A CLT FOR A BAND MATRIX MODEL

GREG ANDERSON AND OFER ZEITOUNI

ABSTRACT. A law of large numbers and a central limit theorem are derived for linear statistics of random symmetric matrices whose on-or-above diagonal entries are independent, but neither necessarily identically distributed, nor necessarily all of the same variance. The derivation is based on systematic combinatorial enumeration, study of generating functions, and concentration inequalities of the Poincaré type. Special cases treated, with an explicit evaluation of limiting variances, are generalized Wigner and Wishart matrices.

1. INTRODUCTION

The interest in the limiting properties of the empirical distribution of eigenvalues of large symmetric random matrices can be traced back to [Wis28] and to the pathbreaking article of Wigner [Wig55]. We refer to [Ba99], [De00], [HP00], [Me91] and [PL03] for partial overview and some of the recent spectacular progress in this field.

In this paper we study both convergence of the empirical distribution and central limit theorems for linear statistics of the empirical distribution of a class of random matrices. To give right away the flavor of our results, consider for each positive integer N the N-by-N symmetric random matrix $X(N)$ with on-or-above-diagonal entries $X(N)_{ij} = N^{-1/2} f(i/N, j/N)^{1/2} \xi_{ij}$, where the ξ_{ij} are zero mean unit variance i.i.d. random variables satisfying the Poincaré inequality with constant c (see §11.7 for the definition), and $f(\cdot,\cdot)$ is a nonnegative function symmetric and continuous on $[0,1]^2$ such that $\int_0^1 f(x,y)dy \equiv 1$. Define the *semicircle distribution* σ_S of zero mean and unit variance to be the measure on $\mathbb{R}$ of compact support with density $\frac{d\sigma_S}{dx} := \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{\{|x|\le 2\}}$. Let $\lambda_1(N) \le \cdots \le \lambda_N(N)$ be the eigenvalues of $X(N)$. Under these assumptions, a corollary of our general results (see Theorem 3.5 below) states that the *empirical distribution* $L(N) := N^{-1}\sum_{i=1}^N \delta_{\lambda_i(N)}$ converges weakly, in probability, to σ_S, and further, for any continuously differentiable function f on $\mathbb{R}$ of polynomial growth, with $\|f'\|_{L^2(\sigma_S)} > 0$, the sequence of random variables

$$\sum_{i=1}^N f(\lambda_i(N)) - E\left[\sum_{i=1}^N f(\lambda_i(N))\right] \tag{1}$$

converges in distribution to a nondegenerate zero mean Gaussian random variable with variance given by an explicit formula. Similar explicit results hold for a class of generalized Wishart matrices, see Theorem 12.7. For polynomial test functions f, such results hold in much greater generality, see Theorem 3.3.

Our approach has two main components. The first, of some interest on its own, is a combinatorial enumeration scheme for the different types of terms that contribute

Date: June 15, 2004. Revised December 1, 2004.
O.Z. was partially supported by NSF grant number DMS-0302230.

to the expectation of products of traces of powers of the matrices under study. This scheme takes the bulk of the paper to develop. The other component, which allows us to move from polynomial test functions to continuously differentiable ones, is based on concentration inequalities of the Poincaré type. The latter component is developed in §11, building on earlier results of concentration for random matrices that can be found in [CB04] and [GZ00].

CLT results related to Theorems 3.3 and 3.5 below have already been stated in the literature. An especially strong inspiration to our study is the work of Jonsson [Jo82], who gives CLT statements for traces of polynomial functions of Gaussian Wishart matrices, based on the method of moments introduced by Wigner in [Wig55]. The method of moments was revisited in the far-reaching work of Sinai and Soshnikov [SS98], where, as an easy by-product of their results, they state a CLT for traces of analytic functions of Wigner-type matrices. Pastur and co-authors, on the one hand, and Bai and co-authors, on the other, have championed an approach based on the evaluation of resolvents. The latter approach has the advantage of allowing one to relax hypotheses on matrix entries; in particular one does not need to have all moments finite. CLT statements based on these techniques and expressions for the resulting variance, for functions of the form $f(x) = \sum a_i/(z_i - x)$ where $z_i \in \mathbb{C} \setminus \mathbb{R}$, and matrices of Wigner type, can be found in [KKP96], with somewhat sketchy proofs. Earlier statements can be found in [Gi90]. A complete treatment for f analytic in a domain including the support of the limit of the empirical distribution of eigenvalues is given in [BY03] for matrices of Wigner type, and in [BS04] for matrices of Wishart type under a certain restriction on fourth moments. Much more is known for restricted classes of matrices: Johansson [Joh98], using an approach based on the explicit joint density of the eigenvalues available in the independent case only in the Gaussian Wigner situation, characterizes completely those functions f for which a CLT holds. Cabanal-Duvillard [CD01] introduces a stochastic calculus approach and proves a CLT for traces of polynomials of Gaussian Wigner and Wishart matrices, as well as for traces of non-commutative polynomials of pairs of independent Gaussian Wigner matrices. Recent extensions and reinterpretation of his work, using the notion of second order freeness, can be found in [MS04]. Still in the Gaussian case, Guionnet [Gu02], using a stochastic calculus approach, gives a CLT (with a somewhat implicit variance computation) for a class of functions f in the case of *band matrices*. Earlier, laws of large numbers for band matrices were derived, see e.g. [MPK92], [Sh96] and the references therein. In comparison with the references mentioned above, our work can be seen as relaxing the structural assumptions on the variance of the entries of the matrix $X(N)$, as well as the Gaussian assumption, while still requiring rather strong moment bounds on the individual entries (if one is interested only in polynomial test functions) or Poincaré type conditions on the entries (if one wants a wider class of test functions).

The structure of the article is as follows. In §2, we introduce the matrix model considered throughout the paper and set basic notations. Our main results for polynomial test functions f are stated in §3. §4 develops the language we use in the combinatorial enumeration mentioned above. §5 is devoted to some preliminary limit calculations which are then immediately applied in §6 to prove our main result concerning limiting spectral measures, Theorem 3.2. §7 is devoted to the derivation of some *a priori* estimates, following [FK81], useful in the study of the support of the empirical distribution $L(\mathcal{N})$. §8 is the heart of our enumeration scheme, and the results are immediately applied in §9 to yield the proof of our main CLT statement, Theorem 3.3. §10 is devoted to the proof of Theorem 3.4, which is a technical result describing how to approximate $E \operatorname{tr} X(\mathcal{N})^n$ at CLT scale; this part of the paper may be skipped without much loss of comprehension of the remainder of the paper. §11 is devoted to concentration of measure results based on the Poincaré inequality. Finally, in §12 we specialize our main results to generalized Wigner and Wishart matrices, and derive explicit representations for the resulting variances.

2. The model

We define in this section the class of random matrices we are going to deal with. Matrices of this class are symmetric, with on-or-above-diagonal entries independent, with all entries possessing moments of all orders, and with off-diagonal entries of mean zero; further and crucially, subject to the constraints of symmetry and vanishing of off-diagonal means, the moments of entries of such matrices are allowed to depend upon position. Now as it turns out, only certain statistical properties of the patterns of first, second and fourth moments of entries figure in our limit formulas. Accordingly, our description of the class is contrived so as to emphasize those statistical properties and to suppress unneeded detail concerning the exact dependence of moments of entries on position. The notion crucial for gaining "statistical control" is that of *color*. The reader interested only in Wigner matrices should take as space of colors a space consisting of a single color.

2.1. **The band matrix model.**

2.1.1. *Colors.* We fix a Polish space, elements of which we call *colors*. We declare the Borel sets of color space to be measurable. We fix a probability measure θ on color space. We fix a bounded measurable real-valued function D on color space. For each positive integer k we fix a bounded measurable nonnegative function $d^{(k)}$ on color space and a symmetric bounded measurable nonnegative function $s^{(k)}$ on the product of two copies of color space. We make the following assumptions:

- $d^{(k)}$ is constant for $k \neq 2$.
- $s^{(k)}$ is constant for $k \notin \{2, 4\}$.
- $s^{(k)}$ has discontinuity set of measure zero with respect to $\theta \otimes \theta$.
- D, $d^{(k)}$ and the diagonal restriction of $s^{(k)}$ have discontinuity sets of measure zero with respect to θ.

For any bounded function f on color space, or on a product of copies of color space, we write $|f|_\infty$ for its supremum norm.

2.1.2. *Letters.* We fix a countably infinite set, elements of which we call *letters*. We fix a function κ_0 from letter space to color space, and we say that $\kappa_0(\alpha)$ is the *color* of the letter α. Given any nonempty finite set $\mathcal{N}$ of letters of cardinality N put

$$\theta_{\mathcal{N}} = N^{-1} \sum_{\alpha \in \mathcal{N}} \delta_{\kappa_0(\alpha)},$$

which is the color distribution of letters belonging to $\mathcal{N}$. We reserve the script letter $\mathcal{N}$ for use in this context and invariably denote the cardinality of $\mathcal{N}$ by the roman letter N. Analogously, given a sequence $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3, \dots$ of finite nonempty sets of letters, $N_1, N_2, N_3, \dots$ denotes the corresponding sequence of cardinalities.

2.1.3. *The family $\{\xi_e\}$ of random variables.* We fix a family $\{\xi_e\}$ of independent real-valued mean zero random variables indexed by unordered pairs e of letters. We assume that for all letters α, β and positive integers k we have

$$E|\xi_{\{\alpha,\beta\}}|^k \leq \begin{cases} s^{(k)}(\kappa_0(\alpha), \kappa_0(\beta)) & \text{if } \alpha \neq \beta, \\ d^{(k)}(\kappa_0(\alpha)) & \text{if } \alpha = \beta, \end{cases} \tag{2}$$

and moreover we assume that equality holds above whenever one of the following conditions holds:

- $k = 2$.
- $\alpha \neq \beta$ and $k = 4$.

In other words, the rule is to enforce equality whenever the not-necessarily-constant functions $d^{(2)}$, $s^{(2)}$ or $s^{(4)}$ are involved, but otherwise merely to impose a bound.

2.1.4. *Random matrices.* Given any nonempty finite set $\mathcal{N}$ of letters, let $X(\mathcal{N})$ be the $N \times N$ real symmetric random matrix with entries

$$X(\mathcal{N})_{\alpha\beta} = D(\kappa_0(\alpha))\delta_{\alpha\beta} + N^{-1/2}\xi_{\{\alpha,\beta\}} \qquad (\alpha, \beta \in \mathcal{N}),$$

denote the eigenvalues of $X(\mathcal{N})$ by $\lambda_1(\mathcal{N}) \leq \cdots \leq \lambda_N(\mathcal{N})$, and let

$$L(\mathcal{N}) = N^{-1} \sum_{i=1}^{N} \delta_{\lambda_i(\mathcal{N})}$$

be the empirical distribution of the spectrum of $X(\mathcal{N})$. Put

$$\overline{L}(\mathcal{N}) = EL(\mathcal{N}).$$

Note that

$$\langle \overline{L}(\mathcal{N}), x^n \rangle = \frac{1}{N} E \operatorname{tr} X(\mathcal{N})^n,$$

where here and often below we employ the abbreviated notation

$$\langle \mu, f \rangle = \int f(x)\mu(dx)$$

for integrals.

2.2. **Generating functions.**

2.2.1. Let σ be any probability measure on color space. Let

$$[\Phi_{n,\sigma}(c)]_{n=1}^{\infty}$$

be the unique sequence of real-valued bounded measurable functions on color space characterized by the generating function identity

$$\Phi_\sigma(c,t) = \left(\frac{t}{1 - D(c)t}\right) \left(1 - \frac{t}{1 - D(c)t} \int s^{(2)}(c, c')\Phi_\sigma(c', t)\sigma(dc')\right)^{-1} \tag{3}$$

where

$$\Phi_\sigma(c,t) = \sum_{n=1}^{\infty} \Phi_{n,\sigma}(c)t^n$$

is the corresponding generating function. We emphasize that we view the power series here formally, i. e., as devices for managing sequences, not as analytic functions. We write (3) as a shorthand for the recursion obtained by formally expanding both sides of (3) in powers of t, and then equating coefficients of like powers of t. When $\sigma = \theta$, we omit it from the notation.

2.2.2. For each positive integer r we define a function

$$K_r(c_1,\dots,c_r)=\begin{cases} s^{(2)}(c_1,c_1) & \text{if } r=1,\\ s^{(2)}(c_1,c_2)^2 & \text{if } r=2,\\ s^{(2)}(c_1,c_2)s^{(2)}(c_2,c_3)\cdots s^{(2)}(c_r,c_1) & \text{if } r\geq 3,\end{cases}$$

on the product of r copies of color space. We define

$$\Theta(x,y)=\sum_{r=1}^{\infty}\frac{1}{r}\int\cdots\int K_r(c_1,\cdots,c_r)\prod_{i=1}^{r}\left(\Phi(c_i,x)\Phi(c_i,y)\theta(dc_i)\right). \tag{4}$$

We view $\Theta(x,y)$ as a formal power series in x and y with real coefficients; in keeping with this point of view, the integrals on the right side of (4) are to be evaluated by first expanding the integrands in powers of x and y and then integrating term by term (and hence, all integrals, being expectations of bounded measurable functions, are well defined).

2.2.3. Put

$$\begin{aligned}\Psi(x,y) &= \int(d^{(2)}(c)-2s^{(2)}(c,c))\Phi(c,x)\Phi(c,y)\theta(dc)\\ &\quad+\frac{1}{2}\int\int(s^{(4)}(c_1,c_2)-3s^{(2)}(c_1,c_2)^2)\\ &\quad\times\Phi(c_1,x)\Phi(c_2,x)\Phi(c_1,y)\Phi(c_2,y)\theta(dc_1)\theta(dc_2).\end{aligned} \tag{5}$$

We view $\Psi(x,y)$ as a formal power series in x and y with real coefficients. As above, the integrals are to be evaluated by first expanding integrands in powers of x and y and then integrating term by term.

2.2.4. In order to gain convenient access to the information coded in the formal power series $\Theta(x,y)$ and $\Psi(x,y)$ we introduce the following (abuse of) notation. We write

$$\left\langle\sum_{i=0}^{\infty}a_it^i,\sum_{i=0}^{\infty}b_jt^i\right\rangle=\sum_{i=0}^{\infty}a_ib_i$$

for any sequences $[a_i]_{i=0}^{\infty}$ and $[b_j]_{j=0}^{\infty}$ of real numbers such that the sum on the right has only finitely many nonzero terms. Similarly we write

$$\left\langle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}a_{ij}x^iy^j,\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}b_{ij}x^iy^j\right\rangle=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}a_{ij}b_{ij}$$

for any doubly infinite sequences $[a_{ij}]_{i,j=0}^{\infty}$ and $[b_{ij}]_{i,j=0}^{\infty}$ of real numbers such that the sum on the right has only finitely many nonzero terms.

The following fact, proved in §6, explains the role of the sequence $\Phi_{n,\sigma}$:

Lemma 2.3. *If $\sigma=\theta$ or $\sigma=\theta_{\mathcal{N}}$ for some finite nonempty set of letters $\mathcal{N}$, then there exists a unique probability measure μ_σ on the real line such that*

$$\langle\mu_\sigma,x^n\rangle=\langle\sigma,\Phi_{n+1,\sigma}\rangle\quad(n=0,1,2,\dots), \tag{6}$$

$$\operatorname{supp}\mu_\sigma\subset[-C,C]\quad(C=2(|D|_\infty+|s^{(2)}|_\infty^{1/2})). \tag{7}$$

In what follows, we write $\mu=\mu_\theta$ and $\mu_{\mathcal{N}}=\mu_{\theta_{\mathcal{N}}}$.

3. Assumptions and main theorem

Throughout, we let $\Rightarrow$ denote the weak convergence of probability measures. All of our results will be obtained under the following basic

Assumption 3.1. *All the assumptions and notations in §2 hold. Further, there exists a sequence $[\mathcal{N}_k]_{k=1}^{\infty}$ of finite nonempty sets of letters such that $N_k \to \infty$ and $\theta_{\mathcal{N}_k} \Rightarrow \theta$.*

Our main results are:

Theorem 3.2. *Let Assumption 3.1 hold. Then: (i) $L(\mathcal{N}_k) \Rightarrow \mu$ in probability. (ii) $\mu_{\mathcal{N}_k} \Rightarrow \mu$.*

Theorem 3.3. *Let Assumption 3.1 hold. Fix a real-valued polynomial function $f(\cdot)$ on the real line. Then the sequence of random variables*

$$Z_{f,k} := \operatorname{tr} f(X(\mathcal{N}_k)) - E \operatorname{tr} f(X(\mathcal{N}_k))$$

converges in distribution to a zero mean Gaussian random variable Z_f of variance

$$EZ_f^2 = \langle 2\Theta(x,y) + \Psi(x,y), xf'(x)yf'(y)\rangle. \tag{8}$$

Theorem 3.4. *In the setting of the preceding theorem, we also have*

$$\lim_{k\to\infty} E \operatorname{tr} f(X(\mathcal{N}_k)) - N_k \cdot \langle \mu_{\mathcal{N}_k}, f\rangle = \frac{1}{2}\langle \Theta(t,t) + \Psi(t,t), tf'(t)\rangle =: E_f. \tag{9}$$

We state the formulas (8) and (9) in separate theorems because their proofs are separated in the main body of the paper. In fact, we have structured the paper so that the reader interested only in (8) and its applications can largely ignore the extra (and somewhat heavy) apparatus needed to prove (9).

The results above can be made more transparent, and their range extended, for certain special cases. Of particular interest is the following:

Theorem 3.5. *Let Assumption 3.1 hold, and further assume that*

$$D \equiv 0, \quad \int s^{(2)}(c,c')\theta(dc') \equiv 1. \tag{10}$$

Then: (i) μ is the semicircle law σ_S of zero mean and unit variance. (ii) For polynomial functions f the random variables $Z_{f,k}$ converge in distribution toward a mean zero Gaussian random variable Z_f with variance given by (67). (iii) If the random variables $\xi_{\{\alpha,\beta\}}$ all satisfy a Poincaré inequality with common constant c (see §11 for definitions), then statement (ii) extends to continuously differentiable functions f with polynomial growth, with variance again given by (67).

We refer to the situation in Theorem 3.5 above as the *generalized Wigner matrix* model, because when $s^{(2)} \equiv 1$ one recovers Wigner matrices. The expression E_f in (9) can also be computed in this case, see (68) below. We note in passing that for Gaussian matrices, the condition (10) has been identified in [NSS02, Corollary 3.4] as sufficient and necessary (if $D = 0$) for μ to equal the semicircle distribution.

Similar considerations apply to the *generalized Wishart matrix* model, see §12.6 for details.

4. Basic spelling, grammar and counting

We introduce in this section the basic language employed throughout the paper for discussing enumeration problems. From letters we build *words*, from words we build *sentences*, and then we distinguish certain classes of words and sentences in terms of properties of naturally associated *graphs*. The classes of words and sentences singled out here for special attention are eventually going to be used to enumerate the terms in sums giving the (mixed and/or centered) moments of traces of powers of our random matrices. In particular, the *Wigner words* enumerate the only terms whose contributions to the law of large numbers for linear statistics do not vanish in the limit, whereas the *CLT word-pairs* take care of the only terms whose contributions to the CLT variance do not vanish in the limit. Further, the *CLT sentences* (which can be built up systematically from the CLT word-pairs) enumerate the nonnegligible terms in sums giving mixed centered moments of traces of powers of our random matrices. Critical weak Wigner words, and marked Wigner words, are needed (only) in the evaluation of the mean shift of linear statistics.

4.1. **Words and sentences.** A *word* is a finite sequence of letters at least one letter long. (Words are never empty!) We denote the length of a word w by $\ell(w)$. We say that a word w is *closed* if the first and last letters of w are the same. (Every one-letter word is automatically closed.) We view letters as one-letter words. A *sentence* is a finite sequence of words at least one word long. (Sentences are never empty, nor do they contain empty words!) We view words as one-word sentences. The *support* $\operatorname{supp} a$ of a sentence a is the set of letters appearing in a, and the *combinatorial weight* $\operatorname{wt} a$ is the cardinality of $\operatorname{supp} a$. We say that sentences a and b are *disjoint* if $\operatorname{supp} a \cap \operatorname{supp} b = \emptyset$. We say that sentences a and b are *equivalent* and write $a \sim b$ if there exists a one-to-one letter-valued function ψ defined on $\operatorname{supp} a$ such that the result of applying ψ letter by letter to a is b. In other words, $a \sim b$ whenever a codes to b under a simple substitution cipher.

