Random Matrices: Theory and Applications Vol. 7, No. 2 (2018) 1850002 (21 pages) © The Author(s) DOI: 10.1142/S2010326318500028



# Evidence of the Poisson/Gaudin–Mehta phase transition for band matrices on global scales

Sheehan Olver

Department of Mathematics, Imperial College, London, UK s.olver@imperial.ac.uk

Andrew Swan

Department of Pure Mathematics and Mathematical Statistics University of Cambridge, UK acks2@cam.ac.uk

> Received 7 March 2017 Revised 19 October 2017 Accepted 5 November 2017 Published 26 January 2018

We prove that the Poisson/Gaudin–Mehta phase transition conjectured to occur when the bandwidth of an  $N \times N$  symmetric band matrix grows like  $b = \sqrt{N}$  is naturally observable in the rate of convergence of the level density to the Wigner semi-circle law. Specifically, we show for periodic and non-periodic band matrices the rate of convergence of the fourth moment of the level density is independent of the boundary conditions in the localized regime  $b \ll \sqrt{N}$  with a rate of  $\mathcal{O}(\frac{1}{b})$  for both cases, whereas in the delocalized regime  $b \gg \sqrt{N}$  where boundary effects become important, the rate of convergence for the two ensembles differs significantly, slowing to  $\mathcal{O}(\frac{b}{N})$  for non-periodic band matrices. Additionally, we examine the case of thick non-periodic band matrices b = cN, showing that the fourth moment is maximally deviated from the Wigner semi-circle law when  $b = \frac{2}{5}N$ , and provide numerical evidence that the eigenvector statistics also exhibit critical behavior at this point.

*Keywords*: Random band matrix; Poisson/Gaudin–Mehta transition; Anderson transition; global statistics.

Mathematics Subject Classification 2010: 60B20, 82B44, 15B52

### 1. Introduction

An important open problem in random matrix theory is the *Poisson/Gaudin–Mehta* conjecture on the existence of a phase transition in the local eigenvalue statistics of random real symmetric band matrices.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

**Conjecture 1.** The limiting local bulk statistics of a random band matrix with independent entries and bandwidth  $b \simeq N^{\alpha}$  are Poisson if  $\alpha < \frac{1}{2}$  and are Gaudin–Mehta if  $\alpha > \frac{1}{2}$ .

This paper concerns the question of whether this phase transition is also observable in the global eigenvalue statistics, e.g. the level density of the eigenvalues.

**Definition 1.** An  $N \times N$  matrix H is a real symmetric random band matrix with bandwidth b if the following four conditions hold:

- (1) H is a real symmetric matrix.
- (2) The upper triangular elements  $h_{ij}$ ,  $i \leq j$ , are jointly independent symmetric real random variables.
- (3) For ||i j|| < b,  $\mathbb{E} h_{ij} = 0$ ,  $\mathbb{E} h_{ij}^2 = 1$ ,  $\mathbb{E} h_{ij}^4 = q$ .
- (4) For  $||i j|| \ge b$ ,  $h_{ij} = 0$ .

The metric on the entries is defined as

$$||i - j|| = \min(|i - j|, N - |i - j|)$$

for *periodic band matrices*, and as

$$||i - j|| = |i - j|$$

for non-periodic band matrices. The matrix ensembles associated to these two cases will be denoted as  $\text{PBE}_{N,b,q}$  and  $\text{NBE}_{N,b,q}$ , respectively.

The limiting level densities of random band matrices were first considered in 1991 by Bogachev, Khorunzhii, Molchanov, and Pastur [3, 12]. In these papers, they obtained the following result:

**Proposition 1.** In the limit  $N \to \infty$ , the normalized level density of a periodic random band matrix with bandwidth  $b = cN^{\alpha}$  is given by the Wigner semi-circle law for any  $0 < \alpha \le 1$  and  $0 < c \le 1$ . For non-periodic band matrices, the limiting level density is given by the Wigner semi-circle law only if  $0 < \alpha < 1$  or  $\alpha = 1$  and c = 1.

This is surprising for two reasons: first, the Wigner semi-circle law extends down to small bandwidths  $b = \mathcal{O}(N^{\varepsilon})$ , but even more surprisingly, it does not hold for very large bandwidths b = cN in the non-periodic case. It would therefore seem that the Poisson/Gaudin–Mehta transition is unobservable at a macroscopic scale, and indeed, the authors of [12] make an interesting remark to this effect:

[The parameter  $\frac{b^2}{N}$ ] does not play a special role in the formation of the density of states of the considered matrices. However, we are not inclined to regard our results as incompatible with the interesting [Poisson/Gaudin–Mehta conjecture], since, as is well known, the integrated density of states has very little sensitivity to the localization properties of the states of random operators.

We contend that this is not the case: "very little sensitivity" is not the same as "none". Our main result is a proof that the critical parameter does indeed play a special role on the global scale, by controlling the rate of convergence of the level density to the Wigner semi-circle law. Under the assumption that  $q \neq 2$ ,<sup>a</sup> we show that the rate of convergence of the fourth moment of the level density in the *localized* regime  $\alpha < \frac{1}{2}$  is the same for *both* periodic and non-periodic ensembles with a rate of  $\mathscr{O}(N^{-\alpha})$ , whilst in the *delocalized* regime  $\alpha > \frac{1}{2}$ , the rate of convergence differs significantly between the two cases, increasing to the much slower rate of  $\mathscr{O}(N^{1-\alpha})$ in the non-periodic case.

**Theorem 1.** In the limit  $N \to \infty$ , the fourth moment of the normalized level density

$$m_4(\sigma_{N,b,q}) = \int_{-\infty}^{\infty} x^4 \sigma_{N,b,q}(dx) \tag{1}$$

of a non-periodic band ensemble with bandwidth  $b = N^{\alpha}$ ,  $0 < \alpha < 1$ , has asymptotic expansion

$$m_4(\sigma_{N,b,q}) = 2 + \left(\frac{1}{3}N^{\alpha-1} + \frac{q-2}{2}N^{-\alpha}\right) + \mathcal{O}_q(N^{2\alpha-2} + N^{-2\alpha}),$$

where q is the fourth moment of the entries.

**Theorem 2.** In the limit  $N \to \infty$ , the fourth moment of the normalized level density

$$m_4(\sigma_{N,b,q}) = \int_{-\infty}^{\infty} x^4 \sigma_{N,b,q}(dx)$$
(2)

of a periodic band ensemble with bandwidth  $b = N^{\alpha}$ ,  $0 < \alpha < 1$ , has asymptotic expansion

$$m_4(\sigma_{N,b,q}) = 2 + \frac{q-2}{2}N^{-\alpha} + \mathcal{O}_q(N^{2\alpha-2} + N^{-2\alpha}),$$

where q is the fourth moment of the entries.

The proof may be found in Sec. 4, and an explanation as to why this macroscopic behavior is a indeed genuine Poisson/Gaudin–Mehta phenomenon in Sec. 5.

