ANNALS OF
MATHEMATICS

The spectral edge of some random band matrices

By Sasha Sodin

SECOND SERIES, VOL. 172, NO. 3
November, 2010
ANMAAH

Annals of Mathematics, **172** (2010), 2223–2251

The spectral edge of some random band matrices

By SASHA SODIN

Abstract

We study the asymptotic distribution of the eigenvalues of random Hermitian periodic band matrices, focusing on the spectral edges. The eigenvalues close to the edges converge in distribution to the Airy point process if (and only if) the band is sufficiently wide ($W \gg N^{5/6}$). Otherwise, a different limiting distribution appears.

1. **Introduction**

In this paper, we study the edge of the spectrum of random Hermitian periodic band matrices. The $N \times N$ Hermitian random matrix $H = H_N$ with rows and columns labeled by elements of $\mathbb{Z}/N\mathbb{Z}$ has independent entries above the main diagonal, and

$$H_{uv} = 0 \quad \text{if} \quad |u-v|_N = \min(|u-v|, N-|u-v|) > W_N \quad \text{or} \quad u = v. \tag{1.1}$$

To simplify the exposition, we assume that either

$$\mathbb{P}\{H_{uv} = 1\} = \mathbb{P}\{H_{uv} = -1\} = 1/2, \quad 0 < |u-v|_N \le W_N \tag{1.2}$$

("random signs"), or

$$H_{uv} = \exp(i U_{uv}), \quad U_{uv} \sim U(0, 2\pi), \quad 0 < |u-v|_N \le W_N \tag{1.3}$$

("random phases"), and defer the discussion of possible generalizations to the last section. For the same reason, we assume that $W = W_N \to \infty$ as $N \to \infty$.

The matrix H_N is closely related to the graph $\mathcal{G} = (\mathbb{Z}/N\mathbb{Z}, \mathcal{E})$, where

$$(u, v) \in \mathcal{E} \Longleftrightarrow 0 < |u-v| \le W_N, \tag{1.4}$$

and can be viewed as a Hamiltonian of quantum evolution in a disordered environment on $\mathcal{G}$ ((1.3) corresponds to broken time-reversal symmetry). We refer the reader to the work of Fyodorov and Mirlin [11] for a thorough discussion of physical motivation.

Supported in part by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities and by the ISF.

Random matrices similar to H_N have been studied in both mathematical and physical literature. We survey some of the results that pertain to the current work, and refer to [11] and to the recent review of Spencer [28] for detailed bibliography.

(A) Bogachev, Molchanov, and Pastur [4] have proved that the empirical spectral measure (or integrated density of states)

$$\mu_N(\alpha) = \#\left\{\text{eigenvalues of } \frac{H_N}{2\sqrt{2W_N}} \text{ in } (-\infty, \alpha]\right\} \tag{1.5}$$

converges (weakly, in distribution, as $N \to \infty$ and $W_N \to \infty$), to the (deterministic) Wigner measure

$$\mu_{\text{Wigner}}(\alpha) = \int_{-\infty}^{\alpha} \frac{2}{\pi}(1-\alpha_1^2)_+^{1/2} d\alpha_1. \tag{1.6}$$

That is, the *global* behavior of the spectrum of H_N is similar to that of Wigner matrices (which correspond to the special case $W_N = N/2$).

(B) One of the interesting questions concerning *local* eigenvalue statistics is the crossover between the Random Matrix regime and the Poisson regime. The Thouless criterion [29], applied to random band matrices by Fyodorov and Mirlin [11], predicts the following.

If the mixing exponent ρ_{mix} of the (classical) random walk on $\mathcal{G}$ is much larger than the eigenvalue spacing near α_0:

$$\rho_{\text{mix}} \gg \frac{\text{mean spacing at } \alpha_0}{\text{mean density at } \alpha_0},$$

then the eigenvalues of H_N near α_0 obey Random Matrix statistics, and the corresponding eigenvectors are extended. If

$$\rho_{\text{mix}} \ll \frac{\text{mean spacing at } \alpha_0}{\text{mean density at } \alpha_0},$$

then the eigenvalues near α_0 obey Poisson statistics, and the corresponding eigenvectors are localised. For our graph $\mathcal{G}$, ρ_{mix} is of order W_N^2/N^2, and in the bulk of the spectrum, $-1 < \alpha_0 < 1$, (1.6) suggests that

$$\frac{\text{mean spacing at } \alpha_0}{\text{mean density at } \alpha_0} \asymp 1/N.$$

Thus the crossover should occur at $W_N \asymp \sqrt{N}$.

On the physical level of rigor, these predictions have been justified by Fyodorov and Mirlin [11], who have derived a detailed description of both asymptotic regimes (and also of the crossover). So far, these results resist mathematical justification (see however the work of Khorunzhiy and Kirsch [16] for some relatively recent developments).

We remark that Spencer and Wang [33, Ch. III] have managed to make one direction of (a slightly different form of) the Thouless criterion rigorous under certain assumptions, in a fairly general setting. However, for now these assumptions have not been verified for the problem under consideration. We refer the reader to the review of Spencer [28] for a discussion of the mathematical approach to the Thouless criterion and its application to band matrices.

(C) We focus on the edge of the spectrum: $\alpha_0 = \pm 1$. There,

$$\frac{\text{mean spacing at } \alpha_0}{\text{mean density at } \alpha_0} \asymp \frac{N^{-2/3}}{N^{-1/3}} = N^{-1/3};$$

therefore the crossover should occur at $W_N \asymp N^{5/6}$. On the physical level of rigor, this has been confirmed by Silvestrov [23]. In fact, one of the aims of the current paper is to put some of the methods and results of [23] on firm mathematical ground.

Now we state the main results.

① For $W_N \gg N^{5/6}$ (i.e. $W_N / N^{5/6} \to +\infty$), set

$$\sigma_R(\lambda) = \# \left\{ \text{eigenvalues of } \frac{H_N}{2\sqrt{2W_N}} \text{ in } \left(1 - \frac{\lambda}{2N^{2/3}},\ +\infty\right] \right\},$$
$$\sigma_L(\lambda) = \# \left\{ \text{eigenvalues of } \frac{H_N}{2\sqrt{2W_N}} \text{ in } \left(-\infty,\ -1 + \frac{\lambda}{2N^{2/3}}\right] \right\}.$$

THEOREM 1.1. *If $W_N/N^{5/6} \to \infty$ as $N \to \infty$, then the measures $\sigma_R(\lambda)$ and $\sigma_L(\lambda)$ converge in distribution*[1] *to $\mathfrak{Ai}_\beta(-\lambda)$, where $\mathfrak{Ai}_\beta$ is the* (*random*) *distribution function corresponding to the Airy point process*, $\beta = 1$ *for* (1.2), *and* $\beta = 2$ *for* (1.3).

Hence, *according the Tracy-Widom theorem* ([30], [31]), *the scaled extreme eigenvalues $2N^{2/3}(\alpha_{\max} - 1)$ and $-2N^{2/3}(\alpha_{\min} + 1)$ of $H_N/(2\sqrt{2W_N})$ converge in distribution to the Tracy-Widom law TW_β.*

② For $1 \ll W_N \ll N^{5/6}$, set

$$\sigma_R(\lambda) = \frac{W_N^{6/5}}{N} \# \left\{ \text{eigenvalues of } \frac{H_N}{2\sqrt{2W_N}} \text{ in } \left(1 - \frac{\lambda}{2W_N^{4/5}},\ +\infty\right] \right\},$$
$$\sigma_L(\lambda) = \frac{W_N^{6/5}}{N} \# \left\{ \text{eigenvalues of } \frac{H_N}{2\sqrt{2W_N}} \text{ in } \left(-\infty,\ -1 + \frac{\lambda}{2W_N^{4/5}}\right] \right\}.$$

[1]on test functions $f \in C(\mathbb{R})$ such that supp $f \cap \mathbb{R}_+$ is compact.

THEOREM 1.2. *If $W_N \to \infty$ and $W_N / N^{5/6} \to 0$, then the measures σ_R and σ_L converge in distribution to the* (*same*) *deterministic measure σ_β. We have*

$$\sigma_\beta(\lambda) = \frac{2}{3\pi}\lambda^{3/2} + O(\lambda), \quad \lambda \to +\infty,$$
$$\sigma_\beta(\lambda) \le C \exp(-C|\lambda|^{5/4}), \quad \lambda \to -\infty.$$

In some sense, the eigenvalues close to the edges behave as independent random samples from σ_β. The method of the current paper yields the following manifestation of this belief:

$$\mathbb{P}\left\{2W_N^{4/5}(1-\alpha_{\max}) \ge \lambda\right\} - \exp\left\{-\frac{N}{W_N^{6/5}}\sigma_\beta(\lambda)\right\} = o(1)$$

uniformly in $\lambda \in \mathbb{R}$. Unfortunately, our description of the left tail of σ_β is not sufficiently precise to deduce convergence to a max-stable law (cf. Gnedenko [12] for the description of domains of attraction of max-stable laws).

Khorunzhiy [15] has proved (for a slightly different class of band matrices) that, if $W_N \gg \log^{3/2} N$,

$$\left\| H_N \big/ (2\sqrt{2W_N}) \right\| \xrightarrow{D} 1 \tag{1.7}$$

as $N \to \infty$ (actually, he has established a stronger form of convergence). He has conjectured (private communication) that the same conclusion holds under the weaker assumption $W_N \gg \log N$. We confirm this conjecture:

THEOREM 1.3. *If $W_N / \log N \to +\infty$, then $\left\| H_N \big/ (2\sqrt{2W_N}) \right\| \xrightarrow{D} 1$.*

As one can see from the argument of Bogachev, Molchanov, and Pastur [4], this result is sharp, meaning that the conclusion fails if $W_N / \log N \to 0$.

The proofs of the three results are based on a modification of the moment method. The moment method has been applied to random matrices since the work of Wigner [34]. Bogachev, Molchanov, and Pastur [4] have applied it to study the spectrum of band matrices (see above).

The moment method appears particularly useful to study the eigenvalue statistics at the spectral edge. In [26], Soshnikov has applied it to derive the limiting distribution of the extreme eigenvalues of Wigner-type random matrices. Extensions of his approach have allowed a number of related problems to be solved; e.g., [27], [21], and [10].

In this paper, we apply a modification of the moment method, which goes back at least to the work of Bai and Yin [2] (see [25] for more detailed references). In [9], it has been used, in particular, to give another proof of Soshnikov's result. The combinatorial technique of [9] is the main ingredient of the current work. In Section 2, we review this technique, and formulate the combinatorial statements needed to prove the main results.

Another ingredient of the proof is an asymptotic description of the (classical) random walk on $\mathscr{G}$. We prove the necessary facts in Section 3. In Section 4, we apply these facts to count subgraphs of $\mathscr{G}$ of a certain form. In Section 5, we specialise to the setting of Section 2, and prove the combinatorial statements formulated there.

Section 6 collects some facts related to Levitan's uniqueness theorem [17] which we use in the sequel.

In Section 7, we conclude the proofs of the main results. Section 8 explains how to modify the proofs written for (1.2) to deal with (1.3). In Section 9, we discuss generalizations and related open problems.

Notation. In this paper, $C, c, \ldots$ stand for positive constants, the value of which may change from line to line. If X, Y are quantities depending on some large parameter, then $X \ll Y \iff Y \gg X \iff X = o(Y) \iff X/Y \to 0$.

2. **Preliminaries**

From this section, we omit the subscript N, and write H for H_N and W for W_N. Also, we consider the matrices with entries (1.2); in Section 8 we shall explain the modifications needed for the case (1.3).