We warn the reader that we distinguish between a sentence a and the word w obtained by concatenating all words in a; the "punctuation" carries information important for our purposes and therefore must not be ignored. For example, taking the set $\{1, 2, 3\}$ temporarily as our alphabet, the word 123123, the two-word sentence $[123, 123]$ and the three-word sentence $[1, 231, 23]$ are distinct objects according to our point of view.

4.2. **Graphs.** We fix terminology concerning graphs in a slightly restrictive but convenient way as follows. A *graph* $G = (V, E)$ is an ordered pair consisting of a finite nonempty set V of letters and a set E (possibly empty), where each element of E is an unordered pair of elements of V, i. e., a subset of V of cardinality 1 or 2. Elements of V are called *vertices* of G, elements of E are called *edges* of G, and edges of cardinality 1 are said to be *degenerate*. We say that a word $w = \alpha_1 \cdots \alpha_n$ of n letters is a *walk* on G provided that $\alpha_i \in V$ for $i = 1, \ldots, n$, and $\{\alpha_i, \alpha_{i+1}\} \in E$ for $i = 1, \ldots, n-1$, in which case we say that each of the vertices α_i and edges $\{\alpha_i, \alpha_{i+1}\}$ of G is *visited* by w. A *geodesic* in G is a walk visiting no vertex more than once. We say that G is *connected* if any two vertices are joined by a walk. If G is connected then $\#E \geq \#V - 1$. We call G a *tree* if G is connected and G has no nontrivial loops (in particular, G has no degenerate edges). Every two vertices of a tree are joined by a unique geodesic. For G to be a tree it is necessary and sufficient that G be connected and $\#E \leq \#V - 1$. A graph $G' = (V', E')$ where

$V' \subset V$ and $E' \subset E$ is called a *subgraph* of G. A *connected component* of G is a connected subgraph of G maximal in the family of connected subgraphs of G. We call G a *forest* if every connected component of G is a tree. A *spanning forest* in G is a graph $G' = (V', E')$ with $V' = V$ and $E' \subset E$ such that G' is a forest having the same number of connected components as does G. Every graph contains at least one spanning forest.

4.3. **Orthographic and grammatical notions.**

4.3.1. *The graph associated to a sentence.* Given a sentence

$$a = [w_i]_{i=1}^n = [[\alpha_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$$

consisting of n words (following a pattern we use often in the sequel, w_i denotes the i^{th} word of the sentence, and α_{ij} denotes the j^{th} letter of the i^{th} word) we define

$$G_a = (V_a, E_a)$$

to be the graph with

$$V_a = \operatorname{supp} a, \quad E_a = \left\{ \{\alpha_{ij}, \alpha_{i,j+1}\} \;\middle|\; \begin{array}{l} i = 1, \dots, n, \\ j = 1, \dots, \ell(a_i) - 1 \end{array} \right\}.$$

We view each word w_i of the sentence a in the natural way as a walk on G_a. We emphasize that $E_a = \emptyset$ if a consists of one-letter words. Note also the difference between the graph associated to the sentence a and the graph associated to the single word consisting of the concatenation of the words of a; in general the former has fewer edges than the latter.

4.3.2. *Weak Wigner words.* A word w is called a *weak Wigner word* under the following two conditions:

- w is closed.
- w visits every edge of G_w at least twice.

Suppose now that w is a weak Wigner word. If $\operatorname{wt} w = (\ell(w) + 1)/2$, then we drop the modifier "weak" and call w a *Wigner word*. (Every single letter word is automatically a Wigner word.) If $\operatorname{wt} w = (\ell(w) - 1)/2$, then we call w a *critical* weak Wigner word. For example, spelling with the alphabet $\{1, 2, 3\}$, we have that $w = 121$ is a Wigner word and that $w = 12121$ is a critical weak Wigner word.

4.3.3. *Weak CLT sentences.* Let $a = [w_i]_{i=1}^n$ be a sentence consisting of n words. We say that a is a *weak CLT sentence* under the following three conditions:

- All the words w_i are closed.
- Jointly the words/walks w_i visit each edge of G_a at least two times.
- For each $i \in \{1, \dots, n\}$ there exists $j \in \{1, \dots, n\} \setminus \{i\}$ such that the graphs G_{w_i} and G_{w_j} (both of which are subgraphs of G_a) have an edge in common.

Suppose now that a is a weak CLT sentence. If $\operatorname{wt} a = \sum_{i=1}^n \frac{\ell(w_i)-1}{2}$, then we drop the modifier "weak" and call a a *CLT sentence*. If $n = 2$ and a is a CLT sentence, then we call a a *CLT word-pair*. For example, again spelling with the alphabet $\{1, 2, 3\}$, we have that $a = [1231, 1321]$ is a CLT word-pair.

4.3.4. *Marked Wigner words.* A *marked Wigner word* is a three-word sentence $[w, \alpha, \beta]$ where w is a Wigner word, and α and β are distinct letters appearing in w.

4.3.5. *Cyclic permutations.* Given a word $w = [\alpha_i]_{i=1}^n$ of length n and a permutation σ of $\{1,\dots,n\}$, we define w^σ to be the word $[\alpha_{\sigma(i)}]_{i=1}^n$. If σ is a power of the cycle $(123\cdots n)$, then we call σ a *cyclic permutation*, and we say that w^σ is a *cyclic permutation* of w.

Lemma 4.4 ("The parity principle")**.** *Let G be a forest. Let e be an edge of G. Let w be a word admitting interpretation as a walk on G. Let w_* be the unique geodesic in G with initial and terminal vertices coinciding with those of w. Then the word/walk w visits the edge e an odd number of times if and only if the geodesic w_* visits e.*

This simple principle is repeatedly applied in the sequel. The proof is elementary and therefore omitted.

Proposition 4.5. *Let w be a weak Wigner word. (i) We have $\mathrm{wt}\, w \leq \frac{\ell(w)+1}{2}$ with equality if and only if G_w is a tree. (ii) If $\mathrm{wt}\, w = \frac{\ell(w)+1}{2}$ then w visits every edge of the tree G_w exactly twice. (iii) w is a Wigner word if and only if there exists a decomposition $w = \alpha w_1 \cdots \alpha w_r \alpha$ where α is the first letter of w and $w_1,\dots,w_r$ are pairwise disjoint Wigner words in which α does not occur. (iv) The inequality $\frac{\ell(w)-1}{2} < \mathrm{wt}\, w < \frac{\ell(w)+1}{2}$ is impossible.*

These are ideas coming up in some proofs of Wigner's semicircle law by the method of moments.

Proof. Put $G = (V,E) = G_w = (V_w, E_w)$. (i) The existence of the walk w makes it clear that G is connected. We have

$$\mathrm{wt}\, w - 1 \leq \#E \leq \frac{\ell(w)-1}{2}, \tag{11}$$

on the left because G is connected, and on the right by the hypothesis that w is a weak Wigner word. The result follows. (ii) Clear. (iii)($\Leftarrow$) Trivial. (iii)($\Rightarrow$) By (i) and (ii) already proved, the parts of the tree G explored by the walk w between successive visits to the vertex α have to be disjoint. (iv) Suppose rather that the inequality in question holds. Then $\ell(w)$ is even and $\#V = \frac{\ell(w)}{2}$, hence by (11) we have $\#E = \frac{\ell(w)}{2} - 1 = \#V - 1$, and hence G is a tree. We now arrive at a contradiction: by the parity principle w cannot be both closed and a walk that takes an odd number of steps. $\square$

As a consequence of Proposition 4.5, one can visualize equivalence classes of Wigner words as rooted planar trees, with the Wigner word determining an exploration path on the tree that visits each vertex at least once and goes over each edge exactly twice, c.f. Figure 1. We do not make explicit use of this correspondence but it does drive much of our intuition.

4.6. **Cross-sections.** We say that a set of sentences A is a *cross-section* of a set of sentences S if $A \subset S$ and for each sentence $s \in S$ there exists exactly one sentence in A equivalent to s. All the cross-sections of S arise by a process of selecting exactly one element from each $\sim$-equivalence class in S.

4.7. **Enumeration of Wigner words by Wigner words.** Fix a letter α. For each positive integer i choose a cross-section W_i of the set of Wigner words so as to achieve the following conditions:

- For all i, the letter α appears in no word belonging to W_i.

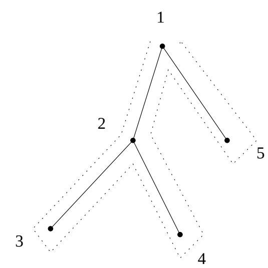

FIGURE 1. The rooted planar tree (solid) and exploration process (dotted) corresponding to the equivalence class of the Wigner word $w = 123242151$. Note the decomposition $w = 1w_1 1w_2 1$ with $w_1 = 23242$ and $w_2 = 5$.

- For all distinct i and j, every word belonging to W_i is disjoint from every word belonging to W_j.

(This is always possible to achieve because the set of letters is countably infinite.) Let φ be any real-valued function of words such that $\varphi(w)$ depends only on the equivalence class of w and vanishes for $\ell(w) \gg 0$, in which case the support of φ consists of only finitely many equivalence classes of words. By Proposition 4.5(iii) we have an enumeration formula

$$\sum_w \varphi(w) = \sum_{r=0}^{\infty} \sum_{w_1 \in W_1} \cdots \sum_{w_r \in W_r} \varphi(\alpha w_1 \cdots \alpha w_r \alpha) \tag{12}$$

where w ranges over any cross-section of the set of Wigner words. Formula (12) leads to many useful recursions. For example, it implies that the number of equivalence classes of Wigner words of length $2n+1$ is the n^{th} *Catalan number* $\frac{1}{n+1}\binom{2n}{n}$. The latter fact is anyhow clear from the rooted planar tree interpretation of equivalence classes of Wigner words.

Proposition 4.8. *Let w be a critical weak Wigner word. Put*

$$G = (V, E) = G_w = (V_w, E_w).$$

The following hold:

1. *G is connected.*
2. *Either $\#V - 1 = \#E$ or $\#V = \#E$.*
3. *If $\#V - 1 = \#E$, then:*
   (a) *G is a tree.*
   (b) *With exactly one exception w visits each edge of G exactly twice.*
   (c) *But w visits the exceptional edge exactly four times.*
4. *If $\#V = \#E$, then:*
   (a) *G is not a tree.*
   (b) *w visits each edge of G exactly twice.*

We state these facts for the sake of convenient reference. We omit the easy proofs.

Proposition 4.9. *Let $a = [w_i]_{i=1}^n$ be a weak CLT sentence consisting of n words. (i) We have $\operatorname{wt} a \leq \sum_{i=1}^n \frac{\ell(w_i)-1}{2}$. (ii) Suppose now that equality holds, i. e., that a is a CLT sentence. Then the words w_i of the sentence a are perfectly matched* *in the sense that for all i there exists unique j distinct from i such that w_i and w_j have a letter in common. In particular, n is even.*

This assertion (without proof) was made in [Jo82].

Proof. Lemma 4.10 below is the essential point of the proof. □

Lemma 4.10. *Let $a = [w_i]_{i=1}^n$ be a weak CLT sentence consisting of n words. Put $G = G_a$. Let k be the number of connected components of G. Then (i) $k \leq \left\lfloor \frac{n}{2} \right\rfloor$ and (ii) $\mathrm{wt}\, a \leq k - n + \left\lfloor \frac{\sum_{i=1}^n \ell(w_i)}{2} \right\rfloor$, where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.*

Proof. Inequality (i) is trivial: by hypothesis every word w_i of the sentence a is "mated" with at least one other word w_j $(j \neq i)$ of the sentence in the sense that the connected subgraphs G_{w_i} and G_{w_j} share an edge and *a fortiori* share a vertex.

Harder work is required to prove inequality (ii). Put $a = [[\alpha_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$, $I = \bigcup_{i=1}^n \{i\} \times \{1, \ldots, \ell(w_i) - 1\}$ and $A = [\{\alpha_{ij}, \alpha_{i,j+1}\}]_{(i,j) \in I}$. We visualize A as a left-justified table of n rows. Let $G' = (V', E')$ be any spanning forest in G. Since every connected component of G' is a tree, we have $\mathrm{wt}\, a = k + \#E'$ and so in order to prove (ii), we just have to bound $\#E'$. Now let $X = \{X_{ij}\}_{(i,j) \in I}$ be a table of the same "shape" as A, but with all entries equal either to 0 or 1. We call X an *edge-bounding table* under the following conditions:

- For all $(i,j) \in I$, if $X_{ij} = 1$, then $A_{ij} \in E'$.
- For each $e \in E'$ there exist distinct $(i_1, j_1), (i_2, j_2) \in I$ such that $X_{i_1 j_1} = X_{i_2 j_2} = 1$ and $A_{i_1 j_1} = A_{i_2 j_2} = e$.
- For each $e \in E'$ and index $i \in \{1, \ldots, n\}$, if e appears in the i^{th} row of A then there exists $(i,j) \in I$ such that $A_{ij} = e$ and $X_{ij} = 1$.

For any edge-bounding table X the corresponding quantity $\frac{1}{2} \sum_{(i,j) \in I} X_{ij}$ bounds $\#E'$, whence the terminology. At least one edge-bounding table exists, namely the table with a 1 in position (i,j) for each $(i,j) \in I$ such that $A_{ij} \in E'$ and 0's elsewhere. Now let X be an edge-bounding table such that for some index i_0 all the entries of X in the i_0^{th} row are equal to 1. Then the closed word w_{i_0} is a walk in G', and hence by the parity principle every entry in the i_0^{th} row of A appears there an even number of times and *a fortiori* at least twice. Now choose $(i_0, j_0) \in I$ such that $A_{i_0 j_0} \in E'$ appears in more than one row of A. Let Y be the table obtained by replacing the entry 1 of X in position (i_0, j_0) by the entry 0. Then it is not difficult to check that Y is again an edge-bounding table. Proceeding in this way we can find an edge-bounding table with 0 appearing at least once in every row, and hence we have $\#E' \leq \lfloor \frac{\#I - n}{2} \rfloor$, which is exactly what we need to prove (ii). □

4.11. **Enumeration of CLT sentences by CLT word-pairs.** Fix an even positive integer n and for $i = 1, \ldots, n/2$ choose a cross-section P_i of the set of CLT word-pairs so as to achieve the following condition:

- For all distinct i and j, every word-pair belonging to P_i is disjoint from every word-pair belonging to P_j.

We declare a permutation σ of $\{1, \ldots, n\}$ to be a *perfect matching* if it satisfies the following conditions:

- $\sigma(2i-1) < \sigma(2i)$ for $i = 1, \ldots, n/2$.
- $\sigma(2i-1) < \sigma(2i+1)$ for $i = 1, \ldots, n/2 - 1$.

Now let φ be any real-valued function of n-word-long sentences $a = [w_i]_{i=1}^n$ such that $\varphi(a)$ depends only on the equivalence class of a and vanishes for $\sum_{i=1}^n \ell(w_i) \gg 0$, in which case the support of φ consists of only finitely many equivalence classes of sentences. By Proposition 4.9 we have an enumeration formula

$$\sum_a \varphi(a) = \sum_{[p_1,p_2]\in P_1} \cdots \sum_{[p_{n-1},p_n]\in P_{n/2}} \sum_{\substack{\sigma\in S_n \\ \sigma\text{: perfect matching}}} \varphi([p_{\sigma^{-1}(i)}]_{i=1}^n) \tag{13}$$

where a ranges over any cross-section of the set of n-word-long CLT sentences.

Proposition 4.12. *Let $a = [w, x]$ be a CLT word-pair and put*

$$G = (V, E) = G_a = (V_a, E_a).$$

For each $e \in E$ let $\nu(e, w)$ (resp., $\nu(e, x)$) denote the number of visits to e by the word/walk w (resp., x). The following hold:

(1) *G is connected.*
(2) *Either $\#V - 1 = \#E$ or $\#V = \#E$.*
(3) *If $\#V - 1 = \#E$, then:*
   (a) *G is a tree.*
   (b) *For all $e \in E$ both $\nu(e, w)$ and $\nu(e, x)$ are even.*
   (c) *For unique $e_0 \in E$ we have $\nu(e_0, w) = \nu(e_0, x) = 2$.*
   (d) *For all $e \in E \setminus \{e_0\}$ we have $\nu(e, w) + \nu(e, x) = 2$.*
   (e) *Both w and x are Wigner words.*
(4) *If $\#V = \#E$, then:*
   (a) *G is not a tree.*
   (b) *For all $e \in E$ we have $\nu(e, w) + \nu(e, x) = 2$.*
   (c) *For some $e \in E$ we have $\nu(e, w) = \nu(e, x) = 1$.*

We state these facts for the sake of convenient reference. We omit the easy proofs.

5. Limit calculations

We work out limits of and estimates for moments needed as "raw material" for the proofs of Theorems 3.2 and 3.3. Assumption 3.1 remains in force throughout these calculations.

5.1. **Random variables indexed by sentences.** Fix a sentence

$$a = [w_i]_{i=1}^n = [[\alpha_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$$

consisting of n words. We attach several random variables to a, as follows.

5.1.1. We define

$$\xi(a) = \prod_{i=1}^n \prod_{j=1}^{\ell(w_i)-1} \xi_{\{\alpha_{ij},\alpha_{i,j+1}\}}.$$

From the independence of the family $\{\xi_e\}$ and assumption (2) concerning the absolute moments of these random variables, we deduce that

$$E|\xi(a)| = \prod_{e\text{: edge of } G_a} E|\xi_e|^{\nu(e)} \le \prod_{\substack{e=\{\alpha,\beta\}, \\ \text{edge of } G_a}} \left\{ \begin{array}{ll} s^{(\nu(e))}(\kappa_0(\alpha), \kappa_0(\beta)) & \text{if } \alpha \neq \beta, \\ d^{(\nu(e))}(\kappa_0(\alpha)) & \text{if } \alpha = \beta, \end{array} \right. \tag{14}$$

where

$$\nu(e) = \left(\begin{array}{l} \text{total number of visits made to } e \\ \text{by all the words/walks } w_i \end{array} \right). \tag{15}$$

(While $\nu(e)$ depends on the sentence a, to avoid unnecessary clutter we omit this dependence from the notation). Further and crucially, since all the random variables of the family $\{\xi_e\}$ are of mean zero, if w is a closed word, then $E\xi(w) = 0$ unless w is a weak Wigner word.

5.1.2. We define

$$\bar{\xi}(a) = \prod_{i=1}^{n} (\xi(w_i) - E\xi(w_i)).$$

Expanding the product on the right in evident fashion we find that

$$\bar{\xi}(a) = \sum_{I \subset \{1,\dots,n\}} (-1)^{\#I} \prod_{j \in \{1,\dots,n\} \setminus I} \xi(w_j) \cdot \prod_{i \in I} E\xi(w_i). \tag{16}$$

Clearly $E|\bar{\xi}(a)|$ is bounded by a constant depending only on $\sum_{i=1}^{n} \ell(w_i)$. Further and crucially, if all the words w_i are closed, then we have $E\bar{\xi}(a) = 0$ unless a is a weak CLT sentence.

5.1.3. *Auxiliary color-valued random variables.* We fix a letter-indexed i.i.d. family $\{\kappa(\alpha)\}$ of color-valued random variables with common distribution θ. These random variables are going to be used only for bookkeeping purposes. They need not be defined on the same probability space as the random variables ξ_e.

5.1.4. Put

$$M(a) = \prod_{\substack{e=\{\alpha,\beta\},\\ \text{edge of } G_a}} \left\{ \begin{array}{cl} 0 & \text{if } \nu(e) = 1, \\ s^{(\nu(e))}(\kappa(\alpha), \kappa(\beta)) & \text{if } \nu(e) > 1 \text{ and } \alpha \neq \beta, \\ d^{(\nu(e))}(\kappa(\alpha)) & \text{if } \nu(e) > 1 \text{ and } \alpha = \beta, \end{array} \right.$$

where $\nu(e)$ is as in (15).

5.1.5. Put

$$\overline{M}(a) = \sum_{I \subset \{1,\dots,n\}} (-1)^{\#I} M(a/I) \cdot \prod_{i \in I} M(w_i),$$

where for $I \neq \{1, \dots, n\}$ we denote by a/I the sentence obtained by striking the i^{th} word of a for all $i \in I$, and for $I = \{1, \dots, n\}$ we agree to put $M(a/I) = 1$. Note the analogy with expansion (16). Note also that $M(a)$, $\overline{M}(a)$ are random variables.