## 2. Previous Results

#### 2.1. Local statistics

The Poisson/Gaudin–Mehta conjecture appears as the final entry in Percy Deift's list of open problems in integrable systems and random matrices [6]. Interest in this

<sup>&</sup>lt;sup>a</sup>The case q = 2 is somewhat degenerate as this is when the fourth moment of the entries coincides with the fourth moment of the Wigner semi-circle law. We expect that a similar relationship holds for higher moments with the degenerate case being when the distribution of the entries is semi-circular.

conjecture can be traced back to the numerical work of Seligman, Verbaarschot, and Zirnbauer [14], and Casati, Izrailev, and Molinari [5] in the late 1980s/early 1990s. Since this time, progress has been rather modest: results concerning local universality are only just beginning to emerge, and only apply to periodic band matrices with large bandwidth. For a specially structured "periodic block band GUE" with large bandwidth b = cN, Shcherbina [15] proved local universality for energies  $|\lambda| < \sqrt{2}$  using a rigorous supersymmetric method; Bourgade *et al.* [4] strengthened this result using heat flow methods combined with a new mean field reduction technique to prove local universality for more general periodic band Wigner ensembles in both the real symmetric/Hermitian cases for all energies in the bulk, again for b = cN.

Outside of numerical simulation and heuristic supersymmetric arguments (see Fyodorov and Mirlin [11]), the only results indicating the critical bandwidth  $b = \sqrt{N}$  have concerned the two-point correlation function of the characteristic polynomial

$$F_2(\lambda_1, \lambda_2) = \mathbb{E}(\det(A - \lambda_1) \det(A - \lambda_2)), \qquad (3)$$

which, for any  $\lambda_0 \in (-2, 2)$  and  $\xi$  varying in any compact set  $C \subset \mathbb{R}$ , in the scaling limit goes to

$$D_2^{-1} F_2\left(\lambda_0 + \frac{\xi}{N\rho(\lambda_0)}, \lambda_0 - \frac{\xi}{N\rho(\lambda_0)}\right) \to \frac{\sin(2\pi\xi)}{2\pi\xi}$$
(4)

if  $\alpha > \frac{1}{2}$  [16], and goes to

$$D_2^{-1}F_2\left(\lambda_0 + \frac{\xi}{N\rho(\lambda_0)}, \lambda_0 - \frac{\xi}{N\rho(\lambda_0)}\right) \to 1$$
(5)

if  $\alpha < \frac{1}{2}$  [17], where

$$D_2 = F_2(\lambda_0, \lambda_0). \tag{6}$$

These results, although consistent with Poisson/Gaudin–Mehta local statistics, do not imply it.

On the other hand, much more is known at the edge of the spectrum: Sodin showed that for real symmetric/Hermitian periodic band matrices with symmetric subgaussian entries, the limiting distribution of the extreme eigenvalues is given by the Tracy–Widom distribution if  $b \gg N^{\frac{5}{6}}$  and by some other distribution if  $b \ll N^{\frac{5}{6}}$  [18].

### 2.2. Mesoscopic statistics

Mesoscopic statistics describe the eigenvalues on scales which are intermediate to the local and global regimes, probing the spectrum in ranges which are much larger than the typical eigenvalue gap size, but still small enough so that macroscopic properties (such as the density) are constant. One such example is the variance of the number of eigenvalues  $\mathcal{N}_{\chi}$  on an energy interval of size  $\frac{1}{N} \ll \chi \ll 1$ . Motivated by their

investigations into small metallic conductors, Altshuler and Shklovskii introduced this statistic in [1], showing that the mesoscopic statistics undergo a phase transition at a scale  $\chi_c = \min\left(\frac{b^2}{N^2}, \frac{1}{b^2}\right)$ . For  $\chi \gg \chi_c$ , they predicted what is now known as the Altshuler–Shklovskii formula for band matrices

$$\operatorname{Var}(\mathscr{N}_{\chi}) \asymp \frac{N}{b} \sqrt{\chi}.$$
 (7)

This expression holds in both Possion and Gaudin–Mehta regimes. On the other hand, for  $\chi \ll \chi_c$ , there are two possible cases: for  $b^2 \gg N$ 

$$\operatorname{Var}(\mathscr{N}_{\chi}) \asymp \log N,\tag{8}$$

whereas for  $b^2 \ll N$ 

$$\operatorname{Var}(\mathscr{N}_{\chi}) \asymp N\chi. \tag{9}$$

These expressions were recently put on rigorous footing by Erdős and Knowles in [8, 9], under the assumption that  $\chi \gg b^{-\frac{1}{3}}$ . These formulas are consistent with the Poisson/Gaudin–Mehta conjecture, but we will defer discussion of this relationship until Sec. 5.

### 2.3. Eigenvector statistics

Compared to the local statistics, the situation has been somewhat better on the eigenvector side of the story. Phrased in terms of the eigenvectors, the analogous conjecture is as follows.

**Conjecture 2.** The bulk eigenvector localization length l for a random band matrix with independent entries and bandwidth b is  $l \simeq \min(b^2, N)$ .

Due to the connection with disordered conductors, this is also known as the Anderson transition conjecture for band matrices. The first result concerning this conjecture was due to Schenker [13], who provided an upper bound  $l \leq b^8$ . A lower bound of  $l \geq b^{\frac{7}{6}}$  was demonstrated by Erdős and Knowles [7]; together with Yau and Yin [10], they improved this to  $l \geq b^{\frac{5}{4}}$ . These lower bounds hold in a weaker sense, in that they only hold for "most" eigenvectors. A strong lower bound of  $l \geq b^{\frac{7}{6}}$  for eigenvectors with corresponding eigenvalues  $|\lambda| < \sqrt{2}$  was shown by Bao and Erdős [2] for block band ensembles with entries matching Gaussians to four moments.

## 3. Correlation Functions

The one-point correlation function or level density of a Hermitian random matrix  $H \sim E_N$  is the unique measure  $\rho_N : \mathscr{B}(\mathbb{R}) \to \mathbb{R}^+$  such that

$$\int_{-\infty}^{\infty} f(x)\rho_N(dx) = \sum_{i=1}^{N} \mathbb{E}f(\lambda_i)$$
(10)

for all continuous compactly supported functions  $f : \mathbb{R} \to \mathbb{R}$ , where  $\lambda_i$  are the N ordered eigenvalues of H. For discrete ensembles,  $\rho_N$  can only be defined as a

### S. Olver & A. Swan

measure on  $\mathbb{R}$ , but for continuous ensembles we can write  $\rho_N(dx) = \rho_N(x)dx$  as a density. In this case, we equivalently have

$$\rho_N(x) = N \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(x_1, \dots, x_N) dx_2 \cdots dx_N,$$
(11)

where  $P(x_1, \ldots, x_N)$  is the joint density of the eigenvalues. From the one point correlation function, we construct the *normalized one-point correlation function* or *normalized level density*  $\sigma_N : \mathscr{B}(\mathbb{R}) \to \mathbb{R}^+$  through the rescaling

$$\sigma_N(x) = \nu \rho_N(\xi x), \tag{12}$$

where the scaling parameters

$$\xi = \sqrt{\frac{m_2(\rho_N)}{m_0(\rho_N)}} \tag{13}$$

and

$$\nu = \sqrt{\frac{m_2(\rho_N)}{m_0(\rho_N)^3}}$$
(14)

are given in terms of the moments of the one point correlation function

$$m_k(\rho_N) = \int_{-\infty}^{\infty} x^k \rho_N(dx) = \mathbb{E}\operatorname{tr}(H^k).$$
(15)