Let $U_n(\cos\theta) = \sin((n+1)\theta)/\sin\theta$ be the Chebyshev polynomials of the second kind; set $U_{-2} \equiv U_{-1} \equiv 0$. Let

$$H^{(n)} = (2W-1)^{n/2}\left\{U_n\left(\frac{H}{2\sqrt{2W-1}}\right) - \frac{1}{2W-1}\,U_{n-2}\left(\frac{H}{2\sqrt{2W-1}}\right)\right\}.$$

The following lemma (see, e.g., [25] for a more general version) is at the basis of our considerations.

LEMMA 2.1. *For any Hermitian $N \times N$ matrix H satisfying*

$$H_{uv} = \begin{cases} \pm 1, & 0 < |u-v|_N \le W \\ 0, & \textit{otherwise} \end{cases},$$

and any $u_0, u_n \in \mathbb{Z}/N\mathbb{Z}$,

$$H^{(n)}_{u_0 u_n} = \sum{}' H_{u_0u_1} H_{u_1u_2} \cdots H_{u_{n-1}u_n},$$

where the sum is over all $(n-1)$-tuples $(u_1, u_2, \ldots, u_{n-1})$ (which we regard as paths $p_n = u_0u_1\cdots u_n$), such that

(a) $0 < |u_j - u_{j+1}|_N \le W$, $0 \le j \le n-1$ (*p_n is a path on $\mathscr{G}$*),

(b) $u_{j+2} \ne u_j$, $0 \le j \le n-2$ (*p_n is* nonbacktracking).

The following corollary is immediate from the lemma and (1.2).

COROLLARY 2.2. *For the random matrix H as defined in the introduction,*

$$\mathbb{E}\,\mathrm{tr}\,H^{(n(1))}\,\mathrm{tr}\,H^{(n(2))}\cdots\mathrm{tr}\,H^{(n(k))}$$

is equal to the number of k-tuples of paths (*shortly*: *k-paths*)

$$p_{n(1),\dots,n(k)} = u_0^1 u_1^1 \cdots u_{n(1)}^1 \; u_0^2 u_1^2 \cdots u_{n(2)}^2 \; \dots \; u_0^k u_1^k \cdots u_{n(k)}^k$$

that satisfy (a), (b), *and also*

(c) $u_{n(j)}^j = u_0^j$, $1 \le j \le k$ (*the paths are closed*);

(d) *the number*

$$\#\left\{(i,j) \mid u_i^j = u,\, u_{i+1}^j = v\right\} - \#\left\{(i,j) \mid u_i^j = v,\, u_{i+1}^j = u\right\}$$

is even, *for any* $u, v \in \mathbb{Z}/N\mathbb{Z}$.

As in [9], we group the k-paths satisfying (a)–(d) into topological equivalence classes, which are in one-to-one correspondence with *k-diagrams*:

Definition 2.3. Let $\beta \in \{1,2\}$. A *k-diagram* is an (undirected) multigraph $\bar{G} = (\bar{V}, \bar{E})$, together with a k-tuple of circuits

$$\bar{p} = \bar{u}_0^1 \bar{u}_1^1 \cdots \bar{u}_0^1, \;\; \bar{u}_0^2 \bar{u}_1^2 \cdots \bar{u}_0^2, \;\; \dots, \;\; \bar{u}_0^k \bar{u}_1^k \cdots \bar{u}_0^k \tag{2.1}$$

on $\bar{G}$, such that

- $\bar{p}$ is *nonbacktracking* (meaning that in every circuit no edge is followed by its reverse, unless the edge is a loop $\bar{u}\bar{u}$);
- For every $(\bar{u}, \bar{v}) \in \bar{E}$,

$$\#\left\{(i,j) \mid \bar{u}_j^i = \bar{u},\, \bar{u}_{j+1}^i = \bar{v}\right\} + \#\left\{j \mid \bar{u}_j^i = \bar{v},\, \bar{u}_{j+1}^i = \bar{u}\right\} = 2;$$

- the degree of u_0^i in $\bar{G}$ is 1; the degrees of all the other vertices are equal to 3.

Remark 2.4. We emphasize that $\bar{G}$ is a multigraph in which the coinciding edges are distinguished. Thus, strictly speaking, a circuit is not uniquely determined by the vertices it passes. Still, we find it convenient to use the notation as in (2.1). Next, we do not distinguish two diagrams which are isomorphic in the natural sense. Thus, by a diagram we actually mean an equivalence class (e.g. in the second part of the following lemma).

The following lemma summarizes some properties of k-diagrams from [9, Part II].

LEMMA 2.5.

1. *For every k-diagram, there exists an integer $s \ge k$* (“*nonorientable genus*”; *cf. Figures* 1 *and* 2), *such that the diagram has* $2s$ *vertices and* $3s - k$ *edges.*
2. *The number $D_k(s)$ of k-diagrams corresponding to a given s satisfies*

$$\frac{(s/C)^{s+k-1}}{(k-1)!} \le D_k(s) \le \frac{(Cs)^{s+k-1}}{(k-1)!}.$$

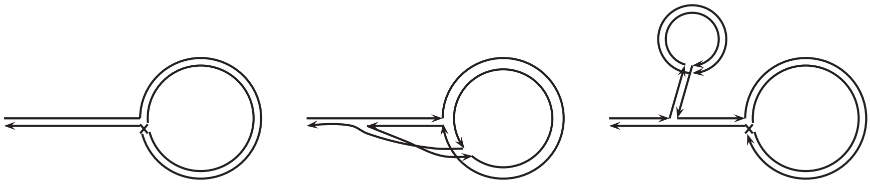

Figure 1. Some 1-diagrams: $s = 1$ (left), $s = 2$ (center, right)

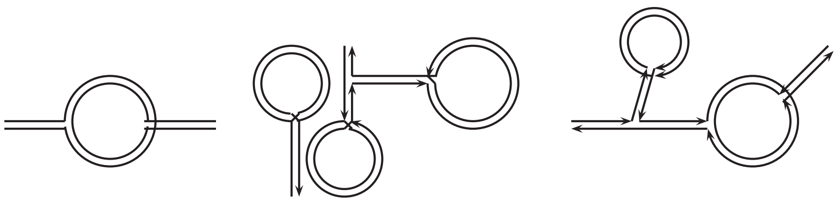

Figure 2. Some 2-diagrams: $s = 2$ (left, center), $s = 3$ (right)

Thus our goal is to compute the number of k-paths corresponding to a given diagram. It will be convenient to consider connected diagrams only. This can be done as follows: set

$$T(n(1)) = \mathbb{E}\,\mathrm{tr}\, H^{(n(1))},$$

$$T(n(1), n(2)) = \mathbb{E}\,\mathrm{tr}\, H^{(n(1))}\,\mathrm{tr}\, H^{(n(2))} - T(n(1))T(n(2)),$$

$$\cdots\cdots$$

$$T(n(1), \dots, n(k)) = \mathbb{E} \prod_{j=1}^{k} H^{(n(j))} - \sum_{\Pi} \prod_{P \in \Pi} T(\{n(j)\}_{j \in P}),$$

where the sum is over nontrivial partitions of $\{1, 2, \dots, k\}$ into disjoint sets. It is not hard to see that indeed $T(n(1), \dots, n(k))$ counts the number of k-paths $p_{n(1),\dots,n(k)}$ that satisfy (a)–(d) and correspond to connected diagrams. Now we formulate the two main technical statements of this paper.

PROPOSITION 2.6. *Let* $W \gg N^{5/6}$, $R \geq 0$. *For any*

$$1 \leq n(1) \leq n(2) \cdots \leq n(k) \leq RN^{1/3}$$

such that $n(1) + \cdots + n(k) = 2n$ *is even*,

$$\frac{T(n(1), \dots, n(k))}{(2W-1)^n} = \left[\prod_{j=1}^{k} n(j)\right] \phi_k(n(1)/N^{1/3}, \dots, n(k)/N^{1/3}) + N^{k/3} \varepsilon_N(n(1), \dots, n(k)),$$

where

1. $\phi_k(z_1, \dots, z_k) = \sum_{s \geq k} g_{k,s}(z_1, \dots, z_k)$, $g_{k,s}$ *being a continuous homogeneous function of degree* $3(s-k)$,

$$g_{k,s}(z_1, \dots, z_k) \leq \frac{(C\|z\|)^{3(s-k)}}{(cs)^{2s-3k+1}}.$$

2. $\varepsilon_N(n(1), \dots, n(k)) = o(1)$, *where the implicit constant depends only on* R *and on* $W/N^{5/6}$, *and*

$$\sum_{1 \leq n(1) \leq \dots \leq n(k) \leq N^{1/3}} \frac{\varepsilon_N(n(1), \dots, n(k))}{n(1) \cdots n(k)} \leq C_k.$$

If $n(1) + \dots + n(k) = 2n+1$ *is odd*, $T(n(1), \dots, n(k)) = 0$.

PROPOSITION 2.7. *Let* $W \ll N^{5/6}$, $R \geq 0$. *For any*

$$1 \leq n(1) \leq \dots \leq n(k) \leq RW^{2/5}$$

such that $n(1) + \dots + n(k) = 2n$ *is even*,

$$\frac{T(n(1), \dots, n(k))}{(2W-1)^n}$$
$$= \frac{N}{W^{6/5}} \left\{ \left[\prod_{j=1}^{k} n(j) \right] \psi_k \left(\frac{n(1)}{W^{2/5}}, \dots, \frac{n(k)}{W^{2/5}} \right) + W^{2k/5} \varepsilon(n(1), \dots, n(k)) \right\},$$

where

1. $\psi_k(z_1, \dots, z_k) = \sum_{s \geq k} h_{k,s}(z_1, \dots, z_k)$, $h_{k,s}$ *being a homogeneous function of degree* $(5s-5k-1)/2$, *continuous outside the origin*,

$$h_{k,s}(z_1, \dots, z_k) \leq \frac{(C\|z\|)^{\frac{5s-5k-1}{2}}}{(cs)^{\frac{3s-5k+1}{2}}}.$$

2. $\varepsilon_N(n(1), \dots, n(k)) = o(1)$, *where the implicit constant depends only on* R *and on* $N^{5/6}/W$, *and*

$$\sum_{1 \leq n(1) \leq \dots \leq n(k) \leq W^{2/5}} \frac{\varepsilon_N(n(1), \dots, n(k))}{n(1) \cdots n(k)} \leq C_k.$$

If $n(1) + \dots + n(k) = 2n+1$ *is odd*, $T(n(1), \dots, n(k)) = 0$.

3. **Random walk on the circle**

Let $\mathcal{G} = (\mathbb{Z}/N\mathbb{Z}, \mathcal{E})$ be defined by (1.4) (in this section we do not assume that $W \gg 1$). Denote by $\mathcal{W}_n(R)$ the number of paths of length n in $\mathcal{G}$ between two (fixed) vertices u, v such that $|u-v|_N = R$. We prove the following statements.

