}^1 \frac{dU}{\sqrt{1-U^2}} \iint_{-2}^2 dx dy \iint_{-2}^2 dx' dy' A(U, x, y) \overline{A(U, x', y')} \\
&\quad \times \int_0^\infty e^{it(x-x')\sigma} \rho(\sigma) \sigma d\sigma \int_0^\infty e^{\pm it(y-y')\sigma'} \rho(\sigma') \sigma' d\sigma' \\
&\sim t^2 \int_{-1}^1 \frac{dU}{\sqrt{1-U^2}} \iint_{-2}^2 dx dy \iint_{-2}^2 dx' dy' \frac{d}{dx} A(U, x, y) \frac{d}{dy'} \overline{A(U, x', y')} \\
&\quad \times \int_0^\infty e^{it(x-x')\sigma} \rho(\sigma) d\sigma \int_0^\infty e^{\pm it(y-y')\sigma'} \rho(\sigma') d\sigma' \\
&\sim t^2 \int_{-1}^1 \frac{dU}{\sqrt{1-U^2}} \iiint_{-2}^2 dx dy dx' dy' \hat{\rho}(t(x-x')) \hat{\rho}(\pm t(y-y')) \frac{d}{dx} A(U, x, y) \frac{d}{dy'} \overline{A(U, x', y')}.
\end{aligned} \tag{5.38}$$

Here we performed two integrations by parts in x and y' and ignored the boundary terms. Estimating the derivative of A as in (5.37), we can continue

$$|(5.38)| \lesssim t^2 \int_{-1}^1 \frac{dU}{\sqrt{1-U^2}} \iint_{-2}^2 \frac{dx dy}{(1-U) + |x-y|} \iint_{-2}^2 \frac{dx' dy'}{(1-U) + |x'-y'|} |\hat{\rho}(t(x-x')) \hat{\rho}(t(y-y'))|.$$

The last two factors essentially restrict the integration to the regime $|x-x'| \lesssim 1/t$, $|y-y'| \lesssim 1/t$ and by scaling we obtain a bound of order $t^{1/2}$ for $|(5.38)|$. This completes the sketch of the proof of (2.28) in the radially symmetric case, the general case is analogous but technically more cumbersome and we omit the details.

Data availability statement. All data generated or analysed are included in this published article.

REFERENCES

- [1] Z. D. Bai and J. Yao, *On the convergence of the spectral empirical process of Wigner matrices*, Bernoulli **11**, 1059–1092 (2005), MR2189081.
- [2] Z. Bao and Y. He, *On Cramér–von Mises statistic for the spectral distribution of random matrices*, Ann. Appl. Probab. **32**, 4315–4355 (2022), MR4522353.
- [3] Z. Bao and Y. He, *Quantitative CLT for linear eigenvalue statistics of Wigner matrices*, preprint (2021), arXiv:2103.05402.
- [4] M. V. Berry, *Semiclassical theory of spectral rigidity*, Proc. Roy. Soc. London Ser. A **400**, 229–251 (1985), MR805089.
- [5] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Isotropic local laws for sample covariance and generalized Wigner matrices*, Electron. J. Probab. **19**, no. 33, 53 (2014), MR3183577.
- [6] O. Bohigas, M.-J. Giannoni, and C. Schmit, *Characterization of chaotic quantum spectra and universality of level fluctuation laws*, Phys. Rev. Lett. **52**, 1–4 (1984), MR730191.
- [7] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin, *Fixed energy universality for generalized Wigner matrices*, Comm. Pure Appl. Math. **69**, 1815–1881 (2016), MR3541852.
- [8] A. Boutet de Monvel and A. Khorunzhy, *Asymptotic distribution of smoothed eigenvalue density. II. Wigner random matrices*, Random Oper. Stochastic Equations **7**, 149–168 (1999), MR1689027.