This scaling is chosen so that

$$m_0(\sigma_N) = m_2(\sigma_N) = 1, \tag{16}$$

i.e. so that  $\sigma_N$  is a probability measure with variance 1. From this scaling, we therefore have

$$m_k(\sigma_N) = \frac{m_0(\rho_N)^{\frac{k}{2}-1}m_k(\rho_N)}{m_2(\rho_N)^{\frac{k}{2}}} = \frac{N^{\frac{k}{2}-1}\mathbb{E}\operatorname{tr}(H^k)}{(\mathbb{E}\operatorname{tr}(H^2))^{\frac{k}{2}}},\tag{17}$$

and in particular, the fourth moment of  $\sigma_N$  is

$$m_4(\sigma_N) = \frac{N \mathbb{E} \operatorname{tr}(H^4)}{(\mathbb{E} \operatorname{tr}(H^2))^2}.$$
(18)

## 4. Critical Behavior on the Global Scale

In terms of moments, Proposition 1 states that in the limit  $N \to \infty$ ,  $b = N^{\alpha}$ , the moments of normalized level density  $m_k(\sigma_{N,b,q})$  of a band ensemble will converge to those of the Wigner semi-circle  $m_k(\mu_{sc})$ . What about the moments for finite N and b? Figure 1 shows a numerical simulation of  $m_4(\sigma_{N,b,q})$ . As can be seen, for  $b \ll \sqrt{N}$ , the rate of convergence to the Wigner semi-circle law is the same for both periodic and non-periodic ensembles, however, for  $b \gg \sqrt{N}$ , the rates of convergence are very different.

**Remark 1.** In the following, it will be convenient to take N = mb as an integer multiple of the bandwidth b, where  $m \in \mathbb{N}$  is not necessarily fixed. In this case, the



Fig. 1. This figure shows  $\log(|m_4(\sigma_{N,b,q}) - 2|)$  against  $\alpha$  with  $b = N^{\alpha}$  for N = 200, 400, q = 3.

non-periodic/periodic band matrices can be represented as

$$H = \begin{bmatrix} A_{1} & L_{1} & & & \\ L_{1}^{\top} & A_{2} & L_{2} & & \\ & L_{2}^{\top} & A_{3} & \ddots & \\ & & \ddots & \ddots & L_{m-1} \\ & & & L_{m-1}^{\top} & A_{m} \end{bmatrix},$$
(19)  
$$H = \begin{bmatrix} A_{1} & L_{1} & & L_{m} \\ L_{1}^{\top} & A_{2} & L_{2} & & \\ & L_{2}^{\top} & A_{3} & \ddots & \\ & & \ddots & \ddots & L_{m-1} \\ L_{m}^{\top} & & & L_{m-1}^{\top} & A_{m} \end{bmatrix},$$
(20)

where the  $A_i$  are  $b \times b$  random matrices drawn from  $\text{NBE}_{b,b,q}$ , and the  $L_i$  are strictly lower triangular  $b \times b$  random matrices with independent entries  $l_{ij}$  below the diagonal. We may consider these lower triangular blocks  $L_i$  as drawn from an ensemble  $\text{LTE}_{b,q}$ , defined in the obvious way.

**Remark 2.** To simplify notation, we drop subscripts to indicate a generic element.

**Lemma 1.** Let  $A \sim \text{NBE}_{b,b,q}$  and  $L_1, L_2 \sim \text{LTE}_{b,q}$  be  $b \times b$  jointly independent random matrices. Then

$$\mathbb{E}\operatorname{tr}(A^4) = 2b^3 + (q-2)b^2, \tag{21}$$

$$\mathbb{E}\operatorname{tr}(A^2 L L^{\top}) = \frac{b^3 - b^2}{2},$$
(22)

$$\mathbb{E}\operatorname{tr}(A^{2}L^{\top}L) = \frac{b^{3} - b^{2}}{2},$$
(23)

$$\mathbb{E}\operatorname{tr}(L_1^{\top}L_1L_2L_2^{\top}) = \frac{b^3 - 3b^2 + 2b}{6},\tag{24}$$

$$\mathbb{E}\operatorname{tr}(LL^{\top}LL^{\top}) = \frac{4b^3 + (3q - 12)b^2 - (3q - 8)b}{6}.$$
 (25)

**Proof.** The proof requires a straight forward, but tedious, moment calculation. Details may be found in the appendix.

We are now ready to prove our main result.

**Proof of Theorem 1.** The second moment of  $H \sim \text{NBE}_{N,b,q}$  is simply equal to the number of entries inside the band

$$\mathbb{E}\operatorname{tr}(H^2) = \mathbb{E}\operatorname{tr}(HH^{\top}) = \sum_{i,j} \mathbb{E}h_{ij}^2 = 2Nb - N - b^2 + b.$$
(26)

The calculation of the fourth moment is more involved. Taking the dimension N = mb as an integer multiple of the bandwidth, H can be represented in block form as in Eq. (20). Assuming that  $b \leq \frac{N}{2}$ , we can then write  $C = H^2$ ,

or, in block coordinates

$$C_{i,i} = L_{i-1}^{\top} L_{i-1} + A_i^2 + L_i L_i^{\top},$$

$$C_{i,i+1} = C_{i+1,i}^{\top} = A_i L_i + L_i A_{i+1},$$

$$C_{i,i+2} = C_{i+2,i}^{\top} = L_i L_{i+1}$$
(28)

with the convention that  $L_0 = L_m = 0$ . The fourth moment of H is hence

$$\mathbb{E} \operatorname{tr} H^{4} = \mathbb{E} \operatorname{tr} C^{2} = \sum_{i,j}^{m} \mathbb{E} \operatorname{tr} (C_{ij} C_{ij}^{\top})$$
$$= \sum_{i}^{m} \mathbb{E} \operatorname{tr} ((L_{i-1}^{\top} L_{i-1} + A_{i}^{2} + L_{i} L_{i}^{\top}) (L_{i-1}^{\top} L_{i-1} + A_{i}^{2} + L_{i} L_{i}^{\top})^{\top})$$

$$+2\sum_{i}^{m-1} \mathbb{E} \operatorname{tr}((A_{i}L_{i} + L_{i}A_{i+1})(A_{i}L_{i} + L_{i}A_{i+1})^{\top}) + 2\sum_{i}^{m-2} \mathbb{E} \operatorname{tr}((L_{i}L_{i+1})(L_{i}L_{i+1})^{\top}).$$
(29)

