

# Rigidity and Edge Universality of Discrete $\beta$ -Ensembles

Alice Guionnet

*CNRS & Université de Lyon, ENSL, UMPA, UMR 5669*  
*E-mail: alice.guionnet@ens-lyon.fr*

Jiaoyang Huang

*Harvard University*  
*E-mail: jiaoyang@math.harvard.edu*

## Abstract

We study discrete  $\beta$ -ensembles as introduced in [17]. We obtain rigidity estimates on the particle locations, i.e. with high probability, the particles are close to their classical locations with an optimal error estimate. We prove the edge universality of the discrete  $\beta$ -ensembles, i.e. for  $\beta \geq 1$ , the distribution of extreme particles converges to the Tracy-Widom  $\beta$  distribution. As far as we know, this is the first proof of general Tracy-Widom  $\beta$  distributions in the discrete setting. A special case of our main results implies that under the Jack deformation of the Plancherel measure, the distribution of the lengths of the first few rows in Young diagrams, converges to the Tracy-Widom  $\beta$  distribution, which answers an open problem in [39]. Our proof relies on Nekrasov's (or loop) equations, a multiscale analysis and a comparison argument with continuous  $\beta$ -ensembles.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Background on Discrete $\beta$ -Ensemble	6
1.2	Main Results	10
<b>2</b>	<b>Rigidity of discrete <math>\beta</math>-Ensembles</b>	<b>11</b>
2.1	Rigidity of discrete $\beta$ -Ensembles	12
2.2	Bootstrap and Proof of Theorem 2.4	22
<b>3</b>	<b>Local Measure</b>	<b>29</b>
<b>4</b>	<b>Edge Universality</b>	<b>45</b>
<b>5</b>	<b>Multi-cut Case</b>	<b>50</b>

<b>6</b>	<b>Examples and Generalizations</b>	<b>58</b>
6.1	$\theta = 1$ Binomial Weight . . . . .	58
6.2	Arbitrary Convex Potential on $\mathbb{R}$ With No Saturation . . . . .	61
6.3	Lozenge Tilings . . . . .	62
6.4	Jack Deformation of Plancherel Measure . . . . .	65
<b>A</b>	<b>Hilbert Transform on a finite interval <math>[a, b]</math>.</b>	<b>71</b>
<b>B</b>	<b>Rigidity and Proof of Proposition 3.2 and 3.3</b>	<b>71</b>
<b>C</b>	<b>Proof of Propositions 3.17, 3.18, 3.19 and 3.20</b>	<b>75</b>
<b>D</b>	<b>Level Repulsion and Proof of Proposition 4.3</b>	<b>85</b>

# 1 Introduction

During the eighteenth century, De Moivre was the first to show that the Gaussian (or Normal) distribution describes the fluctuations of the sums of independent binomial variables. Soon after, Laplace generalized his result to more general independent variables. Gauss advertised this central limit theorem by showing that it allows the evaluation of the errors in a large class of systems characterized by independence, and many other proofs of the central limit theorems were given, see e.g. [10, 42, 73, 97]. During the last twenty years, understanding the fluctuations of much more correlated systems became of central interest. Instead of the Gaussian distribution, some of these systems kept producing another statistical curve, which had become known as the Tracy-Widom distribution. A typical example concerns the eigenvalues of random matrices with independent entries (up to the symmetry constraint). The eigenvalues are far from being independent and are in general ordered. The fluctuations of the extreme eigenvalues are described by the Tracy-Widom distribution with an exotic scaling exponent two-third. Such results were first proven for very specific examples that are integrable, namely the Gaussian ensembles [52, 77, 100, 101]. It was much more recently proven that these results are universal, in the sense that the local fluctuations do not depend too much on the microscopic details of the random matrices, i.e. single entry distribution, sparsity, see e.g. [44, 61, 68, 70, 96, 98]. It is worth noting that this universality is always proven by comparison to integrable models and there is not yet a direct approach to this problem for general models. Besides matrix models, the Tracy-Widom distributions also appear in a large class of models which a priori are very different. In the beautiful article [3], it was shown that the length of the longest increasing subsequence of a uniform random permutation is also asymptotically described by the Tracy-Widom distribution. Other highly correlated systems such as uniform random lozenge tiling [60], interacting particle systems [13, 15, 102], polymers in disordered environments [14, 92] and the corner growth model [56], fluctuate following the Tracy-Widom distribution. Investigating the universality class of the Tracy-Widom distribution has become a key issue in probability theory. This article investigates the universality class of the so-called Tracy-Widom  $\beta$  distributions, which has been introduced and studied in [91], characterizing the fluctuations of the extreme particles of Gaussian  $\beta$ -ensembles. We show that a natural family of discrete models, the discrete  $\beta$ -ensembles as introduced in [17], and the Jack deformation of the Plancherel measure [83] belong to the universality class of the Tracy-Widom  $\beta$  distributions.

---

This work of A.G. is partially supported by the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon.

We fix the parameter  $\theta = \beta/2$  and a sequence of positive real-valued functions  $w(x; N)$ . A discrete  $\beta$ -ensemble is a sequence of distributions

$$\mathbb{P}_N(\ell_1, \ell_2, \dots, \ell_N) = \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} \frac{\Gamma(\ell_j - \ell_i + 1) \Gamma(\ell_j - \ell_i + \theta)}{\Gamma(\ell_j - \ell_i) \Gamma(\ell_j - \ell_i + 1 - \theta)} \prod_{i=1}^N w(\ell_i; N), \quad (1.1)$$

on ordered  $N$ -tuples  $\ell_1 < \ell_2 < \dots < \ell_N$ , such that  $\ell_i = a(N) + \lambda_i + \theta i$  and  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N < b(N) - a(N) - N\theta$  are integers. We refer to  $w(x; N)$  as the weight of the discrete  $\beta$ -ensemble, and  $\ell_i$  as the positions of particles. We denote the configuration space by  $\mathbb{W}_N^\theta$ . When  $\theta$  equals one, the particles live on  $\mathbb{Z}$ , but in general the possible positions of a particle depend on the position of the previous particle, and they do not live on a fixed lattice. Let us note that when  $\theta = 1$  or  $1/2$  the above ratio of gamma functions is simply the usual Coulomb gas interaction  $\prod |\ell_i - \ell_j|^{2\theta}$ .

When  $\theta$  equals one, the distribution (1.1) arises in numerous problems of statistical mechanics and asymptotic representation theory, e.g., the distribution of uniformly random domino tilings of the Aztec diamond (cf. [59]), the distribution of uniformly random lozenge tilings of an abc-hexagon (cf. [59]), last passage percolation (cf. [57]), stochastic systems of non-intersecting paths (cf. [59, 62]), and the representation theory of the infinite dimensional unitary group  $U(\infty)$  (cf. [22, Section 5], [12, Section 4]). One of the most heavily studied examples is the random lozenge tilings of an abc-hexagon. The distribution of the positions of horizontal lozenges on a vertical line is given by such a discrete  $\beta$ -ensemble with  $\theta = 1$  and  $w(x; N)$  certain binomial weight [57, 59]; it is called the Hahn ensemble. In fact, more general domains can be handled provided they can be described in terms of Gelfand-Tsetlin patterns [23, 33, 54, 89]. Choosing uniformly random lozenge tilings of those domains, the distributions of the positions of horizontal lozenges on a vertical line are given by discrete  $\beta$ -ensemble with  $\theta = 1$  and more complicated binomial weights, e.g., [17, Section 9.2]. For general  $\theta > 0$  and special weights, the distribution (1.1) appears in [85], called the  $zw$ -measures. The normalization constants in the  $zw$ -measures are explicit, which could be interpreted as that the special form of the interaction in  $\mathbb{P}_N$ , a ratio of Gamma functions, is a discrete analogue of the Selberg integral. Other families of related measures are the Jack deformation of the Plancherel measures on Young diagrams and, more generally, the  $z$ -measures [21, 39, 53, 74, 75, 83, 86]. We remark that the Jack deformation of the Plancherel measures and  $z$ -measures on Young diagrams are slightly different from the discrete  $\beta$ -ensemble (1.1), where the number of particles is fixed.

Moreover, the discrete  $\beta$ -ensemble (1.1) is a natural discrete analogue of the continuous  $\beta$ -ensemble given by

$$d\mathbb{P}_N^{\text{cont}}(\lambda_1, \dots, \lambda_N) = \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^\beta e^{-N \sum_{i=1}^N V(\lambda_i)} \prod_{i=1}^N d\lambda_i.$$

For classical values of  $\beta = 1, 2, 4$  and  $V(x) = \beta x^2/4$ ,  $\mathbb{P}_N^{\text{cont}}$  corresponds to the joint law of the eigenvalues of the Gaussian Orthogonal (with real entries), Gaussian Unitary (with complex entries) or Gaussian Symplectic (with quaternion entries) Ensembles. When  $V$  is quadratic, Dumitriu and Edelman [41] constructed for any  $\beta > 0$  a real symmetric tri-diagonal matrix with independent chi-square and Gaussian entries, whose eigenvalues are distributed according to  $\mathbb{P}_N^{\text{cont}}$ .

The continuous  $\beta$ -ensembles could be studied in details. It is well known that the particles of  $\beta$ -ensembles are asymptotically distributed according to a so-called equilibrium measure. The fluctuations of the empirical measures could be studied in the case where the equilibrium measure has a connected support [26, 55, 63] and in the multi-cut case [25, 27, 94]. Local fluctuations were first studied in the case  $\beta = 2$  where the

determinantal structure of the law provides specific tools. They have been first analyzed in the pioneering works of Gaudin, Mehta and Dyson [43, 76, 78] (see [77] for a review), where it was shown that the local correlation statistics are governed by the sine kernel for the GUE in the bulk. At the edge, the local fluctuations for the GUE were first identified by Tracy and Widom [100] to be asymptotically given by the Tracy-Widom distribution. Later, bulk universality was proven in [11] for  $V$  polynomial, [35] for  $V$  analytic and [87, 88] for  $V$  locally  $C^3$ . [71] provides the most general result for bulk universality in the case  $\beta = 2$ . The edge universality for analytic  $V$  was proven in [35]. In the case  $\beta = 1, 4$ , the distribution is Pfaffian and more difficult to analyze. For quadratic  $V$ , the local statistics were derived in [43, 76] for the bulk and [101] for the edge. Universality was proven in [37] in the bulk, and [36] at the edge, for monomial potentials  $V$  (see [34] for a review). For more general potentials, universality was shown in [63] ( $\beta = 1, 4$ , one-cut case) and [93] ( $\beta = 1, 4$ , multi-cut case). The local fluctuations of more general  $\beta$ -ensembles were only derived recently [91, 104] for quadratic  $V$ , based on Dumitriu-Edelman matrix representation. It was shown that correlation functions in the bulk are given by the Sine  $\beta$  kernel, and local fluctuations at the edge converge to the Tracy-Widom  $\beta$  law. Universality in  $\beta$ -ensembles was first addressed in the bulk [28, 30] ( $\beta > 0$ ,  $V \in C^4$ ), then at the edge [29] ( $\beta \geq 1$ ,  $V \in C^4$ ) and [64] ( $\beta > 0$ ,  $V$  convex polynomial). Later, the transportation of measure approach was developed to prove the universality of local fluctuations [95] (bulk,  $\beta > 0$ ,  $V$  analytic, both one-cut and multi-cut cases), [6] (both bulk and edge,  $\beta > 0$ ,  $V \in C^{31}$ , one-cut case) and [8] (edge,  $\beta > 0$ ,  $V$  analytic, multi-cut case).

A central point to study mesoscopic or microscopic fluctuations of  $\beta$ -ensembles is to derive rigidity estimates showing that particles are very close to their deterministic limit. They are usually formulated in terms of concentration of the Stieltjes transform of the empirical particle density at short scales. These results were first established for Wigner matrices in a series of breakthrough papers [47–50, 99], then extended to other matrix models, i.e. sparse random matrices [45], deformed Wigner ensembles [66, 69]. Beyond matrix models, rigidity estimates have been established for one-cut and multi-cut  $\beta$ -ensembles [28–30, 72], and two-dimensional Coulomb gas [5, 67]. Another key tool to study fluctuations is the loop (or Dyson-Schwinger) equations. They were introduced to the mathematical community by Johansson [55] to derive macroscopic central limit theorems for general  $\beta$ -ensembles, see also [25, 26, 63]. They were later on used to study fluctuations on microscopic scales [29, 30] and mesoscopic scales [7]. The central idea is to analyze these equations by linearizing them around the limit, hence obtaining linear equations at the first order, and to show that higher order terms are negligible by concentration of measure arguments. To this end, one needs to solve the linear equations, which amounts to invert some linear operators. This can easily be done when the equilibrium measure has a connected support with density vanishing like a square root near the edges. In this case, we say the potential and the equilibrium measure are off-critical. It can be shown (see [65]) that most potentials are off-critical, in the sense that if we multiply the potential by a constant, the equilibrium measure will be off-critical for almost all such constants. Critical potentials lead to different fluctuations at the edge [38].

It is natural to seek for similar results for the discrete  $\beta$ -ensembles. As for continuous models, the convergence of the empirical measure of the particles towards an equilibrium measure is easy to prove, and a large deviation principle can be derived [51]. However, proving global or local fluctuations is much more challenging in the discrete setting than in the continuous one. When  $\beta = 2$ , the ensembles are called the discrete orthogonal polynomial ensembles, which can be analyzed thanks to their determinantal structure. The central limit theorem for the empirical measure was proven in [16, 31, 32, 90]. Local fluctuations were studied for a large family of integrable choices of weights in [18, 20, 57, 59, 60, 84] and it was shown that local fluctuations in the bulk converge to the discrete sine kernel, whereas Tracy-Widom distribution describes the fluctuations at the edge. [1, 2] provide the most general results for the local fluctuations in the case  $\beta = 2$ .

However, these results mentioned above depend on the integrability of the system, and do not include the multi-cut case.

In [17], the first macroscopic central limit theorems for discrete  $\beta$ -ensembles were proved for general  $\beta$  in multi-cut setting. It was shown in particular that the covariance structure of the central limit theorem is the same as that in the continuous case. The proof relies on the Nekrasov's equations, which originated in the work of Nekrasov and his collaborators [79–81] and requires the specific form of the interaction in  $\mathbb{P}_N$  (a ratio of Gamma functions rather than the Coulomb interaction). These equations can be analyzed in a spirit very close to the uses of loop (or Dyson-Schwinger) equations in the continuous setting.

In this paper, we study the local fluctuations of the positions of extreme particles of discrete  $\beta$ -ensembles (also the positions of extreme holes in the case  $\beta = 2$  by duality) in off-critical situations (corresponding to the case where the density behaves like a square root near zero or  $\theta^{-1}$ ). We show that these local fluctuations at the edge are the same as for continuous  $\beta$ -ensembles, i.e. are governed by the Tracy-Widom  $\beta$  distributions. This reflects that the fluctuations of extreme particles are in a much larger space size than the mesh size of the lattice. The central point is thus to obtain rigidity estimates of the particle locations on the optimal scale, or at least on a scale comparable to the scale of the fluctuations. With the rigidity estimates as input, we then can prove universality by a comparison of discrete and continuous local measures, for which the edge universality was proven in [29]. Typically, the continuous models can be studied as the degenerations of discrete models. However, our proof goes in the opposite direction, i.e. we prove the universality of discrete models by comparing with continuous models. We believe that this approach can be further developed to study other discrete models.

Our approach to rigidity estimates relies on the Nekrasov's equations from [17], and a multiscale analysis combining ideas from [5, 28–30, 67]. Indeed, a priori Nekrasov's equation only allow us to obtain fine estimates on the convergence to the equilibrium measure up to a scale of order the square root of the mesh of the lattice. To improve the estimates, we use a multi-scale analysis based on local measures, which are the measures on a small number of consecutive particles that are obtained by fixing all other particles which act as boundary conditions. These measures play a crucial role in the proof of the rigidity of continuous  $\beta$ -ensembles in [30]. However, the analysis in [30] heavily depends on the Brascamp-Lieb inequality and the concentration of measure from the logarithmic Sobolev inequality, which do not exist in our case. Instead, we develop a method to obtain the rigidity estimates using only the loop equations. As far as we know, this is the first rigidity result for the discrete models. Our multiscale iteration approach is as follows. The optimal rigidity estimates at macroscopic scale were proven in [17]. With the rigidity estimates on a larger scale, we derive large deviation estimates of the local measures. Those large deviation estimates all together lead to weak rigidity estimates of the original discrete  $\beta$ -ensemble, but in a finer scale. We then use the Nekrasov's equation to bootstrap the rigidity estimates up to optimal errors. By iterating this procedure, after finite steps, we get the rigidity estimates on the optimal scale.

We generalize rigidity and the edge universality results in section 5 to the case of multi-cut models with fixed filling fractions. We then apply these results to the Krawtchouk ensembles, which correspond to binomial weights in section 6.1. We also generalize to cases where particles are not constrained to stay within a bounded set but are submitted to a uniformly convex potential in section 6.2. We discuss also the case of uniformly random lozenge tilings of a hexagon with a hole in section 6.3. It is similar to multi-cut models but where filling fractions are not fixed. Under the assumption that the filling fractions do not fluctuate too much (that is verified in a work in progress of G. Borot, V. Gorin and A. Guionnet [24]), we show that there is universality of the fluctuations at the band edges. Finally in section 6.4, we use the universality result for the discrete  $\beta$ -ensembles to study the Jack deformation of the Plancherel measure. We prove that

the distribution of the lengths of the first few rows in Young diagrams under the Jack defromation of the Plancherel measure, converges to the Tracy-Widom  $\beta$  distribution for  $\beta \geq 1$ , which answers an open problem in [39].

*Conventions.* We use  $C$  to represent large universal constants, and  $c$  small universal constants, which may be different from line by line. Let  $Y \geq 0$ . We write that  $X = O(Y)$  if there exists some universal constant such that  $|X| \leq CY$ . We write  $X = o(Y)$ , or  $X \ll Y$  if the ratio  $|X|/Y \rightarrow 0$  as  $N$  goes to infinity. We write  $X \asymp Y$  if there exist universal constants such that  $cY \leq |X| \leq CY$ . We write  $X \lesssim Y$  if there exist universal constant such that  $X \leq CY$ . We write  $X \gtrsim Y$  if there exist universal constants such that  $X \geq cY$ . We denote the set  $\{1, 2, \dots, N\}$  by  $\llbracket 1, N \rrbracket$ . We denote  $\mathbb{Z}_{\geq 0}$  the set of non-negative integers, and  $\mathbb{Z}_{>0}$  the set of positive integers. We say an event  $\Omega$  holds with high probability, if there exist universal constant  $c$ , and  $N \geq N_0(c)$  large enough, so that  $\mathbb{P}(\Omega) \geq 1 - \exp(-c(\ln N)^2)$ .

**Acknowledgements:** We thank A. Borodin and E. Dimitrov for useful comments on the draft of this paper. J.H. thanks R. Bauerschmidt and M. Nikula for explaining their insights of [5], and A. Adhikari, A. Borodin, P. Sosoë and H-T. Yau for helpful discussions. This work was partly supported by LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon and École Normale Supérieure de Lyon.

## 1.1 Background on Discrete $\beta$ -Ensemble

In this section we collect several assumptions of the discrete  $\beta$ -ensembles from [17]. For each  $N > 0$ , we take an interval  $(a(N), b(N))$  such that  $b(N) - a(N) - N\theta \in \mathbb{Z}_{>0}$ .

**Definition 1.1.** *The state space  $\mathbb{W}_N^\theta$  consists of  $N$ -tuples  $a(N) < \ell_1 < \ell_2 < \dots < \ell_N < b(N)$ :*

1.  $\ell_1 - a(N) \in \mathbb{Z}_{>0}$ , and  $b(N) - \ell_N \in \mathbb{Z}_{>0}$ .
2. For  $i \in \llbracket 1, N-1 \rrbracket$ ,  $\ell_{i+1} - \ell_i - \theta \in \mathbb{Z}_{\geq 0}$ .

We need to assume that the weights  $w(\ell; N)$  as in (1.1) depend on  $N$  in a regular way.