5.1.6. For each n-tuple $p = [p_i]_{i=1}^n$ of nonnegative integers put

$$H_p(a) = \sum_{\pi} \prod_{i=1}^{n} \prod_{j=1}^{\ell(w_i)} D(\kappa(\alpha_{ij}))^{\pi_{ij}}$$

where $\pi = [[\pi_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$ ranges over families of nonnegative integers subject to the constraints that $\sum_{j=1}^{\ell(w_i)} \pi_{ij} = p_i$ for $i = 1, \dots, n$. Note that $H_p(a) = \prod_{i=1}^n H_{p_i}(w_i)$. It is convenient to set $H_p(a) = 0$ for every n-tuple p of integers such that $p_i < 0$ for some i. We write

$$MH_p(a) = M(a)H_p(a), \quad \overline{M}H_p(a) = \overline{M}(a)H_p(a)$$

in order to abbreviate notation.

5.1.7. Note that $MH_p(a)$ (resp., $\overline{M}H_p(a)$) remains unchanged if for some permutation σ of $\{1,\dots,n\}$ we replace a by the sentence $[w_{\sigma(i)}]_{i=1}^n$ and p by the n-tuple $[p_{\sigma(i)}]_{i=1}^n$. Note further that if the sentence a can be presented as the concatenation of pairwise disjoint sentences $b_1,\dots,b_k$ where b_i is n_i words long, and correspondingly we present the n-tuple p as the concatenation of tuples $q_1,\dots,q_k$ where q_i is an n_i-tuple, then $MH_p(a)=\prod_{i=1}^k MH_{q_i}(b_i)$ (resp., $\overline{M}H_p(a)=\prod_{i=1}^k \overline{M}H_{q_i}(b_i)$), and moreover the factors on the right are independent.

5.2. **Admissibility.** Let $a=[w_i]_{i=1}^n$ be a sentence consisting of n words. For each edge e of the graph G_a, let $\nu(e)$ be the total number of visits to e by the words/walks w_i. We say that a is *weakly admissible* if for all edges e of G_a the following hold:

- $\nu(e)\in\{1,2,4\}$.
- If $\nu(e)=4$, then e is nondegenerate.

We say that a is *admissible* if for every nonempty subset $\{i_1<\cdots<i_\ell\}\subset\{1,\dots,n\}$ the subsentence $[w_{i_\nu}]_{\nu=1}^\ell$ is weakly admissible. For words weak admissibility and admissibility are the same thing. By Proposition 4.5 every Wigner word is admissible. By Proposition 4.8 every critical weak Wigner word is admissible. By Propositions 4.9 and 4.12 every CLT sentence is admissible.

Proposition 5.3. *Let $a=[w_i]_{i=1}^n$ be a weakly admissible sentence consisting of n words. Let $p=[p_i]_{i=1}^n$ be an n-tuple of nonnegative integers. Let $\gamma_1,\dots,\gamma_r$ be distinct letters such that* $\operatorname{supp} a\subset\{\gamma_1,\dots,\gamma_r\}$. *Then there exists a function f on the product of r copies of color space with the following properties:*

- *f is bounded and measurable.*
- *f has discontinuity set of measure zero with respect to $\theta^{\otimes r}$.*
- *$f(\kappa(\gamma_1),\dots,\kappa(\gamma_r))=MH_p(a)$.*
- *For all distinct letters $\delta_1,\dots,\delta_r$, the equivalent word $b=[[\beta_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$ to which a codes by the rule $\gamma_i\mapsto\delta_i$ for $i=1,\dots,r$ satisfies the equation*

$$f(\kappa_0(\delta_1),\dots,\kappa_0(\delta_r))=E\xi(b)\sum_\pi\prod_{i=1}^n\prod_{j=1}^{\ell(w_i)}D(\kappa_0(\beta_{ij}))^{\pi_{ij}}$$

where $\pi=[[\pi_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$ ranges over families of nonnegative integers subject to the constraints $\sum_{j=1}^{\ell(w_i)}\pi_{ij}=p_i$ for $i=1,\dots,n$.

Proof. For simplicity we discuss only the case $p=0$ and leave the remaining details to the reader. Put $G=(V,E)=G_a=(V_a,E_a)$ and as above, for all $e\in E$, let $\nu(e)$ be the total number of visits to e made by the words/walks w_i. Put

$$f(c_1,\dots,c_r)=\prod_{e=\{\alpha,\beta\}\in E}\left\{\begin{array}{cl}0 & \text{if } \nu(e)=1,\\ s^{(\nu(e))}(c_{\gamma^{-1}(\alpha)},c_{\gamma^{-1}(\beta)}) & \text{if } \nu(e)>1 \text{ and } \alpha\neq\beta,\\ d^{(\nu(e))}(c_{\gamma^{-1}(\alpha)}) & \text{if } \nu(e)>1 \text{ and } \alpha=\beta,\end{array}\right.$$

where γ^{-1} is the inverse of the bijection $(i\mapsto\gamma_i):\{1,\dots,r\}\to\{\gamma_1,\dots,\gamma_r\}$. Clearly f has the first three of the desired properties. If $\nu(e)=1$ for some $e\in E$, then the fourth property holds trivially (both sides of the desired equation vanish identically). Otherwise, if $\nu(e)>1$ for all $e\in E$, then f has the fourth property because, under the hypothesis of weak admissibility, we are operating in the regime in which we enforce equality in the moment bound (2). □

5.4. **Limiting behavior of $\langle \overline{L}(\mathcal{N}), x^n \rangle$.** Fix a nonempty finite set $\mathcal{N}$ of letters and a positive integer n.

5.4.1. We have an expansion

$$\operatorname{tr} X(\mathcal{N})^n = \sum_w \sum_J \sum_x N^{\frac{1}{2}(\#J-n)} \xi(x/J) \prod_{j\in J} D(\kappa_0(\beta_j))$$

where:

- $w = [\alpha_i]_{i=1}^{n+1}$ ranges over a cross-section of the set of closed words of length $n+1$;
- J ranges over subsets of the set $\{j \in \{1,\dots,n\} | \alpha_j = \alpha_{j+1}\}$;
- $x = [\beta_i]_{i=1}^{n+1}$ ranges over words such that $x \sim w$ and $\operatorname{supp} x \subset \mathcal{N}$; and
- x/J denotes the word obtained by striking the j^{th} letter of x for each $j \in J$.

Note that x/J arises from x by selective suppression of repeated letters. Note also that $E\xi(x/J) = 0$ unless x/J is a weak Wigner word. By considering how we may insert repetitions of letters into a given weak Wigner word, and after some further algebraic manipulation, we obtain an expansion

$$\begin{aligned}\langle \overline{L}(\mathcal{N}), x^n \rangle &= \sum_w N^{\operatorname{wt} w - \frac{1+\ell(w)}{2}} \sum_x \sum_\pi N^{-\operatorname{wt} w} E\xi(x) \prod_{i=1}^{\ell(w)} D(\kappa_0(\beta_i))^{\pi_i} \\ &=: \sum_w N^{\operatorname{wt} w - \frac{1+\ell(w)}{2}} S(\mathcal{N}, w) \qquad (17)\end{aligned}$$

where:

- w ranges over a cross-section of the set of weak Wigner words of length $\le n+1$;
- $x = [\beta_i]_{i=1}^{\ell(w)}$ ranges over words such that $x \sim w$ and $\operatorname{supp} x \subset \mathcal{N}$;
- $\pi = [\pi_i]_{i=1}^{\ell(w)}$ ranges over $\ell(w)$-tuples of nonnegative integers summing to $n+1-\ell(w)$; and
- $S(\mathcal{N}, w)$ is the result of carrying out the inner summations on x and π.

Note that for n fixed, as $N \to \infty$ and $\theta_{\mathcal{N}} \Rightarrow \theta$, only the part of the sum indexed by Wigner words w contributes nonnegligibly.

5.4.2. In this paragraph fix attention on a Wigner word w such that $\ell(w) \le n+1$. We want to understand the subsum $S(\mathcal{N}, w)$ appearing in formula (17) as a function of $\mathcal{N}$. Let $\gamma_1, \dots, \gamma_r$ be an enumeration of $\operatorname{supp} w$. Since Wigner words are admissible, Proposition 5.3 provides us with a function f defined on the product of r copies of color space with the following properties:

- f is bounded and measurable.
- f has discontinuity set of measure zero with respect to $\theta^{\otimes r}$.
- $MH_{n+1-\ell(w)}(w) = f(\kappa(\gamma_1), \dots, \kappa(\gamma_r))$.

- $S(\mathcal{N},w) = N^{-r} \sum_{\substack{(\beta_1,\dots,\beta_r)\in\mathcal{N}^r \\ \beta_1,\dots,\beta_r:\ \text{distinct}}} f(\kappa_0(\beta_1),\dots,\kappa_0(\beta_r)).$

Now let $[\mathcal{N}_k]_{k=1}^\infty$ be as in Assumption 3.1. We clearly have

$$\lim_{k\to\infty} S(\mathcal{N}_k,w) = \int\cdots\int f(c_1,\dots,c_r)\theta(dc_1)\cdots\theta(dc_r) = EMH_{n+1-\ell(w)}(w).$$

We remark that it is here we make use of the hypothesis that color space is Polish: we need it to guarantee weak convergence $\theta_{\mathcal{N}_k}^{\otimes r} \Rightarrow \theta^{\otimes r}$.

5.4.3. We may now conclude that

$$\lim_{k\to\infty}\langle \overline{L}(\mathcal{N}_k), x^n\rangle = \sum_w EMH_{n+1-\ell(w)}(w) \tag{18}$$

where the sum on the right is extended over a cross-section of the set of Wigner words. Note that only finitely many terms on the right are nonvanishing because $p<0 \Rightarrow H_p \equiv 0$.

Lemma 5.5. *With C as in (7), $\overline{L}(\mathcal{N}_k)$ converges weakly to a limit μ supported in the interval $[-C,C]$, and moreover $\langle \overline{L}(\mathcal{N}_k), x^n\rangle \to \langle \mu, x^n\rangle$ for all integers $n>0$.*

Proof. It is enough to prove that the right side of (18) is $O(C^n)$. There are $\binom{p+n-1}{n-1}$ n-tuples of nonnegative integers summing to p. Consequently we have

$$|MH_p(w)| \le \binom{p+\ell(w)-1}{\ell(w)-1}(|s^{(2)}|_\infty^{1/2})^{\ell(w)-1}|D|_\infty^p$$

for all Wigner words w and nonnegative integers p. There are $\frac{1}{\ell+1}\binom{2\ell}{\ell}$ equivalence classes of Wigner words of length $2\ell+1$ and clearly there are no Wigner words of even length. Consequently there are $O(2^n)$ equivalence classes of Wigner words of length $\le n+1$. The desired $O(C^n)$ bound for the right side of (18) follows. □

5.6. **Limiting behavior of $E\prod_{i=1}^n(\operatorname{tr} X(\mathcal{N})^{\nu_i} - E\operatorname{tr} X(\mathcal{N})^{\nu_i})$.** Again fix a finite non-empty set $\mathcal{N}$ of letters and a positive integer n. Also fix positive integers $\nu_1,\dots,\nu_n$ and put $\nu=[\nu_i]_{i=1}^n$.

5.6.1. We have an expansion

$$\prod_{i=1}^n(\operatorname{tr} X(\mathcal{N})^{\nu_i} - E\operatorname{tr} X(\mathcal{N})^{\nu_i})$$

$$= \sum_a\sum_K\sum_b N^{\frac12\sum_{i=1}^n(\#K_i-\nu_i)}\prod_{i=1}^n\left(\bar{\xi}(x_i/K_i)\prod_{j\in K_i}D(\kappa_0(\beta_{ij}))\right)$$

where:

- $a=[w_i]_{i=1}^n = [[\alpha_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$ ranges over a cross-section of the set of sentences n words long with i^{th} word of length ν_i+1 for $i=1,\dots,n$;
- $K=[K_i]_{i=1}^n$ ranges over n-tuples of sets of positive integers such that K_i is a subset of $\{j\in\{1,\dots,\nu_i\}|\alpha_{ij}=\alpha_{i,j+1}\}$ for $i=1,\dots,n$;
- $b=[x_i]_{i=1}^n = [[\beta_{ij}]_{j=1}^{\nu_i+1}]_{i=1}^n$ ranges over sentences $b\sim a$ such that $\operatorname{supp} b\subset \mathcal{N}$; and

- x_i/K_i denotes the word obtained by striking the k^{th} letter of x_i for all $k \in K_i$.

After some further algebraic manipulation, we obtain an expansion

$$
\begin{aligned}
&E\prod_{i=1}^{n}(\operatorname{tr} X(\mathcal{N})^{\nu_i} - E\operatorname{tr} X(\mathcal{N})^{\nu_i})\\
&= \sum_a N^{\mathrm{wt}\,a - \sum_{i=1}^{n}\frac{\ell(w_i)-1}{2}} \sum_b \sum_\pi N^{-\,\mathrm{wt}\,a} E\bar{\xi}(b) \prod_{i=1}^{n}\prod_{j=1}^{\ell(w_i)} D(\kappa_0(\beta_{ij}))^{\pi_{ij}}
\end{aligned} \tag{19}
$$

where:

- $a = [w_i]_{i=1}^n$ ranges over a cross-section of the set of weak CLT sentences n words long with i^{th} word of length $\leq \nu_i + 1$ for $i = 1, \dots, n$;
- $b = [[\beta_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$ ranges over sentences $b \sim a$ such that $\operatorname{supp} b \subset \mathcal{N}$; and
- $\pi = [[\pi_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$ ranges over families of nonnegative integers subject to the constraints $\sum_{j=1}^{\ell(w_i)} \pi_{ij} = \nu_i + 1 - \ell(w_i)$ for $i = 1, \dots, n$.

Note that for fixed ν, as $N \to \infty$ and $\theta_{\mathcal{N}} \Rightarrow \theta$, only the part of the sum indexed by CLT sentences a contributes nonnegligibly.

5.6.2. Now let $[\mathcal{N}_k]_{k=1}^\infty$ be as in Assumption 3.1. Since CLT words are admissible, an analysis similar to that undertaken in §5.4.2 leads to the conclusion that

$$
\lim_{k\to\infty} E\prod_{i=1}^{n}(\operatorname{tr} X(\mathcal{N}_k)^{\nu_i} - E\operatorname{tr} X(\mathcal{N}_k)^{\nu_i}) = \sum_a E\overline{M}H_{[\nu_i+1-\ell(w_i)]_{i=1}^n}(a) \tag{20}
$$

where $a = [w_i]_{i=1}^n$ ranges over a cross-section of the set of CLT sentences n words long. Since the analysis is straightforward, somewhat long, and very tedious, we omit it. Note that only finitely many nonzero terms appear in the sum on the right.

Lemma 5.7. *There exists a family $[Y_n]_{n=1}^\infty$ of mean zero random variables defined on a common probability space with Gaussian joint distribution such that for all positive integers n and positive integers $\nu_1, \dots, \nu_n$ the right side of limit formula (20) gives the expectation $E\prod_{i=1}^n Y_{\nu_i}$.*

Proof. Let $A(\nu_1, \dots, \nu_n)$ denote the right side of (20). The matrix $[[A(i,j)]_{i=1}^\infty]_{j=1}^\infty$ is symmetric and every finite block $[[A(i,j)]_{i=1}^r]_{j=1}^r$ in the upper left corner is positive semidefinite since it is the limit of such matrices. Consequently there exists a family $[Y_n]_{n=1}^\infty$ of mean zero random variables on a common probability space with Gaussian joint distribution such that $EY_iY_j = A(i,j)$ for all i and j. By the enumeration formula (13) and the relations discussed in §5.1.7, we have

$$
A(\nu_1, \dots, \nu_n) = \begin{cases} \displaystyle\sum_{\substack{\sigma \in S_n \\ \sigma:\ \text{perfect matching}}} \prod_{i=1}^{n/2} A(\nu_{\sigma(2i-1)}, \nu_{\sigma(2i)}) & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd.} \end{cases}
$$

But the expression on the right side is the Wick formula for the expectation $E\prod_{i=1}^n Y_{\nu_i}$, cf. [Ja97, Theorem 1.28]. □

6. Proofs of Lemma 2.3 and Theorem 3.2

Lemma 6.1. *Fix $K > \max(1, C^2)$ with C as in (7). Then we have*

$$\lim_{k\to\infty} \langle \overline{L}(\mathcal{N}_k), |g|\mathbf{1}_{|x|>K}\rangle = 0$$

for every real-valued measurable function g on the real line with polynomial growth at infinity.

Proof. Let n be any nonnegative integer. We have

$$\langle \overline{L}(\mathcal{N}_k), |x|^n\mathbf{1}_{|x|>K}\rangle \le \sqrt{\langle \overline{L}(\mathcal{N}_k), x^{2n}\rangle}\sqrt{\langle \overline{L}(\mathcal{N}_k), \mathbf{1}_{|x|>K}\rangle} \le \langle \overline{L}(\mathcal{N}_k), x^{2n}\rangle/K^n$$

by Cauchy-Schwartz followed by Chebyshev, and hence

$$\limsup_{k\to\infty} \langle \overline{L}(\mathcal{N}_k), |x|^n\mathbf{1}_{|x|>K}\rangle \le \limsup_{k\to\infty} \langle \overline{L}(\mathcal{N}_k), x^{2n}\rangle/K^n. \tag{21}$$

Because $K > 1$, the quantity on the left side of (21) is an increasing function of n, and moreover that quantity bounds $\limsup_{k\to\infty} \langle \overline{L}(\mathcal{N}_k), |g|\mathbf{1}_{|x|>K}\rangle$ for all $n \gg 0$ (because g is of polynomial growth). But by Lemma 5.5, because $K > C^2$, the quantity on the right side of (21) tends to 0 as $n \to \infty$. The result follows. ☐

6.2. **The functions $\Phi^{(w,p)}(c)$.** To each Wigner word w and nonnegative integer p we associate a real-valued bounded measurable function $\Phi^{(w,p)}(c)$ on color space by the following recursive procedure. As in Proposition 4.5, in the unique way possible, write $w = \alpha w_1 \cdots \alpha w_r \alpha$ where α is the first letter of w and the w_i are pairwise disjoint Wigner words in which α does not appear, and then put

$$\Phi^{(w,p)}(c) = \sum_\pi D(c)^{\pi_0+\cdots+\pi_r} \prod_{i=1}^r \int s^{(2)}(c,c')\Phi^{(w_i,\pi_{i+r})}(c')\theta(dc') \tag{22}$$

where $\pi = [\pi_i]_{i=0}^{2r}$ ranges over $(2r+1)$-tuples of nonnegative integers summing to p. By convention, if w is the single letter word α, then $r = 0$ and therefore $\Phi^{(\alpha,p)} = D(c)^p$, which gives a way to initialize the recursions (22). Note that for fixed p and c the quantity $\Phi^{(w,p)}(c)$ depends only on the equivalence class of w. Intuitively, $\Phi^{(w,p)}(c)$ determines the dominant contribution to the expectation of $\operatorname{tr} X(\mathcal{N})^{\ell(w)+p}$ by those terms that use entries from D p times, such that when these are discarded, the resulting word determined by the indices is equivalent to w, and such that the color of the initial letter is c. For example, in the special case that $D(\cdot) = 0$, one must have $p = 0$, hence all π_i vanish, and the contribution, for a given w, can be visualized by writing on each edge (v_1, v_2) of the rooted planar tree the value $(s^{(2)})^{1/2}(\kappa(v_1), \kappa(v_2))$, collecting the product of such values along the exploration path determined by the word w, and averaging over the choices of colors *except* for the choice of the color of the root, which is fixed at c.

Lemma 6.3. *We have the following identity of formal power series in t with coefficients in the space of real-valued bounded measurable functions on color space:*

$$\Phi(c,t) = \sum_w \sum_{p=0}^\infty \Phi^{(w,p)}(c)t^{\ell(w)+p} \tag{23}$$

Here w ranges over a cross-section of the set of Wigner words.