In each of the three sums, the terms being summed over are identical, with the exception of the first and last term in the first sum. We therefore have

$$\mathbb{E} \operatorname{tr} H^{4} = (m-2) \mathbb{E} \operatorname{tr} ((L_{1}^{\top}L_{1} + A^{2} + L_{2}L_{2}^{\top})(L_{1}^{\top}L_{1} + A^{2} + L_{2}L_{2}^{\top})^{\top}) + 2 \mathbb{E} \operatorname{tr} ((A^{2} + L^{\top}L)(A^{2} + L^{\top}L)^{\top}) + 2(m-1) \mathbb{E} \operatorname{tr} ((A_{1}L + LA_{2})(A_{1}L + LA_{2})^{\top}) + 2(m-2) \mathbb{E} \operatorname{tr} ((L_{1}L_{2})(L_{1}L_{2})^{\top}).$$
(30)

Expanding this out and applying Lemma 1, we find

$$\mathbb{E}\operatorname{tr} H^{4} = m \mathbb{E}\operatorname{tr}(A^{4}) + (6m - 8) \mathbb{E}\operatorname{tr}(A^{2}LL^{\top}) + 2m \mathbb{E}\operatorname{tr}(A^{2}L^{\top}L) + (4m - 8) \mathbb{E}\operatorname{tr}(L_{1}^{\top}L_{1}L_{2}L_{2}^{\top}) + (2m - 2) \mathbb{E}\operatorname{tr}(LL^{\top}LL^{\top}) = 8mb^{3} + (2q - 12)mb^{2} - (q - 4)mb - \frac{20}{3}b^{3} - (q - 12)b^{2} + \left(q - \frac{16}{3}\right)b = 8Nb^{2} + (2q - 12)Nb - (q - 4)N - \frac{20}{3}b^{3} - (q - 12)b^{2} + \left(q - \frac{16}{3}\right)b.$$
(31)

Together with Eq. (26), we find the fourth moment of the normalized level density as

$$m_4(\sigma_{N,b,q}) = \frac{N\left(8Nb^2 + (2q-12)Nb - (q-4)N - \frac{20}{3}b^3 - (q-12)b^2 + \left(q - \frac{16}{3}\right)b\right)}{(2Nb - N - b^2 + b)^2} = 2 + \frac{(2q-4)N^2b + \frac{4}{3}Nb^3 - (q-2)N^2 - qNb^2 + \left(q - \frac{4}{3}\right)Nb - 2b^4 + 4b^3 - 2b^2}{(2Nb - N - b^2 + b)^2}.$$
(32)

Setting  $b = N^{\alpha}$ ,  $0 < \alpha < 1$ , we then have

$$m_4(\sigma_{N,b,q}) = 2 + \left(\frac{1}{3}N^{\alpha-1} + \frac{q-2}{2}N^{-\alpha}\right) + \mathcal{O}_q(N^{2\alpha-2} + N^{-2\alpha}), \qquad (33)$$

or, in terms of the critical parameter  $\eta = \frac{b}{\sqrt{N}}$ ,

$$m_4(\sigma_{N,b,q}) = 2 + \frac{1}{\sqrt{N}} \left( \frac{1}{3} \eta + \frac{q-2}{2} \eta^{-1} \right) + \mathcal{O}_q \left( \frac{\eta^2 + \eta^{-2}}{N} \right).$$
(34)

**Remark 3.** Interestingly, although the expression for the fourth moment of the level density (32) is exact for  $b \leq \frac{N}{2}$ , it is false for  $b > \frac{N}{2}$ . This because the combinatorial structure of the problem is different in this region.

**Proof of Theorem 2.** We will omit the details as the calculation is essentially the same as in the non-periodic case. For  $H \sim \text{PBE}_{N,b,q}$ , the second moment is

$$\mathbb{E}\operatorname{tr}(H^2) = 2Nb - N,\tag{35}$$

and the fourth moment is

$$\mathbb{E}\operatorname{tr}(H^{4}) = m \mathbb{E}\operatorname{tr}((L_{1}^{\top}L_{1} + A^{2} + L_{2}L_{2}^{\top})(L_{1}^{\top}L_{1} + A^{2} + L_{2}L_{2}^{\top})^{\top}) + 2m \mathbb{E}\operatorname{tr}((A_{1}L + LA_{2})(A_{1}L + LA_{2})^{\top}) + 2m \mathbb{E}\operatorname{tr}((L_{1}L_{2})(L_{1}L_{2})^{\top}) = 8Nb^{2} + (2q - 12)Nb - (q - 4)N.$$
(36)

This gives the fourth moment of the normalized density as

$$m_4(\sigma_{N,b,q}) = 2 + \frac{q-2}{2b-1}$$
  
=  $2 + \frac{q-2}{2}N^{-\alpha} + \mathcal{O}_q(N^{-2\alpha}).$  (37)

### 5. Relationships Between Scales

The connections across scales in the eigenvalue/eigenvector statistics are most clearly seen from the disordered conductor perspective, where our band matrix is interpreted as the Hamiltonian for a single electron Anderson tight binding model on N sites with hopping range b; the diagonal elements are seen as random on sites potentials, and the off diagonal elements as random hopping amplitudes.

Let us first look at the case where the eigenvectors are localized with length  $l \ll N$ . Under the assumption that the localization positions  $c_i$  of the eigenvectors  $\psi_i$  are independent and uniformly distributed, i.e.

$$|\psi_i|(j) \le C_i e^{\frac{-|j-c_i|}{l}} \tag{38}$$

for some random  $C_i = \mathcal{O}(l^{-\frac{1}{2}})$ , if we choose two consecutive eigenvalues  $\lambda_1$ ,  $\lambda_2$  in the bulk at random, then their associated eigenvectors will not significantly overlap

with high probability  $1 - \frac{l}{N}$ . If we then vary the on-site potential/hopping energies on just the sites on which  $\psi_1$  was supported, leaving those on which  $\psi_2$  was supported unchanged, we would shift  $\lambda_1$  whilst keeping  $\lambda_2$  constant; the eigenvalues are independent and the gap distribution will be Poisson.

Attempting to continue this argument to mesoscopic scales reveals why Poisson statistics break down: in an energy window of size  $\chi$  there will be  $\mathcal{O}(N\chi)$  eigenvalues, and if this is much larger than  $\frac{N}{l}$ , i.e. if  $\chi \gg \frac{1}{l}$ , then a significant number of eigenvectors must overlap and will therefore not display Poisson statistics above this scale. Indeed, the Altshuler–Shklovskii formula predicts that the variance in the number of energy levels in a window will be much smaller than the Poisson case if  $\frac{N}{b}\sqrt{\chi} \ll N\chi$ , i.e. if  $\chi \gg \frac{1}{b^2}$ . Putting these two descriptions of the transition together gives the correct prediction of the localization length as length as  $l \simeq b^2$ , as pointed out by Erdős and Knowles in [8, 9].

Although the variance in the Altshuler–Shklovskii regime is much less than the Poisson case, it is of course much more than typical mean field statistics where  $\operatorname{Var}(\mathcal{N}_{\chi}) \asymp \log N$ : on mesoscopic scales, the spectrum of band matrices displays a sort of soft rigidity.