**Assumption 1.2.** *There exists  $-\infty < \hat{a} < \hat{b} < +\infty$  such that as  $N \rightarrow \infty$ ,*

$$a(N) = N\hat{a} + O(\ln(N)), \quad b(N) = N\hat{b} + O(\ln(N)). \quad (1.2)$$

*We require that  $w(a(N); N) = w(b(N); N) = 0$ , and on  $[a(N) + 1, b(N) - 1]$ , it has the form*

$$w(x; N) = \exp\left(-NV_N\left(\frac{x}{N}\right)\right),$$

*for a function  $V_N$  that is continuous in the intervals  $[(a(N) + 1)/N, (b(N) - 1)/N]$ , and such that*

$$V_N(u) = V(u) + O\left(\frac{\ln(N)}{N}\right),$$

*uniformly over  $u \in [(a(N) + 1)/N, (b(N) - 1)/N]$ . The function  $V(u)$  is twice continuously differentiable and the following bound holds for a constant  $C > 0$ ,*

$$|V'(u)| \leq C(1 + |\ln(u - \hat{a})| + |\ln(u - \hat{b})|).$$

One important tool for the study of measures  $P_N$  is the Nekrasov's equation of Theorem 1.4 for which we will need the following additional assumption.

**Assumption 1.3.** *There exists an open set  $\mathcal{M}$  of the complex plane, which contains the interval  $[\hat{a}, \hat{b}]$ , and functions  $\psi_N^\pm(x)$  such that*

$$\frac{w(x; N)}{w(x-1; N)} = \frac{\psi_N^+(x)}{\psi_N^-(x)},$$

which satisfy

$$\psi_N^\pm(x) = \phi^\pm\left(\frac{x}{N}\right) + \mathcal{O}\left(\frac{1}{N}\right),$$

uniformly over  $x/N$  in compact subsets of  $\mathcal{M}$ . All the aforementioned functions are holomorphic in  $\mathcal{M}$ .

**Theorem 1.4.** *Let  $\mathbb{P}_N$  be a distribution on  $N$ -tuples  $(\ell_1, \ell_2, \dots, \ell_N) \in \mathbb{W}_N^\theta$  as in (1.1), with*

$$\frac{w(x, N)}{w(x-1, N)} = \frac{\psi_N^+(x)}{\psi_N^-(x)}.$$

We define

$$R_N(\xi) = \psi_N^-(\xi) \mathbb{E}_{\mathbb{P}_N} \left[ \prod_{i=1}^N \left(1 - \frac{\theta}{\xi - \ell_i}\right) \right] + \psi_N^+(\xi) \mathbb{E}_{\mathbb{P}_N} \left[ \prod_{i=1}^N \left(1 + \frac{\theta}{\xi - \ell_i - 1}\right) \right].$$

If  $\psi_N^\pm(\xi)$  are holomorphic in a domain  $N\mathcal{M} = \{Nz : z \in \mathcal{M}\} \in \mathbb{C}$ , then so is  $R_N(\xi)$ .

We denote the empirical particle density as  $\mu_N = N^{-1} \sum_{i=1}^N \delta_{\ell_i/N}$ . Let  $\mu(dx) = \rho_V(x)dx$  be the constrained equilibrium measure of the discrete  $\beta$ -ensemble  $\mathbb{P}_N$ , which is given by the minimizer of the following functional

$$-\theta \int_{x \neq y} \ln|x-y| \rho(x) \rho(y) dx dy + \int V(x) \rho(x) dx, \quad (1.3)$$

over all the densities  $\rho(x)$ , supported on  $[\hat{a}, \hat{b}]$ , with  $0 \leq \rho(x) \leq \theta^{-1}$ . We define  $\hat{a}_N := \min\{a(N)/N, \hat{a}\}$  and  $\hat{b}_N := \max\{b(N)/N, \hat{b}\}$ . Then both the empirical particle density  $\mu_N$  and the constrained equilibrium measure  $\mu$  are supported on  $[\hat{a}_N, \hat{b}_N]$ . We have the following characterization of the constrained equilibrium measure  $\mu$ . The restriction  $0 \leq \rho_V(x) \leq \theta^{-1}$  leads to the subdivision of  $[\hat{a}, \hat{b}]$  into three types of regions (we borrow the terminology from [1]):

- Maximal (with respect to inclusion) closed connected intervals where  $\rho_V(x) = 0$  are called voids.
- Maximal open connected intervals where  $0 < \rho_V(x) < \theta^{-1}$  are called bands.
- Maximal closed connected intervals where  $\rho_V(x) = \theta^{-1}$  are called saturated regions.

In a related context of random tilings and periodically-weighted dimers the voids and saturated regions are usually called frozen, while bands are liquid regions. There exists a constant  $f_V$  so that the effective potential  $F_V$  given by

$$F_V(x) := -2\theta \int_{\hat{a}}^{\hat{b}} \ln|x-y| \rho_V(y) dy + V(x) - f_V, \quad (1.4)$$

satisfies:

1.  $F_V(x) \geq 0$ , for all  $x$  in voids in  $[\hat{a}, \hat{b}]$ ;
2.  $F_V(x) \leq 0$ , for all  $x$  in saturated regions in  $[\hat{a}, \hat{b}]$ ;
3.  $F_V(x) = 0$ , for all  $x$  in bands in  $[\hat{a}, \hat{b}]$ .

We denote the classical particle locations  $\gamma_1, \gamma_2, \dots, \gamma_N$  corresponding to the constrained equilibrium measure  $\mu$ , as

$$\frac{i - 1/2}{N} = \int_{\hat{a}}^{\gamma_i} \rho_V(x) dx, \quad i \in \llbracket 1, N \rrbracket. \quad (1.5)$$

A convenient way to study the particle system  $\mathbb{P}_N$  is through the Stieltjes transform  $G_N$  of the empirical particle density  $\mu_N$ ,

$$G_N(z) = \frac{1}{N} \int \frac{\mu_N(x)}{z - x} = \frac{1}{N} \sum_{i=1}^N \frac{1}{z - \ell_i/N}. \quad (1.6)$$

We also define the Stieltjes transform  $G_\mu$  of the equilibrium measure  $\mu$ ,

$$G_\mu(z) = \int \frac{\rho_V(x) dx}{z - x}. \quad (1.7)$$

Notice that  $G_N$  and  $G_\mu$  are well defined outside  $[\hat{a}_N, \hat{b}_N]$ , and are holomorphic there.

We next define the following two functions  $R_\mu$  and  $Q_\mu$ , which are important for our asymptotic study,

$$\begin{aligned} R_\mu(z) &:= \phi^-(z) e^{-\theta G_\mu(z)} + \phi^+(z) e^{\theta G_\mu(z)}, \\ Q_\mu(z) &:= \phi^-(z) e^{-\theta G_\mu(z)} - \phi^+(z) e^{\theta G_\mu(z)}. \end{aligned} \quad (1.8)$$

It is proven in [17] that under Assumptions 1.2 and 1.3,  $R_\mu(z)$  is analytic on  $\mathcal{M}$ . For the function  $Q_\mu(z)$ , we assume

**Assumption 1.5.** *We assume there exists a function  $H(z)$  holomorphic in  $\mathcal{M}$  and numbers  $A, B$  such that*

- $\hat{a} < A < B < \hat{b}$ ;
- $Q_\mu(z) = H(z) \sqrt{(z - A)(z - B)}$ , where the branch of square root is such that  $\sqrt{(z - A)(z - B)} \sim z$  as  $z \rightarrow \infty$ ;
- $H(z) \neq 0$  for all  $z$  in a neighborhood of  $[\hat{a}, \hat{b}]$ .

**Remark 1.6.** *Assumption 1.5 is slightly different from [17, Assumption 4], i.e. we do not allow that  $\hat{a} = A$  or  $\hat{b} = B$ . If this is the case, the equilibrium measure might not have square root behaviors at band edges. For example, for the Krawtchouk orthogonal polynomial ensemble in section 6.1, if  $\mathbf{m} = 2$ , then  $\hat{a} = A = 0$ ,  $\hat{b} = B = 2$  and the equilibrium measure  $\mu_{\mathbf{m}}(x) = \mathbf{1}_{[0,2]}(x)/2$ , which does not have square root behaviors at the edges.*



For any function  $F(z)$  defined on the complex plane  $\mathbb{C}$ , we denote,

$$F(E + 0i) := \lim_{\eta \rightarrow 0^+} F(E + \eta i), \quad F(E - 0i) := \lim_{\eta \rightarrow 0^+} F(E - \eta i).$$

The measure  $\mu$  can be recovered by its Stieltjes transform. And more precisely, we have

$$G_\mu(E + 0i) = P.V. \int \frac{\rho_V(x) dx}{E - x} - \pi \rho_V(E) i. \quad (1.9)$$

We assume Assumption 1.5. It follows from (1.8) and (1.9), the density  $\rho_V$  satisfies

$$\frac{R_\mu(E) + Q_\mu(E + 0i)}{R_\mu(E) + Q_\mu(E - 0i)} = e^{2\pi\theta i \rho_V(E)}. \quad (1.10)$$

For  $E \in (A, B)$ , we have  $Q_\mu(E \pm 0i) = \pm H(E) i \sqrt{(B-E)(E-A)} \neq 0$ . Therefore  $0 < \rho_V(E) < \theta^{-1}$ . If  $R_\mu(A), R_\mu(B) \neq 0$ , then either  $\rho_V(E)$  or  $\theta^{-1} - \rho_V(E)$  has square root behaviors at  $A, B$ . Otherwise, by symmetry, we assume  $R_\mu(A) = 0$ . We also have  $Q_\mu(A) = 0$ , then from (1.8)  $\phi^\pm(A) = 0$ , which is impossible since  $\hat{a} < A < \hat{b}$ . For  $E \notin [A, B]$ , we have  $Q_\mu(E \pm 0i) = H(E) \sqrt{(E-B)(E-A)} \neq 0$ . Therefore  $e^{2\pi\theta i \rho_V(E)} = 1$ , so  $\rho_V(E) \equiv 0$  or  $\rho_V(E) \equiv \theta^{-1}$ . Notice that under Assumptions 1.2 and 1.3, we have  $e^{-V'(x)} = \phi^+(x)/\phi^-(x)$ . For  $E \notin [A, B]$ , we have  $Q_\mu(E \pm 0i) = H(E) \sqrt{(E-A)(E-B)} \neq 0$ . It follows from (1.8) and (1.9), the Stieltjes transform  $G_\mu$  satisfies

$$\frac{R_\mu(E) + Q_\mu(E + 0i)}{R_\mu(E) - Q_\mu(E - 0i)} = e^{V'(E) - 2\theta \operatorname{Re}[G_\mu(E)]} = e^{F'_V(E)}, \quad (1.11)$$

where  $F'_V(E)$  is defined in (1.4). As a result,  $F'_V(E)$  has square root behaviors at  $A, B$ . We remark that, the righthand sides of (1.10) and (1.11) can be rewritten as the same expression,

$$\frac{R_\mu(E) + Q_\mu(E + 0i)}{R_\mu(E) - Q_\mu(E - 0i)} = \begin{cases} e^{2\pi\theta i \rho_V(E)}, & E \in [A, B], \\ e^{F'_V(E)}, & E \notin [A, B]. \end{cases} \quad (1.12)$$

In summary, we have the following four cases. There exists  $\varepsilon > 0$  so that:

1. the constrained equilibrium measure is supported on  $[A, B]$ . There exist analytic functions  $s_A$  and  $s_B$  such that  $s_A(x)\sqrt{x-A} = 2\pi\theta\rho_V(x)$  on  $[A, A+\varepsilon]$ ,  $s_A(x)\sqrt{A-x} = -F'_V(x)$  on  $[A-\varepsilon, A]$ ; and  $s_B(x)\sqrt{B-x} = 2\pi\theta\rho_V(x)$  on  $[B-\varepsilon, B]$ ,  $s_B(x)\sqrt{x-B} = F'_V(x)$  on  $x \in (B, B+\varepsilon)$ .
2. the constrained equilibrium measure is supported on  $[A, \hat{b}]$ , and  $\rho(x) \equiv \theta^{-1}$  on  $[B, \hat{b}]$ . There exist analytic functions  $s_A$  and  $s_B$  such that  $s_A(x)\sqrt{x-A} = 2\pi\theta i \rho_V(x)$  on  $[A, A+\varepsilon]$ ,  $s_A(x)\sqrt{A-x} = -F'_V(x)$  on  $[A-\varepsilon, A]$ ; and  $s_B(x)\sqrt{B-x} = 2\pi - 2\pi\theta\rho_V(x)$  on  $[B-\varepsilon, B]$ ,  $s_B(x)\sqrt{x-B} = -F'_V(x)$  on  $x \in (B, B+\varepsilon)$ .
3. the constrained equilibrium measure is supported on  $[\hat{a}, B]$ , and  $\rho(x) \equiv \theta^{-1}$  on  $[\hat{a}, A]$ . There exist analytic functions  $s_A$  and  $s_B$  such that  $s_A(x)\sqrt{x-A} = 2\pi - 2\pi\theta\rho_V(x)$  on  $[A, A+\varepsilon]$ ,  $s_A(x)\sqrt{A-x} = F'_V(x)$  on  $[A-\varepsilon, A]$ ; and  $s_B(x)\sqrt{B-x} = 2\pi\theta\rho_V(x)$  on  $[B-\varepsilon, B]$ ,  $s_B(x)\sqrt{x-B} = F'_V(x)$  on  $x \in (B, B+\varepsilon)$ .
4. the constrained equilibrium measure is supported on  $[\hat{a}, \hat{b}]$ ,  $\rho(x) \equiv \theta^{-1}$  on  $[\hat{a}, A]$  and  $\rho(x) \equiv \theta^{-1}$  on  $[B, \hat{b}]$ . There exist analytic functions  $s_A$  and  $s_B$  such that  $s_A(x)\sqrt{x-A} = 2\pi - 2\pi\theta\rho_V(x)$  on  $[A, A+\varepsilon]$ ,  $s_A(x)\sqrt{A-x} = F'_V(x)$  on  $[A-\varepsilon, A]$ ; and  $s_B(x)\sqrt{B-x} = 2\pi - 2\pi\theta\rho_V(x)$  on  $[B-\varepsilon, B]$ ,  $s_B(x)\sqrt{x-B} = -F'_V(x)$  on  $x \in (B, B+\varepsilon)$ .

For the proof of edge universality, we need stronger control on the potential  $V_N$  (as in Assumption 1.2), i.e. the difference between the derivative of  $V_N$  and the derivative of its limit is negligible.

**Assumption 1.7.** *For any  $u$  in a small neighborhood of  $[A, B]$ ,  $V_N(u)$  is analytic and the following holds*

$$V'_N(u) = V'(u) + O\left(N^{-1/3}\right).$$

## 1.2 Main Results

The goal of this paper is to prove the edge universality of discrete  $\beta$ -ensemble. In the rest of the paper, we assume that  $\hat{a} < A$  and  $[\hat{a}, A]$  is a void region, i.e.  $\rho_V(x) = 0$  on  $[\hat{a}, A]$ . As discussed in the previous section, the density vanishes like a square root at  $A$ ,

$$\rho_V(u) = (1 + O(u - A))s_A\sqrt{u - A}/\pi, \quad u \rightarrow A^+.$$

The proof of the edge universality consists of two steps. In the first step, we prove the rigidity of the particle configuration, i.e. with high probability, the particle locations are very close to their classical locations up to optimal scale, which occupies the main part of the paper. With the rigidity of the particle configuration as input, the edge universality of discrete  $\beta$ -ensembles follows from a direct comparison with continuous  $\beta$ -ensemble, for which the edge universality was proven in [29].

To state the rigidity theorem, we need some more definitions. We fix a small parameter  $\mathfrak{a} > 0$ , and define the spectral domain  $\mathcal{D}_r$  given for  $r \geq 0$  by

$$\begin{aligned} \mathcal{D}_r &:= \mathcal{D}_r^{\text{int}} \cup \mathcal{D}^{\text{ext}}, \\ \mathcal{D}_r^{\text{int}} &:= \{E + i\eta \in \mathcal{M} \cap \mathbb{C}_+ : E \in [A, B - r], \eta\sqrt{\kappa_E + \eta} \geq N^{\mathfrak{a}}/N\}, \\ \mathcal{D}^{\text{ext}} &:= \{E + i\eta \in \mathcal{M} \cap \mathbb{C}_+ : E \leq A, \eta \geq (N^{\mathfrak{a}}/N)^{2/3}\}, \end{aligned} \tag{1.13}$$

where  $\kappa_E = \text{dist}(E, \{A, B\})$ .

**Theorem 1.8.** *We assume Assumptions (1.2), (1.3), (1.5) and that  $[\hat{a}, A]$  is a void region, i.e.  $\rho_V(x) = 0$  on  $[\hat{a}, A]$ , then the following holds:*

1. Fix  $s \geq (\ln N)^2$  and small  $r > 0$ . With probability at least  $1 - e^{-cs}$ , we have uniformly for any  $z = E + i\eta \in \mathcal{D}_r$ ,

$$|G_N(z) - G_\mu(z)| \leq \frac{s}{N\eta}. \tag{1.14}$$

2. Fix a small  $\mathfrak{c}$  such that  $0 < \mathfrak{c} < \mathfrak{a}/4$ . With probability at least  $1 - \exp(-c(\ln N)^2)$ , we have uniformly for  $z = E + i(N^{\mathfrak{a}}/N)^{2/3} \in \mathcal{M}$  with  $E \leq A - N^{2\mathfrak{c}}(N^{\mathfrak{a}}/N)^{2/3}$ ,

$$|G_N(z) - G_\mu(z)| \ll \frac{1}{N\eta}. \tag{1.15}$$

A similar statement holds if  $[B, \hat{b}]$  is a void region in the vicinity of  $\hat{b}$ .

**Remark 1.9.** We believe that the rigidity estimate (1.14) also holds for saturated regions. However, for technical reasons (see e.g. the proof of Proposition 3.16), the current method does not work in this case.

Following a standard application of the Helffer-Sjöstrand functional calculus along the lines of [46, Lemma B.1], the following result may be deduced from Theorem 1.8.

**Corollary 1.10.** We assume Assumptions (1.2), (1.3), (1.5) and that  $[\hat{a}, A]$  is a void region, i.e.  $\rho_V(x) = 0$  on  $[\hat{a}, A]$ , then the following holds: For any  $\mathfrak{a} > 0$  and  $r > 0$ , with probability at least  $1 - \exp(-c(\ln N)^2)$ , uniformly for  $i < (\mu([A, B]) - r)N$ , we have

$$\left| \frac{\ell_i}{N} - \gamma_i \right| \leq \frac{N^{\mathfrak{a}}}{N^{2/3} \min\{i, N - i\}^{1/3}},$$

where  $\gamma_i$  are the classical particle locations of the equilibrium measure  $\mu$  as defined in (1.5). A similar statement holds if  $[B, \hat{b}]$  is a void region.

With the rigidity of the particle configuration as input, the edge universality of discrete  $\beta$ -ensemble follows from a direct comparison with continuous  $\beta$ -ensembles.

**Theorem 1.11.** Fix  $\beta \geq 1$ ,  $\theta = \beta/2$ ,  $0 < \mathfrak{b} < 1/13$  and  $L = N^{\mathfrak{b}}$ . We assume Assumptions 1.2, 1.3, 1.5, 1.7 and that  $[\hat{a}, A]$  is a void region,  $\rho_V(u) = (1 + O(u - A))s_A\sqrt{u - A}/\pi$ . Take any fixed  $m \geq 1$  and a continuously differentiable compactly supported function  $O : \mathbb{R}^m \rightarrow \mathbb{R}$ . For any index set  $\Lambda \in \llbracket 1, L \rrbracket$  with  $|\Lambda| = m$ , we have

$$\left| \mathbb{E}_{\mathbb{P}_N} \left[ O \left( \left( N^{2/3} k^{1/3} s_A^{2/3} (\ell_k/N - \gamma_k) \right)_{k \in \Lambda} \right) \right] - \mathbb{E}_{G\beta E} \left[ O \left( \left( N^{2/3} k^{1/3} (x_k/N - \tilde{\gamma}_k) \right)_{k \in \Lambda} \right) \right] \right| = O(N^{-\chi}), \quad (1.16)$$

where  $\gamma_k$  are classical eigenvalue locations of  $\rho_V$ , and  $\tilde{\gamma}_k$  are classical eigenvalue locations of semi-circle distribution. A similar statement holds if  $[B, \hat{b}]$  is a void region.

## 2 Rigidity of discrete $\beta$ -Ensembles

Our argument for the rigidity of discrete  $\beta$ -ensemble is based on a multi-scale analysis. In this section we introduce some basic notations, and give an outline of the proof.