- [9] E. Brézin and S. Hikami, *Spectral form factor in a random matrix theory*, Phys. Rev. E (3) **55**, 4067–4083 (1997), [MR1449379](#).
- [10] G. Cipolloni, *Fluctuations in the spectrum of non-Hermitian i.i.d. matrices*, J. Math. Phys. **63**, Paper No. 053503, 7 (2022), [MR4420179](#).
- [11] G. Cipolloni, L. Erdős, and D. Schröder, “Mesoscopic Central Limit Theorem for non-Hermitian Random Matrices”, Preprint, 2022, [arXiv:2210.12060](#).
- [12] G. Cipolloni, L. Erdős, and D. Schröder, *Central Limit Theorem for Linear Eigenvalue Statistics of non-Hermitian Random Matrices*, Comm. Pure Appl. Math. (2019), [arXiv:1912.04100](#).
- [13] G. Cipolloni, L. Erdős, and D. Schröder, *Functional Central Limit Theorems for Wigner Matrices*, preprint (2020), [arXiv:2012.13218](#).
- [14] G. Cipolloni, L. Erdős, and D. Schröder, *Functional central limit theorems for Wigner matrices*, Ann. Appl. Probab. **33**, 447–489 (2023), [MR4551555](#).
- [15] G. Cipolloni, L. Erdős, and D. Schröder, *Quenched universality for deformed Wigner matrices*, Probab. Theory Related Fields (2021), [arXiv:2106.10200](#).
- [16] G. Cipolloni, L. Erdős, and D. Schröder, *Thermalisation for Wigner matrices*, J. Funct. Anal. **282**, Paper No. 109394, 37 (2022), [MR4372147](#).
- [17] J. Cotler, N. Hunter-Jones, J. Liu, and B. Yoshida, *Chaos, Complexity, and Random Matrices*, preprint, JHEP 1711 (2017) 048 (2017), [arXiv:1706.05400](#).
- [18] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker, D. Stanford, A. Streicher, and M. Tezuka, *Black Holes and Random Matrices*, preprint, JHEP 1705:118, 2017 (2016), [arXiv:1611.04650](#).
- [19] L. Erdős, T. Krüger, and D. Schröder, *Random matrices with slow correlation decay*, Forum Math. Sigma **7**, Paper No. e8, 89 (2019), [MR3941370](#).
- [20] L. Erdős, H.-T. Yau, and J. Yin, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math. **229**, 1435–1515 (2012), [MR2871147](#).
- [21] P. J. Forrester, *Differential identities for the structure function of some random matrix ensembles*, J. Stat. Phys. **183**, Paper No. 33, 28 (2021), [MR4257834](#).
- [22] P. J. Forrester, *Quantifying dip-ramp-plateau for the Laguerre unitary ensemble structure function*, Comm. Math. Phys. **387**, 215–235 (2021), [MR4312363](#).
- [23] A. M. García-García, Y. Jia, and J. J. M. Verbaarschot, *Universality and Thouless energy in the supersymmetric Sachdev-Ye-Kitaev model*, Phys. Rev. D **97**, 106003, 13 (2018), [MR3884483](#).
- [24] A. M. García-García and J. J. M. Verbaarschot, *Analytical spectral density of the Sachdev-Ye-Kitaev model at finite N* , Phys. Rev. D **96**, 066012 (2017).
- [25] H. Gharibyan, C. Pattison, S. Shenker, and K. Wells, *Work in preparation*, (2021).
- [26] A. Guionnet, *Large deviations upper bounds and central limit theorems for non-commutative functionals of Gaussian large random matrices*, Ann. Inst. H. PoincarÉ Probab. Statist. **38**, 341–384 (2002), [MR1899457](#).
- [27] M. C. Gutzwiller, *Chaos in classical and quantum mechanics*, Vol. 1, Interdisciplinary Applied Mathematics (Springer-Verlag, New York, 1990), pp. xiv+432, [MR1077246](#).
- [28] Y. He, *Mesoscopic linear statistics of Wigner matrices of mixed symmetry class*, J. Stat. Phys. **175**, 932–959 (2019), [MR3959983](#).
- [29] Y. He and A. Knowles, *Mesoscopic eigenvalue density correlations of Wigner matrices*, Probab. Theory Related Fields **177**, 147–216 (2020), [MR4095015](#).
- [30] Y. He and A. Knowles, *Mesoscopic eigenvalue statistics of Wigner matrices*, Ann. Appl. Probab. **27**, 1510–1550 (2017), [MR3678478](#).
- [31] S. Heusler, S. Müller, A. Altland, P. Braun, and F. Haake, *Periodic-Orbit Theory of Level Correlations*, Phys. Rev. Lett. **98**, 044103 (2007).
- [32] Y. Jia and J. J. M. Verbaarschot, *Spectral fluctuations in the Sachdev-Ye-Kitaev model*, J. High Energy Phys., 193, 57 (2020), [MR4136970](#).
- [33] K. Johansson, *On fluctuations of eigenvalues of random Hermitian matrices*, Duke Math. J. **91**, 151–204 (1998), [MR1487983](#).
- [34] A. M. Khoruzhny, B. A. Khoruzhenko, and L. A. Pastur, *Asymptotic properties of large random matrices with independent entries*, J. Math. Phys. **37**, 5033–5060 (1996), [MR1411619](#).
- [35] A. Knowles and J. Yin, *The isotropic semicircle law and deformation of Wigner matrices*, Comm. Pure Appl. Math. **66**, 1663–1750 (2013), [MR3103909](#).