Proof. Via the enumeration formula (12) it follows from definition (22) that the power series on the right side of (23) satisfies (3), whence the result. ☐

Lemma 6.4. *Let w be a Wigner word. Let α be the first letter of w. Let p be a nonnegative integer. Then we have*

$$E(MH_p(w)|\kappa(\alpha)) = \Phi^{(w,p)}(\kappa(\alpha)), \quad \text{a.s..} \tag{24}$$

Proof. As in definition (22), write $w = \alpha w_1 \cdots \alpha w_r \alpha$ where the w_i are pairwise disjoint Wigner words in which α does not occur and let α_i denote the first letter of w_i. By definition of $M(\cdot)$ and $H_p(\cdot)$ we have

$$MH_p(w) = \sum_\pi D(\kappa(\alpha))^{\pi_0+\cdots+\pi_r} \prod_{i=1}^r s^{(2)}(\kappa(\alpha), \kappa(\alpha_i)) MH_{\pi_{i+r}}(w_i) \tag{25}$$

where $\pi = [\pi_i]_{i=0}^{2r}$ ranges over $(2r+1)$-tuples of nonnegative integers summing to p. Now take conditional expectations on both sides of (25). By induction on $\ell(w)$, and the relations of independence built into the definitions of $M(\cdot)$ and $H_\cdot(\cdot)$, we get (24) after a routine calculation. □

6.5. **Ends of the proofs.**

6.5.1. *Proof of Lemma 2.3.* Uniqueness of a probability measure with moments (6) and support (7) (which is compact) is clear. Only existence requires proof. After enlarging the originally given model in evident fashion we may assume without loss of generality that for every letter there exist infinitely many letters of the same color. And then we may assume without loss of generality that $\sigma = \theta$ because in Assumption 3.1 we may substitute $\theta_\mathcal{N}$ for θ without falsifying it. Now fix any sequence $[\mathcal{N}_k]_{k=1}^\infty$ as in Assumption 3.1. Let μ be the weak limit of $\overline{L}(\mathcal{N}_k)$ provided by Lemma 5.5. By the cited lemma, μ satisfies the support bound (7). Moreover, by the cited lemma combined with limit formula (18), the measure μ has moments

$$\langle \mu, x^n \rangle = \sum_{w \in W} EMH_{n+1-\ell(w)}(w) \tag{26}$$

where w ranges over a cross-section of the set of Wigner words. By Lemmas 6.3 and 6.4 we can evaluate the right side of (26). We find finally that moment formula (6) does indeed hold for μ. □

6.5.2. *Proof of Theorem 3.2.* Fix any real-valued bounded continuous function f on the real line and $\epsilon > 0$. For the convergence $L(\mathcal{N}_k) \Rightarrow \mu$ it is enough to show that

$$\lim_{k\to\infty} P(|\langle L(\mathcal{N}_k), f\rangle - \langle \mu, f\rangle| > \epsilon) = 0. \tag{27}$$

Fix K as in Lemma 6.1. By the Weierstrass approximation theorem write

$$f = g + Q, \quad \sup_{|x|\le K} |g(x)| < \epsilon/4$$

where Q is a polynomial function. We have

$$\begin{aligned}\langle L(\mathcal{N}_k), f\rangle - \langle \mu, f\rangle &= \Big[\langle L(\mathcal{N}_k), \mathbf{1}_{|x|\le K} g\rangle - \langle \mu, \mathbf{1}_{|x|\le K} g\rangle\Big] + \langle L(\mathcal{N}_k), \mathbf{1}_{|x|>K} g\rangle \\ &\quad + \Big[\langle \overline{L}(\mathcal{N}_k), Q\rangle - \langle \mu, Q\rangle\Big] + \Big[\langle L(\mathcal{N}_k), Q)\rangle - \langle \overline{L}(\mathcal{N}_k), Q\rangle\Big]\end{aligned}$$

and therefore have

$$\begin{aligned} P(|\langle L(\mathcal{N}_k), f\rangle - \langle \mu, f\rangle| > \epsilon) &\le P(\langle L(\mathcal{N}_k), \mathbf{1}_{|x|>K}|g|\rangle\rangle > \epsilon/6) \\ &\quad + P(|\langle \overline{L}(\mathcal{N}_k), Q\rangle - \langle \mu, Q\rangle| > \epsilon/6) \\ &\quad + P(|\langle L(\mathcal{N}_k), Q\rangle\rangle - \langle \overline{L}(\mathcal{N}_k), Q\rangle| > \epsilon/6) \\ &:= P_1 + P_2 + P_3. \end{aligned}$$

We have $P_1 \to 0$ by Lemma 6.1. We have $P_2 \to 0$ by Lemma 5.5. We have $P_3 \to 0$ by limit formula (20). Therefore (27) does indeed hold.

We finally turn to proving the convergence $\mu_{\mathcal{N}_k} \Rightarrow \mu$. The proof of Lemma 2.3 shows that the analogue

$$\langle \mu_{\mathcal{N}}, x^n\rangle = \sum_w EM_{\mathcal{N}} H_{n+1-\ell(w),\mathcal{N}}(w) \tag{28}$$

of formula (26) holds for any nonempty finite set of letters $\mathcal{N}$, where the random variables $M_{\mathcal{N}}(w)$ and $H_{p,\mathcal{N}}(w)$ are defined by mimicking the definitions of $M(w)$ and $H_p(w)$, only this time using a letter-indexed family $\{\kappa_{\mathcal{N}}(\alpha)\}$ of color-valued family i.i.d. random variables with common law $\theta_{\mathcal{N}}$. Note that $M_{\mathcal{N}}(w)$ and $H_{p,\mathcal{N}}(w)$ are uniformly bounded in $\mathcal{N}$. Clearly for each Wigner word w and nonnegative integer p we have convergence in distribution $M_{\mathcal{N}_k} H_{p,\mathcal{N}_k}(w) \to MH_p(w)$, which extends to the convergence of expectations by bounded convergence. The sum in (28) being over a finite number of terms, it follows that $\langle \mu_{\mathcal{N}_k}, x^n\rangle \to \langle \mu, x^n\rangle$ for all n, and in turn that $\mu_{\mathcal{N}_k} \Rightarrow \mu$ since the measures in play here have uniformly bounded supports. The proof of Theorem 3.2 is complete. □

7. The Füredi-Komlós circle of ideas

In this section, we describe a (rough) technique which allows us to bound traces of polynomials of our random matrices when the degree of the polynomial is allowed to grow with the dimension of the matrix. The approach we take is inspired by the work of Füredi and Komlós [FK81]. We mention in passing that for Wigner matrices all of whose entries have even distributions, much more detailed information is available in [SS98].

7.1. **FK sentences.** Let $a = [w_i]_{i=1}^n$ be a sentence of n words. We say that a is an *FK sentence* under the following conditions:

- G_a is a tree.
- Jointly the words/walks w_i visit no edge of G_a more than twice.
- For $i = 1, \ldots, n-1$, the first letter of w_{i+1} belongs to $\bigcup_{j=1}^i \operatorname{supp} w_j$.

We say that a is an *FK word* if $n = 1$. Any word admitting interpretation as a walk on a forest visiting no edge of the forest more than twice is automatically an FK word. The constituent words of an FK sentence are FK words. If an FK sentence is at least two words long, then the result of dropping the last word is again an FK sentence. If the last word of an FK sentence is at least two letters long, then the result of dropping the last letter of the last word is again an FK sentence.

7.2. **The graph G_a^1 associated to a sentence.** Given an n-word-long sentence $a = [w_i]_{i=1}^n$, we define $G_a^1 = (V_a^1, E_a^1)$ to be the subgraph of $G_a = (V_a, E_a)$ with $V_a^1 = V_a$ and E_a^1 equal to the set of edges $e \in E_a$ such that the words/walks w_i jointly visit e exactly once.

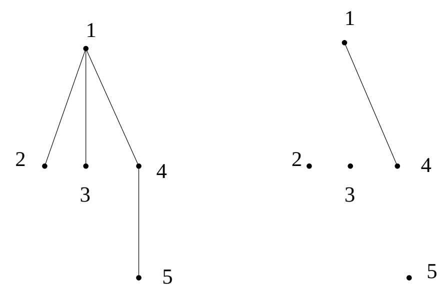

FIGURE 2. The graphs G_w (left) and G_w^1 (right) for the FK word $w = 12131454$

Proposition 7.3. *Let w be an FK word. There is exactly one way to write $w = w_1 \cdots w_r$ where the words w_i are pairwise disjoint Wigner words.*

In this situation, denoting by α_i the first letter of w_i, we declare the word $\alpha_1 \cdots \alpha_r$ to be the *acronym* of the FK word w.

Proof. The only possible decomposition $w = w_1 \cdots w_r$ of the desired type is the one with breaks at the edges of G_w^1. Since the transition from w_{i-1} to w_i is along an edge of the tree G_w never again visited by w, the words w_i must be pairwise disjoint. Since every edge of G_w visited by w_i is visited exactly twice by w, and the w_i are pairwise disjoint, in fact w_i visits every edge of G_w either twice or never, hence by the parity principle w_i is closed, and hence w_i is a Wigner word. □

Lemma 7.4. *There are at most 2^{n-1} equivalence classes of FK words of length n.*

Proof. From the recursion (12) (see also §12.3.3 below) it is easy to deduce that the sum of terms $t^{\ell(w)}$ extended over a cross-section of the set of Wigner words is

$$\Phi(t) = \frac{1 - \sqrt{1 - 4t^2}}{2t}.$$

Via the preceding lemma, it follows that the sum of terms $t^{\ell(w)}$ extended over a cross-section of the set of FK words is

$$\frac{\Phi(t)}{1 - \Phi(t)} = -1/2 + \frac{1}{2}\frac{1 + 2t}{\sqrt{1 - 4t^2}} = t + \left(\frac{1}{2} + t\right)\sum_{n=1}^{\infty} \binom{2n}{n} t^{2n},$$

whence the claimed bound. □

7.5. **FK syllabification.** Let $w = [\alpha_i]_{i=1}^n$ be a word of length n. Roughly speaking, we wish to define a parsing of w into an FK sentence by going sequentially over the letters in w and declaring a new word each time not doing so would prevent the sentence formed up to that point from being an FK sentence. More precisely, we define a sentence w', which we call the *FK syllabification* of w, by the following procedure. We declare an edge e of G_w to be *new* (relative to w) if for some index $1 \leq i < n$ we have $e = \{\alpha_i, \alpha_{i+1}\}$ and $\alpha_{i+1} \notin \{\alpha_1, \ldots, \alpha_i\}$, and otherwise we declare e to be *old*. We define w' to be the sentence obtained by breaking w at all visits to old edges of G_w and at third and subsequent visits to new edges of G_w. For example, temporarily spelling with the alphabet $\{1, 2, 3\}$, the FK syllabification of

$w = 1231$ is the sentence $w' = [123, 1]$ consisting of two words; the FK syllabification process has to "insert a comma" between 3 and 1 because 1231 is not an FK word, whereas 1, 12 and 123 are. It is clear that $G_{w'}$ is a spanning tree in G_w, that w' is an FK sentence, and that w is the concatenation of the constituent words of w'. Moreover, we have $w = w'$ if and only if w is an FK word. Clearly the FK syllabification process preserves equivalence, i. e., $w \sim x \Rightarrow w' \sim x'$.

Lemma 7.6. *Let $a = [w_i]_{i=1}^n$ be a sentence of $n \geq 2$ words. Put $b = [w_i]_{i=1}^{n-1}$ and $c = w_n$. Assume that b is an FK sentence, that c is an FK word, and that the first letter of c belongs to $\operatorname{supp} b$. Let $\gamma_1 \cdots \gamma_r$ be the acronym of c spelled out in full. (Note that by hypothesis $\gamma_1 \in \operatorname{supp} b$.) Let ℓ be the largest index such that $\gamma_\ell \in \operatorname{supp} b$ and write $d = \gamma_1 \cdots \gamma_\ell$. The following conditions are both necessary and sufficient for a to be an FK sentence:*

- *d is a geodesic in the forest G_b^1.*
- $\operatorname{supp} b \cap \operatorname{supp} c = \operatorname{supp} d$.

Consequently there exist at most $(\operatorname{wt} b)^2$ equivalence classes of FK sentences $[x_i]_{i=1}^n$ such that $b \sim [x_i]_{i=1}^{n-1}$ and $c \sim x_n$. See Figure 3 for an example of two such equivalence classes and their pictorial description.

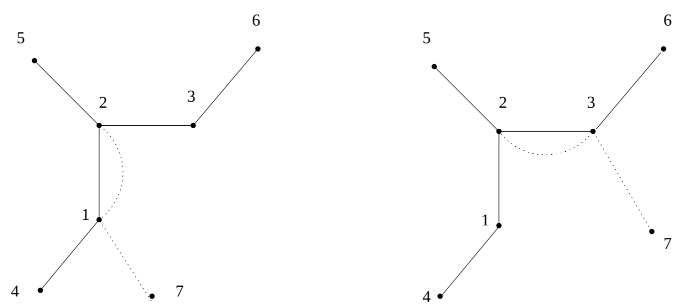

FIGURE 3. Two inequivalent FK sentences $[x_1, x_2]$ corresponding to $b = 141252363$ (solid) and $c = 1712 \sim 3732$ (dashed).

Proof. Sufficiency is easy to check. We omit the details. We turn to the proof of necessity. To begin with, since G_a is a tree, d is the unique geodesic in $G_c \subset G_a$ joining γ_1 to γ_ℓ, and hence is also the unique geodesic in $G_b \subset G_a$ joining γ_1 to γ_ℓ. Now d only visits edges of G_b already visited by the constituent words of b. Therefore we have $E_d \subset E_b^1$, i. e., d is a walk in G_b^1. By Proposition 7.3 we have $E_c^1 = E_{\gamma_1 \cdots \gamma_r}$. By definition of an FK sentence we have $E_b \cap E_c \subset E_b^1 \cap E_c^1$. It follows that $E_b \cap E_c = E_d$. Finally, we have

$$\#V_a = 1 + \#E_a = 1 + \#E_b + 1 + \#E_c - 1 - \#E_d = \#V_b + \#V_c - \#V_d,$$

and hence, since $\#V_b + \#V_c - \#V_b \cap V_c = \#V_a$, the inclusion $V_d \subset V_b \cap V_c$ is in fact an equality. $\square$

Lemma 7.7. *Let $\Gamma(k, \ell, m)$ denote the set of equivalence classes of FK sentences $a = [w_i]_{i=1}^m$ consisting of m words such that $\sum_{i=1}^m \ell(w_i) = \ell$ and $\operatorname{wt} a = k$. We have*

$$\#\Gamma(k, \ell, m) \leq 2^{\ell - m} \binom{\ell - 1}{m - 1} k^{2(m-1)}.$$

Proof. There are exactly $\binom{\ell-1}{m-1}$ m-tuples of positive integers summing to ℓ and hence by Lemma 7.4 there are at most $2^{\ell-m}\binom{\ell-1}{m-1}$ ways to prescribe equivalence classes of FK words $w_1,\dots,w_m$ subject to the constraint $\sum_{i=1}^m \ell(w_i)=\ell$. Now fix FK words $w_1,\dots,w_m$ such that $\sum_{i=1}^m \ell(w_i)=\ell$. By Lemma 7.6 there exist at most $k^{2(m-1)}$ equivalence classes of FK sentences $b=[x_i]_{i=1}^m$ such $k=\mathrm{wt}\, b$ and $w_i\sim x_i$ for $i=1,\dots,m$. The result follows. □

Lemma 7.8. *For any FK sentence $a=[w_i]_{i=1}^m$ consisting of m words we have*

$$m=\#E_a^1-2\,\mathrm{wt}\,a+2+\sum_{i=1}^m \ell(w_i). \tag{29}$$

Proof. Put $M:=\sum_{i=1}^m \ell(w_i)$. Consider the word $[\alpha_i]_{i=1}^M$ obtained by concatenating the words of the sentence a. Consider the list $A=[\{\alpha_i,\alpha_{i+1}\}]_{i=1}^{M-1}$ of unordered pairs of letters. Among the entries of A we find $2\#E_a-\#E_a^1$ of them that are edges of G_a, while the rest correspond to the $m-1$ "commas" in the sentence a; and moreover, since G_a is a tree, we have $\#E_a=\mathrm{wt}\,a-1$. The result follows. □

Proposition 7.9. *For all positive integers n,k satisfying $n\geq 2k-2$ there are at most*

$$N_{\mathrm{FK}}(n,k):=2^n n^{3(n-2k+2)} \tag{30}$$

equivalence classes of weak Wigner words w such that $\ell(w)=n+1$ and $\mathrm{wt}\,w=k$.

This is a crude but easy-to-apply version of the estimate one obtains by exploiting the idea of "coding" introduced by Füredi and Komlós in [FK81].

Proof. Let w be a weak Wigner word. Let w' be the FK syllabification of w. Let m be the number of words in the sentence w'. We must have $E^1_{w'}=\emptyset$ lest there exist an edge of G_w visited only once by w and so we must have $m=\ell(w)-2\,\mathrm{wt}\,w+2$ by the preceding lemma. Therefore $\#\Gamma(k,n+1,n-2k+3)$ bounds the quantity we wish to estimate, whence the desired result by Lemma 7.7, after a short further calculation which we omit. □

7.10. **Companion estimate.** To exploit the preceding proposition we need also to bound $E|\xi(w)|$ for all weak Wigner words w such that $\ell(w)=n+1$ and $k=\mathrm{wt}\,w$. Fix such a word w now. We claim that

$$E|\xi(w)|\leq C(3(n+2-2k))\cdot C(2)^{n/2}\,,\ \text{with } C(q):=1\vee\sup_{\alpha,\beta}\max_{m=1}^{q} E|\xi_{\{\alpha,\beta\}}|^m. \tag{31}$$

Consider the graph $G_w=(V_w,E_w)$ and let ℓ be the number of edges of E_w visited exactly twice by w. We have by (14) and the Hölder inequality that

$$E|\xi(w)|\leq C(n-2\ell)C(2)^\ell.$$

We have $\#E_w\geq \#V_w-1=k-1$ since G is connected, $n\geq 3\cdot(\#E_w-\ell)+2\ell$ by counting, and hence

$$n-2\ell\leq 3(n+2-2k).$$

The desired estimate now follows since $C(q)$ is a nondecreasing function of q bounded below by 1.

8. Bracelets, polarizations and enumeration

We have already seen in Section 5 that limiting variances are determined by the enumeration of CLT word-pairs. In the current section, we study the structure of such word-pairs and their associated graphs. These turn out to be classified by certain "bracelets with pendant trees".

8.1. **Graph-theoretical definitions.**

8.1.1. *Bracelets.* We say that a graph $G = (V, E)$ is a *bracelet* if there exists an enumeration $\alpha_1, \ldots, \alpha_r$ of V such that

$$E = \begin{cases} \{\{\alpha_1, \alpha_1\}\} & \text{if } r = 1, \\ \{\{\alpha_1, \alpha_2\}\} & \text{if } r = 2, \\ \{\{\alpha_1, \alpha_2\}, \{\alpha_2, \alpha_3\}, \{\alpha_3, \alpha_1\}\} & \text{if } r = 3, \\ \{\{\alpha_1, \alpha_2\}, \{\alpha_2, \alpha_3\}, \{\alpha_3, \alpha_4\}, \{\alpha_4, \alpha_1\}\} & \text{if } r = 4, \end{cases}$$

and so on. We call r the *circuit length* of the bracelet G.

8.1.2. *Unicyclic graphs.* We say that a graph $G = (V, E)$ is *unicyclic* if G is connected and $\#V = \#E$. In other words, a unicyclic graph is a connected graph with one too many edges to be a tree. Any bracelet of circuit length $\neq 2$ is unicyclic. However, a bracelet of circuit length 2 is a tree.

Proposition 8.2. *Let $G = (V, E)$ be a unicyclic graph. For each edge $e \in E$ put $G \setminus e = (V, E \setminus \{e\})$. Let Z be the subgraph of G consisting of all $e \in E$ such that $G \setminus e$ is connected, along with all attached vertices. Let r be the number of edges of Z. Let F be the graph obtained from G by deleting all edges of Z. The following statements hold:*

1. *F is a forest with exactly r connected components.*
2. *If G has a degenerate edge, then $r = 1$.*
3. *If G has no degenerate edge, then $r \geq 3$.*
4. *Z meets each connected component of F in exactly one vertex.*
5. *Z is a bracelet of circuit length r.*
6. *For all $e \in E$ the following conditions are equivalent:*
   (a) *$G \setminus e$ is connected.*
   (b) *$G \setminus e$ is a tree.*
   (c) *$G \setminus e$ is a forest.*

We call Z the *bracelet* of G. We call r the *circuit length* of G, and each of the components of F we call a *pendant tree*.