The connection between the local eigenvalue statistics and the eigenvector statistics for the delocalized regime follows essentially as before: when the eigenvectors are delocalized, their supports will overlap significantly; changing the onsite potential and hopping amplitudes to shift a single eigenvalue will of course modify the other, typically in the same direction, as both eigenvectors feel the same change. This manifests as the famous RMT eigenvalue repulsion, resulting in Gaudin–Mehta local statistics. Unfortunately, these simple heuristics do not seem to have an interpretation in the mesoscopic regime.

From the Anderson model perspective, the periodic/non-periodic band matrices correspond to models of periodic/non-periodic conductors. We believe that reason the two models display different rates of convergence for their macroscopic statistics in the delocalized regime is because it is here that a significant quantity of the eigenvectors feels the effect of the boundary. An important difference between the two models is that for the non-periodic band matrices, the sites near the edges are less connected than those in the middle. Specifically, if  $b \leq i \leq N-b$  then  $\deg(u_i) =$ 2(b-1), whereas if i < b or i > N-b then  $b-1 \leq \deg(u_i) < 2(b-1)$ . Heuristically, we expect that the nature of the boundary effect on a given eigenvector will depend on its associated eigenvalue: eigenvectors with smaller eigenvalues should tend to have slightly more mass in the boundary region (see Sec. 5.1 for further discussion on this point). Consider what happens when we "periodize" a non-periodic band matrix: upon introducing the corner blocks, only those eigenvalues whose eigenvectors are significantly supported in the boundary regions will undergo a shift in energy, and as a result, almost all eigenvalues will have the same distribution in both cases. On the other hand, when the eigenvectors are delocalized, all eigenvectors are supported on the boundary; changing the conductor from periodic to non-periodic will then shift all eigenvalues and so will be observable in the macroscopic statistics.

#### 5.1. Thick band matrices

In this final section, we would like to present a few observations on thick nonperiodic band matrices. As discussed in Proposition 1, the level density of nonperiodic band matrices does not converge to the Wigner semi-circle law for large bandwidths b = cN, 0 < c < 1. The question then arises as for what c is the deviation from the semi-circle law maximal? From the Anderson model perspective, one might guess that this should occur when  $c = \frac{1}{2}$ , because this is when there is no homogeneous middle section in the conductor. In other words, this is when the difference between the periodic and non-periodic cases is largest; to periodize the conductor for  $c = \frac{1}{2}$  requires the addition of  $\frac{N^2}{4}$  elements in the corner, which is maximal. It is therefore surprising that maximum deviation from the semi-circle law occurs not when  $c = \frac{1}{2}$ , but when  $c = \frac{2}{5}$ . What is even more surprising is that at this point that the eigenvectors are also maximally deviated from the mean field case in the sense that they are least flat. This provides further evidence that eigenvector delocalization and macroscopic statistics are indeed related.

**Corollary 1.** In the limit  $N \to \infty$ , the fourth moment of the level density of a nonperiodic band matrix with bandwidth b = cN has a local maximum at  $c = \frac{2}{5} + o(1)$ .

**Proof.** Taking b = cN,  $0 < c \leq \frac{1}{2}$ , the fourth moment of the normalized level density is

$$m_{4}(\sigma_{N,b,q}) = \frac{(2q-4)N^{4}c + \frac{4}{3}N^{4}c^{3} - (q-2)N^{2} - qN^{3}c^{2} + \left(q - \frac{4}{3}\right)N^{2}c}{-2N^{4}c^{4} + 4N^{3}c^{3} - 2N^{2}c^{2}} = 2 + \frac{-2N^{4}c^{4} + 4N^{3}c^{3} - 2N^{2}c^{2}}{(2N^{2}c - N - N^{2}c^{2} + Nc)^{2}} = 2 - \frac{2c^{2} - \frac{4}{3}c}{(c-2)^{2}} + \mathcal{O}\left(\frac{1}{N}\right).$$

$$(39)$$

Taking the derivative with respect to c gives

$$\partial_c m_4(\sigma_{N,b,q}) = \frac{4}{3} \frac{5c-2}{(c-2)^3} + \mathcal{O}\left(\frac{1}{N}\right)$$
(40)

and which will go to zero in the limit  $N \to \infty$  only if  $c = \frac{2}{5} + o(1)$ . Checking the second partial derivative shows that this critical point is a maximum.

**Remark 4.** This proof only covers the range  $0 < c \leq \frac{1}{2}$ , but we see numerically (Fig. 2) that this is the only critical point when the bandwidth is thick b = cN. When q > 2, a second critical point exists near  $b \simeq N^{-\frac{1}{2}}$ , which corresponds to the crossover in the convergence rate of the fourth moment.



Fig. 2. This figure shows  $m_4(\sigma_{N,b,q})$  against c with b = cN for N = 100, 200, 400. A local maximum is evident near  $b = \frac{2}{5}N$ , rather than  $b = \frac{N}{2}$  as we might otherwise expect. The local minimum near  $b \approx \sqrt{N}$  corresponds to the bandwidth with fastest convergence to the Wigner semi-circle law.

Turning to the eigenvectors, we would like to examine a statistic that measures their global flatness. An appropriate choice is the total inverse participation ratio

$$I_4(Q) = \sum_j I_4(\psi^{(j)}), \tag{41}$$

where  $I_4(\psi^{(j)}) = \sum_i |\psi_i^{(j)}|^4$  is the inverse participation ratio of the *j*th eigenvector  $\psi^{(j)}$ . Heuristically, this quantity is on the order of  $\frac{N}{l_{av}}$ , where  $l_{av}$  is the average localization length. For large N, we expect  $l_{av} = \mathscr{O}(\min(b^2, N))$ , as contributions from the edge eigenvectors will be negligible. We will therefore have  $I_4(Q) = \mathscr{O}(1)$  for  $b \gg \sqrt{N}$  and  $I_4(Q) = \mathscr{O}(\frac{N}{b^2})$  for  $b \ll \sqrt{N}$  across the localization/delocalization transition, and complete flatness  $I_4(Q) = 3$  when b = N. Numerical evidence (Figs. 3 and 4) confirms these expectations, and importantly, indicates that a local maximum can again be seen around  $b \simeq \frac{2}{5}N$ . This maximum is rather difficult to detect numerically, and is only unequivocally apparent for  $N \gtrsim 4000$ . These figures show that like the macroscopic eigenvalue statistics, the eigenvector localization properties of non-periodic band matrices differ from the mean field case at large bandwidths.