We define the spectral domains on the scale  $\tilde{\eta} = M/N$ , for any  $1 \ll M \leq N$ :

$$\begin{aligned} \mathcal{D}_{M,r} &:= \mathcal{D}_{M,r}^{\text{int}} \cup \mathcal{D}_M^{\text{ext}}, \\ \mathcal{D}_{M,r}^{\text{int}} &:= \{E + i\eta \in \mathcal{M} \cap \mathbb{C}_+ : E \in [A, B - r], \eta\sqrt{\kappa_E + \eta} \geq M/N\}, \\ \mathcal{D}_M^{\text{ext}} &:= \{E + i\eta \in \mathcal{M} \cap \mathbb{C}_+ : E \leq A, \eta \geq (M/N)^{2/3}\}. \end{aligned} \quad (2.1)$$

where  $\kappa_E = \text{dist}(E, \{A, B\})$ . The spectral domain  $\mathcal{D}_{M,r}$  contains information of the particle configuration away from the right endpoint  $B$ . For the multi-scale analysis, we also need certain weak control on the particle configuration around the right endpoint  $B$ . To do this, we fix  $\mathfrak{a} > 0$  and define the following spectral

domain on the scale  $\tilde{\eta} = N^{-(1-a)/2}$ ,

$$\begin{aligned}\mathcal{D}_* &:= \mathcal{D}_*^{\text{int}} \cup \mathcal{D}_*^{\text{ext}}, \\ \mathcal{D}_*^{\text{int}} &:= \{E + i\eta \in \mathcal{M} \cap \mathbb{C}_+ : E \in [A, B], \eta\sqrt{\kappa_E + \eta} \geq N^{-(1-a)/2}\}, \\ \mathcal{D}_*^{\text{ext}} &:= \{E + i\eta \in \mathcal{M} \cap \mathbb{C}_+ : E \notin [A, B], \eta \geq N^{-(1-a)/3}\}.\end{aligned}\tag{2.2}$$

The proof of Theorem 1.8 uses the iteration of the following two-step scheme. If uniformly for any  $z = E + i\eta \in \mathcal{D}_{M_i, r_0} \cup \mathcal{D}_*$ , we have

$$\mathbb{P}_N \left( |G_N(z) - G_\mu(z)| \leq \frac{M_i}{N\eta} \right) \geq 1 - e^{-cM_i},\tag{2.3}$$

then Theorem 2.6 implies that for  $M_{i+1} = M_i^{\mathfrak{b}/2}$ , where  $1 < \mathfrak{b} < 2$ , with probability  $1 - e^{-cM_i}$ ,

$$|G_N(z) - G_\mu(z)| \ll \sqrt{\kappa_E + \eta}$$

uniformly for  $z = E + i\eta \in \mathcal{D}_{M_{i+1}, r_0+r} \cup \mathcal{D}_*$ , where  $\kappa_E = \text{dist}(E, \{A, B\})$ . The hypotheses of Theorem 2.4 are therefore fulfilled and we deduce that uniformly for any  $z = E + i\eta \in \mathcal{D}_{M_{i+1}, r_0+r} \cup \mathcal{D}_*$ , and  $s \geq (\ln N)^2$ , we have

$$\mathbb{P}_N \left( |G_N(z) - G_\mu(z)| \lesssim \frac{s}{N\eta} \right) \geq 1 - e^{-cs}.\tag{2.4}$$

This gives (2.3) at a smaller scale. Then we can use (2.4) as the input, and iterate the above scheme again, the optimal rigidity follows after finite steps.

## 2.1 Rigidity of discrete $\beta$ -Ensembles

For the equilibrium measure of discrete  $\beta$ -ensemble, there are void regions and saturated regions. For the energy level close to void regions, the empirical particle density is a natural object to study. However, for the energy level close to saturated regions, the *dual* of the empirical particle density is a more natural object to estimate. In the following we define the *dual* equilibrium measure and the *dual* empirical particle density, and some control parameters.

We recall the constrained equilibrium measure  $\mu$ , the empirical particle density  $\mu_N$  and their Stieltjes transforms  $G_\mu$  and  $G_N$  as defined in Section 1.1. We define the dual equilibrium measure  $\mu^{\text{dual}}$ ,

$$\mu^{\text{dual}} = \rho_V^{\text{dual}}(x)dx, \quad \rho_V^{\text{dual}}(x) = \mathbf{1}_{[\hat{a}, \hat{b}]} \theta^{-1} - \rho_V(x),\tag{2.5}$$

and the dual empirical particle density  $\mu_N^{\text{dual}}$ ,

$$\mu_N^{\text{dual}} = \frac{1}{\theta N} \sum_{i=0}^N \sum_{\substack{x \in \ell_i + \mathbb{Z}_{>0}, \\ x < \ell_{i+1}}} \delta_{x/N},\tag{2.6}$$

where  $\ell_0 = a(N)$ ,  $\ell_{N+1} = b(N)$ , and  $\mathbb{Z}_{>0}$  is the set of positive integers. When  $\theta = 1$ , this is essentially the empirical particle density of the dual ensemble studied in [1, Section 3.2]. We denote the Stieltjes transforms

of (2.5) and (2.6) by  $G_\mu^{\text{dual}}(z)$  and  $G_N^{\text{dual}}(z)$  respectively. We remark that  $\mu_N + \mu_N^{\text{dual}}$  is an *almost*  $1/N$ -discretization of the measure  $\mathbf{1}_{[a(N)/N, b(N)/N]} \theta^{-1} dx$ . And it implies the following estimates, which are useful later

$$\begin{aligned} \left| \int_{a(N)/N}^{b(N)/N} \frac{1}{\theta(z-x)} dx - \int \frac{1}{z-x} d(\mu_N + \mu_N^{\text{dual}}) \right| &= O\left(\frac{1}{N\eta}\right), \\ \left| \int_{a(N)/N}^{b(N)/N} \frac{1}{\theta(z-x)^2} dx - \int \frac{1}{(z-x)^2} d(\mu_N + \mu_N^{\text{dual}}) \right| &= O\left(\frac{1}{N\eta^2}\right), \end{aligned} \quad (2.7)$$

where  $\eta = \text{Im}[z]$ .

For any  $z = E + i\eta \in \mathbb{C}_+$ , let  $\kappa_E = \text{dist}(E, \{A, B\})$ ,  $\hat{\kappa}_E = \text{dist}(E, \{\hat{a}, \hat{b}\})$ , and we define the control parameter:

$$\Theta_\mu(z) = \min \left\{ \frac{|\text{Im}[G_\mu(z)]|}{N\eta}, \frac{|\text{Im}[G_\mu^{\text{dual}}(z)]|}{N\eta} + \frac{1}{N(\eta + \hat{\kappa}_E)} \right\}. \quad (2.8)$$

Thanks to the square root behaviors of the equilibrium density  $\rho_V(x)$  at  $A, B$ , the asymptotic behavior of  $\Theta_\mu$  depends on  $\eta$ ,  $\kappa_E$  and  $\hat{\kappa}_E$ ,

$$\Theta_\mu(z) \asymp \begin{cases} \sqrt{\kappa_E + \eta}/N\eta, & E \in [A, B], \\ 1/N\sqrt{\kappa_E + \eta}, & E \notin [A, B] \text{ on voids}, \\ 1/N\sqrt{\kappa_E + \eta} + 1/N(\hat{\kappa}_E + \eta), & E \notin [A, B] \text{ on saturated regions}. \end{cases} \quad (2.9)$$

We also define,

$$\Theta_N(z) = \begin{cases} |\text{Im}[G_N(z)]|/N\eta, & \text{if } \Theta_\mu(z) = |\text{Im}[G_\mu(z)]|/N\eta, \\ |\text{Im}[G_N^{\text{dual}}(z)]|/N\eta + 1/N(\eta + \hat{\kappa}_E), & \text{if } \Theta_\mu(z) = |\text{Im}[G_\mu^{\text{dual}}(z)]|/N\eta + 1/N(\eta + \hat{\kappa}_E). \end{cases}$$

In the following we collect some basic estimates on the Green's function  $G_N(z)$ . It turns out that in the discrete setting, the Green's function  $G_N(z)$  behaves much more regularly, compared with the continuous setting.

**Lemma 2.1.** *For any particle configuration  $(\ell_1, \ell_2, \dots, \ell_N) \in \mathbb{W}_N^\theta$ , and  $z = E + i\eta$  with  $\eta \geq \ln N/N$ , we have*

$$|G_N(z)| = O(\ln N), \quad |\text{Im}[G_N(z)]| = O(1), \quad |\partial_z G_N(z)| = O(1/\eta), \quad (2.10)$$

where the implicit constants are universal. Moreover, we have the following relation between densities and their duals

$$|G_N(z) + G_N^{\text{dual}}(z) - G_\mu(z) - G_\mu^{\text{dual}}(z)| = O\left(\frac{\ln N}{N\eta}\right). \quad (2.11)$$

*Proof.* By the definition of  $\mathbb{W}_N^\theta$ , the particles  $\ell_i$  are well separated by distance at least  $\theta$ . We have the estimate,

$$|G_N(E + i\eta)| \leq \frac{1}{N} \sum_{i=1}^N \frac{1}{|E - \ell_i/N + i\eta|} \leq \frac{2\sqrt{2}}{N} \sum_{i=0}^N \frac{1}{i\theta/N + \eta} = O(\ln N).$$

For the imaginary part of  $G_N(z)$ , we have

$$|\operatorname{Im}[G_N(E + i\eta)]| = \frac{1}{N} \sum_{i=1}^N \frac{\eta}{(E - \ell_i/N)^2 + \eta^2} \leq \frac{2}{N} \sum_{i=0}^N \frac{\eta}{(i\theta/N)^2 + \eta^2} = O(1).$$

For the derivative of  $G_N(z)$ , we have

$$|\partial_z G_N(E + i\eta)| \leq \frac{1}{N} \sum_{i=1}^N \frac{1}{(E - \ell_i/N)^2 + \eta^2} \leq \frac{|\operatorname{Im}[G_N(E + i\eta)]|}{\eta} = O(1/\eta).$$

Finally for (2.11), we notice that  $G_\mu(z) + G_\mu^{\text{dual}}(z)$  is the Stieltjes transform of the measure  $\theta^{-1} \mathbf{1}_{[\hat{a}, \hat{b}]}$ , and from (2.7)  $G_N(z) + G_N^{\text{dual}}(z)$  is approximated by the Stieltjes transform of the measure  $\theta^{-1} \mathbf{1}_{[a(N)/N, b(N)/N]}$ . Therefore, we have

$$|G_N(z) + G_N^{\text{dual}}(z) - G_\mu(z) - G_\mu^{\text{dual}}(z)| = O\left(\frac{1}{N\eta}\right) + \left| \left( \int_{\hat{a}}^{\hat{b}} - \int_{a(N)/N}^{b(N)/N} \right) \frac{dx}{\theta(z-x)} \right| = O\left(\frac{\ln N}{N\eta}\right),$$

where we used Assumption 1.2,  $\hat{a} = a(N)/N + O(\ln N/N)$  and  $\hat{b} = b(N)/N + O(\ln N/N)$ .  $\square$

Thanks to (2.11), the following estimate holds,

$$\Theta_N(z) - \Theta_\mu(z) = \pm \frac{\operatorname{Im}[G_N(z)] - \operatorname{Im}[G_\mu(z)]}{N\eta} + O\left(\frac{\ln N}{(N\eta)^2}\right). \quad (2.12)$$

The following lemma implies that, the rigidity estimates of the Green function  $G_N(z)$  are equivalent to the rigidity estimates of the quantity  $\sum_{i=1}^N \ln(1 + 1/N(z - \ell_i/N))$ , which naturally appears in the Nekrasov's equation.

**Lemma 2.2.** *For any particle configuration  $(\ell_1, \ell_2, \dots, \ell_N) \in \mathbb{W}_N^\theta$ ,  $z = E + i\eta$  with  $\eta \geq \ln N/N$ ,  $\alpha = \pm 1, \pm i$  and  $0 \leq t \leq 1$ , we have*

$$\sum_{i=1}^N \ln\left(1 + \frac{\alpha t}{N(z - \ell_i/N)}\right) = \alpha t G_N(z) + O\left(\Theta_N(z) + \frac{\ln N}{(N\eta)^2}\right). \quad (2.13)$$

where the implicit constant is universal.

*Proof.* By Taylor expansion, it follows

$$\sum_{i=1}^N \ln\left(1 + \frac{\alpha t}{N(z - \ell_i/N)}\right) = \alpha t G_N(z) - \frac{(\alpha t)^2}{2N^2} \sum_{i=1}^N \frac{1}{(z - \ell_i/N)^2} + O\left(\frac{1}{N^3} \sum_{i=1}^N \frac{1}{|z - \ell_i/N|^3}\right). \quad (2.14)$$

The last term on the righthand side of (2.14) is  $O((N\eta)^{-2})$ . For the second term on the righthand side of (2.14), we can directly estimate it,

$$\left| \frac{1}{N^2} \sum_{i=1}^N \frac{1}{(z - \ell_i/N)^2} \right| \lesssim \frac{1}{N^2} \sum_{i=1}^N \frac{1}{|z - \ell_i/N|^2} = \frac{|\operatorname{Im}[G_N(z)]|}{N\eta}. \quad (2.15)$$

There is another way to estimate it. Since  $\mu_N + \mu_N^{\text{dual}}$  is an *almost*  $1/N$ -discretization of  $\mathbf{1}_{[a(N)/N, b(N)/N]} \theta^{-1} dx$ , we deduce from (2.7) that

$$\int \frac{1}{(z-x)^2} (d\mu_N(x) + d\mu_N^{\text{dual}}(x) - \mathbf{1}_{[a(N)/N, b(N)/N]} \theta^{-1} dx) = O\left(\frac{1}{N\eta^2}\right).$$

Therefore, we have

$$\begin{aligned} \frac{1}{N^2} \sum_{i=1}^N \frac{1}{(z - \ell_i/N)^2} &= -\frac{1}{N} \int \frac{1}{(z-x)^2} d\mu_N^{\text{dual}}(x) + \frac{1}{\theta N} \left( \frac{1}{z - b(N)/N} - \frac{1}{z - a(N)/N} \right) + O\left(\frac{1}{(N\eta)^2}\right) \\ &= O\left(\frac{|\text{Im}[G_N^{\text{dual}}(z)]|}{N\eta} + \frac{1}{N(\eta + \hat{\kappa}_E)} + \frac{1}{(N\eta)^2}\right). \end{aligned} \tag{2.16}$$

The statement follows from combining (2.15) and (2.16).  $\square$

We recall that  $\mu$  is the limit of the empirical distribution of the particles locations under  $\mathbb{P}_N$  and  $G_\mu(z)$  its Stieltjes transform. For any  $v \in \mathbb{C}_+$ ,  $0 \leq t \leq 1$ ,  $\alpha = \pm 1, \pm i$ , and large number  $K = K(N, \text{Im}[v])$ , we introduce the deformed probability measure

$$\begin{aligned} \mathbb{P}_N^{K,t,v,\alpha} &= \frac{Z_N}{Z_N^{K,t,v,\alpha}} e^{K \left( \sum_{i=1}^N \ln \left( 1 + \frac{\alpha t}{N(v - \ell_i/N)} \right) + \ln \left( 1 + \frac{\bar{\alpha} t}{N(\bar{v} - \bar{\ell}_i/N)} \right) \right)} \mathbb{P}_N \\ &= \frac{Z_N}{Z_N^{K,t,v,\alpha}} e^{2K \text{Re} \left[ \sum_{i=1}^N \ln \left( 1 + \frac{\alpha t}{N(v - \ell_i/N)} \right) \right]} \mathbb{P}_N. \end{aligned} \tag{2.17}$$

The strategy to derive the optimal rigidity consists of two steps. The first step is the large deviation estimate at scale  $\tilde{\eta}$ , (we start with scale  $O(1)$ ), which gives us certain control on  $G_N(z) - G_\mu(z)$  with respect to the measure  $\mathbb{P}_N^{K,t,v,\alpha}$ . However we don't expect it to be optimal. Then we use the loop equation to get improved estimates. By iteration, the optimal rigidity at scale  $\tilde{\eta}$  follows. Then using the optimal rigidity at larger scale, we can repeat the process at a smaller scale. We can reach the optimal scale  $O(1/N)$ , in finite steps.

In the macroscopic scale, for any deformed measure  $\mathbb{P}_N^{K,t,v,\alpha}$ , with  $K \leq N$ , notice that (2.10) and (2.13) implies

$$2K \text{Re} \left[ \sum_{i=1}^N \ln \left( 1 + \frac{\alpha t}{N(v - \ell_i/N)} \right) \right] = O(2K \text{Re}[\alpha t G_N(v)]) = O(N \ln N).$$

Therefore, by a large deviation argument, similar to [17, Proposition 2.15], see also Proposition 3.16, we have

**Proposition 2.3.** *For two probability measures  $\nu, \rho$  set*

$$\mathcal{D}(\nu, \rho) = \left( - \int_{\mathbb{R}} \int_{\mathbb{R}} \ln |x - y| d(\nu(x) - \rho(x)) d(\nu(y) - \rho(y)) \right)^{1/2}. \tag{2.18}$$

Fix a parameter  $p > 2$  and let  $\tilde{\mu}_N$  denote the convolution of the empirical measure  $\mu_N$  with the uniform measure on the interval  $[-N^{-p}/2, N^{-p}/2]$ . Let  $K \leq N$ . Then for any  $\gamma > 0$

$$\mathbb{P}_N^{K,t,v,\alpha} \left( \mathcal{D}(\tilde{\mu}_N, \mu) \geq \gamma \right) \leq \exp(CN \ln(N) - \gamma^2 N^2). \quad (2.19)$$

As a consequence, there exists a constant  $C > 0$ , such that if  $\hat{a}_N = \min\{\hat{a}, a(N)/N\}$  and  $\hat{b}_N = \min\{\hat{b}, b(N)/N\}$ , for any  $z$  such that  $\text{dist}(z, [\hat{a}_N, \hat{b}_N]) \geq N^{-p+1}$ ,

$$\mathbb{P}_N^{K,t,v,\alpha} \left( |G_N(z) - G_\mu(z)| \geq \frac{C(\ln N \ln \text{dist}(z, [\hat{a}_N, \hat{b}_N]))^{1/2}}{N^{1/2} \text{dist}(z, [\hat{a}_N, \hat{b}_N])} \right) \leq e^{-N \ln N}. \quad (2.20)$$

*Proof.* The first point is a direct consequence of [17, Proposition 2.15] as the density of  $\mathbb{P}_N^{K,t,v,\alpha}$  with respect to  $\mathbb{P}_N$  is uniformly bounded by  $e^{cN \ln N}$ . For the second point, notice that there is an alternative formula for  $\mathcal{D}(\nu, \rho)$ , if  $\nu - \rho$  has zero mass, cf. [9],

$$\mathcal{D}(\nu, \rho) = \sqrt{\int_0^\infty \frac{1}{t} \left| \int_{\mathbb{R}} e^{itx} (\nu(x) - \rho(x)) dx \right|^2 dt}. \quad (2.21)$$

Let  $d = \text{dist}(z, [\hat{a}_N, \hat{b}_N])$ . Note that as  $\tilde{\mu}_N$  and  $\mu$  are supported in the finite interval  $[\hat{a}_N, \hat{b}_N]$ ,

$$G_N(z) - G_\mu(z) = \int f_z(x) d(\mu_N - \mu)(x)$$

with  $f_z(x) = (z-x)^{-1} \chi(x)$  where  $\chi$  is nonnegative, equals to one on  $[\hat{a}_N, \hat{b}_N]$  and vanishes outside a slightly larger interval  $[\hat{a}_N - d/2, \hat{b}_N + d/2]$ . We further assume that  $\|\chi'\|_\infty = O(1/d)$ . Therefore, by taking the Fourier transform of  $f_z$ ,

$$\begin{aligned} |G_N(z) - G_\mu(z)| &\leq \left| \int f_z(x) d(\tilde{\mu}_N - \mu)(x) \right| + \left| \int f_z(x) d(\mu_N - \tilde{\mu}_N)(x) \right| \\ &\leq \left| \int \hat{f}_z(s) (\widehat{\tilde{\mu}_N - \mu})(s) ds \right| + N^{-p} \sup_{x \in [\hat{a}_N, \hat{b}_N]} |f'_z(x)| \\ &\leq 2 \left( \int s |\hat{f}_z(s)|^2 ds \right)^{1/2} \mathcal{D}(\tilde{\mu}_N, \mu) + \frac{1}{N^p d^2}. \end{aligned} \quad (2.22)$$

Finally, since  $\int f'_z(x) dx = 0$ ,

$$\begin{aligned} \int s |\hat{f}_z(s)|^2 ds &= \int s^{-1} |\hat{f}'_z(s)|^2 ds \\ &= - \int \int \log|x-y| f'_z(x) dx \bar{f}'_z(y) dy \\ &\leq \int \int |\log|x-y|| \left( \frac{\chi(x)}{|z-x|^2} + \frac{|\chi'(x)|}{|z-x|} \right) dx \left( \frac{\chi(y)}{|z-y|^2} + \frac{|\chi'(y)|}{|z-y|} \right) dy. \end{aligned} \quad (2.23)$$

We can divide the integral domain of (2.23) into two pieces,  $|x-y| \leq d$  and  $|x-y| \geq d$ . It is easy to check that both integrals are bounded by  $O(\ln d/d^2)$ . We can estimate the right hand side of (2.23) by  $O(\ln d/d^2)$ . Hence, (2.22) and (2.19) give (2.20).  $\square$



Once we have the estimate (2.20), we can use the loop equation to improve the estimate. More generally, we have the following theorem. We postpone its proof to the end of this section.