PROPOSITION 3.1. *If* $1 \ll n \ll N^2/W^2$, $R \ll n^{3/4}W$, *and* $R \leq 0.49N$, *then*

$$\frac{\mathcal{W}_n(R)}{(2W)^n} = (1+o(1)) \left[\frac{\pi n}{3}(W+1)(2W+1) \right]^{-1/2} \exp\left\{ -\frac{3R^2}{n(W+1)(2W+1)} \right\}.$$

PROPOSITION 3.2. *If $N^2/W^2 \ll n$, then*

$$\frac{\mathcal{W}_n(R)}{(2W)^n} = (1+o(1))N^{-1}.$$

PROPOSITION 3.3. *Without any assumptions on n, N, R, W,*

$$\frac{\mathcal{W}_n(R)}{(2W)^n} \le C\left[(W\sqrt{n})^{-1}\exp\left\{-\frac{CR^2}{nW^2}\right\} + N^{-1}\right].$$

For fixed W, Proposition 3.1 follows from Richter's local limit theorem for lattice variables (see Ibragimov and Linnik, [13, Ch. VII]). However, we need the asymptotics to be uniform in W, which we have not found in the literature. Proposition 3.2 (perhaps with an extra logarithmic factor in the assumptions) follows easily from the spectral estimates on the mixing time.

Let $\mathcal{A}$ be the adjacency matrix of $\mathcal{G}$. As $\mathcal{G}$ is invariant under cyclic shifts, the discrete Fourier transform diagonalises $\mathcal{A}$. We state this as a lemma:

LEMMA 3.4. $(2W)^{-1}\mathcal{A} = \sum_{k=0}^{N-1} a_k \mathfrak{f}_k \otimes \mathfrak{f}_k$, *where*

$$a_k = \frac{\sin\left(W\frac{\pi k}{N}\right)}{W\sin\frac{\pi k}{N}}\cos\left((W+1)\frac{\pi k}{N}\right),$$

and $\mathfrak{f}_k(\ell) = \exp\left\{\frac{2\pi i k\ell}{N}\right\}$.

Proof. The vectors $\mathfrak{f}_k$ are of unit length; hence we only need to check that $\mathcal{A}\mathfrak{f}_k = 2Wa_k\mathfrak{f}_k$. Let $\omega = \exp\frac{2\pi i}{N}$. Then

$$(\mathcal{A}\mathfrak{f}_k)(m) = \sum_{0<|\ell-m|_N\le W}\omega^{k\ell} = \mathfrak{f}_k(m)\sum_{0<|\ell|_N\le W}\omega^{k\ell}.$$

Now,

$$\begin{aligned}\sum_{0<|\ell|_N\le W}\omega^{k\ell} &= -1+\omega^{-kW}\frac{1-\omega^{(2W+1)k}}{1-\omega^k}\\ &= \frac{\omega^{(W+1/2)k}-\omega^{k/2}+\omega^{-k/2}-\omega^{-(W+1/2)k}}{\omega^{k/2}-\omega^{-k/2}}\\ &= \frac{\omega^{Wk/2}-\omega^{-Wk/2}}{\omega^{k/2}-\omega^{-k/2}}\left(\omega^{(W+1)k/2}+\omega^{-(W+1)k/2}\right) = 2Wa_k.\end{aligned}$$

□

Denote by $\delta_0,\dots,\delta_{N-1}$ the standard basis in $\mathbb{R}^{\mathbb{Z}/N\mathbb{Z}}$. Then

$$\delta_j = N^{-1/2}\sum_{k=0}^{N-1}\exp\frac{-2\pi ijk}{N}\,\mathfrak{f}_k;$$

therefore

$$\frac{\mathcal{W}_n(R)}{(2W)^n} = \langle(\mathcal{A}/2W)^n\delta_R,\delta_0\rangle = N^{-1}\sum a_k^n\exp\frac{2\pi iRk}{N}. \tag{3.1}$$

Thus we have to estimate the above sum. Informally, the argument is as follows: $\max(|a_1|, \dots, |a_{N-1}|) = 1 - \Theta(N^2/W^2)$, whereas $a_0 = 1$. If $n \gg N^2/W^2$, then the sum is dominated by the first addend, which is equal to 1. If $n \ll N^2/W^2$, then the sum can be replaced with an integral, which is then evaluated using the saddle point method. And now to the formal proof.

Denote

$$f(z) = \frac{\sin[W\pi z]}{W\sin[\pi z]} \cos[(W+1)\pi z]\,.$$

The following lemma summarizes some properties of f:

LEMMA 3.5. *The function f is an entire function with period* 1;

$$\frac{|f(x+iy)|}{|f(iy)|} \le \exp\left\{-cW^2x^2\right\}, \quad -1/2 \le x \le 1/2.$$

Proof of Proposition 3.1. Choose $\frac{\sqrt{n}W}{N} \ll \eta \ll 1$, and consider two cases.

(a) $R \le \eta\sqrt{n}W$. Then, according to (3.1) and Lemma 3.5,

$$\begin{aligned}\frac{\mathcal{W}_n(R)}{(2W)^n} &= N^{-1} \sum_{-N/2<k\le N/2} \left[f(k/N)\right]^n \exp\frac{2\pi i Rk}{N}\\ &= N^{-1} \sum_{-N/2<k\le N/2} \left[f(k/N)\right]^n \left[1 + O\left(\frac{\eta|k|\sqrt{n}W}{N}\right)\right].\end{aligned}$$

The contribution of the second addend is negligible:

$$\begin{aligned}&\left|N^{-1} \sum_{-N/2<k\le N/2} \left[f(k/N)\right]^n \frac{\eta|k|\sqrt{n}W}{N}\right|\\ &\quad\le \left|N^{-1} \sum_{-N/2<k\le N/2} \exp\left\{-\frac{cW^2nk^2}{N^2}\right\} \frac{\eta|k|\sqrt{n}W}{N}\right| \le \frac{C\eta}{\sqrt{n}W} \ll \frac{1}{\sqrt{n}W}.\end{aligned}$$

Then,

$$\begin{aligned}N^{-1}\sum\left[f(k/N)\right]^n &= N^{-1}\sum\exp\left\{-\frac{\pi n}{3}(W+1)(2W+1)\frac{k^2}{N^2}\right\}\\ &\quad+N^{-1}\sum\left(\left[f(k/N)\right]^n - \exp\left\{-\frac{\pi n}{3}(W+1)(2W+1)\frac{k^2}{N^2}\right\}\right).\end{aligned}$$

The second sum is negligible, whereas the first one is

$$\begin{aligned}(1+o(1))N^{-1}\sum_{k=-\infty}^{\infty} &\exp\left\{-\frac{\pi n}{3}(W+1)(2W+1)\frac{k^2}{N^2}\right\}\\ &= (1+o(1))\left[\frac{\pi n}{3}(W+1)(2W+1)\right]^{-1/2},\end{aligned}$$

according to the Poisson summation formula.

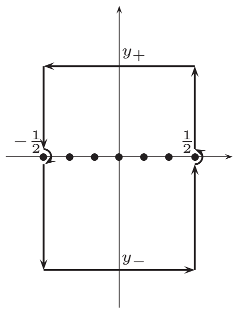

Figure 3. The contour.

(b) $R > \eta\sqrt{n}W$. According to (3.1), Lemma 3.5, and the residue theorem,

$$\begin{aligned}\frac{\mathcal{W}_n(R)}{(2W)^n} &= N^{-1} \sum_{-N/2<k\le N/2} \big[f(k/N)\big]^n \exp\frac{2\pi i R k}{N} \\ &= -\int_C \big[f(z)\big]^n \frac{\exp[2\pi i R z]}{1-\exp[2\pi i N z]} dz,\end{aligned}$$

where the contour C encloses the zeros k/N, $-N/2 < k \le N/2$, of the denominator, as in Figure 3.

The sum of integrals along the vertical parts of C vanishes; hence

$$\frac{\mathcal{W}_n(R)}{(2W)^n} = \int_{iy_+-1/2}^{iy_++1/2} - \int_{iy_--1/2}^{iy_-+1/2}.$$

The first integral:

$$I_+ = \int_{iy_+-1/2}^{iy_++1/2} = \int_{iy_+-1/2}^{iy_++1/2} \exp\phi_+(z) \frac{dz}{1-\exp[2\pi i N z]},$$

where $\phi_+(z) = n \ln f(z) + 2\pi i R z$. We choose y_+ to make ϕ'_+ vanish at $z_+ = iy_+$. Expanding ϕ_+ in Taylor series at 0, we obtain

$$\phi_+(z) = -\frac{n\pi^2}{3}(W+1)(2W+1)z^2\,(1+O(z^2W^2)) + 2\pi i R z.$$

This expansion is of course differentiable, so

$$\begin{aligned}\phi'_+(z) &= -\frac{2n\pi^2}{3}(W+1)(2W+1)z\,(1+O(z^2W^2)) + 2\pi i R, \\ \phi''_+(z) &= -\frac{2n\pi^2}{3}(W+1)(2W+1)\,(1+O(z^2W^2));\end{aligned}$$

therefore there exists a solution $z_+ = iy_+$ of $\phi'_+(z_+) = 0$ such that

$$z_+ = \frac{3iR}{\pi n(W+1)(2W+1)}\Big(1 + O(R^2/(n^2W^2))\Big), \tag{3.2}$$
$$\phi_+(z_+) = \frac{3R^2}{n(W+1)(2W+1)} + O(R^4/(n^3W^4)),$$
$$\phi''_+(z_+) = -\frac{2n\pi^2}{3}(W+1)(2W+1)(1 + O(R^2/(n^2W^2))).$$

Under the assumptions of Proposition 3.1, all the error terms in (3.2) are $o(1)$. Also,

$$\frac{1}{1 - \exp\{2\pi i N z\}} = 1 + O\left(\exp\{-2\pi N y_+\}\right) = 1 + o(1)$$

uniformly in $-1/2 \le x \le 1/2$, since

$$Ny_+ \ge \frac{cNR}{nW^2} \ge \frac{cN\eta}{\sqrt{n}W} \gg 1.$$

These observations and Lemma 3.5 justify the saddle point approximation, which yields

$$I_+ = (1 + o(1))\left[\frac{\pi n}{3}(W+1)(2W+1)\right]^{-1/2} \exp\left\{-\frac{3R^2}{n(W+1)(2W+1)}\right\}.$$

Now consider the second integral:

$$I_- = -\int_{iy_- - 1/2}^{iy_- + 1/2} = \int_{iy_- - 1/2}^{iy_- + 1/2} \exp\phi_-(z)\frac{\exp[2\pi i N z]}{\exp[2\pi i N z] - 1}\,dz,$$

where $\phi_-(z) = n\ln f(z) + 2\pi i(R - N)z$. According to Lemma 3.5,

$$|I_-| \le \exp\phi_-(iy_-)\int_{-1/2}^{1/2}\exp\left\{-cnW^2\right\}dx$$
$$\le \frac{C}{\sqrt{nW}}\exp\phi_-(iy_-).$$

Choosing $y_- = -y_+$, we see that

$$|I_-| \ll \frac{1}{\sqrt{n}W}\exp\left\{-\frac{3R^2}{n(W+1)(2W+1)}\right\}. \qquad \square$$

Proof of Proposition 3.2. According to (3.1),

$$\frac{\mathscr{W}_n(R)}{(2W)^n} = N^{-1}\sum_{-N/2<k\le N/2}\left[f(k/N)\right]^n \exp\frac{2\pi i R k}{N}.$$

We have: $f(0) = 1$, whereas by Lemma 3.5

$$\left|\left[f(k/N)\right]^n\right| \le \exp\left\{-cnW^2k^2/N^2\right\}.$$

Therefore

$$\left| \sum_{|k|>0} \left[f(k/N) \right]^n \exp \frac{2\pi i R k}{N} \right| \le C \exp \left\{ -\frac{cnW^2}{N^2} \right\} \ll 1. \qquad \square$$

Proof of Proposition 3.3 (*sketch*). Consider two cases: $\sqrt{n}W \ge N$, and $\sqrt{n}W < N$. In the former case,

$$\frac{\mathcal{W}_n(R)}{(2W)^n} \le \frac{1}{N} \sum_{-N/2<k\le N/2} \exp\left\{-cnW^2k^2/N^2\right\} \le C/N.$$

In the latter case, we proceed as in the proof of Proposition 3.1, making sure that one can choose $y_\pm$ so that

$$\phi_\pm(iy_\pm) \le -\frac{cR^2}{nW^2}. \qquad \square$$

4. **Embeddings into $\mathcal{G}$**

In this section, we study the number of "topological embeddings" of a graph G into $\mathcal{G}$, satisfying certain restrictions. Let us start with the precise definitions.Let $G = (V, E)$ be a connected multigraph, and let $\overrightarrow{n} = (n_1, \dots, n_E)$ be an E-tuple of numbers[2], $n_e \ge -1$. Construct a new graph $G^{\overrightarrow{n}}$ as follows:

- if $n_e \ge 0$, then replace $e = (u, v)$ with a chain $(u, u_1, u_2, \dots, u_{n_e}, v)$;
- if $n_e = -1$, then contract $e = (u, v)$ into a single vertex.