Why do the eigenvectors also display a local maximum? The reason, again, is due to boundary effects. Until the first transition point  $b \approx \sqrt{N}$ , the localization length grows as  $l \approx b^2$  causing  $I_4(Q)$  to decay as  $\mathscr{O}(\frac{N}{b^2})$ . After this point, the Anderson delocalization mechanism becomes saturated and secondary phenomena affecting  $I_4(Q)$  at the scale of  $\mathscr{O}(1)$  become visible. In the delocalization regime, we expect



Fig. 3. This figure shows the logarithm of the total inverse participation ratio  $\log(I_4(Q))$  for 3 sequences of random matrices drawn from a Gaussian NPE<sub>N,b,q</sub> with N = 2000, 4000, 8000, plotted against  $\alpha$  with  $b = N^{\alpha}$ . This figure highlights the Anderson transition: for  $\alpha < \frac{1}{2}$ , the gradient is approximately  $-2\log(N)$ , which is consistent with  $I_4(Q) = \mathcal{O}(\frac{N}{b^2})$ ; for  $\alpha > \frac{1}{2}$ , the curves are approximately flat, which is again consistent with  $I_4(Q) = \mathcal{O}(1)$ .



Fig. 4. This figure shows the same data as Fig. 3, but now with a linear scale plotting  $I_4(Q)$  against c with b = cN. This highlights the local maximum in  $I_4(Q)$ , which can be seen to form as N increases around  $b \sim \frac{2}{5}N$ .

that the eigenvector component fluctuations should be relatively uniform across the homogeneous central section, only changing<sup>b</sup> significantly in the boundary regions i < b, i > N - b. Further, as the delocalization condition implies that  $|\psi_i^{(j)}|^4 = \mathscr{O}(\frac{1}{N^2})$ , the fluctuations at the edge and in the middle are of the same order, i.e. they are different only up to a constant factor. Putting this all together, we see that boundary effects will become significant when the two regions are of comparable

<sup>b</sup>Whether the boundary fluctuations are smaller or larger than the central fluctuations is irrelevant as  $I_4(\psi)$  will increase in both cases due to the normalization  $\sum_i |\psi_i|^2 = 1$ . size, i.e. when  $b \simeq N$ . Continuing this argument, it is easy to see that  $I_4(Q)$  will attain a local maximum in this regime, because although the boundary effect reduces eigenvector flatness, when b = N we recover a standard Wigner ensemble, whose eigenvectors have uniform fluctuations.

As mentioned in Sec. 5, the nature of the boundary effect on a given eigenvector depends on its associated eigenvalue: eigenvectors with smaller eigenvalues tend to have more mass in the boundary region, those with larger eigenvalues tend to have more mass in the central region. We can justify this as follows: consider the  $2N \times 2N$  matrix

$$H = \begin{bmatrix} A & 0\\ 0 & B \end{bmatrix},\tag{42}$$

where  $A \sim \text{NBE}_{N,2}$  and  $B \sim \text{NBE}_{N,N}$ . As H is block diagonal, the eigenvalues of H will be the same as those of A and B. From the second moment of the level densities, the eigenvalues of A will be of size  $\mathcal{O}(1)$ , whereas those of B will be of size  $\mathcal{O}(\sqrt{N})$ . Hence, in an interval of size  $\mathcal{O}(1)$  around the origin, H will have approximately  $\mathcal{O}(N)$  eigenvalues originally from A, and  $\mathcal{O}(\sqrt{N})$  eigenvalues originally from B. As a result, if we choose an eigenvalue at random in this region, with high probability the associated eigenvector will be localized, and located in the region  $1 \leq i \leq N$ . Now, consider the matrix  $\hat{H} = H + P$ , where P is a matrix of zeros except for a random  $p_{N,N+1} = p_{N+1,N} = p$ . This couples the eigenvalues and eigenvectors of A and B together; as a conductor, this represents a model of a thin wire attached to a conducting "ball". As we have only introduced a small perturbation, the eigenvalues, we have from the Lidskii inequality

$$\sum_{i}^{2N} |\lambda_i - \hat{\lambda}_i|^2 \le \operatorname{tr}(P^2) = 2p^2 = \mathscr{O}(1).$$
(43)

If we again pick an eigenvalue at random from an  $\mathscr{O}(1)$  interval around the origin, in the worst case scenario, we will have

$$|\lambda_i - \hat{\lambda}_i|^2 = \mathscr{O}\left(\frac{1}{N}\right),\tag{44}$$

i.e.  $|\lambda_i - \hat{\lambda}_i| = \mathcal{O}(\frac{1}{\sqrt{N}})$  with high probability.<sup>c</sup> The corresponding eigenvectors are also close to the original. If we assume that the localization positions c of the original (localized) eigenvectors  $\psi$  are uniformly distributed over indices  $1, \ldots, N$ , then

$$|\psi_i| \le C e^{\frac{-|i-c|}{l}} \tag{45}$$

<sup>&</sup>lt;sup>c</sup>Of course, in the worst case scenario for a fixed eigenvalue, we would have  $|\lambda_i - \hat{\lambda}_i| = 2p^2$ , but as we are choosing an eigenvalue at random this would give  $|\lambda_i - \hat{\lambda}_i| = 0$  with high probability, which is better; the worst case is when the deviation is spread uniformly across all eigenvalues in the region of interest.

with  $c \leq N - \log(N)$  with high probability. As the bandwidth is equal to 2 in the upper region, we have  $l = \mathcal{O}(1)$  and the random  $C = \mathcal{O}(1)$ . Then

$$(\hat{H} - \lambda)\psi = H\psi - \lambda\psi + P\psi$$
  
=  $P\psi$ . (46)

Taking norms, we have

$$\begin{aligned} \|P\psi\|_{2}^{2} &= p^{2} |\psi_{N}|^{2} + p^{2} |\psi_{N+1}|^{2} \\ &\leq C^{2} |p|^{2} \left( e^{\frac{-2|N-c|}{l}} + e^{\frac{-2|N+1-c|}{l}} \right) \\ &\leq 2C^{2} |p|^{2} e^{\frac{-2|N-N+\log(N)|}{l}} \\ &= \mathscr{O} \left( N^{\frac{-2}{l}} \right), \end{aligned}$$
(47)

i.e. so  $\|P\psi\|_2 = \mathcal{O}(N^{-\frac{1}{t}})$  with high probability. The end result is that small eigenvalues are likely to have eigenvectors supported in regions of the lattice where conductor is less connected (i.e. the bandwidth is smaller). From this, we can also conclude that larger eigenvalues will have eigenvectors supported in regions where the conductor is thicker, as the thin region is already completely occupied. This "ball and chain" model thus serves as an extreme example of what we should expect to see in the case of interest, as the conductor is thinner at the edges. Importantly, because of this combined spacial/energy structure, we should be able to separate the boundary effects from the stochastic effects by examining

$$Y(Q) = \sum_{i,j} \mathbb{E}(|\psi_i^{(j)}|^2)^2.$$
(48)

Let us first consider the two limiting cases, when b = 1 and b = N. In both cases, Y(Q) will be equal to 1, because  $\mathbb{E}(|\psi_i^{(j)}|^2)^2 = \frac{1}{N^2}$  for all i, j. If our heuristics are correct about the nature of the boundary effect for 1 < b < N, then near the boundary  $\mathbb{E}(|\psi_i^{(j)}|^2)^2 > \frac{1}{N^2}$  for small eigenvalues and  $\mathbb{E}(|\psi_i^{(j)}|^2)^2 < \frac{1}{N^2}$  for large



Fig. 5. This figure shows Y(Q) against c with b = cN for a Gaussian NPE<sub>N,b,q</sub>, N = 100, 200. A maximum is again seen near  $b = \frac{2}{5}N$ .

eigenvalues; the net effect should become significant when  $b \simeq N$ , causing Y(Q) to reach a maximum somewhere in this regime. Figure 5 shows a numerical simulation of Y(Q) as a function of the bandwidth for several choices of N. The maximum is now clear, and remarkably, occurs exactly when the level density has maximum deviation at  $b = \frac{2}{5}N$ .