**Theorem 2.4.** *Let  $v = E + i\eta \in \mathbb{C}_+$ ,  $\kappa_E = \text{dist}(E, \{A, B\})$ , and  $\hat{\kappa}_E = \text{dist}(E, \{\hat{a}, \hat{b}\})$ . We assume Assumptions 1.2, 1.3 and 1.5, and that  $N\eta\sqrt{\kappa_E + \eta} \gg 1$ . For any  $0 \leq t \leq 1$ ,  $\alpha = \pm 1, \pm i$ , and  $K \ll (N\eta)^2\sqrt{\kappa_E + \eta}$ ,  $K \leq N$ , we assume that with high probability w.r.t.  $\mathbb{P}_N^{K,t,v,\alpha}$*

$$G_N(v) = G_\mu(v) + O(\varepsilon), \quad (2.24)$$

then we have

$$\mathbb{E}_{\mathbb{P}_N^{K,t,v,\alpha}} [G_N(v) - G_\mu(v)] = O\left(\frac{\varepsilon^2}{\sqrt{\kappa_E + \eta}} + \varepsilon_0\right), \quad (2.25)$$

where,

$$\varepsilon_0 = \frac{1}{\sqrt{\kappa_E + \eta}} \left( \Theta_\mu(v) + \frac{K \min\{|\text{Im}[G_\mu(v)]|, |\text{Im}[G_\mu^{\text{dual}}(v)]|\}}{(N\eta)^2} + \frac{\ln N}{(N\eta)^2} + \frac{\ln N}{N} + \frac{\ln NK}{(N\eta)^3} + \frac{K^2}{(N\eta)^4} \right).$$

Let  $\tilde{\varepsilon}_0 = \max\{\varepsilon_0, (\ln N)^2 K^{-1}\}$ . If there exists some  $\mathfrak{c} > 0$ , such that  $\varepsilon \leq N^{-\mathfrak{c}}\sqrt{\kappa_E + \eta}$ , then we have

$$\mathbb{P}_N (|G_N(v) - G_\mu(v)| \lesssim s\tilde{\varepsilon}_0) \geq 1 - e^{-csK\tilde{\varepsilon}_0}, \quad (2.26)$$

for any  $s \geq 1$ .

As an easy application of Theorem 2.4, with Proposition 2.3 as input, we have the following optimal rigidity estimate on the scale  $\tilde{\eta} \gg N^{-1/2}$ .

**Corollary 2.5.** *We assume Assumptions 1.2, 1.3 and 1.5. Fix  $s \geq (\ln N)^2$ . With probability at least  $1 - e^{-cs}$ , it holds uniformly for any  $z = E + i\eta \in \mathcal{D}_*$  (as defined in (2.2) for a given  $\mathfrak{a} > 0$ ),*

$$|G_N(z) - G_\mu(z)| \lesssim \frac{s}{N\eta}.$$

*Proof.* For any  $z = E + i\eta \in \mathcal{D}_*$ , with  $\kappa_E = \text{dist}(E, \{A, B\})$  and  $\hat{\kappa}_E = \text{dist}(E, \{\hat{a}, \hat{b}\})$ , we have  $\eta\sqrt{\kappa_E + \eta} \geq N^{-(1-\mathfrak{a})/2}$ . By the large deviation estimates of Proposition 2.3, the following holds with high probability with respect to the deformed measure  $\mathbb{P}_N^{N\eta,t,v,\alpha}$  as defined in (2.17)

$$|G_N(z) - G_\mu(z)| \lesssim \frac{N^{\mathfrak{a}/4}}{N^{1/2}\eta} \lesssim N^{-\mathfrak{a}/4}\sqrt{\kappa_E + \eta}.$$

By taking  $K = N\eta$  and  $\varepsilon = N^{-\mathfrak{a}/4}\sqrt{\kappa_E + \eta}$  in Theorem 2.4, we have

$$\varepsilon_0 \lesssim \frac{1}{\sqrt{\kappa_E + \eta}} \left( \Theta_\mu(z) + \frac{\ln N}{(N\eta)^2} + \frac{\ln N}{N} \right).$$

From (2.9) and  $\hat{\kappa}_E + \kappa_E \gtrsim 1$ , we see that  $\Theta_\mu(z)/\sqrt{\kappa_E + \eta} \lesssim 1/N\eta$ . Moreover, from the definition of the domain  $\mathcal{D}_*$ ,  $N\eta\sqrt{\kappa_E + \eta} \gg 1$ . We have

$$\varepsilon_0 \lesssim \frac{1}{N\eta} + \frac{\ln N}{N\sqrt{\kappa_E + \eta}} + \frac{\ln N}{(N\eta)} \lesssim \frac{\ln N}{N\eta}, \quad \tilde{\varepsilon}_0 = \frac{(\ln N)^2}{N\eta}.$$

Therefore, by (2.26), we deduce that

$$\mathbb{P}_N \left( |G_N(z) - G_\mu(z)| \lesssim \frac{s}{N\eta} \right) \geq 1 - e^{-cs}, \quad (2.27)$$

for any  $s \geq (\ln N)^2$ . By an union bound and the Lipschitz property of  $G_N(z) - G_\mu(z)$  from Lemma 2.1, we find a constant  $c > 0$  such that for any  $s \geq (\ln N)^2$ ,

$$\mathbb{P}_N \left( \sup_{z \in \mathcal{D}_*} \text{Im}[z] |G_N(z) - G_\mu(z)| \lesssim \frac{s}{N} \right) \geq 1 - e^{-cs}. \quad (2.28)$$

□

**Theorem 2.6.** *Fix small constants  $r_0, r > 0$ , and a parameter  $M$  so that  $N^\alpha \leq M \ll N$ . We assume the following holds*

$$\mathbb{P}_N \left( |G_N(z) - G_\mu(z)| \geq \frac{M}{N \text{Im}[v]} \right) \leq e^{-cM}, \quad (2.29)$$

uniformly for any  $z \in \mathcal{D}_{M, r_0} \cup \mathcal{D}_*$ . Fix  $L = M^\flat$ , where  $1 < \flat < 2$ . For any  $z = E + i\eta \in \mathbb{C}_+$ , with  $\kappa_E = \text{dist}(E, \{A, B\})$ , the following holds with probability  $1 - e^{-cM}$ ,

1. if  $E - A \gtrsim (L/N)^{2/3}$  and  $E \leq B - r_0 - r$ , then

$$|G_N(z) - G_\mu(z)| \leq C(\ln N)^2 \left( \frac{\sqrt{L}}{N\eta} + \frac{M}{L} \sqrt{\kappa_E} \right). \quad (2.30)$$

2. If  $E \leq A$  or  $E - A \lesssim (L/N)^{2/3}$ , then

$$|G_N(z) - G_\mu(z)| \leq C(\ln N)^2 \left( \frac{\sqrt{L}}{N\eta} + \left( \frac{M}{L} \right)^{1/2} \left( \frac{L}{N} \right)^{1/3} \wedge \frac{M}{L} \left( \frac{L}{N} \right)^{2/3} \frac{1}{\sqrt{\kappa_E + \eta}} + \frac{M}{L} \left( \frac{L}{N} \right)^{1/3} \right). \quad (2.31)$$

The proof of Theorem 2.6 is given in Section 3.

*Proof of Theorem 1.8.* Let  $M_1 := N^{1/2+\alpha}$ , and an arbitrarily small constant  $r > 0$ . By Corollary 2.5, there exists  $c > 0$  so that for any  $s \geq (\ln N)^2$ ,

$$\mathbb{P}_N \left( |G_N(z) - G_\mu(z)| \lesssim \frac{s}{N\eta} \right) \geq 1 - e^{-cs}, \quad (2.32)$$

uniformly for any  $z \in \mathcal{D}_*$ . Since  $\mathcal{D}_{M_1, r} \cup \mathcal{D}_* = \mathcal{D}_*$ , (2.32) holds on domain  $\mathcal{D}_{M_1, r} \cup \mathcal{D}_*$ . In the following we construct a decreasing sequence:  $M_1 \gg M_2 \gg \dots M_n = N^\alpha$ . Given that (2.32) holds on domain  $\mathcal{D}_{M_k, kr} \cup \mathcal{D}_*$ , we prove that (2.32) holds on the domain  $\mathcal{D}_{M_{k+1}, (k+1)r} \cup \mathcal{D}_*$ . Thus, (2.32) holds on the domain  $\mathcal{D}_{M_n, nr} = \mathcal{D}_{nr}$ , and (1.14) follows, since we can take  $r$  arbitrarily small.

We assume that (2.32) holds for any  $z \in \mathcal{D}_{M_k, kr} \cup \mathcal{D}_*$ . Especially (2.29) holds for any  $M$  such that  $M_k \leq M \leq N$  in the spectral domain  $\mathcal{D}_{M, kr} \cup \mathcal{D}_*$ . We assume that  $M_k > N^\alpha$ , otherwise we have finished the

proof of (1.14). We take  $L_k = M_k^{\mathfrak{b}}$ , ( $1 < \mathfrak{b} < 2$  will be chosen later) and  $z = E + i\eta \in \mathcal{D}_{M_k^{\mathfrak{b}/2}, (k+1)r} \setminus \mathcal{D}_{M_k, kr}$  with  $\kappa_E = \text{dist}(E, \{A, B\})$ . More precisely,  $\mathcal{D}_{M_k^{\mathfrak{b}/2}, (k+1)r} \setminus \mathcal{D}_{M_k, kr}$  is given by  $E \in [A, B - (k+1)r]$ ,  $M_k^{\mathfrak{b}/2} \leq N\eta\sqrt{\kappa_E + \eta} \leq M_k$ ; or  $E \leq A$ , and  $(M_k^{\mathfrak{b}/2}/N)^{2/3} \leq \eta \leq (M_k/N)^{2/3}$ .

Thanks to Theorem 2.6 with  $r_0 = kr$ , for any  $M$  such that  $M_k \leq M \leq N$ , we have

$$\mathbb{P}_N (|G_N(z) - G_\mu(z)| \leq \omega(M, z)) \geq 1 - e^{-cM}, \quad (2.33)$$

where  $\omega(M, z)$  is defined as the right hand side of (2.30) or (2.31): Setting  $L = M^{\mathfrak{b}}$ , if  $E - A \gtrsim (L/N)^{2/3}$  and  $E \leq B - (k+1)r$ , then

$$\omega(M, z) := C(\ln N)^2 \left( \frac{\sqrt{L}}{N\eta} + \frac{M}{L} \sqrt{\kappa_E} \right),$$

whereas if  $E \leq A$  or  $E - A \lesssim (L/N)^{2/3}$ , then

$$\omega(M, z) := C(\ln N)^2 \left( \frac{\sqrt{L}}{N\eta} + \left(\frac{M}{L}\right)^{1/2} \left(\frac{L}{N}\right)^{1/3} \wedge \frac{M}{L} \left(\frac{L}{N}\right)^{2/3} \frac{1}{\sqrt{\kappa_E + \eta}} + \frac{M}{L} \left(\frac{L}{N}\right)^{1/3} \right).$$

In the following we check that for any  $z \in \mathcal{D}_{M_k^{\mathfrak{b}/2}, (k+1)r} \setminus \mathcal{D}_{M_k, kr}$  and  $M \geq N^{\mathfrak{a}}$ ,  $N\eta\omega(M, z)$  is much smaller than  $M$ . For the bulk case, if  $E - A \gtrsim (L/N)^{2/3}$  and  $E \leq B - (k+1)r$ , we have

$$N\eta\omega(M, z) = C(\ln N)^2 \left( \sqrt{L} + \frac{M}{L} N\eta\sqrt{\kappa_E} \right) \leq C(\ln N)^2 \left( M^{\mathfrak{b}/2} + \frac{M}{L} M_k \right) \ll M.$$

where we used that on  $\mathcal{D}_{M_k^{\mathfrak{b}/2}, (k+1)r} \setminus \mathcal{D}_{M_k, kr}$ ,  $N\eta\sqrt{\kappa_E + \eta} \leq M_k$ . For the edge case, if  $E \leq A$  or  $E - A \lesssim (L/N)^{2/3}$ , we have  $\eta \leq (M_k/N)^{2/3}$ , and

$$\begin{aligned} N\eta\omega(M, z) &\leq C(\ln N)^2 \left( \sqrt{L} + N\eta \left(\frac{M}{L}\right)^{1/2} \left(\frac{L}{N}\right)^{1/3} \wedge N\sqrt{\eta} \frac{M}{L} \left(\frac{L}{N}\right)^{2/3} + N\eta \frac{M}{L} \left(\frac{L}{N}\right)^{1/3} \right) \\ &\leq C(\ln N)^2 \left( M^{\mathfrak{b}/2} + \left(\frac{M_k}{L}\right)^{1/3} M + \left(\frac{M_k}{L}\right)^{2/3} M \right) \ll M. \end{aligned}$$

Fix any  $z = E + i\eta \in \mathcal{D}_{M_k^{\mathfrak{b}/2}, (k+1)r} \setminus \mathcal{D}_{M_k, kr}$ , we consider the deformed measure

$$\mathbb{P}_N^{N\eta, t, z, \alpha} = \frac{Z_N}{Z_N^{N\eta, t, z, \alpha}} e^{\tilde{H}} \mathbb{P}_N, \quad \tilde{H}(z) = 2N\eta \text{Re} \left[ \sum_{i=1}^N \ln \left( 1 + \frac{\alpha t}{N(z - \ell_i/N)} \right) \right],$$

for some  $0 \leq t \leq 1$  and  $\alpha = \pm 1, \pm i$ . By Lemma 2.2, we have

$$X(z) := \left| \tilde{H}(z) - 2tN\eta \text{Re} [\alpha G_\mu(z)] \right| = 2N\eta |G_N(z) - G_\mu(z)| + \mathcal{O}(|\text{Im}[G_N(z)]|). \quad (2.34)$$

Thanks to (2.33), for any  $M_k \leq M \leq N$ , we have  $|G_N(z) - G_\mu(z)| \leq \omega(M, z)$ , with probability at least  $1 - e^{-cM}$  with respect to  $\mathbb{P}_N$ . Therefore, we have for any  $z \in \mathbb{C}^+$

$$\begin{aligned}
& \mathbb{P}_N^{N\eta, t, z, \alpha} (|G_N(z) - G_\mu(z)| \geq \omega(M_k, z)) = \frac{\int \mathbf{1}(|G_N(z) - G_\mu(z)| \geq \omega(M_k, z)) e^{\tilde{H}(z)} d\mathbb{P}_N}{\int e^{\tilde{H}(z)} d\mathbb{P}_N} \\
& \leq \frac{\int \mathbf{1}(|G_N(z) - G_\mu(z)| \geq \omega(M_k, z)) e^{\tilde{H}(z)} d\mathbb{P}_N}{\int \mathbf{1}(X(z) \leq N\eta\psi(M_k)) e^{\tilde{H}(z)} d\mathbb{P}_N} \\
& \leq \frac{\int \mathbf{1}(|G_N(z) - G_\mu(z)| \geq \omega(M_k, z)) e^{\tilde{H}(z) - 2tN\eta \operatorname{Re}[\alpha G_\mu(z)]} d\mathbb{P}_N}{\int \mathbf{1}(X(z) \leq N\eta\omega(M_k, z)) e^{\tilde{H}(z) - 2tN\eta \operatorname{Re}[\alpha G_\mu(z)]} d\mathbb{P}_N} \\
& \leq e^{N\eta\omega(M_k, z)} \int \mathbf{1}(|G_N(z) - G_\mu(z)| \geq \omega(M_k, z)) e^{X(z)} d\mathbb{P}_N.
\end{aligned} \tag{2.35}$$

We take  $m = \lfloor \log_2(N/M_k) \rfloor$  and decompose the integral region as  $\{|G_N(z) - G_\mu(z)| \geq N\eta\omega(M_k, z)\} \subset \cup_{i=0}^m A_i$ , where

$$\begin{aligned}
A_i &= \{|G_N(z) - G_\mu(z)| \in [N\eta\omega(2^i M_k, z), N\eta\omega(2^{i+1} M_k, z)]\}, \quad 0 \leq i \leq m-1, \\
A_m &= \{|G_N(z) - G_\mu(z)| \geq N\eta\omega(2^m M_k, z)\},
\end{aligned}$$

and the set above is empty if  $N\eta\omega(2^i M_k, z) > N\eta\omega(2^{i+1} M_k, z)$ . By (2.10),  $|G_N(z) - G_\mu(z)| \lesssim \ln N$ , and combining with (2.34) we have that  $X(z) \lesssim N\eta \ln N$  on  $A_m$ . Similarly for  $0 \leq i \leq m-1$ , on  $A_i$  we have  $X(z) \lesssim N\eta\omega(2^{i+1} M_k, z)$ . On the other hand, by (2.33), we have  $\mathbb{P}_N(A_m) \leq e^{-cN}$  and for  $0 \leq i \leq m-1$ ,  $\mathbb{P}_N(A_i) \leq e^{-c2^i M_k}$ . Hence, We then have

$$\begin{aligned}
\int \mathbf{1}(|G_N(z) - G_\mu(z)| \geq N\eta\omega(M_k, z)) e^{X(z)} d\mathbb{P}_N &\leq \sum_{i=0}^m \int_{A_i} e^{X(z)} d\mathbb{P}_N \\
&\leq \sum_{i=0}^{m-1} e^{CN\eta\omega(2^{i+1} M_k, z)} \mathbb{P}_N(A_i) + e^{CN\eta \ln N} \mathbb{P}_N(A_m) \\
&\leq \sum_{i=0}^m e^{CN\eta\omega(2^{i+1} M_k, z)} e^{-c2^i M_k} + e^{CN\eta \ln N} e^{-cN} \leq e^{-cM_k/2}
\end{aligned}$$

where we used the fact that  $\omega(2M, z) \lesssim \omega(M, z)$  as well as  $N\eta\omega(M, z) \ll M$ . Hence, (2.35) yields

$$\mathbb{P}_N^{N\eta, t, z, \alpha} (|G_N(z) - G_\mu(z)| \geq \omega(M_k, z)) \leq e^{-cM_k/2 + N\eta\omega(M_k, z)} \leq e^{-cM_k/3}.$$

It follows that, with respect to the new measure  $\mathbb{P}_N^{N\eta, t, z, \alpha}$ , the following holds

$$|G_N(z) - G_\mu(z)| \leq \omega(M_k, z) \tag{2.36}$$

with probability  $1 - e^{-cM_k/3}$ , provided  $z \in \mathcal{D}_{M_k^{\mathfrak{b}/2}, (k+1)r} \setminus \mathcal{D}_{M_k, kr}$ .