Finally, consider a system of linear equations $\mathfrak{E}$ in the variables $n_1, \dots, n_E$:

$$\mathfrak{E}: \quad \sum_{e\in E} c_j(e) n_e = n(j), \quad j = 1, \dots, k,$$

and denote by $\mathrm{Emb}(G, \mathfrak{E})$ the number of subgraphs $G' \hookrightarrow \mathcal{G}$ that are isometric to $G^{\overrightarrow{n}}$ for some $\overrightarrow{n}$ satisfying the system $\mathfrak{E}$, and by $\mathrm{Emb}_+(G, \mathfrak{E})$ — the number of such subgraphs with the additional requirement $n_e > 0$, $e \in E$. We impose the following restrictions on $\mathfrak{E}$:

1. $\mathfrak{E}$ is linearly independent;
2. $c_j(e) \ge 0$, $1 \le j \le k$, $e \in E$;
3. $\sum_{j=1}^k c_j(e) = 2$, $e \in E$.

These restrictions are satisfied for the systems of equations that we need, and slightly simplify the notation.

[2]to simplify the notation, we write $V = \#V$, $E = \#E$ throughout this section.

PROPOSITION 4.1. *In the setting above and under the assumptions* 1–3, *if* $\frac{E^2N^2}{W^2} \ll n(1) \le n(2) \le \cdots \le n(k)$ *and* $\sum n(j) = 2n \ll W$, *then*

$$\frac{\mathrm{Emb}(G, \mathfrak{E})}{(2W-1)^n} = (1 + o(1))\, N^{-E+V}\, g_G(n(1), \dots, n(k)),$$

where g_G *is a continuous homogeneous function of degree* $E - k$,

$$g_G(n(1), \dots, n(k)) \le \frac{(Cn)^{E-k}}{(E-k)!}.$$

The same is true for $\mathrm{Emb}_+(G, \mathfrak{E})$. *If* $\sum n(j) = 2n + 1$,

$$\mathrm{Emb}(G, \mathfrak{E}) = \mathrm{Emb}_+(G, \mathfrak{E}) = 0.$$

PROPOSITION 4.2. *In the setting above and under the assumptions* 1–3, *if* $E^2 \ll n(1) \le n(2) \le \cdots \le n(k) \ll \min\left(N^2/W^2,\ W\right)$, $\sum n(j) = 2n$, *then*

$$\frac{\mathrm{Emb}(G, \mathfrak{E})}{(2W-1)^n} = (1 + o(1))\, N\, (2W)^{-E+V-1} h_G(n(1), \dots, n(k)),$$

where h_G *is a homogeneous function of degree* $\frac{E+V-2k-1}{2}$, *continuous outside the origin, and*

$$h_G(n(1), \dots, n(k)) \le \frac{(Cn)^{\frac{E+V-2k-1}{2}}}{\left(\frac{E+V-2k-1}{2}\right)!}.$$

The same is true for $\mathrm{Emb}_+(G, \mathfrak{E})$. *If* $\sum n(j) = 2n + 1$,

$$\mathrm{Emb}(G, \mathfrak{E}) = \mathrm{Emb}_+(G, \mathfrak{E}) = 0.$$

PROPOSITION 4.3. *Without any restrictions on* n, W,

$$\frac{\mathrm{Emb}(G, \mathfrak{E})}{(2W)^n} \le \frac{(C_1 n)^{E-k}}{(E-k)!} (N)^{V-E} + C_2 N\, \frac{(C_3 n)^{\frac{E+V-2k-1}{2}}}{\left(\frac{E+V-2k-1}{2}\right)!} W^{-E+V-1}.$$

To apply the results of Section 3, we need a simple observation. Let $u, v \in \mathbb{Z}/N\mathbb{Z}$, $A, B \subset \mathbb{Z}/N\mathbb{Z}$. Denote by $\widetilde{\mathcal{W}}_n(u, v, A, B)$ the number of paths $p_n = u u_1 u_2 \cdots u_{n-1} v$ from u to v in $\mathcal{G}$, such that

- $u \neq u_2, u_1 \neq u_3, \dots, u_{n-2} \neq v$ (“*nonbacktracking*”),
- $u_1 \notin A$, $u_{n-1} \notin B$.

LEMMA 4.4. *In the notation above*,

$$\mathcal{W}_n(|u-v|_N) \ge \widetilde{\mathcal{W}}_n(u, v, A, B) \ge \left(1 - C\, \frac{n + \#A + \#B}{W}\right) \mathcal{W}_n(|u-v|_N).$$

Proof of Proposition 4.1. Choose $N^2/W^2 \ll n_0 \ll n(1)/E^2$. Let $\Delta_{\mathfrak{E}}$ be the set of real nonnegative solutions of $\mathfrak{E}$; obviously, $\Delta_{\mathfrak{E}}$ is an $(E-k)$-dimensional polytope with $E+k$ faces. It is not hard to see that

$$\mathrm{Vol}_{E-k-1}\,\partial\Delta_{\mathfrak{E}} = O\left(\frac{E^2}{n(1)}\,\mathrm{Vol}\,\Delta_{\mathfrak{E}}\right).$$

Hence, the number $\mathfrak{n}(\mathfrak{E})$ of integer solutions $\overrightarrow{n}$ of $\mathfrak{E}$ with $n_e \geq -1$ satisfies:

$$\mathfrak{n}(\mathfrak{E}) = \gamma_{\mathfrak{E}}\,\mathrm{Vol}_{E-k-1}\,\Delta_{\mathfrak{E}}\left\{1 + O(E^2/n(1))\right\} = \mathrm{Vol}\,\Delta'_{\mathfrak{E}}\left\{1 + O(E^2/n(1))\right\},$$

where $\Delta'_{\mathfrak{E}} = \mathrm{Proj}_{\mathbb{R}^{E-k-1}}\,\Delta_{E}q$ is the projection of $\Delta_{\mathfrak{E}}$ onto a set of $E-k-1$ independent coordinates. Moreover, the number of solutions with at least one coordinate $< n_0$ is at most

$$\mathrm{Vol}\,\Delta'_{\mathfrak{E}} \cdot O(n_0 E^2/n(1)). \tag{4.1}$$

Let $\overrightarrow{n}$ be a solution with

$$n_1, \dots, n_E \geq n_0. \tag{4.2}$$

It corresponds to

$$N^{V-E}(2W-1)^n(1+o(1))$$

different embeddings. Indeed, there are N^V ways to embed the vertices of G, and, according to Proposition 3.2 and Lemma 4.4,

$$(2W-1)^n n^{-E}(1+o(1))$$

ways to embed the edges. Thus the total number of embeddings corresponding to solutions that satisfy (4.2) is

$$\mathrm{Vol}\,\Delta'_{\mathfrak{E}}\, N^{V-E}(2W-1)^n(1+o(1)). \tag{4.3}$$

Applying (4.1), Proposition 3.3 and Lemma 4.4, we see that the number of embeddings that violate (4.2) is negligible with respect to (4.3). Therefore

$$\frac{\mathrm{Emb}(G,\mathfrak{E})}{(2W-1)^n} = (1+o(1))\ \mathrm{Vol}\,\Delta'_{\mathfrak{E}}\, N^{V-E}.$$

Finally, observe that $\mathrm{Vol}\,\Delta'_{\mathfrak{E}}$ is an $(E-k)$-homogeneous function of $n(1), \dots, n(k)$, and that

$$\mathrm{Vol}\,\Delta'_{\mathfrak{E}} \leq \mathrm{Vol}\,\{0 \leq u_1, \dots, u_E,\ u_1 + \cdots + u_E \leq n\} = \frac{n^{E-k}}{(E-k)!}. \qquad \square$$

Proof of Proposition 4.2. First, number the edges $1, \dots, E$ so that $1, \dots, V-1$ is a spanning tree. For $e = (i_e, t_e)$, let $R_e = i'_e - t'_e$, where $i'_e, t'_e \in \mathbb{Z}/N\mathbb{Z}$ are the images of i_e, t_e in G'. Note that all the R_e are linear combinations of $R_1, \dots, R_{V-1}$. To every $e \in E$ we correspond a vector $v_e \in \mathbb{R}^{V-1}$ such that

$$R_e = \sum_{f=1}^{V-1} v_e(f) R_f.$$

By Proposition 3.1 and Lemma 4.4,

$$\begin{aligned}\frac{\mathrm{Emb}(G,\mathfrak{E})}{(2W-1)^n} &= (1+o(1))\frac{\mathrm{Emb}(G,\mathfrak{E})}{(2W)^n} \\ &= (1+o(1)) \sum_{R_1,\dots,R_{V-1}} \sum_{\vec{n}}{}^{*} \prod_{e\in E} \left[\frac{2\pi}{3} n_e W^2\right]^{-1/2} \exp\left\{-\frac{3}{2}\frac{R_e^2}{n_e W^2}\right\},\end{aligned} \tag{4.4}$$

where the sum is over n_e satisfying $\mathfrak{E}$. Here we have disregarded the contribution of $n_e \leq n_0$, where $1 \ll n_0 \ll n(1)$; this can be easily justified using Proposition 3.3 and the asymptotic estimates below. Now, applying the Poisson summation formula and discarding the negligible terms, we have

$$\begin{aligned}&\sum_{R_1,\dots,R_{V-1}\in\mathbb{Z}} \exp\left\{-\frac{3}{2}\frac{R_e^2}{n_e W^2}\right\} \\ &\qquad = (1+o(1)) \left[\frac{2\pi W^2}{3}\right]^{\frac{V-1}{2}} \left\{\det\left[\sum_e n_e^{-1} v_e \otimes v_e\right]\right\}^{-1/2}.\end{aligned}$$

According to the Cauchy-Binet formula,

$$\det \sum_e c_e v_e \otimes v_e = \sum_{T\in\binom{E}{V-1}} \prod_{e\in E} c_e \det{}^2 (v_e)_{e\in T}.$$

If T is not a spanning tree of G, then $(v_e)_{e\in T}$ is not of full rank and hence has zero determinant. If T is a spanning tree, then $(v_e)_{e\in T}$ and its inverse are integer matrices, hence the squared determinant is equal to 1; therefore

$$\det \sum_e c_e v_e \otimes v_e = \sum_T \sum_{e\in T} c_e,$$

where now the sum is over spanning trees $T \subset E$. Going back to (4.4), we deduce

$$\begin{aligned}\frac{\mathrm{Emb}(G,\mathfrak{E})}{(2W-1)^n} &= (1+o(1)) \\ &\times \left[\frac{2\pi W^2}{3}\right]^{\frac{-E+V-1}{2}} \sum_{\vec{n}}{}^{*} \left(\prod_{e\in E}\frac{1}{\sqrt{n_e}}\right)\left(\sum_T \prod_{e\in T}\frac{1}{n_e}\right)^{-1/2}.\end{aligned} \tag{4.5}$$