## 6. Conclusion

We have shown that the rate of convergence of fourth moment of the normalized level density for band matrices is independent of the boundary condition in the localized regime, but in the delocalized regime the rate of convergence for non-periodic band ensembles differs significantly from the periodic case. We also have examined the relationship between the boundary effect and the level density in the thick bandwidth case, finding the surprising result that when  $b = \frac{2}{5}N$  both the eigenvalue and eigenvector statistics are maximally deviated from their Wigner matrix counterparts. These results raise several questions:

- (1) Is there evidence of a phase transition in higher order moments? Numerical results suggest yes, see Fig. 6.
- (2) Is there evidence of a phase transition in the local statistics at  $b = \frac{2}{5}N$ ? We were unable to find numerical evidence of this, but given the lack of flatness of the eigenvectors in this regime, the existence of such a transition is plausible.
- (3) Is there evidence in the global statistics of the phase transition of the largest eigenvalue at  $N^{5/6}$  [18]? We have not detected any sign of this transition in



Fig. 6. Higher order moments for a Gaussian NPE with N = 100. Again, local maxima and minima are seen near  $b \approx \sqrt{N}$  and  $b = \frac{2}{5}N$ .

numerical experiments, and we suspect that this transition is too localized to have an observable effect on the global statistics.

## Acknowledgment

Andrew Swan is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) Grant 1648831, Random Matrices and Isospectral Flows.

## Appendix A

**Proof of Lemma 1.** We first note that the trace and expectation operators commute

$$\mathbb{E}(\operatorname{tr}(A)) = \operatorname{tr}(\mathbb{E}(A)) \tag{A.1}$$

for all matrices A. This is easy to see as a result of the linearity of expectation:

$$\mathbb{E}(\operatorname{tr}(A)) = \mathbb{E}\left(\sum_{i} a_{ii}\right) = \sum_{i} \mathbb{E}(a_{ii}) = \operatorname{tr}(\mathbb{E}(A)).$$

This result will be useful as it allows us to compute the expectation of independent matrices before taking their product, e.g.

$$\mathbb{E}(\operatorname{tr}(AB)) = \operatorname{tr}(\mathbb{E}(AB)) = \operatorname{tr}(\mathbb{E}(A)\mathbb{E}(B)).$$
(A.2)

The left-hand side of Eq. (21) may be expanded as

$$\mathbb{E}\operatorname{tr}(A^4) = \sum_{i,j}^b \mathbb{E}\,\tilde{a}_{ij}^2 = \sum_i^b \mathbb{E}\,\tilde{a}_{ii}^2 + \sum_{i\neq j}^b \mathbb{E}\,\tilde{a}_{ij}^2, \tag{A.3}$$

where  $\tilde{a}_{ij}$  are the elements of the matrix  $A^2$ . For the diagonal elements, we have

$$\mathbb{E}(\tilde{a}_{ii}^2) = \mathbb{E}\sum_{k=1}^{b} a_{ik}^4 + \mathbb{E}\sum_{k\neq l}^{b} a_{ik}^2 a_{il}^2$$
  
=  $b \mathbb{E} a^4 + b(b-1)(\mathbb{E} a^2)^2$   
=  $qb + b(b-1).$  (A.4)

For the off diagonal elements, we have

$$\mathbb{E}(\tilde{a}_{ij}^2) = \mathbb{E}\sum_{k=1}^{b} a_{ik}^2 a_{jk}^2 + \mathbb{E}\sum_{k\neq l}^{b} \sum_{l=1}^{b} a_{ik} a_{jk} a_{il} a_{jl}.$$
 (A.5)

As  $i \neq j$  and  $k \neq l$ , every term of second sum has at least two elements which are different and thus has expectation zero. Therefore,

$$\mathbb{E}(\tilde{a}_{ij}^2) = b \,\mathbb{E}\,a_1^2 a_2^2 = b. \tag{A.6}$$

The fourth moment of A is hence

$$\mathbb{E}\operatorname{tr}(A^4) = \sum_{i}^{b} qb + b(b-1) + \sum_{i \neq j}^{b} b$$
  
=  $b(qb + b(b-1)) + b^2(b-1)$   
=  $2b^3 + (q-2)b^2$ . (A.7)

For Eq. (22), we use the commutativity of the trace and expectation operators together with the independence of A and L to give

$$\mathbb{E}\operatorname{tr}(A^2LL^{\top}) = \operatorname{tr}(\mathbb{E}A^2\mathbb{E}(LL^{\top})).$$
(A.8)

The expectation of  $A^2$  is diagonal as

$$\mathbb{E} \,\tilde{a}_{ii} = \sum_{k=1}^{b} \mathbb{E} \,a_{ik}^2 = b$$

$$\mathbb{E} \,\tilde{a}_{ij} = \sum_{k=1}^{b} \mathbb{E} \,a_{ik} \,\mathbb{E} \,a_{jk} = 0.$$
(A.9)

Denoting the entries of  $LL^{\top}$  by  $\tilde{l}$  and using the fact that  $l_{ij}$  is zero if  $i \leq j$ , the diagonal elements are

$$\tilde{l}_{ii} = \sum_{k=1}^{b} l_{ik}^2 = \sum_{k=1}^{i-1} l_{ik}^2, \qquad (A.10)$$

the upper triangular elements i < j are

$$\tilde{l}_{ij} = \sum_{k=1}^{b} l_{ik} l_{jk} = \sum_{k=1}^{i-1} l_{ik} l_{jk}, \qquad (A.11)$$

and the lower triangular elements i > j are

$$\tilde{l}_{ij} = \sum_{k=1}^{b} l_{ik} l_{ij} = \sum_{k=1}^{j-1} l_{ik} l_{jk} = \tilde{l}_{ji}.$$
(A.12)

Taking expectations gives

$$\mathbb{E} \tilde{l}_{ii} = i - 1 \tag{A.13}$$
$$\mathbb{E} \tilde{l}_{ij} = 0$$

and thus  $\mathbb{E} LL^{\top}$  is also diagonal. We hence have

$$\mathbb{E}\operatorname{tr}(A^{2}LL^{\top}) = \sum_{i=1}^{b} \mathbb{E}\,\tilde{a}_{ii} \,\mathbb{E}\,\tilde{l}_{ii} = \sum_{i=1}^{b} b(i-1) = \frac{b^{3} - b^{2}}{2}.$$
 (A.14)