Proof. The proposition is well-known in principle. We just explain how to prove statement 5 and omit the remaining details. Pick an edge $e = \{\alpha, \beta\}$ of G so that $G \setminus e$ is a spanning tree. Then e is an edge of Z, and it is not difficult to verify that the edges of Z distinct from e are the edges of the tree $G \setminus e$ visited by the unique geodesic in $G \setminus e$ joining α to β. So it is clear that Z is a bracelet. $\square$

8.3. **The bracelet of a CLT word-pair.**

Fix a CLT word-pair $[w, x]$. Let $G = G_{[w,x]}$ be the associated graph.

8.3.1. By Proposition 4.12, either G is unicyclic or G is a tree. If G is unicyclic, we define the *bracelet*, *circuit length*, and *pendant trees* of $[w,x]$ to be the same as those defined for G by Proposition 8.2. Suppose now that G is a tree. Then there exists by Proposition 4.12(3)(c) a unique edge of G visited exactly twice by w and twice by x; this edge and attached vertices we declare to be the *bracelet* of $[w,x]$, and we declare the *circuit length* of $[w,x]$ to be the circuit length of its bracelet, namely 2. Erasing the edge of the bracelet from G leaves a forest of two components; as before, we call the components *pendant trees* of $[w,x]$. Note that in all cases the circuit length of $[w,x]$ depends only on the equivalence class of the word-pair $[w,x]$.

8.3.2. Let Z and r be the bracelet and circuit length of $[w,x]$, respectively. Note that G is unicyclic or a tree according to whether $r \neq 2$ or $r = 2$.

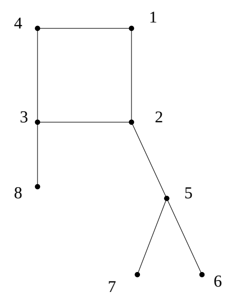

FIGURE 4. The bracelet 1234 of circuit length 4, and the pendant trees, associated with the CLT word-pair $[12565752341, 2383412]$

8.3.3. Now write $w = [\alpha_i]_{i=1}^{\ell(w)}$ and $x = [\beta_j]_{j=1}^{\ell(x)}$. Let $\breve{w}$ and $\breve{x}$ be the words obtained by dropping the last letters of w and x, respectively. Let σ be any cyclic permutation of $\{1,\ldots,\ell(\breve{w})\}$ and let τ be any cyclic permutation of $\{1,\ldots,\ell(\breve{x})\}$. Then $[\breve{w}^\sigma\alpha_{\sigma(1)}, \breve{x}^\tau\beta_{\tau(1)}]$ is again a CLT word-pair with associated graph, bracelet and circuit length the same as for $[w,x]$. (The "exponential notation" used here was defined in §4.3.5.) We declare the ordered pair (σ,τ) to be a *polarization* of $[w,x]$ if the last edge of G visited by the walk $\breve{w}^\sigma\alpha_{\sigma(1)}$ equals the last edge of G visited by the walk $\breve{x}^\tau\beta_{\tau(1)}$. Note that the set of polarizations of a CLT word-pair depends only on its equivalence class. The notions of bracelet and polarization are linked by the following result.

Lemma 8.4. *Let $[w,x]$ be a CLT word-pair. Put $G = G_{[w,x]}$. Let Z and r denote the bracelet and circuit length of $[w,x]$, respectively. Let e be an edge of G. Then: (i) e is an edge of Z if and only if both words/walks w and x visit e. (ii) Unless $r = 2$, there exist exactly r polarizations of $[w,x]$; but if $r = 2$, there exist exactly 4 polarizations of $[w,x]$.*

In the example of Figure 4, the four polarizations lead to the CLT word-pairs

$$[12565752341, 1238341]\,, \quad [25657523412, 2383412]\,,$$
$$[34125657523, 3834123]\,, \quad [41256575234, 4123834]\,.$$

Proof. If G is a tree, part (i) of the lemma holds by definition of Z, while part (ii) is a consequence of Proposition 4.12(3). So assume for the rest of the proof that G is unicyclic. By Proposition 4.12(4)(b) each edge of G is visited a total of exactly two times by w and x, and so part (ii) of the proposition follows immediately from part (i). We have only to prove part (i). ($\Rightarrow$) The graph $G \setminus e$ obtained by deleting e is by hypothesis a tree. If one of the words/walks w or x fails to visit e, say the former, then by the parity principle w must visit every edge of G an even number of times. But then, due to Proposition 4.12(4)(b), it is impossible for w to visit any edge of G visited by x, which is a contradiction. ($\Leftarrow$) By hypothesis and Proposition 4.12(4)(b) the walk w visits e exactly once, hence some cyclic permutation of $\breve{w}$ is a walk on $G \setminus e$ the set of endpoints of which equals e, hence $G \setminus e$ is connected, and hence e is an edge of Z. □

Lemma 8.5. *Let G be a forest. Let w and x be words both admitting interpretation as walks on G. Assume that jointly w and x visit every edge of G either exactly twice or never. (Necessarily then, both w and x are FK words.) Assume further that w and x have at least one letter in common. Then exactly one of the following conditions holds:*

1. *w and x have acronyms which are either equal or mirror images, but have no letters in common apart from those shared by their acronyms.*
2. *w and x are Wigner words with exactly one letter in common, but this common letter does not appear as the first letter of both words.*

Proof. Since any subgraph of a forest is again a forest, we may assume without loss of generality that

$$G = (V, E) = (V_w \cup V_x, E_w \cup E_x).$$

Since w and x have at least one letter in common, in fact G is a tree. Let w_* and x_* be the acronyms of w and x, respectively. Note that w_* (resp., x_*) is the unique geodesic in G with the same initial and terminal vertices as w (resp., x). By the parity principle and the hypotheses we have

$$\begin{aligned} & \{e \in E | w \text{ visits } e \text{ exactly once}\} \\ = & \{e \in E | w_* \text{ visits } e\} = E_w \cap E_x = \{e \in E | x_* \text{ visits } e\} \\ = & \{e \in E | x \text{ visits } e \text{ exactly once}\}, \end{aligned}$$

hence w_* and x_* are words of the same length, say ℓ, and we have

$$\ell = 1 + \#E_w \cap E_x.$$

If $\ell > 1$, then the words w_* and x_* must either be equal or mirror images of each other. If $\ell = 1$, then w and x are Wigner words since each visits every edge of G either exactly twice or never, but note that we need not in this case have equality of w_* and x_*. Finally, since G, G_w and G_x are trees, we have

$$\#V = 1 + \#E = 1 + \#E_w + \#E_x - \#E_w \cap E_x = \#V_w + \#V_x - \ell,$$

which finishes the proof. □

Proposition 8.6. *Fix closed words w and x each of length ≥ 2. Put $k = \ell(\breve{w})$ and $\ell = \ell(\breve{x})$. Let σ (resp., τ) be a cyclic permutation of $\{1, \ldots, k\}$ (resp., $\{1, \ldots, \ell\}$). The following statements are equivalent:*

1. *$[w, x]$ is a CLT word-pair of which (σ, τ) is a polarization.*
2. *$\breve{w}^\sigma$ and $\breve{x}^\tau$ are FK words with acronyms either equal or mirror images, and with no letters in common apart from those shared by their acronyms.*

We remark that under the equivalent conditions above, the common length of the acronyms of $\breve{w}^\sigma$ and $\breve{x}^\tau$ equals the circuit length of $[w,x]$.

Proof. The implication 2⇒1 is easy to check. We omit the details. We turn directly to the proof of the implication 1⇒2. Write $w=[\alpha_i]_{i=1}^{k+1}$ and $x=[\beta_j]_{j=1}^{\ell+1}$. Let e be the last edge of $G=G_{[w,x]}$ visited by the walks $\breve{w}^\sigma\alpha_{\sigma(1)}$ and $\breve{x}^\tau\beta_{\tau(1)}$. Note that e by Lemma 8.4 is automatically an edge of the bracelet of $[w,x]$. Unless $r=2$, let G' be the graph obtained by deleting e from G, but if $r=2$ put $G'=G$. Then in all cases G' is a tree, and the words $\breve{w}^\sigma$ and $\breve{x}^\tau$ are walks on G' satisfying the hypotheses of Lemma 8.5. Were $\breve{w}^\sigma$ and $\breve{x}^\tau$ to be Wigner words with exactly one letter in common not appearing as the first letter of both words, the graph G would have two degenerate edges, which by Proposition 4.12 is impossible. □

8.7. **Enumeration of CLT word-pairs by Wigner words.** We are now ready to state an enumeration formula for CLT word-pairs similar to the enumeration formulas (12) and (13), albeit rather more complicated.

8.7.1. *Enumerative apparatus.* Let $[\gamma_i]_{i=1}^\infty$ be a sequence of distinct letters. For each positive integer i choose cross sections U_i and V_i of the set of Wigner words. Make these choices so as to achieve the following conditions:

- For all i, every word belonging to $U_i\cup V_i$ begins with γ_i, but no word belonging to U_i has a letter other than γ_i in common with any word belonging to V_i.
- For all distinct i and j, every word belonging to $U_i\cup V_i$ is disjoint from every word belonging to $U_j\cup V_j$.

Let φ be a real-valued function defined for all sentences. Assume that $\varphi(a)$ depends only on the equivalence class of a and vanishes when the sum of the lengths of the constituent words of a is sufficiently large, in which case the support of φ consists of only finitely many equivalence classes of sentences.

8.7.2. *Enumeration of CLT word-pairs.* We have

$$\sum_a \varphi(a)$$

$$= \sum_{r=1}^\infty \sum_{u_1\in U_1}\cdots\sum_{u_r\in U_r}\sum_{v_1\in V_1}\cdots\sum_{v_r\in V_r}\sum_\sigma\sum_\tau \tag{32}$$

$$\left\{\begin{array}{ll} \varphi([u^\sigma\alpha_{\sigma(1)},v^\tau\beta_{\tau(1)}]) & \text{if } r=1,\\ \left(\varphi([u^\sigma\alpha_{\sigma(1)},v^\tau\beta_{\tau(1)}])+\varphi([u^\sigma\alpha_{\sigma(1)},\bar{v}^\tau\bar{\beta}_{\tau(1)}])\right)/4 & \text{if } r=2,\\ \left(\varphi([u^\sigma\alpha_{\sigma(1)},v^\tau\beta_{\tau(1)}])+\varphi([u^\sigma\alpha_{\sigma(1)},\bar{v}^\tau\bar{\beta}_{\tau(1)}])\right)/r & \text{if } r\geq 3,\end{array}\right.$$

where:

- a ranges over any cross-section of the set of CLT word-pairs;
- $u=u_1\cdots u_r=[\alpha_i]_{i=1}^{\ell(u)}$;
- $v=v_1\cdots v_r=[\beta_i]_{i=1}^{\ell(v)}$ and $\bar{v}=v_r\cdots v_1=[\bar{\beta}_i]_{i=1}^{\ell(v)}$;
- σ ranges over cyclic permutations of $\{1,\ldots,\ell(u)\}$; and
- τ ranges over cyclic permutations of $\{1,\ldots,\ell(v)\}$.

One verifies that there is neither under- nor over-counting by applying Proposition 7.3 (which gives the structure of FK words) and Proposition 8.6 (which gives the structure of CLT word-pairs) in a straightforward way. We omit further details.

9. Proof of Theorem 3.3

9.1. **Further generating functions.** Fix a sentence

$$a = [w_i]_{i=1}^n = [[\alpha_{ij}]_{j=1}^{\ell(w_i)}]_{i=1}^n$$

consisting of n words.

9.1.1. Let $t = [t_i]_{i=1}^n$ be an n-tuple of independent (algebraic) variables and put

$$H(a,t) = \sum_p H_p(a) \prod_{i=1}^n t_i^{p_i+\ell(w_i)}$$

where $p = [p_i]_{i=1}^n$ ranges over n-tuples of (nonnegative) integers. We view $H(a,t)$ as a formal power series in $t_1, \ldots, t_n$ with random variable coefficients, not as an analytic function of t. In other words, $H(a,t)$ is just a device for manipulating the infinite array $[H_p(a)]$ of random variables. We write

$$MH(a,t) = M(a)H(a,t), \quad \overline{M}H(a,t) = \overline{M}(a)H(a,t)$$

in order to abbreviate notation.

9.1.2. Unraveling the definition of $H(\cdot,\cdot)$ in the case of a single word $w = [\alpha_j]_{j=1}^{\ell(w)}$, we find that

$$H(w,t) = t^{\ell(w)} \sum_\pi \prod_{j=1}^{\ell(w)} [D(\kappa(\alpha_j))t]^{\pi_j} = \prod_{j=1}^{\ell(w)} \frac{t}{1 - tD(\kappa(\alpha_j))}\,, \tag{33}$$

where $\pi = [\pi_j]_{j=1}^{\ell(w)}$ ranges over $\ell(w)$-tuples of nonnegative integers. From (33), it follows that

$$\begin{aligned} H(w\alpha_1, t) &= \left(\frac{t}{1 - tD(\kappa(\alpha_1))}\right)^2 \cdot \prod_{j=2}^{\ell(w)} \frac{t}{1 - tD(\kappa(\alpha_j))} \\ &= \left(t^2 \frac{d}{dt} \frac{t}{1 - tD(\kappa(\alpha_1))}\right) \cdot \prod_{j=2}^{\ell(w)} \frac{t}{1 - tD(\kappa(\alpha_j))}. \end{aligned}$$

Taking the sum over all cyclic permutations σ of $\{1, \ldots, \ell(w)\}$, and arguing similarly, we find that

$$\sum_\sigma H(w^\sigma \alpha_{\sigma(1)}, t) = t^2 \frac{\partial}{\partial t} H(w,t)\,. \tag{34}$$

9.1.3. Returning now to the general situation, from (33) we get the identity

$$H(a,t) = \prod_{i=1}^{n} \prod_{j=1}^{\ell(w_i)} \frac{t_i}{1 - D(\kappa(\alpha_{ij}))t_i} = \prod_{i=1}^{n} H(w_i, t_i)\,. \tag{35}$$

From (34) and (35) we get the differentiation formula

$$t_1^2 \frac{\partial}{\partial t_1} \cdots t_n^2 \frac{\partial}{\partial t_n} H(a,t) = \sum_{\sigma_1} \cdots \sum_{\sigma_n} H([w_i^{\sigma_i} \alpha_{i,\sigma_i(1)}]_{i=1}^n, t), \tag{36}$$

where in the sum σ_i ranges over cyclic permutations of $\{1, \ldots, \ell(a_i)\}$. We emphasize that these identities are to be interpreted formally, i. e., all the expressions are to be expanded as power series in $t_1, \ldots, t_n$ in evident fashion and then coefficients of like monomials in the t_i are to be equated.

9.1.4. For each Wigner word w we define

$$\Phi^{(w)}(c,t) = \sum_{p=0}^{\infty} \Phi^{(w,p)}(c) t^{\ell(w)+p}.$$

As with the generating functions introduced above, this, too, is to be viewed as formal power series in t. By Lemma 6.4 we have

$$E(MH(w,t)|\kappa(\alpha)) = \Phi^{(w)}(\kappa(\alpha), t) \quad \text{a.s.} \tag{37}$$

where to make sense of formula, both sides are expanded in powers of t, the integrals on the left are computed term by term, and then coefficients of like powers of t are to be set equal a.s. By Lemma 6.3 we have

$$\Phi(c,t) = \sum_{w} \Phi^{(w)}(c,t) \tag{38}$$

where w ranges over a cross-section of the set of Wigner words. Note that in the sum on the right, for every fixed degree n, there are only finitely many terms in which the coefficient of t^n is nonvanishing.

Lemma 9.2. *We have an identity*

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} EY_i Y_j \cdot x^i y^j = x \frac{\partial}{\partial x} y \frac{\partial}{\partial y} \Big(2\Theta(x,y) + \Psi(x,y) \Big) \tag{39}$$

of formal power series, where $[Y_i]_{i=1}^{\infty}$ is the Gaussian family defined in Lemma 5.7.

Proof. Let $[\gamma_i, U_i, V_i]_{i=1}^{\infty}$ be the enumerative apparatus introduced in §8.7. In anticipation of applying enumeration formula (32) we temporarily "freeze" data specifying a single term on the right side of that formula:

- Let r be a positive integer.
- Let $u_1 \in U_1, \ldots, u_r \in U_r$ and $v_1 \in V_1, \ldots, v_r \in V_r$.
- Let $u = [\alpha_i]_{i=1}^{\ell(u)} = u_1 \cdots u_r$.
- Let $v = [\beta_j]_{j=1}^{\ell(v)}$ be equal either to $v_1 \cdots v_r$ or to $v_r \cdots v_1$.
- Let σ be a cyclic permutation of $\{1, \ldots, \ell(u)\}$.
- Let τ be a cyclic permutation of $\{1, \ldots, \ell(v)\}$.

By direct appeal to the definitions we have

(40)

$$\overline{M}([u\alpha_1, v\beta_1]) = \overline{M}([u^\sigma \alpha_{\sigma(1)}, v^\tau \beta_{\tau(1)}])$$

$$= \prod_{i=1}^{r} (M(u_i)M(v_i)) \cdot \begin{cases} d^{(2)}(\kappa(\gamma_1)) & \text{if } r = 1, \\ s^{(4)}(\kappa(\gamma_1), \kappa(\gamma_2)) - s^{(2)}(\kappa(\gamma_1), \kappa(\gamma_2))^2 & \text{if } r = 2, \\ K_r(\kappa(\gamma_1), \dots, \kappa(\gamma_r)) & \text{if } r \geq 3. \end{cases}$$

To understand this formula, notice that the right side is a product of factors associated to pendant trees times a factor arising from the bracelet. Put

$$\mathcal{F} = \sigma([\kappa(\gamma_i)]_{i=1}^{\infty}).$$

We then have the following identities:

$$\sum_{\sigma} \sum_{\tau} E(\overline{M}H([u^\sigma \alpha_{\sigma(1)}, v^\tau \beta_{\tau(1)}], [x, y])|\mathcal{F})$$

$$= \sum_{\sigma} \sum_{\tau} E(\overline{M}([u\alpha_1, v\beta_1])H(u^\sigma \alpha_{\sigma(1)}, x)H(v^\tau \beta_{\tau(1)}, y)|\mathcal{F})$$

(41)

$$= x^2 \frac{\partial}{\partial x} y^2 \frac{\partial}{\partial y} E(\overline{M}([u\alpha_1, v\beta_1])H(u, x)H(v, y)|\mathcal{F})$$

$$= x^2 \frac{\partial}{\partial x} \prod_{i=1}^{r} \Phi^{(u_i)}(\kappa(\gamma_i), x) \cdot y^2 \frac{\partial}{\partial y} \prod_{i=1}^{r} \Phi^{(v_i)}(\kappa(\gamma_i), y)$$

$$\cdot \begin{cases} d^{(2)}(\kappa(\gamma_1)) & \text{if } r = 1, \\ s^{(4)}(\kappa(\gamma_1), \kappa(\gamma_2)) - s^{(2)}(\kappa(\gamma_1), \kappa(\gamma_2))^2 & \text{if } r = 2, \\ K_r(\kappa(\gamma_1), \dots, \kappa(\gamma_r)) & \text{if } r \geq 3. \end{cases}$$

Here all the conditional expectations are to be calculated by expanding formally in powers of x and y and then integrating term by term; in the same spirit the equal signs are to be interpreted as a.s. equality term by term between formal power series. The preceding holds at the second equality by the differentiation formula (36), and at the third equality by (37). Now take expectations (again, integrating term by term), and then apply identity (38) and enumeration formula (32) to find that

(42)
$$\sum_{[u,v]} \overline{M}H([u, v], [x, y]) = x^2 \frac{\partial}{\partial x} y^2 \frac{\partial}{\partial y} \left(2\Theta(x, y) + \Psi(x, y) \right)$$

where on the left $[u\ v]$ ranges over a cross-section of the set of CLT word-pairs. The result now follows by definition of the random variables Y_i. □

9.3. **End of the proof of Theorem 3.3.** By limit formula (20), Lemma 5.7 and Lemma 9.2, we have for every nonnegative integer n that

$$\lim_{k \to \infty} EZ_{f,k}^n = EY_f^n$$

where

$$Y_f = \sum_{i=1}^{\infty} \langle t^i, f(t) \rangle Y_i.$$

So the method of moments gives the result. □

10. Proof of Theorem 3.4

Because of the strong similarity between the proofs of Theorem 3.3 and Theorem 3.4, and because of the tedious nature of latter proof (the necessary enumerations are rather involved), we proceed quite rapidly, omitting many details. But we strive to provide all the important “landmarks” so that the reader won't get lost.