Finally, replace the sum $\sum^*$ with an integral over the polytope

$$\Delta'_{\mathfrak{E}} = \mathrm{Proj}_{\mathbb{R}^{E-k}} \Delta_{\mathfrak{E}},$$

where $\Delta_{\mathfrak{E}}$ is the set of nonnegative real solutions of $\mathfrak{E}$, and the projection is onto an independent subset of coordinates. This can be done, for example, using the

Poisson summation formula. We deduce

$$\frac{\mathrm{Emb}(G,\mathfrak{E})}{(2W-1)^n} = (1+o(1))\left[\frac{2\pi W^2}{3}\right]^{\frac{-E+V-1}{2}} \times \underset{\Delta'_{\mathfrak{E}}}{\int\cdots\int}\left(\prod_{e\in E}\frac{1}{\sqrt{u_e}}\right)\left(\sum_T\prod_{e\in T}\frac{1}{u_e}\right)^{-1/2} du_1\cdots du_{E-k} \tag{4.6}$$

(where all the other u_e are determined from $\mathfrak{E}$). It is easy to see that the integral in the right-hand side is $\frac{E+V-2k-1}{2}$-homogeneous in $n(1),\dots,n(k)$. Replacing Σ_T with a single spanning tree T_0, we deduce the upper bound. ☐

Example 4.5. The reader may find the following example illustrative: for $k=1$,

$$\mathfrak{E}:\quad \sum 2n_e = 2n,$$

we have

$$\underset{\Delta'_{\mathfrak{E}}}{\int\cdots\int} = n^{\frac{E+V-3}{2}}\underset{\tilde{\Delta}'_{\mathfrak{E}}}{\int\cdots\int},$$

where the set $\tilde{\Delta}'_{\mathfrak{E}} = n^{-1}\Delta'_{\mathfrak{E}} = \{\sum x_e \le 1, x_e \ge 0\}$ does not actually depend of n.

We omit the proof of Proposition 4.3, which is very similar to the proofs of the previous two propositions.

5. **Proofs of Propositions** 2.6 **and** 2.7

According to Lemma 2.5 and the discussion preceding and following it,

$$T(n(1),\dots,n(k)) = \sum_{\mathcal{D}} \mathrm{Paths}(n(1),\dots,n(k);\mathcal{D}),$$

where the sum is over connected diagrams $\mathcal{D} = (\bar{G} = (\bar{V},\bar{E}),\bar{p})$, and

$$\mathrm{Paths}(n(1),\dots,n(k);\mathcal{D})$$

is the number of k-paths $p_{n(1),\dots,n(k)}$ corresponding to $\mathcal{D}$ and satisfying (a)–(d) from Section 2. Let

$$c_j(\bar{e}) = \#\left\{\begin{array}{c}\text{times that the } j\text{-th part}\\ \text{of } p \text{ passes through } \bar{e}\end{array}\right\} \in \{0,1,2\},\quad 1\le j\le k,\ \bar{e}\in\bar{E},$$

and consider the system of equations

$$\mathfrak{E}_{\mathcal{D}}:\quad \sum_{\bar{e}} c_j(\bar{e})n_{\bar{e}} = n(j),\quad 1\le j\le k.$$

Then

$$\mathrm{Emb}_+(\bar{G},\mathfrak{E}_{\mathcal{D}}) \le \mathrm{Paths}(n(1),\dots,n(k);\mathcal{D}) \le \mathrm{Emb}(\bar{G},\mathfrak{E}_{\mathcal{D}}); \tag{5.1}$$

here $\#\bar{V} = 2s$ and $\#\bar{E} = 3s-k$. Therefore we can apply the results of Section 4.

Proof of Proposition 2.6. Choose n_0 and s_0 so that

$$N^{1/3} \gg n_0 \gg \frac{s_0 N^2}{W^2} \gg \frac{N^2}{W^2}.$$

If $n_0 \leq n(1) \leq \cdots \leq n(k) \leq RN^{1/3}$, and $\mathcal{D}$ is a connected diagram with $2s$ vertices and $3s-k$ edges, $1 \leq s \leq s_0$, then Proposition 4.1 yields

$$\begin{aligned}\frac{\operatorname{Paths}(n(1),\dots,n(k);\mathcal{D})}{(2W-1)^n} &= (1+o(1))\, g_{\mathcal{D}}(n(1),\dots,n(k)) N^{-s+k} \\ &= (1+o(1))\, n(1)\cdots n(k)\, \tilde{g}_{\mathcal{D}}(n(1)/N^{1/3},\dots,n(k)/N^{1/3}),\end{aligned}$$

where $g_{\mathcal{D}}$ is homogeneous of degree $3s-k$, and

$$\tilde{g}_{\mathcal{D}}(x_1,\dots,x_k) = (x_1\cdots x_k)^{-1} g_{\mathcal{D}}(x_1,\dots,x_k).$$

If $s > s_0$, Proposition 4.3 yields

$$\frac{\operatorname{Paths}(n(1),\dots,n(k);\mathcal{D})}{(2W-1)^n} \leq \frac{(C'n)^{3s-2k}}{(3s-2k)!}\,(C'N)^{-s+k};$$

hence the total contribution of all diagrams with $s > s_0$ is at most

$$\begin{aligned}&C_k \sum_{s>s_0} \frac{(C'n)^{3s-2k}}{(3s-2k)!}\,(C'N)^{-s+k}\,(Cs)^{s+k-1} \\ &\qquad \leq C_1(k) n^k \sum_{s>s_0} \frac{(C_1 n^3/N)^{s-k}}{s^{2s-3k+1}} = o(n^k) = o(N^{k/3}).\end{aligned}$$

As the same bound holds for the sum of

$$n(1)\cdots n(k)\,\tilde{g}_{\mathcal{D}}(n(1)/N^{1/3},\dots,n(k)/N^{1/3})$$

over diagrams with $s > s_0$, we deduce

$$\begin{aligned}\frac{T(n(1),\dots,n(k))}{(2W-1)^n} &= \Big[\prod n(j)\Big] \sum_{\mathcal{D}} \tilde{g}_{\mathcal{D}}(n(1)/N^{1/3},\dots,n(k)/N^{1/3}) + o(N^{k/3}),\end{aligned}$$

where the sum is now over all connected k-diagrams $\mathcal{D}$. To prove the bound for $n(1) < n_0$, apply Proposition 4.1 in a similar fashion. $\square$

Proof of Proposition 2.7 (*sketch*). Choose s_0, n_0, n_0' so that

$$1 \ll s_0^2 \ll n_0 \ll W^{2/5} \ll n_0' \ll \frac{N^2}{W^2}$$

and proceed as in the previous proof, using Proposition 4.2 instead of Proposition 4.1. $\square$

6. **Digression**

Let $\{\mu_N\}_{N=1}^{\infty}$ be a sequence of probability measures on $\mathbb{R}$, and let $s_N \to +\infty$ be a sequence of (real) positive numbers. We shall study the scaled measures

$$\left\{ \begin{array}{l} \sigma_{R,N}(\lambda) = s_N^3 \left(1 - \mu_N(1 - 2s_N^{-2}\lambda)\right) \\ \sigma_{L,N}(\lambda) = s_N^3 \left(\mu_N(-1 + 2s_N^{-2}\lambda)\right) \end{array} \right. , \quad \lambda \in \mathbb{R}. \tag{6.1}$$

This scaling is meaningful if μ_N are close (in some sense) to the Wigner measure μ_{Wigner}. Let

$$\widehat{\mu_N}(n) = \int U_n(\alpha) d\mu_N(\alpha), \tag{6.2}$$

where as before U_n are the Chebyshev polynomials of the second kind, and assume that

$$\widehat{\mu_N}(n) = \frac{n}{s_N^3} \Big[\phi_R(n/s_N) + (-1)^n \phi_L(n/s_N)\Big] + \frac{\varepsilon_N(n)}{s_N^2}, \tag{6.3}$$

where

$$\phi_L, \phi_R \in C(0, +\infty) \bigcap L_1(\exp\left\{-x^{-2+\delta}\right\} dx) \tag{6.4}$$

(for some $\delta > 0$), and $\varepsilon_N(n)$ are "small" (in a sense made precise below), for $n = O(s_N)$. Then $\sigma_{R,N}$ and $\sigma_{L,N}$ converge to limits that can be expressed in terms of ϕ_R, ϕ_L. We state this as a proposition; the main ingredient of the proof is a variant of Levitan's uniqueness theorem [17]. The assumptions can be definitely relaxed: thus, the integrability condition (6.4) can be replaced with a weaker one using the methods of Levitan and Meiman [18] and Vul [32].

PROPOSITION 6.1. *Let $\{\mu_N\}$ be a sequence of measures on $\mathbb{R}$, and assume that the coefficients $\widehat{\mu_N}$ (defined in* (6.2)*) satisfy* (6.3), *with ϕ_L, ϕ_R as in* (6.4), *and*

1. $\sum_{n=1}^{s_N} \frac{|\varepsilon_N(n)|}{n} \le C$;
2. *for any $R \in \mathbb{N}$, $\varepsilon_N(n) = o(1)$ for $n \in \{1, 2, \ldots, Rs_N\}$, with the implicit constant depending only on R.*

Then the measures $\sigma_{R,N}$ and $\sigma_{L,N}$ (defined in (6.1)*) converge to limiting measures σ_R and σ_L* (*respectively*), *where σ_R and σ_L are uniquely defined by ϕ_R and ϕ_L* (*respectively*). *The limiting measures $\sigma = \sigma_R, \sigma_L$ share the following properties*:

$$\left\{ \begin{array}{ll} \left|\sigma(\lambda) - \frac{2}{3\pi}\lambda_+^{3/2}\right| = O(\lambda), & \lambda \to +\infty; \\ \sigma(\lambda) = O\left[\exp\left\{-C'|\lambda|^{\frac{1-\delta/2}{1-\delta}}\right\}\right], & \lambda \to -\infty. \end{array} \right. \tag{6.5}$$

Remark 6.2. One can show that

$$\int_{-\infty}^{+\infty} \frac{\sin x\sqrt{\lambda}}{x\sqrt{\lambda}} d\tau_R(\lambda) = \phi_R(x), \qquad \sigma_R(\lambda) = \tau_R(\lambda) + \frac{2}{3\pi}\lambda_+^{3/2},$$
$$\int_{-\infty}^{+\infty} \frac{\sin x\sqrt{\lambda}}{x\sqrt{\lambda}} d\tau_L(\lambda) = \phi_L(x) \qquad \sigma_L(\lambda) = \tau_L(\lambda) + \frac{2}{3\pi}\lambda_+^{3/2}. \tag{6.6}$$

We do not use this in the sequel, and therefore omit the proof.

To prove Proposition 6.1, we shall need the following Erdős-Turán type inequality:

PROPOSITION 6.3 ([8, Prop. 5]). *Let μ be a probability measure on $\mathbb{R}$. Then, for any $s \geq 1$ and any $\alpha \in \mathbb{R}$,*

$$\left|\mu(\alpha) - \mu_{\mathrm{Wigner}}(\alpha)\right| \leq C \left\{ \frac{\rho(\alpha; s)}{s} + \sqrt{\rho(\alpha; s)} \sum_{n=1}^{s} \frac{|\widehat{\mu}(n)|}{n} \right\},$$

where $\rho(\alpha; s) = \max(1 - |\alpha|, s^{-2})$.