We take  $\mathfrak{b} = 3/2$ , and  $M_{k+1} := \max\{N^\alpha, M_k^{\mathfrak{b}/2} N^{\alpha/5}\}$ . Notice that  $\mathcal{D}_{M_{k+1}, (k+1)r} \subset \mathcal{D}_{M_k^{\mathfrak{b}/2}, (k+1)r}$ . Then for any  $z = E + i\eta \in \mathcal{D}_{M_{k+1}, (k+1)r} \setminus \mathcal{D}_{M_k, kr}$  with  $\kappa_E = \operatorname{dist}(E, \{A, B\})$ , (2.36) holds. We have for the bulk

case,

$$\omega(M_k, z) = C(\ln N)^2 \left( \frac{M_k^{\mathfrak{b}/2}}{N\eta} + M_k^{-(\mathfrak{b}-1)} \sqrt{\kappa_E} \right) \leq N^{-\mathfrak{a}/6} \sqrt{\kappa_E + \eta}, \quad (2.37)$$

where we used  $M_k \geq N^{\mathfrak{a}}$  and  $N\eta\sqrt{\kappa_E + \eta} \geq M_{k+1}$ . For the edge case, since  $N\eta\sqrt{\kappa_E + \eta} \geq M_{k+1}$ ,  $\kappa_E + \eta \gtrsim (M_{k+1}/N)^{2/3}$  and  $\mathfrak{b} = 3/2$ , we get

$$\begin{aligned} \omega(M_k, z) &= C(\ln N)^2 \left( \frac{M_k^{\mathfrak{b}/2}}{N\eta} + \frac{M_k^{1/2-\mathfrak{b}/6}}{N^{1/3}} \wedge \frac{M_k^{1-\mathfrak{b}/3}}{N^{2/3}\sqrt{\kappa_E + \eta}} + \frac{M_k^{1-2\mathfrak{b}/3}}{N^{1/3}} \right) \\ &\leq C(\ln N)^2 \left( \frac{M_k^{\mathfrak{b}/2}}{N\eta} + \frac{M_k^{1/2-\mathfrak{b}/6}}{N^{1/3}} \right) \leq N^{-\mathfrak{a}/16} \sqrt{\kappa_E + \eta}. \end{aligned} \quad (2.38)$$

It follows from (2.36), (2.37) and (2.38), that for any  $z \in \mathcal{D}_{M_{k+1}, (k+1)r} \setminus \mathcal{D}_{M_k, kr}$ , the assumptions in Theorem 2.4 are satisfied. By the same argument as in Corollary 2.5, we deduce that for any  $s \geq (\ln N)^2$ ,

$$\mathbb{P}_N \left( |G_N(z) - G_\mu(z)| \lesssim \frac{s}{N\eta} \right) \geq 1 - e^{-cs},$$

uniformly for any  $z \in \mathcal{D}_{M_{k+1}, (k+1)r}$ . From our choices of  $M_j$ , we have

$$M_j = \max\{N^{4\mathfrak{a}/5 + (3/4)^{j-1}(1/2 + \mathfrak{a}/5)}, N^{\mathfrak{a}}\}, \quad j = 1, 2, 3, \dots$$

And  $M_n = N^{\mathfrak{a}}$  where  $n = 1 + \lceil \log_{3/4}(\mathfrak{a}/5) \rceil$ . It follows that, by repeating the above process  $n$  times, we have for any  $s \geq (\ln N)^2$ ,

$$\mathbb{P}_N \left( |G_N(z) - G_\mu(z)| \lesssim \frac{s}{N\eta} \right) \geq 1 - e^{-cs},$$

uniformly for any  $z \in \mathcal{D}_{M_n, nr} = \mathcal{D}_{nr}$ . Since we can take  $r > 0$  arbitrarily small, this finishes the proof of (1.14).

In the following we prove (1.15). Fix a small constant  $\mathfrak{c} > 0$  (which will be chosen later),  $\eta = (N^{\mathfrak{a}}/N)^{2/3}$ ,  $E \leq A - N^{2\mathfrak{c}}(N^{\mathfrak{a}}/N)^{2/3}$  and  $z = E + i\eta$ . We set  $\mathfrak{b} = 3/2$ . For any  $M$  such that  $N^{\mathfrak{a}} \leq M \leq N$ , let  $L = M^{\mathfrak{b}}$ . Since  $z \in \mathcal{D}_r$ , (1.14) holds. Thanks to Theorem 2.6, we have

$$\mathbb{P}_N (|G_N(z) - G_\mu(z)| \leq \omega(M, z)) \geq 1 - e^{-cM}.$$

Moreover, notice that for  $N^{\mathfrak{a}} \leq M \leq N$

$$N^{1+\mathfrak{c}}\eta\omega(M, z) = N^{\mathfrak{c}}M^{3/4} + N^{\mathfrak{c}+2\mathfrak{a}/3}M^{1/4} \wedge N^{\mathfrak{a}/3}M^{1/2} + N^{\mathfrak{c}+2\mathfrak{a}/3} \ll M.$$

provided  $\mathfrak{c} < \mathfrak{a}/4$ . By the same argument as for (2.36), we have

$$|G_N(z) - G_\mu(z)| \lesssim \omega(N^{\mathfrak{a}}, z) \leq N^{-\mathfrak{c}}\sqrt{\kappa_E + \eta},$$

holds with high probability, with respect to the deformed measure,  $\mathbb{P}_N^{N^{1+\mathfrak{c}}\eta, v, t, \alpha}$ .

We take  $K = N^{1+\epsilon\eta}$  in Theorem 2.4. Since  $\kappa_E \geq N^{2\epsilon}\eta$ , we have

$$\varepsilon_0 \lesssim \frac{N^\epsilon}{N\kappa_E} + \frac{N^{2\epsilon}}{(N\eta)^2\sqrt{\kappa_E + \eta}} \lesssim \frac{1}{N^{1+\epsilon\eta}}, \quad \tilde{\varepsilon}_0 = \frac{(\ln N)^2}{N^{1+\epsilon\eta}}$$

and by (2.26),

$$\mathbb{P}_N \left( |G_N(z) - G_\mu(z)| \lesssim \frac{s}{N^{1+\epsilon\eta}} \right) \geq 1 - e^{-cs}, \quad (2.39)$$

for any  $s \geq (\ln N)^2$ . By an union bound and the Lipschitz property of  $G_N(z) - G_\mu(z)$  from Lemma 2.1, we conclude that

$$\mathbb{P}_N \left( \sup_{\substack{z = E + i(N^\alpha/N)^{2/3} \\ E \leq A - N^{2\epsilon}(N^\alpha/N)^{2/3}}} |G_N(z) - G_\mu(z)| \leq \frac{1}{N^{1+\epsilon(N^\alpha/N)^{2/3}}} \right) \geq 1 - \exp(-c(\ln N)^2), \quad (2.40)$$

provided  $0 < \epsilon < \alpha/4$ .

□

## 2.2 Bootstrap and Proof of Theorem 2.4

In this section, we give the proof of Theorem 2.4. The main ingredient is Nekrasov's equation, which plays the role of Dyson-Schwinger (or loop) equation in continuous  $\beta$ -ensembles. Heuristically, if the weak estimate  $G_N(v) = G_\mu(v) + O(\varepsilon)$  holds with high probability w.r.t. the deformed measure  $\mathbb{P}_N^{K,t,v,\alpha}$ , then a careful analysis of Nekrasov's equation implies that

$$\mathbb{E}_{\mathbb{P}_N^{K,t,v,\alpha}} [G_N(v) - G_\mu(v)] \lesssim O(|G_N(v) - G_\mu(v)|^2) = O(\varepsilon^2).$$

The error is improved to  $O(\varepsilon^2)$ . By integrating over  $t$ , we deduce an upper bound on the Laplace transform of  $G_N(v) - G_\mu(v)$ . As a result, it follows that  $G_N(v) = G_\mu(v) + O(\varepsilon^2)$  holds with high probability w.r.t. the deformed measure. (2.26) is proven by iterating this procedure.

*Proof of Theorem 2.4.* The perturbed measure (defined in (2.17)) is given by

$$\mathbb{P}_N^{K,t,v,\alpha} = \frac{Z_N}{Z_N^{K,t,v,\alpha}} \prod_{i=1}^N \left( 1 + \frac{\alpha t}{N(v - \ell_i/N)} \right)^K \left( 1 + \frac{\bar{\alpha} t}{N(\bar{v} - \ell_i/N)} \right)^K \mathbb{P}_N.$$

For simplicity of notations, in the following proof, we write  $\mathbb{E}_{\mathbb{P}_N^{K,t,v,\alpha}}$  as  $\mathbb{E}$  and  $\alpha t$  as  $t$  (which is therefore complex). We construct two analytic functions  $\phi_N^+(x)$  and  $\phi_N^-(x)$  such that

$$\frac{\phi_N^+(x)}{\phi_N^-(x)} = \frac{w(x; N)}{w(x-1; N)} \frac{\left( 1 + \frac{t}{N(v-x/N)} \right)^K \left( 1 + \frac{\bar{t}}{N(\bar{v}-x/N)} \right)^K}{\left( 1 + \frac{t}{N(v-(x-1)/N)} \right)^K \left( 1 + \frac{\bar{t}}{N(\bar{v}-(x-1)/N)} \right)^K}.$$

In fact, by our Assumption 1.3, we can take

$$\begin{aligned}\phi_N^+(x) &= \psi_N^+(x) \left(v - \frac{x}{N} + \frac{t}{N}\right)^K \left(v - \frac{x-1}{N}\right)^K \left(\bar{v} - \frac{x}{N} + \frac{\bar{t}}{N}\right)^K \left(\bar{v} - \frac{x-1}{N}\right)^K, \\ \phi_N^-(x) &= \psi_N^-(x) \left(v + \frac{t}{N} - \frac{x-1}{N}\right)^K \left(v - \frac{x}{N}\right)^K \left(\bar{v} + \frac{\bar{t}}{N} - \frac{x-1}{N}\right)^K \left(\bar{v} - \frac{x}{N}\right)^K.\end{aligned}\tag{2.41}$$

By Theorem 1.4,  $R_N(zN)$  defined by

$$R_N(zN) := \phi_N^-(zN) \mathbb{E} \left[ \prod_{i=1}^N \left(1 - \frac{\theta}{N(z - \ell_i/N)}\right) \right] + \phi_N^+(zN) \mathbb{E} \left[ \prod_{i=1}^N \left(1 + \frac{\theta}{N(z - (\ell_i + 1)/N)}\right) \right].\tag{2.42}$$

is analytic in  $\mathcal{M}$ . We rearrange the above expression of  $R_N(zN)$ ,

$$\begin{aligned}& \frac{R_N(zN)}{(v - z + t/N)^K (v + 1/N - z)^K (\bar{v} - z + \bar{t}/N)^K (\bar{v} + 1/N - z)^K} \\ &= \psi_N^-(zN) \mathbb{E} \left[ \prod_{i=1}^N \left(1 - \frac{\theta}{N(z - \ell_i/N)}\right) - e^{-\theta G_\mu(z)} \right] + \psi_N^+(zN) \mathbb{E} \left[ \prod_{i=1}^N \left(1 + \frac{\theta}{N(z - (\ell_i + 1)/N)}\right) - e^{\theta G_\mu(z)} \right] \\ &+ \left( \psi_N^-(zN) e^{-\theta G_\mu(z)} + \psi_N^+(zN) e^{\theta G_\mu(z)} \right) + \mathcal{E}(z) \psi_N^-(zN) \mathbb{E} \left[ \prod_{i=1}^N \left(1 - \frac{\theta}{N(z - \ell_i/N)}\right) \right],\end{aligned}\tag{2.43}$$

where

$$\begin{aligned}\mathcal{E}(z) &:= \frac{(v + \frac{t}{N} + \frac{1}{N} - z)^K (v - z)^K (\bar{v} + \frac{\bar{t}}{N} + \frac{1}{N} - z)^K (\bar{v} - z)^K}{(v - z + \frac{t}{N})^K (v + \frac{1}{N} - z)^K (\bar{v} - z + \frac{\bar{t}}{N})^K (\bar{v} + \frac{1}{N} - z)^K} - 1 \\ &= \left(1 - \frac{t}{N^2} \frac{1}{(v - z + t/N)(v - z + 1/N)}\right)^K \left(1 - \frac{\bar{t}}{N^2} \frac{1}{(\bar{v} - z + \bar{t}/N)(\bar{v} - z + 1/N)}\right)^K - 1 \\ &= -\frac{tK}{N^2} \frac{1}{(v - z)^2} - \frac{\bar{t}K}{N^2} \frac{1}{(\bar{v} - z)^2} + \mathcal{O}\left(\frac{K}{(N\eta)^3} + \frac{K^2}{(N\eta)^4}\right),\end{aligned}\tag{2.44}$$

provided  $\eta \gg 1/N$ ,  $|z - v| \gtrsim \eta$  and  $K \ll (N\eta)^2$ .

We take two set of oriented contours:  $\mathcal{C}_1$  consists of two clockwise oriented circles, one is centered at  $v$  with radius  $\eta/2$ , and the other is centered at  $\bar{v}$  with radius  $\eta/2$ .  $\mathcal{C}_2 \subset \mathcal{M}$  is a contour which encloses a small neighborhood of  $[\hat{a}, \hat{b}]$  (especially, for any point on  $z \in \mathcal{C}_2$ ,  $\text{dist}(z, [\hat{a}_N, \hat{b}_N]) \gtrsim 1$  and  $H(z) \neq 0$ ) and has counterclockwise orientation. Notice that  $\eta \gg 1/N$ , so  $v, v + t/N, v + 1/N$  are all inside the contour  $\mathcal{C}_1$ . By dividing (2.43) by  $2\pi i(z - v)H(z)$  (as defined in Assumption 1.5), and integrating along the union of contours  $\mathcal{C}_1 \cup \mathcal{C}_2$ , we get

$$0 = I_1 + I_2 + I_3 + I_4,$$

where we used Theorem 1.4 that the lefthand side of (2.43) is analytic outside the contour  $\mathcal{C}_1$ , and

$$\begin{aligned}
I_1 &= \frac{1}{2\pi i} \int_{\mathcal{C}_1} \left\{ \psi_N^-(zN) \mathbb{E} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{N(z - \ell_i/N)} \right) - e^{-\theta G_\mu(z)} \right] \right. \\
&\quad \left. + \psi_N^+(zN) \mathbb{E} \left[ \prod_{i=1}^N \left( 1 + \frac{\theta}{N(z - (\ell_i + 1)/N)} \right) - e^{\theta G_\mu(z)} \right] \right\} \frac{dz}{(z - v)H(z)}, \\
I_2 &= \frac{1}{2\pi i} \int_{\mathcal{C}_2} \left\{ \psi_N^-(zN) \mathbb{E} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{N(z - \ell_i/N)} \right) - e^{-\theta G_\mu(z)} \right] \right. \\
&\quad \left. + \psi_N^+(zN) \mathbb{E} \left[ \prod_{i=1}^N \left( 1 + \frac{\theta}{N(z - (\ell_i + 1)/N)} \right) - e^{\theta G_\mu(z)} \right] \right\} \frac{dz}{(z - v)H(z)}, \\
I_3 &= \frac{1}{2\pi i} \int_{\mathcal{C}_1 \cup \mathcal{C}_2} \left( \psi_N^-(zN) e^{-\theta G_\mu(z)} + \psi_N^+(zN) e^{\theta G_\mu(z)} \right) \frac{dz}{(z - v)H(z)} \\
I_4 &= \frac{1}{2\pi i} \int_{\mathcal{C}_1 \cup \mathcal{C}_2} \mathcal{E}(z) \psi_N^-(zN) \mathbb{E} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{N(z - \ell_i/N)} \right) \right] \frac{dz}{(z - v)H(z)}.
\end{aligned} \tag{2.45}$$

From (2.10), we have  $|\partial_z G_N(z)| \leq 1/\eta$ . As a consequence, if  $|z - v| \leq \eta/2$ ,  $|G_N(z) - G_N(v)| \leq 1/2$  and if  $|z - \bar{v}| \leq \eta/2$ ,  $|G_N(z) - G_N(\bar{v})| \leq 1/2$ . By our assumption, with high probability,  $G_N(v) = G_\mu(v) + O(\epsilon) = O(1)$ , and  $G_N(\bar{v}) = G_N(v) = O(1)$ . It follows that, uniformly for  $z$  on or inside  $\mathcal{C}_1$ , with high probability,  $G_N(z) = O(1)$ . Moreover, by (2.13) and (2.10),  $\prod_{i=1}^N \left( 1 - \frac{\theta}{N(z - \ell_i/N)} \right)$  is uniformly bounded by  $C \ln N$  for all particle configuration and  $\eta N \gg 1$ . Hence, we can neglect in the expectation the set where  $G_N(z)$  is eventually big (but smaller than  $\ln N$ ). As a consequence, we find that

$$\begin{aligned}
\mathbb{E} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{N(z - \ell_i/N)} \right) \right] &= \mathbb{E} \left[ e^{-\theta G_N(z) + O(1)} \right] = O(1), \\
\mathbb{E} \left[ \prod_{i=1}^N \left( 1 + \frac{\theta}{N(z - (\ell_i + 1)/N)} \right) \right] &= \mathbb{E} \left[ e^{\theta G_N(z - 1/N) + O(1)} \right] = O(1).
\end{aligned} \tag{2.46}$$

In the following, we analyze the contour integrals for  $I_1, I_2, I_3, I_4$ . We show that the leading term of  $I_1$  gives the expectation of  $G_N(v) - G_\mu(v)$  and prove that  $I_2, I_3, I_4$  are all very small, which implies the upper bound (2.25).

For  $I_1$ , the integrand has a single pole at  $z = v$  inside the contour  $\mathcal{C}_1$ . Therefore,  $I_1$  equals

$$\frac{\psi_N^-(vN)}{H(v)} \mathbb{E} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{N(v - \ell_i/N)} \right) - e^{-\theta G_\mu(v)} \right] + \frac{\psi_N^+(vN)}{H(v)} \mathbb{E} \left[ \prod_{i=1}^N \left( 1 + \frac{\theta}{N(v - (\ell_i + 1)/N)} \right) - e^{\theta G_\mu(v)} \right]. \tag{2.47}$$

We notice that from (2.46) and Assumption 1.5, the expectation terms and  $1/H(v)$  are all bounded uniformly (independent of  $N$ ). Thus, by Assumption 1.3, we can replace  $\psi_N^\pm(vN)$  by  $\phi^\pm(v)$ , which gives an error



$O(1/N)$ . The remaining part of the first term in (2.47) simplifies as follows:

$$\begin{aligned}
& \frac{\phi^-(v)}{H(v)} \mathbb{E} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{N(v - \ell_i/N)} \right) - e^{-\theta G_\mu(v)} \right] \\
&= \frac{\phi^-(v) e^{-\theta G_\mu(v)}}{H(v)} \mathbb{E} \left[ e^{-\theta(G_N(v) - G_\mu(v)) + O(\Theta_N(v) + \frac{\ln N}{(N\eta)^2})} - 1 \right] \\
&= - \frac{\phi^-(v) e^{-\theta G_\mu(v)}}{H(v)} \mathbb{E} [\theta(G_N(v) - G_\mu(v))] + O \left( \mathbb{E} [|G_N(v) - G_\mu(v)|^2 + \Theta_N(v)] + \frac{\ln N}{(N\eta)^2} \right) \\
&= - \frac{\phi^-(v) e^{-\theta G_\mu(v)}}{H(v)} \mathbb{E} [\theta(G_N(v) - G_\mu(v))] + O \left( \varepsilon^2 + \frac{|\mathbb{E} [\text{Im}[G_N(v) - G_\mu(v)]]|}{N\eta} + \Theta_\mu(v) + \frac{\ln N}{(N\eta)^2} \right),
\end{aligned} \tag{2.48}$$

where we used Lemma 2.2 in the second line, and (2.12) in the last line. We have a similar estimate for the second term in (2.47):

$$\begin{aligned}
& \frac{\phi^+(v)}{H(v)} \mathbb{E} \left[ \prod_{i=1}^N \left( 1 + \frac{\theta}{N(v - (\ell_i + 1)/N)} \right) - e^{\theta G_\mu(v)} \right] \\
&= \frac{\phi^+(v) e^{\theta G_\mu(v)}}{H(v)} \mathbb{E} [\theta(G_N(v) - G_\mu(v))] + O \left( \varepsilon^2 + \frac{|\mathbb{E} [\text{Im}[G_N(v) - G_\mu(v)]]|}{N\eta} + \Theta_\mu(v) + \frac{\ln N}{(N\eta)^2} \right).
\end{aligned} \tag{2.49}$$

Putting (2.48) and (2.49) together, and recalling Assumption 1.5, we get the following estimate of  $I_1$ :

$$I_1 = \theta \sqrt{(v-A)(v-B)} \mathbb{E} [G_N(v) - G_\mu(v)] + O \left( \varepsilon^2 + \frac{|\mathbb{E} [\text{Im}[G_N(v) - G_\mu(v)]]|}{N\eta} + \Theta_\mu(v) + \frac{1}{N} + \frac{\ln N}{(N\eta)^2} \right). \tag{2.50}$$

Next we estimate  $I_2$ . For  $z \in \mathcal{C}_2$ , we have  $\text{dist}(z, [\hat{a}_N, \hat{b}_N]) \gtrsim 1$ . So from Theorem 2.3,  $|G_N(z) - G_\mu(z)| \lesssim (\ln N/N)^{1/2}$  holds with high probability. Taylor expanding the exponential as  $e^x = 1 + x + O(x^2)$ , we deduce

$$\begin{aligned}
\psi_N^-(zN) \mathbb{E} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{N(z - \ell_i/N)} \right) - e^{-\theta G_\mu(z)} \right] &= \phi^-(z) \mathbb{E} \left[ e^{-\theta G_N(z) + O(1/N)} - e^{-\theta G_\mu(z)} \right] + O \left( \frac{1}{N} \right) \\
&= -\theta \phi^-(z) e^{-\theta G_\mu(z)} (G_N(z) - G_\mu(z)) + O \left( \frac{\ln N}{N} \right).
\end{aligned}$$

We have a similar estimate for the second term in  $I_2$ . Combining them we obtain

$$\begin{aligned}
I_2 &= \frac{1}{z\pi i} \int_{\mathcal{C}_2} \frac{\theta(G_N(z) - G_\mu(z)) (\phi^+(z) e^{\theta G_\mu(z)} - \phi^-(z) e^{-\theta G_\mu(z)})}{(z-v)H(z)} dz + O \left( \frac{\ln N}{N} \right) \\
&= \frac{\theta}{z\pi i} \int_{\mathcal{C}_2} \frac{(G_N(z) - G_\mu(z)) \sqrt{(z-A)(z-B)}}{z-v} dz + O \left( \frac{\ln N}{N} \right) = O \left( \frac{\ln N}{N} \right).
\end{aligned} \tag{2.51}$$

where the integral vanishes, since the integrand is analytic outside  $\mathcal{C}_2$ , and behaves like  $1/z^2$  as  $z \rightarrow \infty$ .