- [36] B. Landon, P. Lopatto, and P. Sosoë, *Single eigenvalue fluctuations of general Wigner-type matrices*, preprint (2021), [arXiv:2105.01178](#).
- [37] B. Landon and P. Sosoë, “Almost-optimal bulk regularity conditions in the CLT for Wigner matrices”, Preprint, 2022, [arXiv:2204.03419](#).
- [38] B. Landon and P. Sosoë, *Applications of mesoscopic CLTs in random matrix theory*, Ann. Appl. Probab. **30**, 2769–2795 (2020), [MR4187127](#).
- [39] B. Landon, P. Sosoë, and H.-T. Yau, *Fixed energy universality for Dyson Brownian motion*, preprint (2016), [arXiv:1609.09011](#).
- [40] L. Leviandier, M. Lombardi, R. Jost, and J. Pique, *Fourier transform: A tool to measure statistical level properties in very complex spectra*, Phys. Rev. Lett. **56**, 2449–2452 (1986), [PMID10032995](#).
- [41] A. Lytova and L. Pastur, *Fluctuations of matrix elements of regular functions of Gaussian random matrices*, J. Stat. Phys. **134**, 147–159 (2009), [MR2489497](#).
- [42] M. L. Mehta, *Random matrices*, Third, Vol. 142, Pure and Applied Mathematics (Amsterdam) (Elsevier/Academic Press, Amsterdam, 2004), pp. xviii+688, [MR2129906](#).
- [43] S. Müller, S. Heusler, P. Braun, F. Haake, and A. Altland, *Semiclassical Foundation of Universality in Quantum Chaos*, Phys. Rev. Lett. **93**, 014103 (2004).
- [44] K. Okuyama, *Spectral form factor and semi-circle law in the time direction*, J. High Energy Phys., 161, front matter + 15 (2019), [MR3925257](#).
- [45] R. E. Prange, *The Spectral Form Factor Is Not Self-Averaging*, Phys. Rev. Lett. **78**, 2280–2283 (1997).
- [46] P. Saad, S. H. Shenker, and D. Stanford, *A semiclassical ramp in SYK and in gravity*, preprint (2018), [arXiv:1806.06840](#).
- [47] M. Shcherbina, *Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices*, Zh. Mat. Fiz. Anal. Geom. **7**, 176–192, 197, 199 (2011), [MR2829615](#).
- [48] M. Sieber and K. Richter, *Correlations between Periodic Orbits and their Role in Spectral Statistics*, Physica Scripta **T90**, 128 (2001).
- [49] P. Sosoë and P. Wong, *Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices*, Adv. Math. **249**, 37–87 (2013), [MR3116567](#).
- [50] G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge Mathematical Library, Reprint of the second (1944) edition (Cambridge University Press, Cambridge, 1995), pp. viii+804, [MR1349110](#).

PRINCETON CENTER FOR THEORETICAL SCIENCE, PRINCETON UNIVERSITY, PRINCETON, NJ 08544, USA

IST AUSTRIA, AM CAMPUS 1, 3400 KLOSTERNEUBURG, AUSTRIA

INSTITUTE FOR THEORETICAL STUDIES, ETH ZURICH, CLAUSIUSSTR. 47, 8092 ZURICH, SWITZERLAND

Email address: gc4233@princeton.edu

Email address: lerdos@ist.ac.at

Email address: dschroeder@ethz.ch