Proof of Proposition 6.1. (**a**) According to assumption 1 and Proposition 6.3,

$$|\tau_{R,N}(\lambda)| \leq C \max(\lambda, 1), \tag{6.7}$$

where

$$\tau_{R,N}(\lambda) = \sigma_{R,N}(\lambda) - \frac{2}{\pi} \int_{1-2s_N^{-2}\lambda_+}^{1} \sqrt{1-\alpha^2}\, d\alpha.$$

Therefore the sequence $\{\sigma_{R,N}\}$ is precompact. The same is of course true for $\{\sigma_{L,N}\}$. If $\sigma_{R,N_j} \to \sigma_R$, then $\tau_{R,N_j} \to \tau_R$, defined by

$$\sigma_R(\lambda) = \tau_R(\lambda) + \frac{2}{3\pi}\lambda_+^{3/2}.$$

Therefore we need to prove that σ_R, τ_R are uniquely determined.

(**b**) Let

$$U_n^4(\alpha) = \sum_{k=0}^{4n} c_{n,k}^+ U_k(\alpha),$$
$$U_{n+1}(\alpha) U_n^3(\alpha) = \sum_{k=0}^{4n+1} c_{n,k}^- U_k(\alpha). \tag{6.8}$$

The explicit expressions for $c_{n,k}^{\pm}$ can be easily derived from the identity

$$U_k(\alpha) U_\ell(\alpha) = \sum_{m=0}^{\min(k,\ell)} U_{|\ell-k|+2m}(\alpha).$$

We shall only need the following simple properties:

$$\begin{aligned} c^+_{n,k} &= 2\mathbb{1}_{2\mathbb{Z}}(k)\,(c(k/n)+o(1))\,n^2, \\ c^+_{n,0} &= (c_0+o(1))\,n, \\ c^-_{n,k} &= 2\mathbb{1}_{2\mathbb{Z}+1}(k)\,(c(k/n)+o(1))\,n^2, \end{aligned}$$

where $c \in C[0, 4.01]$ and the $o(1)$ terms are uniform. Thus we have:

$$\begin{aligned} &\frac{s_N^3}{n^4}\int U_n^4(\alpha)d\mu_N(\alpha) \\ &\qquad = \int_0^4 xc(x)\,[\phi_R(xn/s_N)+\phi_L(xn/s_N)]\,dx + c_0 s_N^3/n^3 + o(1), \\ &\frac{s_N^3}{n^4}\int U_{n+1}(\alpha)U_n^3(\alpha)d\mu_N(\alpha) \\ &\qquad = \int_0^4 xc(x)\,[\phi_R(xn/s_N)-\phi_L(xn/s_N)]\,dx + o(1). \end{aligned}$$

Passing to the limit along a subsequence $N_j \to \infty$ and applying (6.7) and the dominated convergence theorem, we deduce

$$\int \frac{\sin^4 x\sqrt{\lambda}}{(x\sqrt{\lambda})^4}d\sigma_R(\lambda) = \int_0^4 yc(y)\phi_R(xy)dy + c_0x^{-3}, \tag{6.9}$$

$$\int \frac{\sin^4 x\sqrt{\lambda}}{(x\sqrt{\lambda})^4}d\sigma_L(\lambda) = \int_0^4 yc(y)\phi_L(xy)dy + c_0x^{-3}. \tag{6.10}$$

(**c**) The relations (6.9) determine σ_R, σ_L uniquely. Indeed, according to (6.7) and (6.4), the measure σ_R must satisfy

$$\int_0^\infty \frac{d\sigma_R(\lambda)}{1+\lambda^2} < +\infty, \qquad \int_{-\infty}^0 \exp\Big[x\sqrt{|\lambda|}\Big]\,d\sigma_R(\lambda) \le C_1\exp\Big[C_1x^{2-\delta}\Big]; \tag{6.11}$$

hence one may apply the argument of Levitan [17]. The latter works as follows. Suppose $\widetilde{\sigma}_R$ is another measure for which (6.9) also holds. Then $\widetilde{\sigma}_R$ also satisfies (6.11); $\nu = \sigma_R - \widetilde{\sigma}_R$ satisfies

$$\int \frac{\sin^4 x\sqrt{\lambda}}{(x\sqrt{\lambda})^4}d\sigma(\lambda) = 0, \quad x \in \mathbb{R}.$$

Set

$$f(x) = \int_{-\infty}^0 \frac{\sin^4 x\sqrt{\lambda}}{\lambda^2}d\nu(\lambda) = -\int_0^{+\infty}\frac{\sin^4 x\sqrt{\lambda}}{\lambda^2}d\nu(\lambda)$$

and apply the Phragmén-Lindelöf principle to $f(z)$ in every quadrant. We see that f is bounded, hence constant. Therefore $\widetilde{\sigma}_R = \sigma_R$.

We have proved that the limiting measure is unique; in particular, the sequence $\{\sigma_R, N\}$ converges (to this limit).

(**d**) Returning to (6.11),

$$\int_{-\infty}^{\infty} \exp\left\{x\sqrt{|\lambda|}\right\} d\sigma_R(\lambda) \le C_1 \exp\left[C_1 x^{2-\delta}\right];$$

hence

$$\sigma_R(\lambda) \le C_1 \exp\left[-x\sqrt{|\lambda|} + C_1 x^{2-\delta}\right], \quad \lambda \le 0, x \ge 0.$$

Taking $x = \left[\frac{\sqrt{|\lambda|}}{2C_1}\right]^{\frac{1}{1-\delta}}$, we obtain the second part of (6.5). The first part follows from (6.7). $\square$

Proposition 6.1 can be extended to measures on $\mathbb{R}^\ell$ (for any *fixed* ℓ). Namely, let $\{\mu_N\}$ be a sequence of measures on $\mathbb{R}^\ell$. For simplicity, we assume that μ_N are symmetric (= invariant under permutation of coordinates). For $k \le \ell$, set

$$\widehat{\mu_N}(n(1), \ldots, n(k)) = \int \prod_{j=1}^{k} U_{n(j)}(\alpha_j) d\mu_N(\alpha).$$

PROPOSITION 6.4. *Assume that, for* $1 \le k \le \ell$,

$$\begin{aligned}
&\widehat{\mu_N}(n(1), \ldots, n(k)) \\
&\quad = \frac{\prod_{j=1}^{k} n(j)}{s_N^{3k}} \sum_{I \subset \{1,\ldots,k\}} (-1)^{\sum_{j\in I} n(j)} \phi_{R,k}\left(\{n(j)\}_{j\notin I}\right) \phi_{L,k}\left(\{n(j)\}_{j\in I}\right) \\
&\qquad\qquad + \frac{\varepsilon_N(n(1), \ldots, n(k))}{s_N^{2k}},
\end{aligned}$$

where

$$\phi_{R,k}, \phi_{L,k} \in C\left((0, +\infty)^k\right) \bigcap L_1(\exp\left[-\|x\|^{2-\delta}\right] dx);$$

the coefficients ε_N *tend to zero uniformly on* $\vec{n} \in \{1, \ldots, Rs_N\}^k$, *and*

$$\sum_{1 \le n(1) \le \cdots \le n(k) \le s_N} \frac{\varepsilon_N(n(1), \ldots, n(k))}{n(1) \cdots n(k)} \le C.$$

Then the scaled measures

$$\text{(6.12)} \quad \begin{cases} \sigma_{R,N}(\lambda_1, \ldots, \lambda_\ell) = s_N^3 \left(1 - \mu_N(1 - 2s_N^{-2}\lambda_1, \ldots, 1 - 2s_N^{-2}\lambda_\ell)\right) \\ \sigma_{L,N}(\lambda_1, \ldots, \lambda_\ell) = s_N^3\, \mu_N(-1 + 2s_N^{-2}\lambda_1, \ldots, -1 + 2s_N^{-2}\lambda_\ell) \end{cases}$$

converge to limiting measures σ_R, σ_L, *which are uniquely determined by* $\{\phi_{k,R}\}$, $\{\phi_{k,L}\}$ (*respectively*). *Moreover*, $\sigma = \sigma_L, \sigma_R$ *satisfy*:

$$\begin{cases} \left|\sigma(\lambda_1, \ldots, \lambda_\ell) - \left(\frac{2}{3\pi}\right)^\ell \prod_{j=1}^{\ell} \lambda_{j+}^{3/2}\right| = O(\|\lambda\|^{\frac{3k-1}{2}}), & \lambda_j \ge 0,\ \|\lambda\| \to \infty; \\ \sigma(\lambda_1, \ldots, \lambda_\ell) = O\left[\exp\left\{-C'_\ell \|\lambda\|^{\frac{1-\delta/2}{1-\delta}}\right\}\right], & \lambda_j \le 0,\ \|\lambda\| \to \infty. \end{cases}$$

The proof is similar to the one-dimensional case (Proposition 6.1); we omit it.

7. **Proof of the main results**

Proof of Theorem 1.1. To show that the random counting measures $\sigma_R(\lambda)$, $\sigma_L(\lambda)$ converge in distribution to $\mathfrak{Ai}_1(-\lambda)$, it is sufficient to prove the convergence of the correlation measures

$$\rho_{\ell,R}(\lambda_1,\dots,\lambda_\ell) = \mathbb{E}\prod_{j=1}^{\ell}\sigma_R(\lambda_j), \quad \rho_{\ell,L}(\lambda_1,\dots,\lambda_\ell) = \mathbb{E}\prod_{j=1}^{\ell}\sigma_L(\lambda_j)$$

to

$$\rho_\ell(\lambda_1,\dots,\lambda_\ell) = \mathbb{E}\prod_{j=1}^{\ell}\mathfrak{Ai}_1(-\lambda_j).$$

It will be convenient to denote

$$\mu_N(\alpha) = \#\left\{\text{eigenvalues of } \frac{H_N}{2\sqrt{2W_N-1}} \text{ in } (-\infty,\alpha]\right\} \tag{7.1}$$

(this differs slightly from (1.5)).

Let us first consider $\ell = 1$. According to Proposition 2.6,

$$\begin{aligned}
\widehat{\mathbb{E}\mu_N}(2n) &= \frac{n}{N}\phi_1(n/N^{1/3}) + \frac{\varepsilon_N^{(1)}(n)}{N^{2/3}}\\
&= \frac{2n}{N}\Big[\widetilde{\phi}_1(2n/N^{1/3}) + (-1)^{2n}\widetilde{\phi}_1(2n/N^{1/3})\Big] + \frac{\varepsilon_N^{(2)}(2n)}{N^{2/3}},\\
\widehat{\mathbb{E}\mu_N}(2n) &= 0\\
&= \frac{2n+1}{N}\Big[\widetilde{\phi}_1((2n+1)/N^{1/3}) + (-1)^{2n+1}\widetilde{\phi}_1((2n+1)/N^{1/3})\Big],
\end{aligned}$$

where $\varepsilon_N^{(1)}$ absorb the difference between the matrices $H_N^{(2n)}/(2W_N-1)^n$ and $U_{2n}(H_N/(2\sqrt{2W_N}))$, and $\varepsilon_N^{(2)}$, $\widetilde{\phi}_1$ are introduced to make the notation compatible with Section 6. The sequence of measures $\{\mu_N\}$ satisfies the assumptions of Proposition 6.1 with $s_N = N^{1/3}$; hence

$$\rho_{1,R}, \rho_{1,L} \to \widetilde{\rho}_1,$$

where $\widetilde{\rho}_1$ is a measure determined by $\widetilde{\phi}_1$ (and in particular independent of W_N). For $W_N = N/2$, $\widetilde{\rho}_1 = \rho_1$ according to the result of Soshnikov [26]; hence the same is true for any $W_N \gg N^{5/6}$.