For  $I_3$ , since  $e^{\pm\theta G_\mu(z)} = O(1)$ , we can replace  $\psi_N^\pm(zN)$  by  $\phi^\pm(z)$ , which gives an error  $O(1/N)$ . We can rewrite the integrand as

$$\psi_N^-(zN)e^{-\theta G_\mu(z)} + \psi_N^+(zN)e^{\theta G_\mu(z)} = \phi^-(z)e^{-\theta G_\mu(z)} + \phi^+(z)e^{\theta G_\mu(z)} + O\left(\frac{1}{N}\right) = R_\mu(z) + O\left(\frac{1}{N}\right).$$

Since  $R_\mu(z)/H(z)$  is bounded and analytic inside  $\mathcal{C}_2$ , we have

$$I_3 = \frac{1}{2\pi i} \int_{\mathcal{C}_1 \cup \mathcal{C}_2} \frac{R_\mu(z) + O(1/N)}{(z-v)H(z)} dz = O\left(\frac{\ln N}{N}\right). \quad (2.52)$$

For  $I_4$ , if  $z \in \mathcal{C}_2$ , we have  $|z-v|, |z-\bar{v}| \gtrsim 1$ , and

$$|\mathcal{E}(z)| \lesssim \frac{K}{N^2}.$$

Therefore, it follows,

$$\left| \frac{1}{2\pi i} \int_{\mathcal{C}_2} \mathcal{E}(z) \psi_N^-(zN) \mathbb{E} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{N(z-\ell_i/N)} \right) \right] \frac{dz}{(z-v)H(z)} \right| \lesssim \frac{K}{N^2}. \quad (2.53)$$

The estimates are more involved for  $z \in \mathcal{C}_1$ . We recall that  $\mathcal{C}_1$  consists of two circles, one is centered at  $v$  with radius  $\eta/2$ , and the other is centered at  $\bar{v}$  with radius  $\eta/2$ . Similar to Lemma 2.2, for  $z \in \mathcal{C}_1$ , we have with high probability, since  $G_N(z) = O(1)$

$$\prod_{i=1}^N \left( 1 - \frac{\theta}{N(z-\ell_i/N)} \right) = e^{-\theta G_N(z) + O(1/N\eta)} = e^{-\theta G_N(z)} + O\left(\frac{1}{N\eta}\right),$$

and the same estimate holds for the expectation. Thanks to Assumption 1.3, (2.44) and (2.46), we have

$$\begin{aligned} & \frac{1}{2\pi i} \int_{\mathcal{C}_1} \mathcal{E}(z) \psi_N^-(zN) \mathbb{E} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{N(z-\ell_i/N)} \right) \right] \frac{dz}{z-v} = O\left(\frac{1}{N\eta} + \frac{K}{(N\eta)^3} + \frac{K^2}{(N\eta)^4}\right) \\ & - \frac{1}{2\pi i} \int_{\mathcal{C}_1} \left( \frac{tK}{N^2} \frac{1}{(v-z)^2} + \frac{\bar{t}K}{N^2} \frac{1}{(\bar{v}-z)^2} \right) \phi^-(z) \mathbb{E} \left[ e^{-\theta G_N(z)} \right] \frac{dz}{(z-v)H(z)}. \end{aligned} \quad (2.54)$$

Let

$$f(z) := \frac{\phi^-(z) \mathbb{E} [e^{-\theta G_N(z)}]}{H(z)}.$$

$f$  is analytic inside  $\mathcal{C}_1$  so that the last integral in (2.59) can be estimated by

$$\begin{aligned} I'_4 &= -\frac{1}{2\pi i} \frac{tK}{N^2} \int_{\mathcal{C}_1} \frac{f(z)}{(z-v)^3} dz - \frac{1}{2\pi i} \frac{\bar{t}K}{N^2} \int_{\mathcal{C}_1} \frac{f(z)}{(z-\bar{v})^2(z-v)} dz \\ &= \frac{tK}{2N^2} f''(v) - \frac{1}{2\pi i} \frac{\bar{t}K}{N^2} \int_{\mathcal{C}_1} f(z) \left( \frac{1}{(v-\bar{v})^2} \left( \frac{1}{z-v} - \frac{1}{z-\bar{v}} \right) - \frac{1}{(v-\bar{v})(z-\bar{v})^2} \right) dz \\ &= \frac{tK}{2N^2} f''(v) + \frac{\bar{t}K}{N^2} \left( \frac{1}{(v-\bar{v})^2} (f(v) - f(\bar{v})) - \frac{1}{v-\bar{v}} f'(\bar{v}) \right), \end{aligned} \quad (2.55)$$

Since  $\phi^-(z)$  and  $H(z)$  are both analytic, similar to the proof of Lemma 2.1, we have

$$|f(v) - f(\bar{v})| \lesssim (v - \bar{v}) + \mathbb{E}[|\operatorname{Im}[G_N(v)]|], \quad |f'(\bar{v})| \lesssim \frac{\mathbb{E}[|\operatorname{Im}[G_N(v)]|]}{\eta}, \quad |f''(v)| \lesssim \frac{\mathbb{E}[|\operatorname{Im}[G_N(v)]|]}{\eta^2}, \quad (2.56)$$

which gives an upper bound for (2.55),

$$I'_4 \lesssim \frac{K}{N^2\eta} + \frac{K\mathbb{E}[|\operatorname{Im}[G_N(v)]|]}{(N\eta)^2}. \quad (2.57)$$

Thanks to (2.11), we can rewrite  $f(z)$  in the following way:

$$\begin{aligned} f(z) &= \frac{\phi^-(z)\mathbb{E}[e^{-\theta G_N(z)}]}{H(z)} = \frac{\phi^-(z)\mathbb{E}\left[e^{-\int_{\hat{a}}^{\hat{b}} \frac{dx}{z-x} + \theta G_N^{\text{dual}}(z) + \mathcal{O}(\ln N/N\eta)}\right]}{H(z)} \\ &= \frac{(z - \hat{b})\phi^-(z)\mathbb{E}[e^{\theta G_N^{\text{dual}}(z)}]}{(z - \hat{a})H(z)} + \mathcal{O}\left(\frac{\ln N}{N\eta}\right). \end{aligned}$$

Since  $\phi^-(z)$  vanishes at  $\hat{a}$ ,  $\phi^-(z)/(z - \hat{a})$  is an analytic function on  $\mathcal{M}$ . We have similar estimates as in (2.56), but replacing  $G_N(v)$  by  $G_N^{\text{dual}}(v)$ , which gives another upper bound for  $I'_4$  in (2.55),

$$I'_4 \lesssim \frac{K}{N^2\eta} + \frac{K\mathbb{E}[|\operatorname{Im}[G_N^{\text{dual}}(v)]|]}{(N\eta)^2} + \frac{\ln NK}{(N\eta)^3}. \quad (2.58)$$

Therefore, combining (2.53), (2.57) and (2.58) together, we have the following estimate of  $I_4$ :

$$I_4 \lesssim \frac{K \min\{\mathbb{E}[|\operatorname{Im}[G_N(v)]|], \mathbb{E}[|\operatorname{Im}[G_N^{\text{dual}}(v)]|]\}}{(N\eta)^2} + \mathcal{O}\left(\frac{1}{N} + \frac{K}{N^2\eta} + \frac{\ln NK}{(N\eta)^3} + \frac{K^2}{(N\eta)^4}\right). \quad (2.59)$$

It follows by combining the estimates (2.50), (2.51), (2.52) and (2.59) all together, we have

$$\begin{aligned} \theta\sqrt{(v-A)(v-B)}\mathbb{E}[G_N(v) - G_\mu(v)] &= \mathcal{O}\left(\frac{1}{N\eta} + \frac{K}{(N\eta)^2}\right) |\mathbb{E}[\operatorname{Im}[G_N(v) - G_\mu(v)]]| + \\ &+ \mathcal{O}\left(\varepsilon^2 + \Theta_\mu(v) + \frac{K \min\{|\operatorname{Im}[G_\mu(v)]|, |\operatorname{Im}[G_\mu^{\text{dual}}(v)]|\}}{(N\eta)^2} + \frac{\ln N}{(N\eta)^2} + \frac{\ln N}{N} + \frac{\ln NK}{(N\eta)^3} + \frac{K^2}{(N\eta)^4}\right). \end{aligned} \quad (2.60)$$

Notice that  $\sqrt{(v-A)(v-B)} \asymp \sqrt{\kappa_E + \eta}$ , and by our assumption  $N\eta\sqrt{\kappa_E + \eta} \gg 1$ , and  $(N\eta)^2\sqrt{\kappa_E + \eta} \gg K$ , it follows by rearranging (2.60), that  $\mathbb{E}[G_N(v) - G_\mu(v)]$  is bounded by

$$\frac{1}{\sqrt{\kappa_E + \eta}} \mathcal{O}\left(\varepsilon^2 + \Theta_\mu(v) + \frac{K \min\{|\operatorname{Im}[G_\mu(v)]|, |\operatorname{Im}[G_\mu^{\text{dual}}(v)]|\}}{(N\eta)^2} + \frac{\ln N}{(N\eta)^2} + \frac{\ln N}{N} + \frac{\ln NK}{(N\eta)^3} + \frac{K^2}{(N\eta)^4}\right).$$

This finishes the proof of (2.25). In the following we prove (2.26). Set

$$\varepsilon_2 := \frac{\varepsilon^2}{\sqrt{\kappa_E + \eta}} + \tilde{\varepsilon}_0.$$

If  $\varepsilon \leq N^{-c} \sqrt{\kappa_E + \eta}$ , and  $\varepsilon \geq \tilde{\varepsilon}_0$ , we have

$$\varepsilon_2 \leq \varepsilon N^{-c} + \tilde{\varepsilon}_0, \quad \varepsilon_2 \leq 2\varepsilon.$$

Notice that

$$\partial_t \ln \mathbb{E}_{\mathbb{P}_N} \left[ e^{2K \operatorname{Re} \left[ \sum_{i=1}^N \ln \left( 1 + \frac{\alpha t}{N(v - \ell_i/N)} \right) \right]} \right] = 2K \mathbb{E}_{\mathbb{P}_N^{K,t,v,\alpha}} \left[ \operatorname{Re} \left[ \alpha G_N \left( v + \frac{\alpha t}{N} \right) \right] \right]. \quad (2.61)$$

For the righthand side of (2.61), we use (2.15) and (2.16) to find

$$G_N \left( v + \frac{\alpha t}{N} \right) - G_N(v) = -\frac{\alpha t}{N^2} \frac{1}{(v - \ell_i/N)^2} + O \left( \frac{1}{(N\eta)^2} \right) = O \left( \Theta_N(z) + \frac{1}{(N\eta)^2} \right). \quad (2.62)$$

(2.25) and (2.62) give the following bound of the righthand side of (2.61),

$$\begin{aligned} & 2K \mathbb{E}_{\mathbb{P}_N^{K,t,v,\alpha}} \left[ \operatorname{Re} \left[ \alpha G_N \left( v + \frac{\alpha t}{N} \right) \right] \right] \\ &= 2K \mathbb{E}_{\mathbb{P}_N^{K,t,v,\alpha}} \left[ \operatorname{Re}[\alpha(G_N(v) - G_\mu(v))] \right] + 2K \operatorname{Re}[\alpha G_\mu(v)] + K O \left( \mathbb{E}_{\mathbb{P}_N^{K,t,v,\alpha}} [\Theta_N(v)] + \frac{1}{(N\eta)^2} \right) \\ &= 2K \operatorname{Re}[\alpha G_\mu(v)] + K O \left( \varepsilon_2 + \frac{|\mathbb{E}_{\mathbb{P}_N^{K,t,v,\alpha}} [\operatorname{Im}[G_N(v) - G_\mu(v)]]|}{N\eta} + \Theta_\mu(v) + \frac{1}{(N\eta)^2} \right) \\ &= 2K \operatorname{Re}[\alpha G_\mu(v)] + O(K\varepsilon_2), \end{aligned}$$

where in the last two lines we used (2.12) and (2.25). It follows by integrating over  $t$  that for any  $\alpha = \pm 1, \pm i$ ,

$$\mathbb{E}_{\mathbb{P}_N} \left[ e^{2K \operatorname{Re} \left[ \sum_{i=1}^N \ln \left( 1 + \frac{\alpha}{N(v - \ell_i/N)} \right) \right]} \right] = e^{2K \operatorname{Re}[\alpha G_\mu(v)] + O(K\varepsilon_2)}.$$

Let  $K_2 = K/2$ , for any  $\alpha_2 = \pm 1, \pm i$ , by Cauchy-Schwarz inequality, we have

$$\mathbb{E}_{\mathbb{P}_N^{K_2,t,v,\alpha}} \left[ e^{2K_2 \operatorname{Re} \left[ \sum_{i=1}^N \ln \left( 1 + \frac{\alpha_2}{N(v - \ell_i/N)} \right) \right]} \right] = e^{2K_2 \operatorname{Re}[\alpha_2 G_\mu(v)] + O(K\varepsilon_2)}.$$

By the Markov inequality, the above implies that, with probability  $1 - e^{-csK\varepsilon_2}$  with respect to the measure  $\mathbb{P}_N^{K_2,t,v,\alpha}$ , the following holds

$$\operatorname{Re} \left[ \sum_{i=1}^N \ln \left( 1 + \frac{\alpha_2}{N(v - \ell_i/N)} \right) \right] - \operatorname{Re}[\alpha_2 G_\mu(v)] \lesssim s\varepsilon_2, \quad (2.63)$$

where  $s \geq 1$ . Thanks to (2.12) and Lemma 2.2, (2.63) leads to

$$\operatorname{Re}[\alpha_2(G_N(v) - G_\mu(v))] + O \left( \frac{|\operatorname{Im}[G_N(v) - G_\mu(v)]|}{N\eta} + \Theta_\mu(v) + \frac{\ln N}{(N\eta)^2} \right) \lesssim s\varepsilon_2. \quad (2.64)$$

Since  $N\eta \gg 1$ , and  $\varepsilon_2 \geq \Theta_\mu(v) + \ln N/(N\eta)^2$ , we deduce by taking  $\alpha_2 = \pm i$  in (2.64) that

$$|\operatorname{Im}[G_N(v) - G_\mu(v)]| \lesssim s\varepsilon_2. \quad (2.65)$$

By taking  $\alpha_2 = \pm 1$  in (2.64) and using (2.65) for the error, we get

$$|\operatorname{Re}[G_N(v) - G_\mu(v)]| \lesssim s\varepsilon_2. \quad (2.66)$$

Finally (2.65) and (2.66) together imply that

$$|G_N(v) - G_\mu(v)| \lesssim s\varepsilon_2.$$

w.r.t.  $\mathbb{P}_N^{K_2, t, v, \alpha}$  for any  $0 \leq t \leq 1$  and  $\alpha = \pm 1, \pm i$ . Especially, by taking  $s = 1$ , with high probability, w.r.t.  $\mathbb{P}_N^{K_2, t, v, \alpha}$ ,

$$G_N(v) = G_\mu(v) + O(\varepsilon_2). \quad (2.67)$$

(2.67) is in the same form as (2.24). However, we pay a price here, the moment  $K$  drops by half. If we repeat the above procedure for  $\lceil 1/\mathfrak{c} \rceil$  times, We will reach the optimal rigidity, with probability  $1 - e^{-\mathfrak{c}sK\tilde{\varepsilon}_0}$

$$|G_N(v) - G_\mu(v)| \lesssim s\tilde{\varepsilon}_0,$$

with respect to the measure  $\mathbb{P}_N^{K', t, v, \alpha}$  where  $K' = 2^{-\lceil 1/\mathfrak{c} \rceil} K \asymp K$ . The statement (2.26) follows by taking  $t = 0$ .

□

### 3 Local Measure

In this Section we prove Theorem 2.6. The main idea is to localize our measure by conditioning, and therefore get estimates on finer scales. We fix small constants  $r_0, r > 0$ , the parameters  $N^a \leq M \ll N$  and  $L = M^b$ , as in the statement of Theorem 2.6. We recall that  $d\mu(x) = \rho_V(x)dx$  denotes the constrained equilibrium measure and  $\gamma_1, \gamma_2, \dots, \gamma_N$  the corresponding classical particle locations defined in (1.5).

**Definition 3.1.** Fix a scale  $\tilde{\eta} \geq \ln N/N$ . We say that  $(\ell_1, \ell_2, \dots, \ell_N) \in \mathbb{W}_N^\theta$  satisfies the rigidity estimates on the scale  $\tilde{\eta}$  in the spectral domain  $\tilde{\mathcal{D}}$ , if

$$|G_N(z) - G_\mu(z)| \leq \frac{\tilde{\eta}}{\operatorname{Im}[z]}, \quad z \in \tilde{\mathcal{D}}. \quad (3.1)$$

The above estimates (3.1) combining with (2.11) lead to

$$|G_N^{\text{dual}}(z) - G_\mu^{\text{dual}}(z)| \leq \frac{2\tilde{\eta}}{\operatorname{Im}[z]}, \quad z \in \tilde{\mathcal{D}}.$$

In the following propositions, we collect some consequences of rigidity estimates in the spectral domains  $\mathcal{D}_{M, r_0}$  and  $\mathcal{D}_*$  (as defined in (2.1) and (2.2)) The proofs follow an application of the Helffer-Sjöstrand functional calculus along the lines of [46, Lemma B.1]. We postpone them to the appendix B.

**Proposition 3.2.** We assume  $(\ell_1, \ell_2, \dots, \ell_N) \in \mathbb{W}_N^\theta$  satisfies the rigidity estimates on the scale  $\tilde{\eta} = M/N$  in the spectral domains  $\mathcal{D}_{M, r_0}$  (as in Definition 3.1). For any interval  $I = [a, b] \subset \mathcal{M}$  with length  $|I| = \eta$ , we denote  $\kappa_I = \operatorname{dist}\{I, \{A, B\}\}$ . There are two cases:

1. the bulk case, where  $A \leq a$  and  $b \leq B - r_0$ . We assume that  $\kappa_I \geq \tilde{\eta}^{2/3}$ ,  $\eta \geq \tilde{\eta}/\sqrt{\kappa_I}$ , and

$$\kappa_I \asymp \sup\{\text{dist}(x, \{A, B\}) : x \in I\};$$

2. the edge case, where  $a \leq A$  and  $b \leq B - r_0$ . We assume  $\eta \geq \tilde{\eta}^{2/3}$ .

Then for any  $C^2$  function  $f$  supported on  $I$ , with  $f(x) = O(1)$ ,  $f'(x) = O(1/\eta)$  and  $f''(x) = O(1/\eta^2)$ , we have

$$\left| \frac{1}{N} \sum_{i=1}^N f(\ell_i/N) - \int f(x) d\mu(x) \right| \lesssim \tilde{\eta} \ln N. \quad (3.2)$$

Moreover, there exists a constant  $C$ , such that uniformly for any index  $i \in \llbracket 1, N \rrbracket$  such that  $\gamma_i \leq B - r_0 - r/2$ ,

$$\gamma_{i-CN\tilde{\eta} \ln N} \leq \ell_i/N \leq \gamma_{i+CN\tilde{\eta} \ln N}, \quad (3.3)$$

where we make the convention that  $\gamma_i = a(N)/N$  if  $i \leq 0$ .