10.1. **Further random variables indexed by sentences.** We enlarge the supply of random variables introduced in §5.1, as follows.

10.1.1. Given any finite nonempty set $\mathcal{N}$ of letters, let $\{\kappa_{\mathcal{N}}(\alpha)\}$ be a letter-indexed color-valued family of i.i.d. random variables with common law $\theta_{\mathcal{N}}$. Then, with $\mathcal{N}$ as above, for any word w and integer p we define $M_{\mathcal{N}}(w)$ and $H_{p,\mathcal{N}}(w)$ by repeating the definitions of $M(w)$ and $H_p(w)$, see §5.1.4 and §5.1.6, with $\theta_{\mathcal{N}}$ in place of θ. In fact, these random variables were already considered in the course of the proof of Theorem 3.2, see equation (28).

10.1.2. Given distinct letters α and β, let $[\beta \mapsto \alpha]$ be the unique map of letter space to itself sending β to α but fixing all other letters. Given also a word w, let $[\beta \mapsto \alpha]_* w$ be the word obtained by applying $[\beta \mapsto \alpha]$ letter by letter to w.

10.1.3. Let ψ be any map of letter space to itself. Let $w = [\alpha_i]_{i=1}^n$ be any closed word. Put

$$M(w,\psi) = \prod_{\substack{e=\{\alpha,\beta\},\\ \text{edge of } G_w}} \left\{ \begin{array}{cl} 0 & \text{if } \nu(e)=1 \\ s^{(\nu(e))}(\kappa(\psi(\alpha)),\kappa(\psi(\beta))) & \text{if } \nu(e)>1 \text{ and } \alpha\neq\beta, \\ d^{(\nu(e))}(\kappa(\psi(\alpha))) & \text{if } \nu(e)>1 \text{ and } \alpha=\beta, \end{array}\right.$$

where $\nu(e)$ is the number of visits made by w to e. Note that if ψ is the identity map, then $M(w,\psi) = M(w)$. The only case of the generalization $M(\cdot,\cdot)$ of $M(\cdot)$ figuring in our limit formulas is that in which w is a Wigner word and $\psi = [\beta \mapsto \alpha]$ for some distinct letters α and β appearing in w. Note that in that case $M(w,\psi)$ depends only on $s^{(2)}$, not on $\{s^{(k)}\}_{k\neq 2} \cup \{d^{(k)}\}$.

10.2. **Approximation of $\langle \overline{L}(\mathcal{N}), x^n\rangle$ at CLT scale.** Fix a positive integer n. Let $[\mathcal{N}_k]_{k=1}^\infty$ be as in Assumption 3.1. Starting again with formula (17), it is possible to obtain the formula

$$\lim_{k\to\infty} N_k \cdot \left(\langle \overline{L}(\mathcal{N}_k), x^n\rangle - \sum_w EM_{\mathcal{N}_k} H_{n+1-\ell(w),\mathcal{N}_k}(w)\right)$$

$$= -\frac{1}{2}\sum_{[u,\alpha,\beta]} EM(u,[\beta\mapsto\alpha])H_{n+1-\ell(u)}([\beta\mapsto\alpha]_* u) \tag{43}$$

$$+\sum_v EMH_{n+1-\ell(v)}(v)$$

where:

- w ranges over a cross-section of the set of Wigner words;
- $[u,\alpha,\beta]$ ranges over a cross-section of the set of marked Wigner words; and
- v ranges over a cross-section of the set of critical weak Wigner words.

Note that only finitely many nonzero terms appear in the sums. Since the proof of (43) is quite similar to that of (18), if rather more complicated, we omit the details. We only give the following hint to the reader. Let us return for a moment to the set up of §5.4.2. We have

$$N(S(\mathcal{N},w) - EM_{\mathcal{N}}H_{n+1-\ell(w),\mathcal{N}}(w)) = -N^{1-r} \sum_{\substack{(\beta_1,\dots,\beta_r)\in\mathcal{N}^r \\ \#\{\beta_1,\dots,\beta_r\}<r}} f(\kappa_0(\beta_1),\dots,\kappa_0(\beta_r))$$

and up to an $O(N^{-1})$ error the right side equals

$$-N^{-r} \sum_{1\leq i<j\leq r} \sum_{(\beta_1,\dots,\beta_r)\in\mathcal{N}^r} f(\kappa_0(\beta_{[j\mapsto i](1)}),\dots,\kappa_0(\beta_{[j\mapsto i](r)}))$$

where $[j \mapsto i]$ denotes the map of $\{1,\dots,r\}$ to itself sending j to i and fixing all other elements. In the case that color space consists of a single color, the preceding remark boils down to the observation that

$$N(N-1)\cdots(N-r+1) - N^r = -\begin{pmatrix} r \\ 2 \end{pmatrix} N^{r-1} + \cdots$$

where the omitted terms are $O(N^{r-2})$.

10.3. **Enumeration of marked Wigner words by Wigner words.** We use the enumerative apparatus introduced in §8.7.1. We have
(44)

$$\sum_{[w,\alpha,\beta]} \varphi([w,\alpha,\beta]) = \sum_{r=1}^{\infty} \sum_{u_1\in U_1} \cdots \sum_{u_r\in U_r} \sum_{v_1\in V_1} \cdots \sum_{v_r\in V_r} \sum_{\sigma} \varphi([u^\sigma \alpha_{\sigma(1)}, \gamma_1, \gamma_{r+1}])$$

where:

- $[w,\alpha,\beta]$ ranges over any cross-section of the set of marked Wigner words;
- $u = u_1 \cdots u_r([\gamma_1 \mapsto \gamma_{r+1}]_* v_1) v_r \cdots v_2 = [\alpha_i]_{i=1}^{\ell(u)}$ and
- σ ranges over cyclic permutations of $\{1,\dots,\ell(u)\}$.

Note that in this setting

$$M(u^\sigma \alpha_{\sigma(1)}, [\gamma_{r+1} \mapsto \gamma_1]) = \prod_{i=1}^{r} (M(u_i)M(v_i)) \cdot K_r(\kappa(\gamma_1),\dots,\kappa(\gamma_r)). \tag{45}$$

The intuition behind (44) is as follows. Let $[w,\alpha,\beta]$ be a marked Wigner word, write $w = [\alpha_i]_{i=1}^{\ell(w)}$, and let $\breve{w}$ be the result of dropping the last letter of w. After replacing w by $\breve{w}^\sigma \alpha_{\sigma(1)}$ for a certain uniquely determined cyclic permutation σ of $\{1,\dots,\ell(\breve{w})\}$, we may assume that α is the first letter of w and that every appearance of α in $\breve{w}$ precedes every appearance of β. We may then view w as a walk out and back on the geodesic connecting α to β in the tree G_w punctuated by sidetrips on the trees hanging from that geodesic. More precisely, an argument employing Proposition 4.5 (which gives the structure of Wigner words), Proposition 7.3 (which gives the structure of FK words) and Lemma 8.5 shows that there is neither under- nor over-counting in (44). We omit the details.

10.4. **The bracelet of a critical weak Wigner word.** Let $w = [\alpha_i]_{i=1}^{\ell(w)}$ be a critical weak Wigner word. Put $G = (V,E) = G_w = (V_w, E_w)$.

10.4.1. According to Proposition 4.8, either G is unicyclic or G is a tree. If G is unicyclic, then we define the *bracelet* and *circuit length* of w to be the same as defined for G in Proposition 8.2. If G is a tree, then there exists a unique edge e of G visited exactly 4 times by w; this edge and attached vertices we declare to be the *bracelet* of w, and we declare the *circuit length* of w to be that of its bracelet, namely 2.

10.4.2. Let Z and r be the bracelet and circuit length of w, respectively. Note that $r \neq 2$ or $r = 2$ according to whether G is unicyclic or a tree. Note that in all cases the graph obtained from G by deleting the edges of Z is a forest with exactly r connected components each of which meets the bracelet in exactly one vertex; again we have a picture of "bracelet with pendant trees". Note that in all cases w makes a total of $2r$ visits to edges of Z. Moreover the walk w visits each edge of the bracelet exactly twice, unless $r = 2$, in which case w visits the unique edge of the bracelet exactly 4 times.

10.4.3. As in §8.3.3, let $\breve{w}$ be the result of dropping the last letter of w and let σ be a cyclic permutation of $\{1, \ldots, \ell(\breve{w})\}$. Then $\breve{w}^\sigma \alpha_{\sigma(1)}$ is a critical weak Wigner word with graph, bracelet and circuit length the same as for w. We say that σ is a *polarization* of w if the last edge of G visited by the walk $\breve{w}^\sigma \alpha_{\sigma(1)}$ is an edge of Z. Clearly:

- There exist exactly $2r$ polarizations of w.

Note that the set of polarizations of w depends only on the equivalence class of w.

10.4.4. Suppose now that we are given a polarization σ of w. We define the *canonical decomposition*

$$\breve{w}^\sigma = p_1 p_2 \cdots p_{2r-1} p_{2r}$$

associated to σ to be the unique decomposition with breaks at visits of the walk $\breve{w}^\sigma$ to edges of the bracelet. From the bracelet-and-pendant-trees picture it is not difficult to deduce that each p_i is a Wigner word and that no two of the p_i have letters in common with the exception that first letters may coincide. Let $s = \alpha_1 \cdots \alpha_{2r}$ be the sequence of first letters of the p_i. We call s the *signature* associated to the critical weak Wigner word w and its polarization σ. Necessarily $s\alpha_1$ is a walk on the bracelet of w visiting every edge of the bracelet exactly twice unless $r = 2$, in which case $s\alpha_1$ visits the unique edge of the bracelet exactly 4 times. Up to equivalence of words there are very few possibilities for s. In fact, the following possibilities are mutually exclusive and exhaustive:

- $r \geq 3$ and $s \sim 123 \cdots r123 \cdots r$.
- $r = 1$ and $s \sim 11$.
- $r = 2$ and $s \sim 1212$.
- $r \geq 3$ and $s^\tau \sim 123 \cdots r1r \cdots 2$ for some cyclic permutation τ of $\{1, \ldots, 2r\}$.

In the first case we say that the signature is *unidirectional*, whereas in the remaining cases we say that the signature is *backtracking*. Notice that if s is unidirectional (resp., backtracking) for some polarization σ, then s is unidirectional (resp., backtracking) for all polarizations σ. Thus it makes sense to say that w itself is either unidirectional or backtracking.

10.4.5. If w is backtracking, then for some polarization σ the associated signature is of the form 11 if $r=1$, 1212 if $r=2$, or $123\cdots r1r\cdots 2$ if $r\geq 3$, in which case we say that σ is a *strong polarization* of w. It is not difficult to verify that:

- If w is backtracking, there exist exactly 2 strong polarizations of w unless $r=2$, in which case every polarization is strong (and so there exist exactly 4 strong polarizations).

Note that the set of strong polarizations of w depends only on the equivalence class of w.

10.5. **Enumeration of critical weak Wigner words by Wigner words.** We again use the enumerative apparatus introduced in §8.7.1. We have

$$\begin{aligned}&\sum_w \varphi(w)\\ (46)\qquad =\ &\sum_{r=1}^{\infty}\sum_{u_1\in U_1}\cdots\sum_{u_r\in U_r}\sum_{v_1\in V_1}\cdots\sum_{v_r\in V_r}\sum_{\sigma}\varphi(u^\sigma\alpha_{\sigma(1)})\Big/\left\{\begin{array}{ll}4 & \text{if } r=2\\ 2 & \text{if } r\neq 2\end{array}\right.\\ &+\sum_{r=3}^{\infty}\sum_{x_1\in U_1}\cdots\sum_{x_r\in U_r}\sum_{y_1\in V_1}\cdots\sum_{y_r\in V_r}\sum_{\tau}\varphi(v^\tau\beta_{\tau(1)})/2r\end{aligned}$$

where:

- w ranges over any cross-section of the set of critical weak Wigner words;
- $u=u_1\cdots u_r v_1 v_r\cdots v_2=[\alpha_i]_{i=1}^{\ell(u)}$;
- σ ranges over cyclic permutations of $\{1,\dots,\ell(u)\}$;
- $v=x_1\cdots x_r y_1\cdots y_r=[\beta_i]_{i=1}^{\ell(v)}$; and
- τ ranges over cyclic permutations of $\{1,\dots,\ell(v)\}$.

In this setting we have

$$(47)\qquad M(u^\sigma\alpha_{\sigma(1)})=\prod_{i=1}^{r}(M(u_i)M(v_i))\cdot\left\{\begin{array}{ll}d^{(2)}(\kappa(\gamma_1)) & \text{if } r=1,\\ s^{(4)}(\kappa(\gamma_1),\kappa(\gamma_2)) & \text{if } r=2,\\ K_r(\kappa(\gamma_1),\dots,\kappa(\gamma_r)) & \text{if } r\geq 3,\end{array}\right.$$

and we have an analogous expression for $M(v^\tau\beta_{\tau(1)})$. Formula (46) may be derived from the preceding discussion of the bracelet of a critical weak Wigner word in a straightforward way. We omit the details.

10.6. **End of the proof.** The left sides of (9) and (43) coincide by formula (28) coming up in the proof of Theorem 3.2. So we can rewrite (43) as an identity of formal power series

$$\begin{aligned}&\sum_{n=1}^{\infty}\left(\lim_{k\to\infty}N_k\cdot\left(\langle\overline{L}(\mathcal{N}_k),x^n\rangle-\langle\mu_{\mathcal{N}_k},x^n\rangle\right)\right)t^{n+1}\\ (48)\qquad =\ &-\frac{1}{2}\sum_{[w,\alpha,\beta]}EM(w,[\beta\mapsto\alpha])H([\beta\mapsto\alpha]_*w,t)+\sum_u EMH(u,t)\end{aligned}$$

where:

- $[w,\alpha,\beta]$ ranges over a cross-section of the set of marked Wigner words; and
- u ranges over a cross-section of the set of critical weak Wigner words.

To finish the proof of the theorem we have just to make the right side of (48) explicit. This can be done by exploiting (44), (45), (46), and (47). Note that many of the terms in the sum on $[w, \alpha, \beta]$ are cancelled by terms in the sum on u due to the parallel structure of formulas (45) and (47). We omit the remaining details of the proof because the calculations are very similar to those undertaken to prove Lemma 9.2. The proof of Theorem 3.4 is complete. □

11. Concentration

In this section we work out sufficient conditions allowing one to prove a CLT for test functions more general than polynomials. Toward this end, we define for random matrices a notion of concentration and a notion of CLT for polynomial test functions. Then, assuming concentration, a polynomial-type CLT, and a further condition on the limiting covariance for polynomial test functions, we prove a CLT for continuously differentiable test functions with polynomial growth (Proposition 11.6). Furthermore, we establish the concentration property for the matrices $X(\mathcal{N}_k)$ studied in Theorems 3.2 and 3.3 when the random variables $\xi_{\{\alpha,\beta\}}$ satisfy the Poincaré inequality with the same constant (Proposition 11.8). The main result of this section (Theorem 11.10) summarizes the preceding considerations in a fashion convenient for applications in §12.

11.1. **The concentration property.** Throughout this section $\{Y_k\}_{k=1}^{\infty}$ denotes a sequence of random symmetric matrices. For such a general sequence we are going to define and study a concentration property. Eventually we are going to take $Y_k = X(\mathcal{N}_k)$, but in anticipation of applications of the concentration idea beyond the scope of this paper, we work in a general setting until the end of the proof of Proposition 11.6. For any Lipshitz function $g : \mathbb{R}^n \to \mathbb{R}$ set

$$\|g\|_{\mathrm{Lip}} := \sup_{\substack{x,y\in\mathbb{R}^n \\ x\neq y}} \frac{|g(x) - g(y)|}{|x - y|}\,, \tag{49}$$

where $|x - y|$ is the Euclidean distance between x and y.

Definition 11.2. *We say that the sequence of matrices* $\{Y_k\}_{k=1}^{\infty}$ *satisfies the* concentration property *under the following conditions:*

(50) *There exists a constant* $c > 0$ *such that for any Lipschitz function* $g : \mathbb{R} \to \mathbb{R}$, *it holds that* $\sup_k \mathrm{Var}\,\mathrm{tr}\, g(Y_k) \le c\|g\|^2_{\mathrm{Lip}}$.

(51) *There exists a compact set* $S \subset \mathbb{R}$ *such that for any function* $f : \mathbb{R} \to \mathbb{R}$ *supported in* S^c *of polynomial growth, it holds that* $E\left([\mathrm{tr}\, f(Y_k)]^2\right) \to_{k\to\infty} 0$.

The next lemma deduces from (50) and (51) a single statement convenient for applications:

Lemma 11.3. *Suppose that* $\{Y_k\}_{k=1}^{\infty}$ *satisfies the concentration property. Then there exists a constant* $\bar{c} > 0$ *and a compact interval* T *such that for any function* f *continuously differentiable on* T *and of polynomial growth one has*

$$\limsup_{k\to\infty} \mathrm{Var}\,\mathrm{tr}\, f(Y_k) \le \bar{c} \sup_{x\in T} |f'(x)|^2.$$

Proof. Let S be as in (51). Choose a compact interval I with interior containing the set S, and then choose a compact interval T with interior containing I. Let $g : \mathbb{R} \to [0, 1]$ be a continuously differentiable function identically equal to 1 on I and identically vanishing in the complement of T. Let ℓ be the length of T. Without loss of generality we may assume that f vanishes at some point of T. Then

$$\|fg\|_{\mathrm{Lip}} \leq \left(1 + \ell \sup_{t\in T} |g'(t)|\right) \sup_{t\in T} |f'(t)|, \quad \operatorname{supp} f(1-g) \subset S^c,$$

and

$$[\operatorname{Var} \operatorname{tr} f(Y_k)]^{1/2} \leq [\operatorname{Var} \operatorname{tr}(fg)(Y_k)]^{1/2} + (E[\operatorname{tr}(f(1-g))(Y_k)]^2)^{1/2},$$

whence the result by definition of the concentration property. □

11.4. **CLT's for differentiable test functions.** Our goal is to prove under suitable hypotheses a central limit theorem for random variables of the form

$$Z_{f,k} := \operatorname{tr} f(Y_k) - E \operatorname{tr} f(Y_k)$$

where f is continuously differentiable on a large enough compact set and of polynomial growth.

Definition 11.5. *We say that the sequence $\{Y_k\}_{k=1}^{\infty}$ satisfies a* polynomial-type CLT *if there exists a mean zero Gaussian family $\{W_n\}_{n=0}^{\infty}$ of random variables such that for every polynomial function $f(x) = \sum_{i=0}^{m} a_i x^i$ it holds that $Z_{f,k}$ converges in distribution as $k \to \infty$ to $W_f := \sum_{i=0}^{m} a_i W_i$.*

The next proposition gives hypotheses under which one can extend a CLT statement from polynomial test functions to differentiable test functions of polynomial growth. After proving the proposition, verification of its hypotheses for $Y_k = X(\mathcal{N}_k)$ under the assumptions of Theorems 3.2 and 3.3, along with further structural assumptions concerning the functions $d^{(2)}$, $s^{(2)}$ and $s^{(4)}$ will be our task for the rest of the paper.

Proposition 11.6. *Assume that the sequence of matrices $\{Y_k\}_{k=1}^{\infty}$ satisfies both the concentration property and a polynomial-type CLT. Assume further the existence of a sequence $\{q_n\}_{n=1}^{\infty}$ of polynomial functions with the following properties:*

- *For some compactly supported finite measure ν on $\mathbb{R}$ the sequence $\{q_n\}_{n=1}^{\infty}$ is an orthonormal system in $L^2(\nu)$.*
- *Every polynomial in x is a finite linear combination of the $q_n(x)$.*
- *With $\bar{q}_n(x) := \int_0^x q_n(y)dy$, the covariance matrix $K(m, n) := EW_{\bar{q}_m} W_{\bar{q}_n}$ of the mean zero Gaussian family $\{W_{\bar{q}_n}\}_{n=1}^{\infty}$ is diagonal.*

Fix T and $\bar{c}$ as in Lemma 11.3, with $T \supset \operatorname{supp} \nu$. Then, for any function f of polynomial growth which is continuously differentiable on T, the random variables $Z_{f,k}$ converge in distribution to a mean zero Gaussian random variable Z_f with variance

$$EZ_f^2 = \|f'\|_K^2 \leq \bar{c} \sup_{t\in T} |f'(t)|^2, \tag{52}$$

where for any function h continuous on T we set

$$\|h\|_K^2 := \sum_{n=1}^{\infty} K(n, n) \langle \nu, hq_n \rangle^2.$$

Proof. Consider at first the case in which f is a polynomial. The polynomial-type CLT implies that the variables $Z_{f,k}$ converge in distribution to W_f, and since f differs by a constant from a finite linear combination of the $\bar{q}_n$, the variance EW_f^2 takes the value asserted in (52), namely $\|f'\|_K^2$. Furthermore, by Lemma 11.3 and the Fatou Lemma, the estimate for $\|f'\|_K^2$ asserted in (52) holds. Thus all assertions are proved if f is a polynomial function.