Similarly, for Eq. (23), we have

$$\mathbb{E}\operatorname{tr}(A^2L^{\top}L) = \operatorname{tr}(\mathbb{E}A^2\mathbb{E}(LL^{\top})).$$
(A.15)

The diagonal entries of  $LL^\top$  are

$$\tilde{l}_{ii} = \sum_{k}^{b} l_{ki}^2 = \sum_{k=i+1}^{b} l_{ki}^2$$
(A.16)

and the upper/lower triangular elements are

$$\tilde{l}_{ij} = \sum_{k=1}^{b} l_{ki} l_{kj} = \sum_{k=i+1}^{b} l_{ki} l_{kj} = \tilde{l}_{ji}.$$
(A.17)

Taking expectations gives

$$\mathbb{E} \tilde{l}_{ii} = b - i$$

$$\mathbb{E} \tilde{l}_{ij} = 0$$
(A.18)

so  $\mathbb{E} \, L L^\top$  is diagonal. Hence

$$\mathbb{E}\operatorname{tr}(A^{2}L^{\top}L) = \sum_{i=1}^{b} \mathbb{E}\,\tilde{a}_{ii} \,\mathbb{E}\,\tilde{l}_{ij} = \sum_{i=1}^{b} b(b-i) = \frac{b^{3}-b^{2}}{2}.$$
 (A.19)

Using (A.13) and (A.18), Eq. (24) is

$$\mathbb{E} \operatorname{tr}(L_1^{\top} L_1 L_2 L_2^{\top}) = \operatorname{tr}(\mathbb{E} L^{\top} L \mathbb{E} L L^{\top})$$
$$= \sum_{i=1}^{b} (i-1)(b-i)$$
$$= \frac{b^3 - 3b^2 + 2b}{6}.$$
(A.20)

Finally, Eq. (25) may be expressed as

$$\mathbb{E}\operatorname{tr}(LL^{\top}LL^{\top}) = \mathbb{E}\operatorname{tr}(LL^{\top}(LL^{\top})^{\top}) = \sum_{i,j}^{b} \mathbb{E}\tilde{l}_{ij}^{2} = \sum_{i}^{b} \mathbb{E}\tilde{l}_{ii}^{2} + \sum_{i\neq j}^{b} \mathbb{E}\tilde{l}_{ij}^{2}, \quad (A.21)$$

where the  $\tilde{l}_{ij}$  are the elements of  $LL^{\top}$ . The terms in the first sum are

$$\mathbb{E}\,\tilde{l}_{ii}^2 = \mathbb{E}\sum_{k=1}^{i-1} l_{ik}^4 + \mathbb{E}\sum_{k\neq m}^{i-1} l_{ik}^2 l_{im}^2$$
  
=  $(i-1) \mathbb{E}\,l^4 + (i-1)(i-2)(\mathbb{E}\,l^2)^2$   
=  $q(i-1) + (i-1)(i-2),$  (A.22)

and those in the second sum are

$$\mathbb{E}\,\tilde{l}_{ij}^2 = \mathbb{E}\sum_{k=1}^{i-1} l_{ik}^2 l_{ij}^2 + \mathbb{E}\sum_{k\neq m}^{i-1} \sum_m^{i-1} l_{ik} l_{jk} l_{im} l_{jm}$$
  
= (i-1). (A.23)

Thus,

$$\mathbb{E}\operatorname{tr}(LL^{\top}LL^{\top}) = \sum_{i}^{b} q(i-1) + (i-1)(i-2) + \sum_{i\neq j}^{b} (i-1)$$
$$= \frac{4b^3 - (3q-12)b^2 - (3q-8)b}{6}.$$
(A.24)

### References

- B. L. Altshuler and B. I. Shklovskii, Repulsion of energy levels and conductivity of small metal samples, *Zh. Eksp. Teor. Fiz.* 91 (1986) 220–234.
- [2] Z. Bao and L. Erdős, Delocalization for a class of random block band matrices, Probab. Theory Relat. Fields 167(3–4) (2017) 673–776.
- [3] L. Bogachev, S. Molchanov and L. Pastur, On the level density of random band matrices, Math. Notes Acad. Sci. USSR 50(6) (1991) 1232–1242.
- [4] P. Bourgade, L. Erdős, H.-T. Yau and J. Yin, Universality for a class of random band matrices, Adv. Theor. Math. Phys. 21(3) (2017) 739–800.
- [5] G. Casati, F. Izrailev and L. Molinari, Scaling properties of the eigenvalue spacing distribution for band random matrices, J. Phys. A, Math. Gen. 4755(20) (1991) 4755–4762.
- [6] P. Deift, Some open problems in random matrix theory and the theory of integrable systems, *Integrable Systems and Random Matrices in Honor of Percy Deift*, Contemporary Mathematics, Vol. 458 (American Mathematical Society, 2007) pp. 1–12.
- [7] L. Erdős and A. Knowles, Quantum diffusion and eigenfunction delocalization in a random band matrix model, *Comm. Math. Phys.* **303**(2) (2011) 509–554.
- [8] L. Erdős and A. Knowles, The Altshuler–Shklovskii formulas for random band matrices I: The unimodular case, Comm. Math. Phys. 333(3) (2015) 1365–1416.
- [9] L. Erdős and A. Knowles, The Altshuler–Shklovskii formulas for random band matrices II: The general case, Ann. Henri Poincaré 16(3) (2015) 709–799.
- [10] L. Erdős, A. Knowles, H. T. Yau and J. Yin, Delocalization and diffusion profile for random band matrices, *Comm. Math. Phys.* **323**(1) (2013) 367–416.
- [11] Y. V. Fyodorov and A. D. Mirlin, Scaling properties of localization in random band matrices: A σ-model approach, *Phys. Rev. Lett.* 67(18) (1991) 2405–2409.
- [12] S. A. Molchanov, L. A. Pastur and A. M. Khorunzhii, Limiting eigenvalue distribution for band random matrices, *Theoret. and Math. Phys.* **90**(2) (1992) 108–118.
- [13] J. Schenker, Eigenvector localization for random band matrices with power law band width, Comm. Math. Phys. 290(3) (2009) 1065–1097.
- [14] T. H. Seligman, J. J. M. Verbaarschot and M. R. Zirnbauer, Quantum spectra and transition from regular to chaotic classical motion, *Phys. Rev. Lett.* 53(3) (1984) 215–217.
- [15] T. Shcherbina, Universality of the local regime for the block band matrices with a finite number of blocks, J. Statist. Phys. 155(3) (2014) 466–499.
- [16] T. Shcherbina, On the second mixed moment of the characteristic polynomials of 1D band matrices, *Comm. Math. Phys.* **328**(1) (2014) 45–82.
- [17] M. Shcherbina and T. Shcherbina, Characteristic polynomials for 1D random band matrices from the localization side, *Comm. Math. Phys.* 351(3) (2017) 1009–1044.
- [18] S. Sodin, The spectral edge of some random band matrices, Ann. Math. 172(3) (2010) 2223–2251.