**Proposition 3.3.** We assume that  $(\ell_1, \ell_2, \dots, \ell_N) \in \mathbb{W}_N^\theta$  satisfies the rigidity estimates on the scale  $\tilde{\eta} = M/N$  in the spectral domains  $\mathcal{D}_*$  (as in Definition 3.1). Then for any  $C^2$  function  $f$  supported in  $\mathcal{M}$ , with  $f(x) = O(1)$ ,  $f'(x) = O(1)$  and  $f''(x) = O(1)$ , we have

$$\left| \frac{1}{N} \sum_{i=1}^N f(\ell_i/N) - \int f(x) d\mu(x) \right| \lesssim \tilde{\eta}. \quad (3.4)$$

**Remark 3.4.** Given any  $b \in [A, B]$ , with  $\kappa_b = \text{dist}(b, \{A, B\}) \gtrsim (M/N)^{2/3}$ . Fix a scale  $\tilde{\eta} \gtrsim M/N$ . Let  $b_0 := b$ . Given  $b_k < B$ , we define  $b_{k+1}$  in the following way: if  $B - b_k \geq 2^{k+1}\tilde{\eta}/\sqrt{\kappa_b}$ , we define  $b_{k+1} := b_k + 2^k\tilde{\eta}/\sqrt{\kappa_b}$ ; otherwise, we take  $b_{k+1} := b_k + 2^{k+2}\tilde{\eta}/\sqrt{\kappa_b}$ . Let  $\hat{b}_N = \max\{\hat{b}, b(N)/N\}$ . Given  $B < b_k < \hat{b}$ , we define  $b_{k+1} := b_k + 2^{k+2}\tilde{\eta}/\sqrt{\kappa_b}$ . In this way, we get a sequence of points  $b_0 < b_1 < b_2 < \dots < b_n < B < b_{n+1} < b_{n+2} < \dots < b_{n'} < \hat{b}_N < b_{n'+1}$ . It is easy to check that

- For  $k \in \llbracket 0, n \rrbracket$ , we have  $b_k - A \asymp (b_0 - A) + 2^k\tilde{\eta}/\sqrt{\kappa_b}$  and  $B - b_k \asymp 2^n\tilde{\eta}/\sqrt{\kappa_b}$ .
- For  $k \in \llbracket n, n' \rrbracket$ , we have  $b_{k+1} - b_k = 2^{k+2}\tilde{\eta}/\sqrt{\kappa_b} \gtrsim (M/N)^{2/3}$ .

We denote  $I_k = [b_{k-1}, b_{k+1}]$  for  $k \in \llbracket 1, n' \rrbracket$ . With the properties of  $b_k$  listed above, it is easy to check  $I_k$ , for  $k \in \llbracket 1, n-1 \rrbracket$ , are bulk intervals, and  $I_k$  for  $k \geq n$ , are edge intervals, and they satisfy the conditions in Proposition 3.2. We can take bump functions  $\chi_k$  supported on  $I_k$ , with values in  $[0, 1]$ ,  $\chi'_k(x) \asymp 1/|I_k|$  and  $\chi'_k(x) \asymp 1/|I_k|^2$ , such that  $\chi_1(x) + \chi_2(x) + \dots + \chi_{n'}(x) = 1$  on  $[b_1, \hat{b}_N]$ . We call  $\chi_1, \chi_2, \dots, \chi_{n'}$  a dyadic partition of unity on the scale  $\tilde{\eta}$ .

For any  $C^2$  function  $f$ , by combining Proposition 3.2 with a dyadic decomposition given above, we can estimate the integral

$$\int_b^{\hat{b}_N} f(x) (d\mu_N(x) - d\mu(x)), \quad (3.5)$$

for some  $b \in [A, B]$  with  $\kappa_b = \text{dist}(b, \{A, B\}) \gtrsim (M/N)^{2/3}$  by decomposing it as the sum

$$\int_{b_0}^{b_1} (1 - \chi_1(x))f(x) (d\mu_N(x) - d\mu(x)) + \sum_{k=1}^{n'} \int \chi_k(x)f(x) (d\mu_N(x) - d\mu(x)). \quad (3.6)$$

Then we can estimate (3.6) term by term using Proposition 3.2.

By symmetry, for the integral

$$\int_{\hat{a}_N}^a f(x) (d\mu_N(x) - d\mu(x)),$$

where  $\hat{a}_N = \min\{\hat{a}, a(N)/N\}$ , we can construct a sequence of points  $a = a_0 > a_1 > a_2 > \dots > a_n > A > a_{n+1} > a_{n+2} > \dots > a_{n'} > \hat{a}_N > a_{n'+1}$  and bump functions  $\chi_1, \chi_2, \dots, \chi_{n'}$ .

We assume that with high probability, the rigidity estimates hold on the scale  $M/N$  in the spectral domain  $\mathcal{D}_{M,r_0} \cup \mathcal{D}_*$ , as in Definition 3.1. We fix  $M \ll L \ll N$  and consider the conditioned discrete  $\beta$  ensembles. More precisely, there are two cases:

1. the bulk case: We fix a bulk index  $K \in \llbracket [L/2] + 1, N_0 \rrbracket$ , where  $N_0 := \max\{i : \gamma_i \leq B - r_0 - r\}$ . We condition on the particles  $\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N$  and resample the particles  $\ell_{K+1}, \ell_{K+2}, \dots, \ell_{K+L}$ . Then  $\ell_{K+1}/N, \ell_{K+2}/N, \dots, \ell_{K+L}/N \in I$ , where

$$I = [a, b], \quad a = (\ell_K + \theta)/N, \quad b = (\ell_{K+L+1} - \theta)/N. \quad (3.7)$$

We denote  $\kappa_I = \text{dist}(I, \{A, B\})$ .

2. the left edge case: We condition on the particles  $\ell_{L+1}, \ell_{L+2}, \dots, \ell_N$  and resample the particles  $\ell_1, \ell_2, \dots, \ell_L$ . Then  $\ell_1/N, \ell_2/N, \dots, \ell_L/N \in I$ , where

$$I = [a, b], \quad a = a(N)/N, \quad b = (\ell_{L+1} - \theta)/N. \quad (3.8)$$

We denote  $\kappa_I = \text{dist}(b, \{A, B\})$ .

The conditioned discrete  $\beta$ -ensemble is given by

$$\mathbb{P}_L(\ell_{K+1}, \ell_{K+2}, \dots, \ell_{K+L}) = \frac{1}{Z_L} \prod_{K+1 \leq i < j \leq K+L} \frac{\Gamma(\ell_j - \ell_i + 1)\Gamma(\ell_j - \ell_i + \theta)}{\Gamma(\ell_j - \ell_i)\Gamma(\ell_j - \ell_i + 1 - \theta)} \prod_{i=K+1}^{K+L} \hat{w}(\ell_i; L), \quad (3.9)$$

where

$$\hat{w}(Nu; L) = e^{-NV_N(u)} \prod_{\substack{j \leq K, \text{ or} \\ j \geq K+L+1}} \frac{\Gamma(\ell_j - Nu + 1)\Gamma(\ell_j - Nu + \theta)}{\Gamma(\ell_j - Nu)\Gamma(\ell_j - Nu + 1 - \theta)},$$

if  $u \in I$ , otherwise  $\hat{w}(Nu; L) = 0$ . Notice that  $K \in \llbracket [L/2] + 1, N_0 \rrbracket$  (where  $N_0 = \max\{i : \gamma_i \leq B - r_0 - r\}$ ) corresponds to the bulk case,  $K = 0$  corresponds to the left edge case. In the rest of this section, We assume that  $K \in \llbracket [L/2] + 1, N_0 \rrbracket$  or  $K = 0$ .

**Definition 3.5.** We define the set  $\mathcal{G}_{M,L,K}$  of “good” boundary conditions as the particle configurations

$$(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N)$$

such that with high probability with respect to the conditioned discrete  $\beta$ -ensemble  $\mathbb{P}_L$ , the particle configuration  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+1}, \dots, \ell_{K+L}, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N)$  satisfies the rigidity estimates on the scale  $M/N$  in the spectral domain  $\mathcal{D}_{M,r_0} \cup \mathcal{D}_*$ , as in Definition 3.1.

The following proposition is an easy consequence of Proposition 3.2.

**Proposition 3.6.** We assume that  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of “good” boundary conditions, and  $I$  is as defined in (3.7) or (3.8). In the bulk case, we have  $\kappa_I \gtrsim (L/N)^{2/3}$ ,  $|I| \asymp L/N \sqrt{\kappa_I} \lesssim \kappa_I$ , and  $\rho_V(x) \asymp \sqrt{\kappa_I}$ . And for the left edge case  $\kappa_I = b - A \asymp (L/N)^{2/3}$ , and  $\rho_V(x) \asymp \sqrt{[x - A]_+}$ . In all cases, we have  $I \subset [a(N)/N, B - r_0 - r/2]$ ,

$$\left| N \int_I \rho_V(x) dx - L \right| = O(M \ln N).$$

We can sample the particle configurations  $(\ell_1, \ell_2, \dots, \ell_N)$  using the following two-step sampling process: we first sample the boundary  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N)$  according to the marginal distribution; then we sample  $(\ell_{K+1}, \ell_{K+2}, \dots, \ell_{K+L})$  according to  $\mathbb{P}_L$ . As a consequence of this point of view, the following proposition holds.

**Proposition 3.7.** We assume the rigidity estimates (as in Definition 3.1) hold on the scale  $M/N$  with probability  $1 - p_1$  for the discrete  $\beta$ -ensemble. Then with probability at least  $1 - p_1 e^{c(\log N)^2}$ , we have  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of “good” boundary conditions.

**Remark 3.8.** If the rigidity estimates (as in Definition 3.1) hold on the scale  $M/N$  with high probability for the discrete  $\beta$ -ensemble, then with high probability,  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N)$  belongs to  $\mathcal{G}_{M,L,K}$ .

It follows from Stirling’s formula that

$$\frac{\Gamma(h + \theta)}{\Gamma(h)} = h^\theta (1 + O(h^{-1})) = h^\theta e^{O(h^{-1})}.$$

We can simplify the expression of the conditioned discrete  $\beta$ -ensemble (3.9)

$$\begin{aligned} \prod_{K+1 \leq i < j \leq K+L} \frac{\Gamma(\ell_j - \ell_i + 1) \Gamma(\ell_j - \ell_i + \theta)}{\Gamma(\ell_j - \ell_i) \Gamma(\ell_j - \ell_i + 1 - \theta)} &= \prod_{K+1 \leq i < j \leq K+L} |\ell_j - \ell_i|^{2\theta} e^{O(|\ell_j - \ell_i|^{-1})} \\ &= e^{O(L \ln N)} \prod_{K+1 \leq i < j \leq K+L} |\ell_j - \ell_i|^{2\theta}, \end{aligned} \quad (3.10)$$

and

$$\begin{aligned} \hat{w}(Nu; L) &= e^{-NV(u) + O(\ln N)} \prod_{\substack{j \leq K, \text{ or} \\ j \geq K+L+1}} |\ell_j - Nu|^{2\theta} e^{O(|\ell_j - Nu|^{-1})} \\ &= e^{-NV(u) + O(\ln N)} \prod_{\substack{j \leq K, \text{ or} \\ j \geq K+L+1}} |\ell_j - Nu|^{2\theta}, \end{aligned} \quad (3.11)$$



where we used  $|\ell_j - \ell_i| \geq |j - i|\theta$  and Assumption 1.2. We can rewrite the conditioned discrete  $\beta$ -ensemble (3.9) by plugging in (3.10) and (3.11)

$$\mathbb{P}_L(\ell_{K+1}, \ell_{K+2}, \dots, \ell_{K+L}) = \frac{1}{Z'_L} e^{-H_L(\ell_{K+1}, \ell_{K+2}, \dots, \ell_{K+L}) + O(L \ln N)} \quad (3.12)$$

where the Hamiltonian is given by

$$H_L = -\theta \sum_{K+1 \leq i \neq j \leq K+L} \ln \left| \frac{\ell_i}{N} - \frac{\ell_j}{N} \right| + L \sum_{i=K+1}^{K+L} W \left( \frac{\ell_i}{N} \right), \quad (3.13)$$

where

$$W(x) = \begin{cases} \frac{N}{L} \left( V(x) - \frac{2\theta}{N} \sum_{\substack{j \leq K, \text{ or} \\ j \geq K+L+1}} \ln \left| x - \frac{\ell_j}{N} \right| - f_V \right), & x \in I; \\ -\infty, & x \notin I, \end{cases} \quad (3.14)$$

where  $f_V$  is as defined in (1.4). For the new potential  $W(x)$ , we can rewrite it as

$$W(x) = \frac{N}{L} \left( F_V(x) + 2\theta \int_I \ln |x - y| \rho_V(y) dy + 2\theta \int_{I^c} \ln |x - y| (\rho_V(y) dy - d\mu_N(y)) \right), \quad (3.15)$$

where  $F_V(x) \geq 0$  vanishes inside  $[A, B]$ , and  $F'_V(x)$  has square root behavior outside  $[A, B]$ .

**Proposition 3.9.** *We assume  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of "good" boundary conditions, and  $I$  is as defined in (3.7) or (3.8). Then for  $x \in I$ , we have  $W(x) = NF_V(x)/L + O(\ln N)$  and  $W'(x) = O(\ln NN/L)$ .*

*Proof.* For the second term in (3.15), if  $I$  is a bulk interval, let  $\kappa_I = \text{dist}(I, \{A, B\})$ . Take  $x \in I$ . Then from Proposition 3.6,  $|I| \asymp L/N\sqrt{\kappa_I}$  and  $\rho_V(x) \asymp \sqrt{\kappa_I}$ , we deduce that

$$\frac{2\theta N}{L} \int_I \ln |x - y| \rho_V(y) dy \asymp \frac{N\sqrt{\kappa_I}}{L} \int_I \ln |x - y| dy = O(\ln N).$$

If  $I$  is the left edge interval, then  $b - A \asymp (L/N)^{2/3}$ , where  $b$  is the right endpoint of  $I$ ,  $\rho_V(x) \asymp \sqrt{|x - A|_+}$  and

$$\frac{2\theta N}{L} \int_I \ln |x - y| \rho_V(y) dy \asymp \frac{N}{L} \int_A^b \ln |x - y| \sqrt{|y - A|_+} dy = O(\ln N).$$

For the last term in (3.15),

$$\frac{2\theta N}{L} \int_{y \geq b} \ln |x - y| (\rho_V(y) dy - d\mu_N(y)) + \frac{2\theta N}{L} \int_{y \leq a} \ln |x - y| (\rho_V(y) dy - d\mu_N(y)). \quad (3.16)$$

We first show the bound for the first term. We recall the construction from Remark 3.4. Let  $\kappa_b = \text{dist}(b, \{A, B\})$ ,  $\kappa_b \gtrsim (L/N)^{2/3}$ , and construct the bump functions  $\chi_1, \chi_2, \dots, \chi_{n'}$  with scale  $\tilde{\eta} = M/N$ .

$$\begin{aligned} & \int_{y \geq b} \ln |x - y| (\rho_V(y) dy - d\mu_N(y)) \\ &= \int_{b_0}^{b_1} (1 - \chi_1(y)) \ln |x - y| (\rho_V(y) dy - d\mu_N(y)) + \sum_{k=1}^{n'} \int_{y \geq b} \chi_k(y) \ln |x - y| (\rho_V(y) dy - d\mu_N(y)). \end{aligned} \quad (3.17)$$

The first term in (3.17) can be bounded by  $\ln NM/N$ . For the sum, from our construction of  $\chi_1, \chi_2, \dots, \chi_{n'}$ , for each  $k \in \llbracket 1, n' \rrbracket$ , we have

$$\chi_k(y) \ln |x - y| = O(\ln N), \quad \partial_y (\chi_k(y) \ln |x - y|) = O(\ln N/|I_k|), \quad \partial_y^2 (\chi_k(y) \ln |x - y|) = O(\ln N/|I_k|^2),$$

where  $\chi_k(y)$  is supported on  $I_k$ . Moreover, either  $I_k \subset [b, B - r_0]$ , or  $I_k$  is not completely contained in  $[b, B - r_0]$ , in which case  $|I_k| \gtrsim 1$ . Therefore, it follows from Proposition 3.2 and 3.3, we have

$$\sum_{k=1}^{n'} \int_{y \geq b} \chi_k(y) \ln |x - y| (\rho_V(y) dy - d\mu_N(y)) \lesssim \sum_{k=1}^{n'} \frac{(\ln N)^2 M}{N} \lesssim \frac{(\ln N)^3 M}{N}.$$

The second term in (3.16) can be estimated in the same way, and the first claim  $W(x) = NF_V(x)/L + O(\ln N)$  follows.

For  $W'(x)$ , using (3.14), we have for  $x \in I$

$$\begin{aligned} W'(x) &= \frac{N}{L} \left( V'(x) - \frac{2\theta}{N} \sum_{j \leq K} \frac{1}{x - \ell_j/N} - \frac{2\theta}{N} \sum_{j \geq K+L+1} \frac{1}{x - \ell_j/N} \right) \\ &= \frac{N}{L} \left( O(\ln N) + \frac{4\theta}{N} \sum_{j \geq 1} \frac{1}{j\theta/N} \right) = O\left(\frac{N \ln N}{L}\right). \end{aligned}$$

□

We denote the constrained equilibrium measure of the conditioned discrete  $\beta$ -ensemble (3.12) as the unique probability measure on  $I$  characterized by

$$W(x) - 2\theta \int_a^b \ln |x - y| \rho_W(y) dy - f_W = F_W(x) = \begin{cases} \leq 0, & \text{on } \rho_W(x) = N/L\theta, \\ = 0, & \text{on } 0 < \rho_W(x) < N/L\theta, \\ \geq 0, & \text{on } \rho_W(x) = 0. \end{cases} \quad (3.18)$$

We will need some properties of the constrained equilibrium measure. Given an interval  $I \subset \mathbb{R}$ , a positive measure  $\sigma$  supported on  $I$  with total mass  $\|\sigma\| > 1$ , and a potential function  $Q$ . We denote  $\rho_Q$  the unconstrained equilibrium measure, i.e the unique probability measure such that there exists a constant  $f_Q$  so that

$$Q(x) - 2\theta \int_I \ln |x - y| \rho_Q(y) dy - f_Q = \begin{cases} = 0, & \text{on } \rho_Q(x) > 0, \\ \geq 0, & \text{on } \rho_Q(x) = 0. \end{cases} \quad (3.19)$$

The equilibrium measure constrained by being bounded above by  $\sigma$  is denoted  $\rho_Q^\sigma$ .

**Theorem 3.10.** [Theorem 2.6 [40]]  $\text{supp}(\rho_Q) \subset \text{supp}(\rho_Q^\sigma)$  and

$$\mathbf{1}_{\{\rho_Q^\sigma < \sigma\}} \rho_Q^\sigma \geq \mathbf{1}_{\{\rho_Q^\sigma < \sigma\}} \rho_Q.$$

**Theorem 3.11.** [Theorem 2.17 [40]] If  $Q$  is convex on  $I$ , then the support of  $\rho_Q^\sigma$  is a single interval.

**Theorem 3.12.** [Corollary 2.11 [40]] Let

$$s = \|\sigma\| - 1, \quad U = \frac{1}{s} \left( 2\theta \int_I \ln|x-y|\sigma(y)dy - Q(x) \right).$$

Then we have

$$\sigma = \rho_W^\sigma + s\rho_U^{\sigma/s}.$$

**Proposition 3.13.** We assume  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of “good” boundary conditions, and  $I$  is as defined in (3.7) or (3.8). Then both  $\rho_W$  and  $N/L\theta - \rho_W$  are supported on single intervals. Especially, there exists a subinterval  $[\alpha, \beta] \subset I = [a, b]$ , such that

1. For  $x \in (\alpha, \beta)$ ,  $0 < \rho_W(x) < N/L\theta$ ;
2. In the bulk case, if  $a < \alpha$ , then  $\rho_W(x) \equiv 0$  or  $\rho_W(x) \equiv N/L\theta$  on  $x \in [a, \alpha]$ ; if  $\beta < b$ , then  $\rho_W(x) \equiv 0$  or  $\rho_W(x) \equiv N/L\theta$  on  $x \in [\beta, b]$ . In the left edge case:  $a < \alpha$  and  $\rho_W(x) \equiv 0$  on  $x \in [a, \alpha]$ ; if  $\beta < b$ , then  $\rho_W(x) \equiv 0$  or  $\rho_W(x) \equiv N/L\theta$  for  $x \in [\beta, b]$ .

**Remark 3.14.** The previous proposition uses crucially that  $1/\theta - \rho_V$  is uniformly bounded below by a positive constant on  $I$ .