We turn to consideration of the general case. Let $\{Q_m\}_{m=1}^\infty$ be a sequence of polynomials tending uniformly on T to f' (such is provided by the Stone-Weierstrass theorem) and put $\overline{Q}_m(x) := \int_0^x Q_m(y)dy$. Clearly, the sequence $\{Q_m\}_{m=1}^\infty$ is $\|\cdot\|_{L^2(\nu)}$-Cauchy. But by (52) the sequence $\{Q_m\}_{m=1}^\infty$ is also $\|\cdot\|_K$-Cauchy. A dominated convergence argument now shows that $\|f'\|_K^2 = \lim_{m\to\infty} \|Q_m\|_K^2$. It follows that the estimate for $\|f'\|_K^2$ asserted in (52) holds. By Lemma 11.3 the family of random variables $Z_{f,k}$ is tight; let Y be any subsequential limit-in-distribution. For any $t \in \mathbb{R}$ one has

$$|Ee^{itY} - Ee^{itW_{\overline{Q}_m}}| \le \limsup_{k\to\infty} E|e^{itZ_{f-\overline{Q}_m,k}} - 1| \le |t| \limsup_{k\to\infty} (EZ^2_{f-\overline{Q}_m,k})^{1/2}.$$

The quantity on the right by Lemma 11.3 tends to 0 as $m \to \infty$, and clearly

$$Ee^{itW_{\overline{Q}_m}} = e^{-t^2\|Q_m\|_K^2/2} \to_{m\to\infty} e^{-t^2\|f'\|_K^2/2}.$$

Therefore (the characteristic function of) Y is (that of) a mean zero Gaussian random variable of variance $\|f'\|_K^2$. Since all subsequential limits are the same we get convergence-in-distribution of $Z_{f,k}$ to a mean zero Gaussian random variable of variance $\|f'\|_K^2$. All assertions have been proved. □

11.7. **Poincaré inequalities for matrices.** We say that a probability distribution η on $\mathbb{R}$ satisfies a *Poincaré inequality* if there exists a constant c_η such that for any $f : \mathbb{R} \to \mathbb{R}$ smooth, it holds that

$$\mathrm{Var}_\eta(f) := \int \left(f(x) - \int f(x)\eta(dx)\right)^2 \eta(dx) \le c_\eta \int |f'(x)|^2 \eta(dx)\,.$$

For such a distribution η one has

$$E\exp\left(\frac{|Y - EY|}{12\sqrt{c_\eta}}\right) \le 2 \quad (Y:\ \text{random variable with law } \eta), \tag{53}$$

see [BU83, Theorem 2] (or [Bo99] for optimal constants).

It is well known (see, e.g., [Le01, Pg. 49]) that if $\eta_i, i = 1, \ldots, K$ satisfy Poincaré inequalities with constants c_{η_i}, then for any smooth function $g : \mathbb{R}^K \to \mathbb{R}$, and with $\eta = \otimes_{i=1}^K \eta_i$ and $c_\eta = \max_{i=1}^K c_{\eta_i}$, one has

$$\mathrm{Var}_\eta(g) =: \int \left(g(x) - \int g(x)\eta(dx)\right)^2 \eta(dx) \le c_\eta \int |\nabla g(x)|^2 \eta(dx)\,. \tag{54}$$

We recall (see e.g. [GZ00, Lemma 1.2]) that if $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz with Lipschitz constant $\|f\|_{\mathrm{Lip}}$, then the function $f_N : \mathbb{R}^{N(N+1)/2} \to \mathbb{R}$ on N-by-N symmetric matrices given by $f_N(X) = \mathrm{tr}\, f(X)$ is Lipschitz with Lipschitz constant $\|f_N\|_{\mathrm{Lip}} \le \sqrt{N}\|f\|_{\mathrm{Lip}}$. It follows that if X is an N-by-N symmetric random matrix with on-or-above-diagonal entries independent and satisfying the Poincaré inequality with the same constant c/N, then for any Lipshitz $f : \mathbb{R} \to \mathbb{R}$ one has

$$\mathrm{Var}\,\mathrm{tr}\, f(X) \le c\|f\|_{\mathrm{Lip}}^2\,.$$

See [CB04] for a systematic use of this fact, and [GZ00] for other concentration inequalities for random matrices. In particular, in the setting of Theorems 3.2 and 3.3, if the random variables $\xi_{\{\alpha,\beta\}}$ satisfy the Poincaré inequality with the same constant c, we see that (50) above holds true, with $Y_k = X(\mathcal{N}_k)$.

Proposition 11.8. *In the setting and under the hypotheses of Theorems 3.2 and 3.3, suppose that the random variables $\xi_{\{\alpha,\beta\}}$ satisfy the Poincaré inequality with the same constant c. Then the sequence $\{X(\mathcal{N}_k)\}_{k=1}^{\infty}$ has the concentration property.*

Before proving the proposition, we state an auxiliary estimate, which may be of interest in its own right. We get the estimate by combining the ideas of [FK81] as summarized in Proposition 7.9 above with the moment bound (53). We remark in passing that under somewhat stronger assumptions, a considerably stronger assertion could be obtained by the methods of [SS98].

Lemma 11.9. *Under the assumptions of Proposition 11.8, there exist constants $C > 0$ and $\epsilon > 0$ such that with $r(N) := \lfloor N^{\epsilon} \rfloor$ one has*

$$\frac{1}{N} E \operatorname{tr} X(\mathcal{N}_k)^{2r(N_k)} \leq C^{2r(N_k)} \tag{55}$$

for all sufficiently large k.

Proof. By an obvious rescaling, we may assume without loss of generality that $C(2) = 1$, where $C(2)$ is as defined in (31). We may further assume, so we claim, that $D = 0$. To see that this is so, set $(\overline{X}(\mathcal{N}_k))_{\alpha\beta} = N_k^{-1/2}\xi_{\{\alpha,\beta\}}$ and suppose that the lemma holds with $\overline{X}(\mathcal{N}_k)$ in place of $X(\mathcal{N}_k)$. Then

$$\begin{aligned}
E \operatorname{tr} X(\mathcal{N}_k)^{2r(N_k)} &\leq N_k|\lambda|_{\max}(X(\mathcal{N}_k))^{2r(N_k)} \leq N_k[|\lambda|_{\max}(\overline{X}(\mathcal{N}_k)) + |D|_\infty]^{2r(N_k)} \\
&\leq 2^{2r(N_k)-1}N_k[|\lambda|_{\max}(\overline{X}(\mathcal{N}_k))^{2r(N_k)} + |D|_\infty^{2r(N_k)}] \\
&\leq 2^{2r(N_k)-1}N_k[E \operatorname{tr} \overline{X}(\mathcal{N}_k)^{2r(N_k)} + |D|_\infty^{2r(N_k)}] \\
&\leq 2^{2r(N_k)-1}N_k^2[C^{2r(N_k)} + |D|_\infty^{2r(N_k)}] \leq (3(C + |D|_\infty))^{2r(N_k)},
\end{aligned}$$

for all k large enough. The claim is proved. We assume for the rest of the proof that $D = 0$. By (17), for any positive integer n,

$$\langle \overline{L}(\mathcal{N}_k), x^{2n} \rangle \leq \sum_{q=1}^{n+1} N_{\mathrm{FK}}(2n,q) N_k^{q-(n+1)} \max_{b \in \mathrm{FK}(2n,q)} E|\xi(b)| \tag{56}$$

with $\mathrm{FK}(2n,q)$ denoting the collection of weak Wigner words of length $2n+1$ and weight q, and N_{FK} as in (30). Note that

$$\begin{aligned}
&\max_{b \in \mathrm{FK}(2n,q)} E|\xi(b)| \leq C(3(2n+2-2q)) \\
\leq\ & \sup_{\alpha,\beta} E\left(\exp\left(|\xi(\alpha,\beta)|/12\sqrt{c}\right)\right)(1 \vee (12\sqrt{c}))^{3(2n+2-2q)}[3(2n+2-2q)]! \\
\leq\ & 2(1 \vee (12\sqrt{c}))^{3(2n+2-2q)}[3(2n+2-2q)]! =: 2C_0^{3(2n+2-2q)}[3(2n+2-2q)]!,
\end{aligned}$$

where the first inequality is due to (31) and the second to (53). Thus

$$\begin{aligned}
\langle \overline{L}(\mathcal{N}_k), x^{2n} \rangle &\leq 2^{n+1} \sum_{q=1}^{n+1} N_k^{q-(n+1)}[3(2n+2-2q)]!(C_0 n)^{3(2n+2-2q)} \\
&\leq 2^{n+1} \sum_{j=0}^{n} N_k^{-j}(6C_0 n)^{12j} \leq 2^{n+2}
\end{aligned}$$

as long as $(6C_0 n)^{12}/N_k \le 1/2$. This completes the proof. $\square$

Proof of Proposition 11.8. In view of Theorem 3.2 and the discussion in §11.7, it only remains to check (51). This is based on Lemma 11.9. Fix C as in the statement of that lemma. Define the compact set $S = [-C-1, C+1]$. Suppose that $|f(x)| \le c_1|x|^{c_2}$ and f is supported on S^c. Then, using that

$$(|x|/(C+1/2))^{r(N_k)} \ge |x|^{2c_2} \quad \text{for } |x| \ge C+1 \text{ and } k \text{ large},$$

one has

$$\begin{aligned}
E\left([\operatorname{tr} f(X(\mathcal{N}_k))]^2\right) &\le N_k E \operatorname{tr} f^2(X(\mathcal{N}_k)) \\
&\le N_k c_1^2 E \sum_{i=1}^{N_k} \lambda_i(\mathcal{N}_k)^{2c_2} \mathbf{1}_{|\lambda_i(\mathcal{N}_k)| \ge (C+1)} \\
&\le N_k c_1^2 E \sum_{i=1}^{N_k} \left(\frac{\lambda_i(\mathcal{N}_k)}{C+1/2}\right)^{r(N_k)} \\
&\le N_k^2 c_1^2 \left(\frac{C}{C+1/2}\right)^{r(N_k)} \to_{k\to\infty} 0\,.
\end{aligned} \tag{57}$$

$\square$

By combining Propositions 11.6 and 11.8, we immediately get the following theorem, which is the main result of this section. Recall that under the assumptions of Theorem 3.3, the sequence $\{X(\mathcal{N}_k)\}_{k=1}^\infty$ satisfies a polynomial-type CLT, i. e., there exists a mean zero Gaussian family $\{W_n\}_{n=0}^\infty$ of random variables such that for every polynomial function $f(x) = \sum_{i=0}^m a_i x^i$ the random variables $\operatorname{tr} f(X(\mathcal{N}_k)) - E \operatorname{tr} f(X(\mathcal{N}_k))$ converge in distribution as $k \to \infty$ to $W_f := \sum_{i=0}^m a_i W_i$.

Theorem 11.10. *We work in the setting and under the hypotheses of Theorems 3.2 and 3.3. We make the following further assumptions:*

- *The random variables $\xi_{\{\alpha,\beta\}}$ satisfy the Poincaré inequality with the same constant c (and hence $\{X(\mathcal{N}_k)\}_{k=1}^\infty$ has the concentration property).*
- *There exists a sequence $\{q_n\}_{n=1}^\infty$ of polynomial functions with the following properties:*
  - *For some compactly supported finite measure ν on $\mathbb{R}$, the sequence $\{q_n\}_{n=1}^\infty$ is an orthonormal system in $L^2(\nu)$.*
  - *Every polynomial in x is a finite linear combination of the $q_n(x)$.*
  - *With $\bar{q}_n(x) := \int_0^x q_n(y)dy$, the covariance matrix $K(m,n) := EW_{\bar{q}_m} W_{\bar{q}_n}$ of the mean zero Gaussian family $\{W_{\bar{q}_n}\}_{n=1}^\infty$ is diagonal.*

Then there exists a compact interval $T \supset \operatorname{supp} \nu$ and a constant $\bar{c} > 0$ such that for any function f of polynomial growth which is continuously differentiable on T, the random variables

$$Z_{f,k} := \operatorname{tr} f(X(\mathcal{N}_k)) - E \operatorname{tr} f(X(\mathcal{N}_k))$$

converge in distribution to a mean zero Gaussian random variable Z_f with variance

$$EZ_f^2 = \sum_{n=1}^\infty K(n,n)\langle \nu, f' q_n\rangle^2 \le \bar{c} \sup_{t\in T} |f'(t)|^2\,. \tag{58}$$

12. Diagonalization by Chebyshev polynomials

We discuss two specializations of the band matrix model in which we can make μ, $\Phi(c,t)$, $\Theta(x,y)$, $\Psi(x,y)$, $\operatorname{Var} Z_f$ and E_f as appearing in Theorems 3.2 and 3.3 much more explicit and moreover apply Theorem 11.10. This will be possible because in these specializations (slight variants of) Chebyshev polynomials diagonalize the covariance matrix of the limiting mean zero Gaussian random variables.

12.1. **Inversion of power series and p-Chebyshev polynomials.** Our computation involves the inversion of formal power series. Fix a sequence of real numbers $\{a_i\}$ and define the formal power series

$$p = p(t) := t + \sum_{i=2}^{\infty} a_i t^i \,. \tag{59}$$

(For the proof of Theorem 3.5 concerning the generalized Wigner matrix model, it will be enough simply to take $p(t) = \Phi(t)$, where $\Phi(t)$ is the generating function for the Catalan numbers defined in (62) below.) For each positive integer n, define the n^{th} *p-Chebyshev polynomial* $T_{n,p}(x)$ as the unique polynomial in x of degree n with real coefficients such that $T_{n,p}(1/t)$ is the principal part of the Laurent series $p(t)^{-n}$. Finally, define the matrix P with rows and columns indexed by the positive integers by setting P_{ij} equal to the coefficient of t^j in p^i, i. e.,

$$P_{ij} := \operatorname{Res}_{t=0}\left(t^{-j} p^i \frac{dt}{t}\right) , \tag{60}$$

where for any sequence $[c_i]_{i=-\infty}^{\infty}$ of constants such that $c_i = 0$ for $i \ll 0$ we set

$$\operatorname{Res}_{t=0} \sum_{i=-\infty}^{\infty} c_i t^i \, dt := c_{-1} \,.$$

Lemma 12.2. *Fix $p(t)$ as in (59) with its associated p-Chebyshev polynomials $T_{n,p}(x)$ and matrix P as in (60). Identify power series in x without constant term in the obvious way with column vectors having entries indexed by the positive integers (thus identifying polynomials in x without constant term with finitely supported infinite column vectors). Then, the n^{th} column of P^{-1} equals $\frac{1}{n} x T'_{n,p}(x)$.*

Proof. Let $r = r(t)$ be the formal power series inverse of $p(t)$, i. e., the unique power series without constant term such that

$$p(r(t)) = r(p(t)) = t.$$

By the *Lagrange inversion formula*, c.f. [St99, §5.4],

$$(P^{-1})_{ij} = \operatorname{Res}_{t=0}\left(t^{-j} r^i \frac{dt}{t}\right) = \frac{i}{j} \operatorname{Res}_{t=0}\left(p^{-j} t^i \frac{dt}{t}\right) .$$

The last expression is by definition exactly the coefficient of x^i in $\frac{1}{j} x T'_{j,p}(x)$. □

12.3. **Chebyshev polynomials.**

12.3.1. *Definition.* For each positive integer n we define the n^{th} *Chebyshev polynomial* $T_n(x)$ of the *first kind* to be the unique polynomial in x such that

$$T_n(z+1/z) = z^n + 1/z^n \quad \text{(equivalently: } T_n(2\cos\theta) = 2\cos n\theta)$$

and we define the n^{th} *Chebyshev polynomial* $U_n(x)$ of the *second kind* by the rule

$$U_n(x) := \frac{1}{n} T_n'(x).$$

We have orthogonality relations

$$\frac{1}{2\pi}\int_{-2}^{2} U_m(x)U_n(x)\sqrt{4-x^2}\,dx = \delta_{mn} \quad (m,n = 1,2,3,\dots) \tag{61}$$

as can be verified directly by the trigonometric substitution $x = 2\cos\theta$. Note that the weight figuring in these orthogonality relations is the semicircle law σ_S of mean 0 and variance 1. These relations say that the family $\{U_n(x)\}_{n=1}^{\infty}$ is the Gram-Schmidt orthogonalization in $L^2(\sigma_S)$ of the family $\{x^{n-1}\}_{n=1}^{\infty}$ of powers of x. We have analogous orthogonality relations for Chebyshev polynomials of the first kind (with a different weight), but these we omit because we have no use for them.

12.3.2. *Warning.* Our definitions are not quite the standard ones. One usually defines $T_n(\cos\theta) = \cos n\theta$ and $U_n(x) = \frac{1}{n+1}T_{n+1}'(x)$. We have rescaled and re-indexed in order to obviate many annoying factors of 2 and shifts of 1. It is also worth pointing out that in our set up the polynomial $xU_n(x)$ is monic of degree n, and moreover even or odd according as n is even or odd.

12.3.3. *Reinterpretation of the Chebyshev polynomials.* Consider the odd power series

$$\Phi(t) = \frac{1-\sqrt{1-4t^2}}{2t} = \sum_{n=0}^{\infty} \frac{1}{n+1}\binom{2n}{n} t^{2n+1} = t + t^3 + 2t^5 + 5t^7 + \cdots \tag{62}$$

having the n^{th} Catalan number as the coefficient of t^{2n+1}. Clearly $\Phi(t)$ satisfies the functional equation

$$1/t = \Phi(t) + 1/\Phi(t)$$

and hence more generally the functional equation

$$T_n(1/t) = \Phi(t)^n + 1/\Phi(t)^n \tag{63}$$

for all positive integers n. In other words, for each positive integer n, the n^{th} Chebyshev polynomial $T_n(x)$ (with its constant term dropped) may be reinterpreted as the n^{th} Φ-Chebyshev polynomial $T_{n,\Phi}(x)$, in the sense of Lemma 12.2.

12.3.4. *Diagonalization formulas.* In Lemma 12.2 let us now take $p(t) = \Phi\left(\frac{t}{1-\gamma t}\right)$ where γ is any real constant. (For the proof of Theorem 3.5 concerning the generalized Wigner matrix model it will be enough to consider just the case $\gamma = 0$.) Note that $T_{n,p}(x)$ and $T_n(x-\gamma)$ differ by a constant and hence $\frac{1}{n}xT_{n,p}'(x) = xU_n(x-\gamma)$. Using the obvious identification between power series in x without constant term and *row* vectors with entries indexed by the positive integers, one may think of $p^i(x)$ as $e_i^T P$, where P is the matrix from (60) and e_i is the (infinite) column vector whose j^{th} entry is δ_{ij}. On the other hand, by the remark following (63), and Lemma 12.2, using the obvious identification between power series in x without constant term and *column* vectors with entries indexed by the positive integers, one can identify $xU_n(x-\gamma)$ with $P^{-1}e_n$. Hence, for any sequence $\{\eta_i\}_{i=1}^\infty$ of real constants,

$$\left\langle \sum_{i=1}^\infty \eta_i \Phi\left(\frac{t}{1-\gamma t}\right)^i, tU_n(t-\gamma)\right\rangle = \eta_n \tag{64}$$

and similarly

$$\left\langle \sum_{i=1}^\infty \eta_i \Phi\left(\frac{x}{1-\gamma x}\right)^i \Phi\left(\frac{y}{1-\gamma y}\right)^i, xU_m(x-\gamma)yU_n(y-\gamma)\right\rangle = \eta_n\delta_{mn}, \tag{65}$$

for all positive integers m and n.