The same argument works for $\ell > 1$. Indeed,

$$\begin{aligned}
\mathbb{E}\prod_{j=1}^{\ell}\operatorname{tr}\frac{H_N^{n(j)}}{(2W_N-1)^n} &= \sum_{\Pi}\prod_{P\in\Pi} T(\{n(j)\}_{j\in P}\\
&= \sum_{\Pi}\prod_{P\in\Pi}\frac{1+(-1)^{n(j)}}{2} T(\{n(j)\}_{j\in P},
\end{aligned}$$

where the sum is over all partitions Π of $\{1, \dots, \ell\}$. For a subset $I \subset \{1, \dots, \ell\}$, write $\Pi \prec I$ if

$$\forall P \in \Pi \;\; P \cap I \in \{P, \varnothing\}.$$

Then

$$\mathbb{E} \prod_{j=1}^{\ell} \operatorname{tr} \frac{H_N^{n(j)}}{(2W_N - 1)^n} = \sum_{I \subset \{1,\dots,\ell\}} (-1)^{\sum_{j \in I} n(j)} \prod_{\Pi \prec I} \frac{T(\{n(j)\}_{j \in P})}{2}.$$

Now apply Proposition 2.6 and then Proposition 6.4. □

Remark 7.1. Another (perhaps, slightly simpler) way to prove the convergence of the correlation measures is to follow the arguments of [9, §I.5], and then use the uniqueness theorem for Laplace transform instead of the arguments of Section 6, as in [26].

Proof of Theorem 1.2. According to Proposition 2.7 with $k = 1$,

$$\widehat{\mathbb{E}\mu_N}(2n) = \frac{1}{W_N^{6/5}} 2n \cdot 2\psi_1^{(1)}(2n/W_N^{2/5}) + \frac{\varepsilon_N^{(1)}}{W_N^{4/5}},$$

where again $\psi_1^{(1)}, \varepsilon_N^{(1)}$ are introduced to make the notation consistent with Section 6. Now apply Proposition 6.1 with $s_N = W_N^{2/5}$, and deduce that

$$\mathbb{E}\sigma_{R,N}, \; \mathbb{E}\sigma_{L,N} \longrightarrow \sigma_1,$$

where

$$\sigma_1(\lambda) = \tau_1(\lambda) + \frac{2}{3\pi}\lambda_+^{3/2}, \qquad \int_{-\infty}^{+\infty} \frac{\sin x\sqrt{\lambda}}{x\sqrt{\lambda}}\, d\tau_1(\lambda) = \psi(x).$$

Applying Proposition 2.7 with $k = 2$ and Proposition 6.4, it is not hard to see that

$$\mathbb{E}\sigma_{R,N} \otimes \sigma_{R,N} - (\mathbb{E}\sigma_{R,N}) \otimes (\mathbb{E}\sigma_{R,N}) \longrightarrow 0,$$
$$\mathbb{E}\sigma_{L,N} \otimes \sigma_{L,N} - (\mathbb{E}\sigma_{L,N}) \otimes (\mathbb{E}\sigma_{L,N}) \longrightarrow 0;$$

hence also

$$\sigma_{R,N}, \sigma_{L,N} \xrightarrow{D} \sigma_1. \qquad \square$$

Remark 7.2. Staring at the asymptotics of ψ near zero, it seems natural to conjecture that

$$\sigma_1(\lambda) = \frac{2}{3\pi}\lambda_+^{3/2} + \sqrt{\frac{3}{32\pi^2}}\,\lambda_+^{1/4} + O(1), \quad \lambda \to +\infty. \tag{7.2}$$

We have not been able to prove this as stated. Applying Marchenko's Tauberian theorem [19], one can show that (7.2) holds in a weak sense (say, after integrating both sides with a compactly supported twice differentiable kernel).

Proof of Theorem 1.3. First, for $n \le W_N$, Proposition 4.3 yields

$$\frac{\mathbb{E}\,\mathrm{tr}\, H_N^{(2n)}}{(2W_N-1)^N} = \sum_{\mathcal{D}} \frac{\mathrm{Paths}(2n;\mathcal{D})}{(2W_N-1)^n}$$
$$= \sum_{\mathcal{D}} \left[\frac{(Cn)^{3s-2}}{(3s-2)!}(cN)^{-s+1} + N\,\frac{(Cn)^{\frac{5s-4}{2}}}{(\frac{5s-4}{2})!}(cW_N)^{-s} \right].$$

Rearranging the sum and using Lemma 2.5, we continue:

$$\frac{\mathbb{E}\,\mathrm{tr}\, H_N^{(2n)}}{(2W_N-1)^N} \le \sum_{s} \left[\frac{(Cn)^{3s-2}}{(3s-2)!}(cN)^{-s+1} + N\,\frac{(Cn)^{\frac{5s-4}{2}}}{(\frac{5s-4}{2})!}(cW_N)^{-s} \right] (Cs)^s$$
$$\le Cn \left\{ \exp\left[\frac{Cn^{3/2}}{N^{1/2}}\right] + \frac{N}{W_N^{6/5}} \exp\left[\frac{Cn^{5/3}}{W_N^{2/3}}\right] \right\};$$

hence also

$$\mathbb{E}\,\mathrm{tr}\, U_{2n}\left(\frac{H_N}{2\sqrt{2W_n}}\right) \le Cn \left\{ \exp\left[\frac{Cn^{3/2}}{N^{1/2}}\right] + \frac{N}{W_N^{6/5}} \exp\left[\frac{Cn^{5/3}}{W_N^{2/3}}\right] \right\}$$

(perhaps, with a different constant C). Applying the identities (6.8), we have

$$\mathbb{E}\,\mathrm{tr}\, U_{2n}^4\left(\frac{H_N}{2\sqrt{2W_n}}\right) \le C \left\{ nN + n^4 \exp\left[\frac{Cn^{3/2}}{N^{1/2}}\right] + \frac{Nn^4}{W_N^{6/5}} \exp\left[\frac{Cn^{5/3}}{W_N^{2/3}}\right] \right\}.$$

If $N^{5/6} \le W_N$, then the right-hand side is bounded by Cn^4 for $n = \lfloor N^{1/3} \rfloor$. As

$$\frac{U_{2n}(\pm(1+\varepsilon))}{2n} \ge c\exp(cn\sqrt{\varepsilon}),$$

we deduce that $\|H_N/(2\sqrt{2W_N})\| \xrightarrow{D} 1$ (and in fact,

$$\left\{ \left(\|H_N/(2\sqrt{2W_N}) - 1 \right) N^{2/3} \right\}_N$$

is stochastically bounded). If $W_N \le N^{5/6}$, then take $n = \lfloor W_N^{3/5} \log^{2/5} N \rfloor$. Then

$$\mathbb{E}\,\mathrm{tr}\, U_{2n}^4\left(\frac{H_N}{2\sqrt{2W_n}}\right) \le Cn^4 N^{C'}/W_N^{6/5}$$

and hence again $\|H_N/(2\sqrt{2W_N})\| \xrightarrow{D} 1$. □

8. **Random phases**

The steps of the proof for the matrices with entries (1.3) are very similar to those for (1.2). We indicate the necessary modifications.

- Section 2: Lemma 2.5 remains valid *verbatim*. Condition (d) in Corollary 2.2 should be replaced with

$$\#\left\{(i,j) \mid u_i^j = u,\, u_{i+1}^j = v\right\} = \#\left\{(i,j) \mid u_i^j = v,\, u_{i+1}^j = u\right\}.$$

In Definition 2.3, loops are not allowed, and the third condition should be also replaced with

$$\#\left\{(i,j) \mid \bar{u}_i^j = \bar{u},\, \bar{u}_{i+1}^j = \bar{v}\right\} = \#\left\{(i,j) \mid \bar{u}_i^j = \bar{v},\, \bar{u}_{i+1}^j = \bar{u}\right\} = 1.$$

Thus, the diagrams in the new sense are a subset of diagrams in the old sense. In Lemma 2.5, s is now always even, and the estimate is valid for even s. In Propositions 2.6,2.7 the sums are now over even s (and the functions ϕ, ψ are different than before).

- In Section 5, one should only consider the diagrams that are valid according to the new definition, and only even values of s.

- The argument in Section 7 is still valid. In Remark 7.2, we would now conjecture that

$$\sigma_2(\lambda) = \frac{2}{3\pi}\lambda_+^{3/2} + O(1), \quad \lambda \to +\infty. \tag{8.1}$$

9. **Concluding remarks**

I. To simplify the exposition, we have only considered the simplest random variables (1.2) and (1.3). Assume that the entries of H_N above the diagonal are independent, and have symmetric distribution with (uniformly) subgaussian tails. Applying the methods of [9, Part III], one can assume that

$$\mathbb{E}H_{uv}^2 = 1, \quad H_{uv} \in \mathbb{R} \text{ a.s.} \quad (0 < |u-v|_N \le W)$$

or

$$\mathbb{E}H_{uv}^2 = 0, \quad \mathbb{E}|H_{uv}|^2 = 1 \quad (0 < |u-v|_N \le W)$$

instead of (1.2) or (1.3) (respectively), and prove analogues of Theorems 1.1 and 1.2. In particular, this extension covers the frequently considered case of matrices with Gaussian elements.

II. The restriction $W_N \gg 1$ is also an artifact of the proof. For $W_N = O(1)$, Lemma 4.4 is no longer applicable, hence one should take into account the difference between nonbacktracking and usual random walk. Thus, we need an analogue of Propositions 3.1–3.3 for nonbacktracking walks. This can probably be proved using either a trace formula for the representation of nonbacktracking random walk as a Markov chain on the space of directed edges (see Smilansky [24]), or the connection to Chebyshev polynomials (see [1]).

III. It would be interesting to obtain a more detailed description of the measure σ_β from Theorem 1.2. In particular, a more precise description of the left tail would allow to find the limiting distribution of the maximal eigenvalue (cf. the remark after Theorem 1.3); as to the right tail, it would be interesting to justify the asymptotics (7.2) (and perhaps derive the next terms in the asymptotic series).

IV. Usual (i.e. nonperiodic) random band matrices have nonzero elements H_{uv} for

$$0 < |u - v| \le W_N .$$

We expect that the results of this paper hold, perhaps in modified form, for these matrices as well. Following Bogachev, Molchanov, and Pastur [4], we note however that even the limiting spectral measure coincides (1.6) only if $1 \ll W_N \ll N$ or $W_N = (1 - o(1))N$. The limiting spectral measure in the complementary regimes has been described by Khorunzhiy, Molchanov, and Pastur in [20].

V. It would also be interesting to study the crossover regime $W_N \asymp N^{5/6}$. We refer the reader to the works of Johansson [14] and Bender [3] for the description of the crossover regime at the spectral edge for different kinds of random matrices.