12.4. **First specialization: generalized Wigner matrices.**

12.4.1. *Specialization of the model.* As in Theorem 3.5, assume now that

$$D \equiv 0, \quad \int s^{(2)}(c,c')\theta(dc') \equiv 1.$$

This specialization of the band matrix model we call the *generalized Wigner matrix* model. In the case that $s^{(2)} \equiv 1$ this is more or less the standard Wigner matrix model, whence the terminology.

12.4.2. *Calculation of* $\Phi(c,t)$ *and* μ. In the case at hand $\Phi(c,t)$ must be independent of c, and hence by functional equation (3) we have $\Phi(c,t) = \Phi(t)$ where the latter is as defined in (62). From $\Phi(t)$ we can read the moments of μ. We conclude that μ is the semicircle law σ_S of mean 0 and variance 1.

12.4.3. *Calculation of* $\Theta(x,y)$ *and* $\Psi(x,y)$. Because $\Phi(c,t) = \Phi(t)$, the integrals figuring in the definitions of $\Theta(x,y)$ and $\Psi(x,y)$ greatly simplify. We find that

$$\Theta(x,y) = \sum_{r=1}^\infty \lambda_r \Phi(x)^r\Phi(y)^r, \quad \Psi(x,y) = \sum_{r=1}^\infty \epsilon_r \Phi(x)^r\Phi(y)^r \tag{66}$$

where

$$\lambda_r = \frac{1}{r}\int\cdots\int K_r(c_1,\cdots,c_r)\theta(dc_1)\cdots\theta(dc_r),$$

$$\epsilon_r = \begin{cases} \int (d^{(2)}(c) - 2s^{(2)}(c,c))\theta(dc) & \text{if } r=1, \\ \frac{1}{2}\int\int (s^{(4)}(c_1,c_2) - 3s^{(2)}(c_1,c_2)^2)\theta(dc_1)\theta(dc_2) & \text{if } r=2, \\ 0 & \text{if } r \geq 3. \end{cases}$$

12.4.4. *Calculation of* $\operatorname{Var} Z_f$ *and* E_f. Using the orthogonality relations (61), for any polynomial function f we can write

$$f'(x) = \sum_{n=1}^\infty U_n(x)\, Ef'(S)U_n(S)$$

where S is a random variable with standard semicircular law σ_S. (Only finitely many nonzero terms appear in the sum.) Then, using the diagonalization formulas

(65) (with $\eta_i = \lambda_i$) and (64) (with $\eta_{2i} = \lambda_i$ and $\eta_{2i+1} = 0$), we can in the present specialization of the band matrix model rewrite (8) and (9) in the form

$$\mathrm{Var}\, Z_f = \sum_{r=1}^{\infty}(2\lambda_r + \epsilon_r)(Ef'(S)U_r(S))^2, \tag{67}$$

$$E_f = \sum_{r=1}^{\infty}\frac{1}{2}\left(\lambda_r + \epsilon_r\right)Ef'(S)U_{2r}(S). \tag{68}$$

12.5. **Proof of Theorem 3.5.** In view of Theorem 3.3, the discussion in §12.4 immediately above, and Theorem 11.10, all that remains to be done is to take in the latter Theorem $\nu = \sigma_S$ and $q_n(x) = U_n(x)$. □

12.6. **Second specialization: generalized Wishart matrices.**

12.6.1. *Square-root generalized Wishart matrices.* Assume that color space is decomposed as a disjoint union $A \cup B$ and that

$$0 < \theta(A) \leq \theta(B).$$

Put

$$\alpha = \sqrt{\frac{\theta(B)}{\theta(A)}} \geq 1, \quad \beta = \sqrt{\frac{\theta(A)}{\theta(B)}} \leq 1, \quad \gamma = \alpha + \beta \geq 2.$$

Assume that $s^{(2)}$ and $s^{(4)}$ vanish identically on $(A \times A) \cup (B \times B)$ and that

$$\int_A s^{(2)}(\cdot, c')\theta(dc') = \beta \mathbf{1}_B, \quad \int_B s^{(2)}(\cdot, c')\theta(dc') = \alpha \mathbf{1}_A.$$

Assume that $D \equiv 0$. Assume that $d^{(k)} \equiv 0$ for all $k > 0$. This specialization of the band matrix model we call the *generalized square-root Wishart matrix* model.

12.6.2. *Generalized Wishart matrices.* As we are about to see, the machinery we developed is well-suited to deal with square-root generalized Wishart matrices. In applications, however, one is often interested in a slight variant. Write

$$\mathcal{N} = \mathcal{N}^A \cup \mathcal{N}^B$$

where $\mathcal{N}^A$ (resp., $\mathcal{N}^B$) is the set of letters in $\mathcal{N}$ with color in A (resp., B). As usual let N, N^A and N^B denote the corresponding cardinalities. By re-arranging coordinates, $X(\mathcal{N})$ can be written in the form

$$X(\mathcal{N}) = \begin{bmatrix} 0 & Y(\mathcal{N}) \\ Y^T(\mathcal{N}) & 0 \end{bmatrix},$$

where the matrix $Y(\mathcal{N})$ has rows indexed by $\mathcal{N}^A$ and columns indexed by $\mathcal{N}^B$. In this situation, we call the symmetric random matrices

$$W(\mathcal{N}) = Y(\mathcal{N})Y(\mathcal{N})^T$$

(rows and columns indexed by $\mathcal{N}^A$) *generalized Wishart matrices*, and we are interested in the empirical distribution of their eigenvalues $\{\lambda_i(W(\mathcal{N}))\}_{i=1}^{N^A}$ (all non-negative):

$$L_W(\mathcal{N}) := \frac{1}{N^A}\sum_{i=1}^{N^A}\delta_{\lambda_i(W(\mathcal{N}))}\,.$$

In the case that $s^{(2)}$ is constant on $(A\times B)\cup(B\times A)$, the spectrum of $W(\mathcal{N})$ is the same as that of standard Wishart matrices, whence the terminology. Note that for any function $g(t)$ on $\mathbb{R}_+$ with $g(0)=0$, and setting $\tilde{g}(t)=g(t^2)$, one has

$$N\langle L(\mathcal{N}),\tilde{g}\rangle = \operatorname{tr}\tilde{g}(X(\mathcal{N})) = 2\operatorname{tr} g(W(\mathcal{N})) = 2N^A\langle L_W(\mathcal{N}),g\rangle. \tag{69}$$

Hence, once results (either LLN or CLT) are derived for $X(\mathcal{N})$, it is a simple exercise in book-keeping to transform them to statements about $W(\mathcal{N})$.

12.6.3. *Calculation of* $\Phi(c,t)$. Under our additional assumptions in §12.6.1, we can write

$$\Phi(\cdot,t) = \mathbf{1}_A\Phi_A(t)+\mathbf{1}_B\Phi_B(t)$$

where $\Phi_A(t)$ and $\Phi_B(t)$ are color-independent. Functional equation (3) in the case at hand specializes to the functional equation

$$\mathbf{1}_A\Phi_A(t)+\mathbf{1}_B\Phi_B(t) = t\left(1-\beta\mathbf{1}_B\Phi_A(t)-\alpha\mathbf{1}_A\Phi_B(t)\right)^{-1},$$

which in turn can be rewritten as the pair of functional equations

$$\Phi_A(t)=t(1-t\alpha\Phi_B(t))^{-1},\quad \Phi_B(t)=t(1-t\beta\Phi_A(t))^{-1}.$$

After a straightforward calculation with formal power series we find that

$$\begin{aligned}\int\Phi(c,t)\theta(dc) &= \theta(A)\Phi_A(t)+\theta(B)\Phi_B(t)\\ &= \frac{1-\sqrt{(1-\gamma t^2)^2-4t^4}}{\gamma t} = t\left(1+\frac{2}{\gamma}\Phi\left(\frac{t^2}{1-\gamma t^2}\right)\right),\end{aligned} \tag{70}$$

$$\Phi_A(t)\Phi_B(t) = \frac{1-\gamma t^2-\sqrt{(1-\gamma t^2)^2-4t^4}}{2t^2} = \Phi\left(\frac{t^2}{1-\gamma t^2}\right), \tag{71}$$

where $\Phi(t)$ is as in §12.3.3.

12.6.4. *Calculation of* μ. From (70) we know the moments of the measure μ and moreover we can compare these moments to those of the semicircle distribution. We find that

$$\langle\mu,f\rangle = \sqrt{1-\frac{4}{\gamma^2}}\,f(0)+\frac{1}{\gamma\pi}\int_{|x^2-\gamma|<2}\frac{f(x)\sqrt{4-(x^2-\gamma)^2}}{|x|}dx. \tag{72}$$

Now μ is the weak limit in probability of the empirical distributions $L(\mathcal{N}_k)$. To calculate the corresponding limit μ_W of the empirical distributions $L_W(\mathcal{N}_k)$ we use the "bookkeeping principle" (69) to find that

$$\langle\mu_W,f\rangle = \frac{\gamma+\sqrt{\gamma^2-4}}{4\pi}\int_{\gamma-2}^{\gamma+2}\frac{f(x)\sqrt{4-(x-\gamma)^2}}{x}dx. \tag{73}$$

See e.g. [PM67] for the latter result in the case of Wishart matrices. Note that if $\theta(A)<\theta(B)$ and hence $\gamma>2$, the measure μ has some mass concentrated at the origin. Notice also that if $\gamma=2$, then μ is the semicircle distribution.

12.6.5. *Calculation of* $\Theta(x,y)$ *and* $\Psi(x,y)$. With λ_{2r} and ϵ_{2r} as defined in (66), we have

$$\begin{aligned}\Theta(x,y) &= \sum_{r=1}^{\infty}\lambda_{2r}\Phi\left(\frac{x^2}{1-\gamma x^2}\right)^r\Phi\left(\frac{y^2}{1-\gamma y^2}\right)^r,\\ \Psi(x,y) &= \sum_{r=1}^{\infty}\epsilon_{2r}\Phi\left(\frac{x^2}{1-\gamma x^2}\right)^r\Phi\left(\frac{y^2}{1-\gamma y^2}\right)^r.\end{aligned}$$

To verify these formulas the main thing to note is that $K_r(c_1,\dots,c_r)=0$ unless colors along the sequence $c_1,\dots,c_r,c_1$ alternate between A and B. It follows in particular that $K_r\equiv 0$ for odd r.

12.6.6. *The measure* ν *and associated orthogonal polynomials.* Let ν be the measure with density

$$\frac{d\nu}{dx}=\mathbf{1}_{|x^2-\gamma|\le 2}\frac{\sqrt{4-(x^2-\gamma)^2}}{2\pi|x|}$$

with respect to Lebesgue measure. Note that μ is a convex combination of ν and a unit mass at the origin. Note that if $\gamma=2$ then $\nu=\sigma_S$. Put

$$V_n(x):=xU_n(x^2-\gamma).$$

By a straightforward calculation one verifies that the system of polynomial functions $\{V_n\}_{n=1}^{\infty}$ is orthonormal in $L^2(\nu)$, and moreover forms the "odd part" of the family of orthogonal polynomials naturally associated to the weight ν. By another straightforward calculation one verifies that for any continuously differentiable function g, setting $\tilde{g}(x)=g(x^2)$, one has

$$\langle\nu,(\tilde{g})'V_n\rangle=2Eg'(S+\gamma)U_n(S), \tag{74}$$

where S is a random variable with standard semicircular law σ_S.

12.6.7. *Calculation of* $\operatorname{Var} Z_f$ *and* E_f. For any even polynomial function f, by the orthogonality relations noted above, we can write

$$f'(x)=\sum_{n=1}^{\infty}V_n(x)\,\langle\nu,f'V_n\rangle,$$

with only finitely many nonzero terms in the sum. Then, by formulas (65) and (64), we can in the present specialization of the band matrix model rewrite (8) and (9) in the form

$$\operatorname{Var} Z_f=\sum_{r=1}^{\infty}(2\lambda_{2r}+\epsilon_{2r})\langle\nu,f'V_r\rangle^2, \tag{75}$$

$$E_f=\sum_{r=1}^{\infty}\frac{1}{2}(\lambda_{2r}+\epsilon_{2r})\langle\nu,f'V_{2r}\rangle, \tag{76}$$

at least when f is an even polynomial function. But then these formulas must remain valid for any polynomial f even or not since $\operatorname{tr} f(X(\mathcal{N}_k))$ vanishes identically for odd f.

12.6.8. *Calculation of* $\operatorname{Var} Z_{W,g}$. Given a polynomial function vanishing at the origin consider the random variables

$$Z_{g,W,k} := \operatorname{tr} g(W(\mathcal{N}_k)) - E \operatorname{tr} g(W(\mathcal{N}_k))$$

where g is any continuously differentiable function of polynomial growth. By the bookkeeping principle (69), with $\tilde{g}(x) = g(x^2)$, we have $Z_{g,W,k} = \frac{1}{2} Z_{\tilde{g},k}$, hence when g is a polynomial function, the random variables $Z_{g,W,k}$ converge in distribution to a mean zero Gaussian random variable $Z_{g,W}$ with variance

$$\operatorname{Var} Z_{g,W} = \sum_{r=1}^{\infty} (2\lambda_{2r} + \epsilon_{2r})(Eg'(S+\gamma)U_n(S))^2, \tag{77}$$

where, as in formula (74), S is a random variable with law σ_S. An analogous evaluation of the shift in mean can also be provided, but to avoid repetitions, we do not state it here.

Our final result is

Theorem 12.7. *We work in the setting and under the hypotheses of Theorems 3.2 and 3.3, and in the specialization of the band matrix model discussed in §12.6. If the random variables* $\xi_{\{\alpha,\beta\}}$ *satisfy a Poincaré inequality with the same constant* c*, then for any continuously differentiable function* g *with polynomial growth, the random variables* $Z_{g,W,k}$ *converge in distribution to a mean zero Gaussian random variable* $Z_{g,W}$*, with variance once again given by (77).*

Proof. Use Theorem 11.10, for the square-root generalized Wishart matrices $X(\mathcal{N}_k)$, taking ν to be as defined in §12.6.6, $\{q_n(x)\}_{n=1}^{\infty}$ to be the Gram-Schmidt orthogonalization in $L^2(\nu)$ of the family $\{x^{n-1}\}_{n=1}^{\infty}$ of powers of x, and $f(x) = g(x^2)$. □

13. Concluding remark

We have chosen to concentrate in this paper on CLT's for symmetric matrices. Similar techniques work also with Hermitian matrices, the main difference being that with $\xi_{\alpha,\beta} = \xi^1_{\alpha,\beta} + i\xi^2_{\alpha,\beta} = \xi^*_{\beta,\alpha}$, when $\alpha \neq \beta$, and $\xi^1_{\alpha,\beta}$ and $\xi^2_{\alpha,\beta}$ independent, identically distributed, zero mean real-valued random variables, it holds that $E[\xi_{\alpha,\beta}]^2 = 0$, and hence in the combinatorial evaluation of the contribution of various terms in expansions similar to (17), the contribution of words in which an edge is traversed twice *in the same direction* vanishes. (In particular, when computing variances for linear statistics of polynomial type, some of the bracelet contributions vanish.) None of the modifications needed to handle the Hermitian case are difficult. However, there are sufficiently many such modifications needed so that to give a careful accounting of them would add a nontrivial number of pages to an already long paper. So we think it best to omit further discussion.

Acknowledgments We owe the idea to look at spanning forests when proving Lemma 4.10 to Victor Reiner. We also thank Sergey Bobkov for a useful discussion concerning Poincaré inequalities.

References

[Ba99] Z. D. Bai, Methodologies in spectral analysis of large-dimensional random matrices, a review, *Statist. Sinica* **9** (1999), pp. 611–677.

[BY03] Z. D. Bai and J.-F. Yao, On the convergence of the spectral empirical process of Wigner matrices, *Preprint* (2003).

[BS04] Z. D. Bai and J. W. Silverstein, CLT for linear spectral statistics of large-dimensional sample covariance matrices, *Annals Probab.* **32** (2004), pp. 553–605.

[Bo99] S. G. Bobkov, Remarks on Gromov-Milman's inequality, *V. of Syktyvkar Univ.* **3** (1999), pp. 15–22 (in Russian).

[BU83] A. A. Borovkov and S. A. Utev, An inequality and a characterization of the normal distribution connected with it, *Theor. Prob. Appl.* **28** (1983), pp. 209–218.

[CD01] T. Cabanal-Duvillard, Fluctuations de la loi empirique de grande matrices aléatoires, *Ann. Inst. H. Poincaré - Probab. Statist.* **37** (2001), pp. 373–402.

[CB04] S. Chatterjee and A. Bose, A new method of bounding rate of convergence of empirical spectral distributions, *preprint* (2004). Available at www-stat.stanford.edu/∼souravc/rateofconv.pdf.

[De00] P. Deift, *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach*, New York University-Courant Institute of Mathematical Sciences, AMS (2000).

[Du89] R. M. Dudley, *Real Analysis and Probability*, Wadsworth and Brooks/Cole, Pacific Grove (1989).

[FK81] Z. Füredi and J. Komlós, The eigenvalues of random symmetric matrices, *Combinatorica* **1** (1981), pp. 233-241.

[Gi90] V. L. Girko, *Theory of Random Determinants*, Kluwer, 1990.

[Gu02] A. Guionnet, Large deviation upper bounds and central limit theorems for band matrices, *Ann. Inst. H. Poincaré Probab. Statist* **38** , (2002), pp. 341-384.

[GZ00] A. Guionnet and O. Zeitouni, Concentration of the spectral measure for large matrices, *Elec. Comm. in Probab.* **5** (2000), 119–136.

[HP00] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, AMS (2000).

[Ja97] S. Janson, *Gaussian Hilbert Spaces*, Cambridge University Press, (1997).

[Jo82] D. Jonsson, Some limit theorems for the eigenvalues of a sample covariance matrix, *J. Mult. Anal.* **12** (1982), pp. 1–38.

[Joh98] K. Johansson, On fluctuations of eigenvalues of random Hermitian matrices, *Duke J. Math.* **91**, (1998), pp. 151–204.

[KKP96] A. M. Khorunzhy, B. A. Khoruzhenko, and L. A. Pastur, Asymptotic properties of large random matrices with independent entries, *J. Math. Phys.* **37** (1996), pp. 5033–5060.

[Le01] M. Ledoux, *The Concentration of Measure Phenomenon*, American Mathematical Society, Providence (2001).

[Me91] M. L. Mehta, *Random Matrices*, 2nd ed. Academic Press (1991).

[MS04] J. A. Mingo and R. Speicher, Second Order Freeness and Fluctuations of Random Matrices: I. Gaussian and Wishart matrices and Cyclic Fock spaces, *preprint*, available at arXiv:math.OA/0405191 (2004).

[MPK92] S. A. Molchanov, L. A. Pastur and A. M. Khorunzhii, Distribution of the eigenvalues of random band matrices in the limit of their infinite order, *Theoret. and Math. Phys.* **90** (1992), pp. 108–118.

[NSS02] A. Nica, D. Shlyakhtenko, and R. Speicher, Operator-valued distributions. I. Characterizations of freeness, *Int. Math. Res. Notices* **29** (2002), pp. 1509–1538.

[PL03] L. Pastur and A. Lejay, Matrices aléatoires: statistique asymptotique des valeurs propres, in *Lecture Notes in Mathematics* **1801** (2003), pp. 135–164.

[PM67] L. A. Pastur and V.A Martchenko, The distribution of eigenvalues in certain sets of random matrices, *Math. USSR-Sbornik* **1** (1967), pp. 457-483.

[Sh96] D. Shlyakhtenko, Random Gaussian band matrices and freeness with amalgamation, *Int. Math. Res. Notices* **20** (1996), pp. 1013–1025.

[SS98] Ya. Sinai and A. Soshnikov, Central limit theorem for traces of large random symmetric matrices with independent matrix elements, *Bol. Soc. Brasil. Mat. (N.S.)* **29** (1998), pp. 1–24.

[St99] R. Stanley, *Enumerative Combinatorics, vol. II*, Cambridge University press (1999).

[Wig55] E. P. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, *Ann. Math.* **62** (1955), pp. 548–564.

[Wis28] J. Wishart, The generalized product moment distribution in samples from a Normal multivariate population, *Biometrika* **20A** (1928), pp. 32–52.

School of Mathematics, University of Minnesota, 206 Church St. SE, Minneapolis, MN 55455

School of Mathematics, University of Minnesota, 206 Church St. SE, Minneapolis, MN 55455, and Departments of Mathematics and of EE, Technion, Haifa 32000, Israel