VI. The method of this paper can be used to study the eigenvectors of

$$H_N / (2\sqrt{2W_N})$$

that correspond to eigenvalues close to the edge, and, in particular, their *inverse participation ratio*

$$\sum_{u=1}^{N} |\mathrm{v}(u)|^4 \Big/ \left(\sum_{u=1}^{N} |\mathrm{v}(u)|^2 \right)^2 .$$

If $W_N \gg N^{5/6}$, then the inverse participation ratio of eigenvectors corresponding to eigenvalues $\alpha = 1 + O(N^{-2/3})$ is, with high probability, of order N^{-1}. If $W_N \ll N^{5/6}$, then the inverse participation ratio averaged over eigenvalues in a window $[1 + a/W_N^{4/5}, 1 + b/W_N^{4/5}]$ is, with high probability, of order $W_N^{-6/5}$.

VII. We remark that Schenker [22] proved a lower bound $\frac{1}{CW^8}$ on the inverse participation ratio of the eigenvalues in the bulk of the spectrum (for a slightly different class of band matrices). In the opposite direction, Erdős and Knowles [6], [7] recently proved an upper bound $W^{-1/3+o(1)}$ for a wide class of band matrices; their argument uses in particular the expansion in Chebyshev polynomials developed in the current paper.

VIII. Finally, there is a natural extension of band matrices to higher-dimensional lattices: the rows and columns of H_N are indexed by elements of $(\mathbb{Z}/N\mathbb{Z})^d$, and $H_N(u, v) = 0$ unless $0 < \|u - v\| \le W_N$. Similar random matrices have been also studied in physical and mathematical literature; cf., Silvestrov [23], Disertori, Pinson, and Spencer [5]. We hope to consider the spectral edges of such matrices in a forthcoming work.

Acknowledgment. I am grateful to my supervisor, Vitali Milman, for his encouragement and support.

Thomas Spencer has shared with me his interest in band matrices, and encouraged me to apply the method of [9] to study their spectral edges. His comments on a preliminary version of this paper have been of great help. Yan Fyodorov has explained to me what is the Thouless criterion, and why its predictions are coherent with the results of the current paper. Bo'az Klartag has suggested to use the Cauchy-Binet formula to simplify the expressions in Section 4. My father has referred me to the works [17], [18], [19] and [32]. Mark Rudelson and Valentin Vengerovsky have helped fix the ambiguities in the definition of diagram; see Remark 2.4. The discussions with Alexei Khorunzhiy, Leonid Pastur, Mariya Shcherbina, and Uzy Smilansky on related topics have been of great benefit to me.

I thank them all very much.

References

[1] N. ALON, I. BENJAMINI, E. LUBETZKY, and S. SODIN, Non-backtracking random walks mix faster, *Commun. Contemp. Math.* **9** (2007), 585–603. MR 2008k:60017 Zbl 1140.60301

[2] Z. D. BAI and Y. Q. YIN, Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix, *Ann. Probab.* **21** (1993), 1275–1294. MR 94j:60060 Zbl 0779.60026

[3] M. BENDER, Edge scaling limits for a family of non-Hermitian random matrix ensembles, *Probab. Theory Related Fields* **147** (2010), 241–271. MR 2594353 Zbl 1188.60003

[4] L. V. BOGACHEV, S. A. MOLCHANOV, and L. A. PASTUR, On the density of states of random band matrices, *Mat. Zametki* **50** (1991), 31–42, 157. MR 93b:82032 Zbl 0762.60015

[5] M. DISERTORI, H. PINSON, and T. SPENCER, Density of states for random band matrices, *Comm. Math. Phys.* **232** (2002), 83–124. MR 2004c:82052 Zbl 1019.15014

[6] L. ERDŐS and A. KNOWLES, Diffusion and eigenfunction delocalization in a random band matrix model. arXiv 1002.1695

[7] ———, Quantum diffusion and delocalization for band matrices with general distribution. arXiv 1005.1838

[8] O. N. FELDHEIM and S. SODIN, One more proof of the Erdős-Turán inequality, and an error estimate in Wigner's law. arXiv 0901.1620

[9] ———, A universality result for the smallest eigenvalues of certain sample covariance matrices, *Geom. Funct. Anal.*, to appear. arXiv 0812.1961

[10] D. FÉRAL and S. PÉCHÉ, The largest eigenvalues of sample covariance matrices for a spiked population: diagonal case, *J. Math. Phys.* **50** (2009), 073302, 33. MR 2010j:62167

[11] Y. V. FYODOROV and A. D. MIRLIN, Statistical properties of eigenfunctions of random quasi 1D one-particle Hamiltonians, *Internat. J. Math. Phys. B* **8** (1994), 3795–3842.

[12] B. GNEDENKO, Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. of Math.* **44** (1943), 423–453. MR 5,41b Zbl 0063.01643

[13] I. A. IBRAGIMOV and Y. V. LINNIK, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff Publishing, Groningen, 1971, With a supplementary chapter by I. A. Ibragimov and V. V. Petrov. MR 48 #1287 Zbl 0219.60027

[14] K. JOHANSSON, From Gumbel to Tracy-Widom, *Probab. Theory Related Fields* **138** (2007), 75–112. MR 2008h:60203 Zbl 1116.60020

[15] O. KHORUNZHIY, Estimates for moments of random matrices with Gaussian elements, in *Séminaire de Probabilités* XLI, *Lecture Notes in Math.* **1934**, Springer-Verlag, New York, 2008, pp. 51–92. MR 2010e:60013 Zbl 1149.15302

[16] A. KHORUNZHY and W. KIRSCH, On asymptotic expansions and scales of spectral universality in band random matrix ensembles, *Comm. Math. Phys.* **231** (2002), 223–255. MR 2004g:82049 Zbl 1034.82022

[17] B. M. LEVITAN, On a uniqueness theorem, *Doklady Akad. Nauk SSSR* **76** (1951), 485–488. MR 12,605i

[18] B. M. LEVITAN and N. N. MEĬMAN, On a uniqueness theorem, *Doklady Akad. Nauk SSSR* **81** (1951), 729–731. MR 13,551d

[19] V. A. MARČENKO, Theorems of Tauberian type in spectral analysis of differential operators, *Izv. Akad. Nauk SSSR. Ser. Mat.* **19** (1955), 381–422. MR 17,852f

[20] S. A. MOLCHANOV, L. A. PASTUR, and A. M. KHORUNZHIĬ, Distribution of the eigenvalues of random band matrices in the limit of their infinite order, *Teoret. Mat. Fiz.* **90** (1992), 163–178. MR 93g:82056

[21] S. PÉCHÉ, Universality results for the largest eigenvalues of some sample covariance matrix ensembles, *Probab. Theory Related Fields* **143** (2009), 481–516. MR 2009m:60013 Zbl 1167.62019

[22] J. SCHENKER, Eigenvector localization for random band matrices with power law band width, *Comm. Math. Phys.* **290** (2009), 1065–1097. MR 2010i:60024 Zbl 1179.82079

[23] P. G. SILVESTROV, Summing graphs for random band matrices, *Phys. Rev. E* **55** (1997), 6419–6432.

[24] U. SMILANSKY, Quantum chaos on discrete graphs, *J. Phys. A* **40** (2007), F621–F630. MR 2009d:81119 Zbl 1124.81024

[25] S. SODIN, Random matrices, nonbacktracking walks, and orthogonal polynomials, *J. Math. Phys.* **48** (2007), 123503, 21. MR 2008m:60082 Zbl 1153.81436

[26] A. SOSHNIKOV, Universality at the edge of the spectrum in Wigner random matrices, *Comm. Math. Phys.* **207** (1999), 697–733. MR 2001i:82037 Zbl 1062.82502

[27] ———, A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices, *J. Statist. Phys.* **108** (2002), 1033–1056. MR 2003h:62108 Zbl 1018.62042

[28] T. SPENCER, Random banded and sparse matrices, in *Oxford Handbook of Random Matrix Theory*, to appear.

[29] D. J. THOULESS, Maximum metallic resistance in thin wires, *Phys. Rev. Lett.* **39** (1977), 1167–1169.

[30] C. A. TRACY and H. WIDOM, Level-spacing distributions and the Airy kernel, *Comm. Math. Phys.* **159** (1994), 151–174. MR 95e:82003 Zbl 0789.35152

[31] ———, On orthogonal and symplectic matrix ensembles, *Comm. Math. Phys.* **177** (1996), 727–754. MR 97a:82055 Zbl 0851.60101

[32] E. B. VUL, Uniqueness theorems for a certain class of functions represented by integrals, *Dokl. Akad. Nauk SSSR* **129** (1959), 722–725. MR 22 #8093

[33] W. M. WANG, On localization and density of states for the random Schrödinger operator, Ph.D. thesis, Princeton University, 1992.

[34] E. P. WIGNER, On the distribution of the roots of certain symmetric matrices, *Ann. of Math.* **67** (1958), 325–327. MR 20 #2029 Zbl 0085.13203

(Received August 27, 2009)

E-mail address: sodinale@post.tau.ac.il
SCHOOL OF MATHEMATICAL SCIENCES, TEL AVIV UNIVERSITY, RAMAV AVIV, TEL AVIV 69978, ISRAEL

ISSN 0003-486X

ANNALS OF MATHEMATICS

This periodical is published bimonthly by the Department of Mathematics at Princeton University with the cooperation of the Institute for Advanced Study. Annals is typeset in TEX by Sarah R. Warren and produced by Mathematical Sciences Publishers. The six numbers each year are divided into two volumes of three numbers each.

Editorial correspondence

Papers submitted for publication and editorial correspondence should be addressed to Maureen Schupsky, Annals of Mathematics, Fine Hall-Washington Road, Princeton University, Princeton, NJ, 08544-1000 U.S.A. The e-mail address is annals@math.princeton.edu.

Preparing and submitting papers

The Annals requests that all papers include an abstract of about 150 words which explains to the nonspecialist mathematician what the paper is about. It should not make any reference to the bibliography. Authors are encouraged to initially submit their papers electronically and in PDF format. Please send the file to: annals@math.princeton.edu or to the Mathematics e-print arXiv: front.math.ucdavis.edu/submissions. If a paper is submitted through the arXiv, then please e-mail us with the arXiv number of the paper.

Proofs

A PDF file of the galley proof will be sent to the corresponding author for correction. If requested, a paper copy will also be sent to the author.

Offprints

Authors of single-authored papers will receive 30 offprints. (Authors of papers with one co-author will receive 15 offprints, and authors of papers with two or more co-authors will receive 10 offprints.) Extra offprints may be purchased through the editorial office.

Subscriptions

The price for a print and online subscription, or an online-only subscription, is \$390 per year for institutions. In addition, there is a postage surcharge of \$40 for print subscriptions that are mailed to countries outside of the United States. Individuals interested in subscriptions for their own personal use should contact the publisher at the address below. Subscriptions and changes of address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 (e-mail: contact@mathscipub.org; phone: 1-510-643-8638; fax: 1-510-295-2608). (Checks should be made payable to “Mathematical Sciences Publishers”.)

Back issues and reprints

Orders for missing issues and back issues should be sent to Mathematical Sciences Publishers at the above address. Claims for missing issues must be made within 12 months of the publication date. Online versions of papers published five or more years ago are available through JSTOR (www.jstor.org).

Microfilm

Beginning with Volume 1, microfilm may be purchased from NA Publishing, Inc., 4750 Venture Drive, Suite 400, PO Box 998, Ann Arbor, MI 48106-0998; phone: 1-800-420-6272 or 734-302-6500; email: info@napubco.com, website: www.napubco.com/contact.html.

ALL RIGHTS RESERVED UNDER THE BERNE CONVENTION AND
THE UNIVERSAL COPYRIGHT CONVENTION

Copyright © 2010 by Princeton University (Mathematics Department)

Printed in U.S.A. by Sheridan Printing Company, Inc., Alpha, NJ

TABLE OF CONTENTS