*Proof.* Thanks to Theorems 3.11 and 3.12, we only need to show that  $W$  and  $2(N/L) \int_I \ln|x-y|dy - W$  are convex on  $I$ . We first check that  $W$  is convex,

$$W''(x) = \frac{N}{L} \left( V''(x) + \frac{2\theta}{N} \sum_{j \geq K+L+1} \frac{1}{(x - \ell_j/N)^2} + \frac{2\theta}{N} \sum_{j \leq K} \frac{1}{(x - \ell_j/N)^2} \right). \quad (3.20)$$

We study the first sum, the second sum is analogous. Let  $\kappa_b = \text{dist}(b, \{A, B\})$ , notice that if  $I$  is a bulk interval  $\kappa_b \gtrsim (L/N)^{2/3}$ , and if  $I$  is the left edge interval  $\kappa_b \lesssim (L/N)^{2/3}$ . We take an interval  $J = [b, b']$ , where  $b' := b + c(L/N\sqrt{\kappa_b})$ , for some small  $c > 0$ , such that  $J$  satisfies the conditions in Proposition 3.2. Then we have

$$\int_J d\mu_N(x) = \int_J \rho_V(x)dx + O(\ln NM/N) \asymp L/N.$$

Therefore, we can lower bound the first sum in (3.20) for all  $x \in I = [a, b]$ :

$$\frac{1}{N} \sum_{j=K+L+1}^N \frac{1}{(x - \ell_j/N)^2} \gtrsim \frac{1}{N} \sum_{\substack{j \geq K+L+1: \\ \ell_j/N \leq b'}} \frac{1}{(x - b')^2} \gtrsim \frac{L}{N} \frac{1}{(x - b')^2} \gtrsim \frac{L}{N} \frac{1}{(L/N\sqrt{\kappa_b})^2} \gtrsim \frac{N}{L\kappa_b} \gtrsim \left(\frac{N}{L}\right)^{1/3}.$$

We deduce that  $W''(x) \gtrsim (N/L)^{4/3}$ , and  $W$  is convex on  $I$ .

The second derivative of  $2(N/L) \int_I \ln|x-y|dy - W$  is given by

$$\frac{2N}{L} \left( \frac{1}{x-a} + \frac{1}{b-x} - \frac{V''(x)}{2} - \frac{\theta}{N} \sum_{j \geq K+L+1} \frac{1}{(x - \ell_j/N)^2} - \frac{\theta}{N} \sum_{j \leq K} \frac{1}{(x - \ell_j/N)^2} \right).$$

Notice that for any  $\ell > 0$

$$\frac{\theta}{N} \sum_{j=1}^{\infty} \frac{1}{(\ell + j\theta/N)^2} \leq \theta \int_0^{\infty} \frac{dx}{(\ell + \theta x)^2} = \frac{1}{\ell}.$$

Let  $J$  be the interval as defined above. The same argument gives

$$\begin{aligned} \frac{1}{b-x} - \frac{\theta}{N} \sum_{j=K+L+1}^N \frac{1}{(x - \ell_j/N)^2} &\geq \frac{\theta}{N} \sum_{j=1}^{\infty} \frac{1}{(b-x + j\theta/N)^2} - \frac{\theta}{N} \sum_{j=K+L+1}^N \frac{1}{(x - \ell_j/N)^2} \\ &\geq \frac{\theta}{(b'-x)^2} \left( |J|/\theta - \int_J d\mu_N(x) \right) \\ &\geq \frac{\theta}{(b'-x)^2} \left( |J|/\theta - \int_J \rho_V(x) dx - O(M/N) \right) \\ &\gtrsim \frac{|J|}{(b'-x)^2} \gtrsim \frac{N\sqrt{\kappa b}}{L} \gtrsim \left( \frac{N}{L} \right)^{2/3}, \end{aligned}$$

where we used that  $\ell_{K+L+1}/N - x \geq \theta/N$ ,  $|\ell_{j+1} - \ell_i| \geq \theta$  and  $\rho_V(x) < \theta^{-1}$ . Similarly if  $I$  is a bulk interval (otherwise this sum is empty),

$$\frac{1}{x-a} - \frac{\theta}{N} \sum_{j=1}^K \frac{1}{(x - \ell_i/N)^2} \gtrsim \left( \frac{N}{L} \right)^{2/3}.$$

It follows that  $((N/L) \int_I \ln|x-y| dy - W)'' \gtrsim (N/L)^{5/3}$ , and  $(N/L) \int_I \ln|x-y| dy - W$  is convex. □

**Proposition 3.15.** *If  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of “good” boundary conditions, and  $I$  is as defined in (3.7) or (3.8), then for  $x \in I$ , we have  $F_W(x) = NF_V(x)/L + O(\ln N)$ . Moreover for any  $p > 0$ , on  $[a + N^{-p}, \alpha - N^{-p}] \cup [\beta + N^{-p}, b - N^{-p}]$  ( $\alpha$  and  $\beta$  are as in Proposition 3.13), we have  $F'_W(x) = O(\ln NN/L)$ .*

*Proof.* From the defining relation (3.18) of  $F_W$ , and  $W(x) = NF_V(x)/L + O(\ln N)$  from Proposition 3.9, we have

$$F_W(x) = \frac{NF_V(x)}{L} + O(\ln N) + \int_I \ln|x-y| \rho_W(y) dy + f_W = \frac{NF_V(x)}{L} + f_W + O(\ln N),$$

where we used that  $\rho_W$  has total mass 1 and  $\rho_W \leq N/L\theta$ . If  $I$  is a bulk interval, then for any  $x \in \text{supp } \rho_W$ , we have  $x \in [A, B]$ , and  $F_W(x) = F_V(x) = 0$ . Therefore  $f_W = O(\ln N)$ . If  $I$  is the left edge interval, there are two cases. If  $\rho_W \equiv 1$  on  $x \in [\beta, b]$  (where  $\beta$  is as in Proposition 3.13), then  $b - \beta \lesssim L/N \ll b - A \asymp (L/N)^{2/3}$ . As a consequence  $\beta \in [A, B]$ ,  $F_W(\beta) = 0$  and  $F_V(\beta) = 0$ . We have  $f_W = O(\ln N)$ . Otherwise, we have  $F_W(a) \geq 0$  and  $F_V(a) = 0$ . So  $f_W \geq -C \ln N$ . On the other hand, for any  $x \in \text{supp } \rho_W$ , we have  $F_W(x) = 0$  and  $F_V(x) \geq 0$ . So  $f_W \leq C \ln N$ . In any case, we have  $f_W = O(\ln N)$ , and  $F_W(x) = NF_V(x)/L + O(\ln N)$ .

From Proposition (3.9),  $W'(x) = O(\ln NN/L)$ . Thus,

$$F'_W(x) = O(\ln NN/L) + P.V. \int_I \frac{\rho_W(y)}{x-y} dy.$$

By symmetry, we assume  $x \in [\beta + N^{-p}, b - N^{-p}]$  (the case when  $x \in [a + N^{-p}, \alpha - N^{-p}]$  is analogous). Since  $\rho_W \leq N/L\theta$  and on  $[\beta, b]$   $\rho_W \equiv 0$  or  $\rho_W \equiv N/L\theta$ ,

$$P.V. \int_I \frac{\rho_W(y)}{x-y} dy = \int_a^\beta \frac{\rho_W(y)}{x-y} dy + P.V. \int_\beta^b \frac{\rho_W(y)}{x-y} dy = O(\ln NN/L).$$

□

For any particle configuration  $\ell_{K+1} < \ell_{K+2} < \dots < \ell_{K+L} \in \mathbb{W}_N^\theta$ , we denote  $\mu_L = 1/L \sum_{i=K+1}^{K+L} \delta_{\ell_i/N}$ , and  $\rho_L(x)dx$  the convolution of the empirical measure  $\mu_L$  with uniform measure on the interval  $[-N^{-p}/2, N^{-p}/2]$ . For any measure  $\rho(x)dx$ , we denote

$$\mathcal{D}^2[\rho] = - \int \ln |x-y| \rho(x)\rho(y) dx dy. \quad (3.21)$$

We have the following large deviation estimate:

**Proposition 3.16.** *Given  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of “good” boundary conditions, and  $I$  is as defined in (3.7) or (3.8). Let  $\rho_W$  be the density defined by (3.19). Then there exists  $C \in \mathbb{R}$  such that for all  $\gamma > 0$  we have*

$$\mathbb{P}_L(\mathcal{D}[\rho_W - \rho_L] \geq \gamma) \leq \exp(C \ln NL - \gamma^2 L^2), \quad (3.22)$$

and for any Lipschitz function  $g(x)$  on  $I$ , if  $\|g\|_{1/2} = (\int_0^\infty s|\hat{g}|^2 ds)^{1/2}$ , we have

$$\mathbb{P}_L\left(\left|\int g(x)(d\mu_L(x) - \rho_W(x)dx)\right| \geq \gamma\|g\|_{1/2} + N^{-p}\|g\|_{Lip}\right) \leq \exp(C \ln NL - \gamma^2 L^2). \quad (3.23)$$

*Proof.* The Hamiltonian  $H_L$  from (3.13) can be written as

$$\begin{aligned} L^{-2}H_L(\ell_{K+1}, \ell_{K+2}, \dots, \ell_{K+L}) &= -\theta \int_{x \neq y} \ln |x-y| d\mu_L(x)d\mu_L(y) + \int W(x)d\mu_L(x) \\ &= -\theta \int \ln |x-y| \rho_L(x)\rho_L(y) dx dy + \int W(x)\rho_L(x) dx + O\left(\frac{\ln N}{L}\right) \\ &= \theta \mathcal{D}^2[\rho_W - \rho_L] + \theta \int \ln |x-y| \rho_W(x)\rho_W(y) dx dy + \int F_W(x)\rho_L(x) dx + f_W + O\left(\frac{\ln N}{L}\right), \end{aligned} \quad (3.24)$$

where we used

$$W(x) = 2\theta \int \ln |x-y| \rho_W(y) dy + F_W(x) + f_W.$$

From the relation (3.18),  $F_W(x) < 0$  happens only if  $x \in [a, \alpha]$  and  $\rho_W \equiv N/L\theta$ , or  $x \in [\beta, b]$  and  $\rho_W \equiv N/L\theta$ , where  $\alpha, \beta$  are as in Proposition (3.13).

$$\int F_W(x)\rho_L(x) dx \geq \int_{F_W < 0} F_W(x)\rho_L(x) dx. \quad (3.25)$$

In the following, we estimate the right hand side of (3.25). Say  $F_W < 0$  on  $[a, \alpha]$  as in Proposition (3.13). Since from Proposition 3.15, we know  $F_W(x) = F_V(x)N/L + O(\ln N)$  and  $F'_W(x) = O(\ln NN/L)$ . Especially, we have  $1_{F_W < 0} F_W(x) = O(\ln N)$ . It follows

$$\int_a^\alpha F_W(x) \rho_L(x) dx = \frac{N}{\theta L} \sum_{\ell_i \in [a, \alpha]} \int_{\ell_i - \theta/2N}^{\ell_i + \theta/2N} F_W(x) dx + O\left(\frac{\ln N}{L}\right) \geq \frac{N}{\theta L} \int_a^\alpha F_W(x) dx + O\left(\frac{\ln N}{L}\right).$$

If  $F_W < 0$  on  $[\beta, b]$ , the same estimate holds. And it follows,

$$\int F_W(x) \rho_L(x) dx \geq \frac{N}{\theta L} \int_{F_W < 0} F_W(x) dx + O\left(\frac{\ln N}{L}\right). \quad (3.26)$$

Thus we can rewrite the distribution  $\mathbb{P}_L$  as

$$\begin{aligned} \mathbb{P}_L(\ell_{K+1}, \ell_{K+2}, \dots, \ell_{K+L}) &= \frac{1}{Z'_L} e^{-H_L(\ell_{K+1}, \dots, \ell_{K+L}) + O(L \ln N)} \\ &\leq \frac{1}{Z'_L} \exp\left(L^2 \left( \theta \mathcal{D}^2[\rho_W] - \theta \mathcal{D}^2[\rho_W - \rho_L] - \frac{N}{\theta L} \int_{F_W < 0} F_W(x) dx - f_W + O\left(\frac{\ln N}{L}\right) \right)\right). \end{aligned} \quad (3.27)$$

In the following we obtain a lower bound for the partition function  $Z'_L$ . Let  $x_i$ ,  $i = 1, 2, \dots, L$  be quantiles of  $\rho_W$  defined through

$$\int_a^{x_i} \rho_W(x) dx = \frac{i-1}{L}, \quad 1 \leq i \leq L.$$

Since  $\rho_W \leq N/L\theta$ ,  $\theta(x_{i+1} - x_i) \geq 1/N$ , and there exists an universal constant  $C$  and a particle configuration  $(\tilde{\ell}_{K+1}, \tilde{\ell}_{K+2}, \dots, \tilde{\ell}_{K+L}) \in \mathbb{W}_N^\theta$  such that

$$|\tilde{\ell}_{K+i}/N - x_i| \leq C/N. \quad (3.28)$$

We denote the empirical particle density  $d\tilde{\mu}_L = 1/L \sum_{i=1}^L \delta_{x_i}$ . We can approximate the Hamiltonian  $H_L(\tilde{\ell}_{K+1}, \tilde{\ell}_{K+2}, \dots, \tilde{\ell}_{K+L})$  by  $H_L(Nx_1, Nx_2, \dots, Nx_L)$ , and their difference is negligible,

$$\begin{aligned} H_L(\tilde{\ell}_{K+1}, \tilde{\ell}_{K+2}, \dots, \tilde{\ell}_{K+L}) &= -\theta \sum_{i \neq j} \ln \left| \frac{\tilde{\ell}_{K+i}/N - \tilde{\ell}_{K+j}/N}{x_i - x_j} \right| + L \sum_i W(\tilde{\ell}_{K+i}/N) \\ &= H_L(Nx_1, Nx_2, \dots, Nx_L) - \theta \sum_{i \neq j} \ln \left| \frac{\tilde{\ell}_{K+i}/N - \tilde{\ell}_{K+j}/N}{x_i - x_j} \right| + L \sum_i \left( W(\tilde{\ell}_{K+i}/N) - W(x_i) \right) \\ &= H_L(Nx_1, Nx_2, \dots, Nx_L) + O\left( \sum_{i \neq j} \left| \frac{(\tilde{\ell}_{K+i}/N - x_i) - (\tilde{\ell}_{K+j}/N - x_j)}{x_i - x_j} \right| + L \sum_i \|W'\|_\infty |\tilde{\ell}_{K+i}/N - x_i| \right) \\ &= H_L(Nx_1, Nx_2, \dots, Nx_L) + O(L \ln N), \end{aligned}$$

where we used  $|\tilde{\ell}_i/N - x_i| \leq C/N$  from (3.28),  $|x_i - x_j| \geq \theta|i - j|/N$  and  $\|W'\|_\infty = O(N \ln N/L)$  from Proposition 3.9. Similar to the derivation in (3.24), we have

$$\begin{aligned} L^{-2} H_L(Nx_1, Nx_2, \dots, Nx_L) &= -\theta \int_{x \neq y} \ln|x-y| d\tilde{\mu}_L d\tilde{\mu}_L + \int W(x) d\tilde{\mu}_L \\ &= -\theta \mathcal{D}^2[\rho_W] - \theta \int_{x \neq y} \ln|x-y| (\rho_W(x) dx - d\tilde{\mu}_L(x)) (\rho_W(y) dy - d\tilde{\mu}_L(x)) + \int F_W(x) d\tilde{\mu}_L(x) + f_W. \end{aligned} \quad (3.29)$$

For the second term on the righthand side of (3.29), we can divide the integral region into small rectangles,  $[x_i, x_{i+1}] \times [x_j, x_{j+1}]$ , for any  $1 \leq i, j \leq L$ . We have the trivial upper bound for  $i < j$

$$\int_{\substack{x \in [x_i, x_{i+1}], \\ y \in [x_j, x_{j+1}]}} \mathbf{1}_{x \neq y} \ln |x - y| (\rho_W(x) dx - d\tilde{\mu}_L(x)) (\rho_W(y) dy - d\tilde{\mu}_L(x)) = O\left(\frac{\ln N}{L^2}\right). \quad (3.30)$$

We can directly bound the integrals over rectangles close to diagonal by (3.30),

$$\begin{aligned} & \int_{x \neq y} \ln |x - y| (\rho_W(x) dx - d\tilde{\mu}_L(x)) (\rho_W(y) dy - d\tilde{\mu}_L(x)) \\ &= 2 \sum_{i+2 \leq j} \int_{\substack{x \in [x_i, x_{i+1}], \\ y \in [x_j, x_{j+1}]}} \ln |x - y| (\rho_W(x) dx - d\tilde{\mu}_L(x)) (\rho_W(y) dy - d\tilde{\mu}_L(x)) + O\left(\frac{\ln N}{L}\right) \\ &= 2 \sum_{i+2 \leq j} \int_{\substack{x \in [x_i, x_{i+1}], \\ y \in [x_j, x_{j+1}]}} \ln \frac{(y-x)(x_j-x_i)}{(y-x_i)(x_j-x)} \rho_W(x) \rho_W(y) dx dy + O\left(\frac{\ln N}{L}\right). \end{aligned}$$

For  $x \in [x_i, x_{i+1}]$ ,  $y \in [x_j, x_{j+1}]$  and  $j \geq i+2$ , we have

$$0 \leq \ln \frac{(y-x)(x_j-x_i)}{(y-x_i)(x_j-x)} \leq \ln \frac{(x_{j+1}-x_{i+1})(x_j-x_i)}{(x_{j+1}-x_i)(x_j-x_{i+1})}.$$

Therefore, it follows

$$\begin{aligned} 0 &\leq 2 \sum_{i+2 \leq j} \int_{\substack{x \in [x_i, x_{i+1}], \\ y \in [x_j, x_{j+1}]}} \ln \frac{(y-x)(x_j-x_i)}{(y-x_i)(x_j-x)} \rho_W(x) \rho_W(y) dx dy \\ &\leq \frac{2}{L^2} \ln \prod_{i+2 \leq j} \frac{(x_{j+1}-x_{i+1})(x_j-x_i)}{(x_{j+1}-x_i)(x_j-x_{i+1})} \\ &\leq \frac{2}{L^2} \ln \prod_{j=3}^L \frac{(x_{j+1}-x_{j-1})(x_j-x_{j-1})}{(x_3-x_1)(x_{L+1}-x_1)} = O\left(\frac{\ln N}{L}\right), \end{aligned}$$

where we used that  $\theta/N \leq x_j - x_i \leq 1$  for  $i < j$ . The second term on the righthand side of (3.29) can be bounded,

$$\theta \int_{x \neq y} \ln |x - y| (\rho_W(x) dx - d\tilde{\mu}_L(x)) (\rho_W(y) dy - d\tilde{\mu}_L(x)) = O\left(\frac{\ln N}{L}\right). \quad (3.31)$$

In the following we estimate the third term on the righthand side of (3.29). We recall the defining relations (3.18) of  $\rho_W$ .  $\tilde{\mu}_L(x)$  is supported on the support of  $\rho_W$ , so on the support of  $\tilde{\mu}_L(x)$ , we have  $F_W(x) \leq 0$ , and

$$\int F_W(x) d\tilde{\mu}_L(x) = \int_{F_W < 0} F_W(x) d\tilde{\mu}_L(x)$$

Moreover,  $F_W(x) < 0$  only if  $x \in [a, \alpha]$  and  $\rho_W \equiv N/L\theta$  on  $[a, \alpha]$ , or  $x \in [\beta, b]$  and  $\rho_W \equiv N/L\theta$  on  $[\beta, b]$ . Say  $F_W(x) < 0$  on  $[a, \alpha]$ . From our choices of  $x_i$ , if  $x_i, x_{i+1} \in [a, \alpha]$ , then  $x_{i+1} - x_i = \theta/N$ . The same argument as for (3.26) implies

$$\int_a^\alpha F_W(x) d\tilde{\mu}_L(x) dx = \frac{N}{\theta L} \int_a^\alpha F_W(x) dx + \left(\frac{\ln N}{L}\right),$$

and the same estimate holds for the integral over  $[\beta, b]$  if  $\rho_W \equiv N/L\theta$  on  $[\beta, b]$ . It follows,

$$\int F_W(x) d\tilde{\mu}_L(x) = \frac{N}{\theta L} \int_{F_W < 0} F_W(x) dx + \left( \frac{\ln N}{L} \right). \quad (3.32)$$

Combining (3.31) and (3.32) together, we conclude that

$$Z'_L \geq \mathbb{P}_L(\tilde{\ell}_1, \tilde{\ell}_2, \dots, \tilde{\ell}_L) = \exp \left( L^2 \left( \theta \mathcal{D}^2[\rho_W] - \frac{N}{\theta L} \int_{F_W < 0} F_W(x) dx - f_W + O \left( \frac{\ln N}{L} \right) \right) \right). \quad (3.33)$$

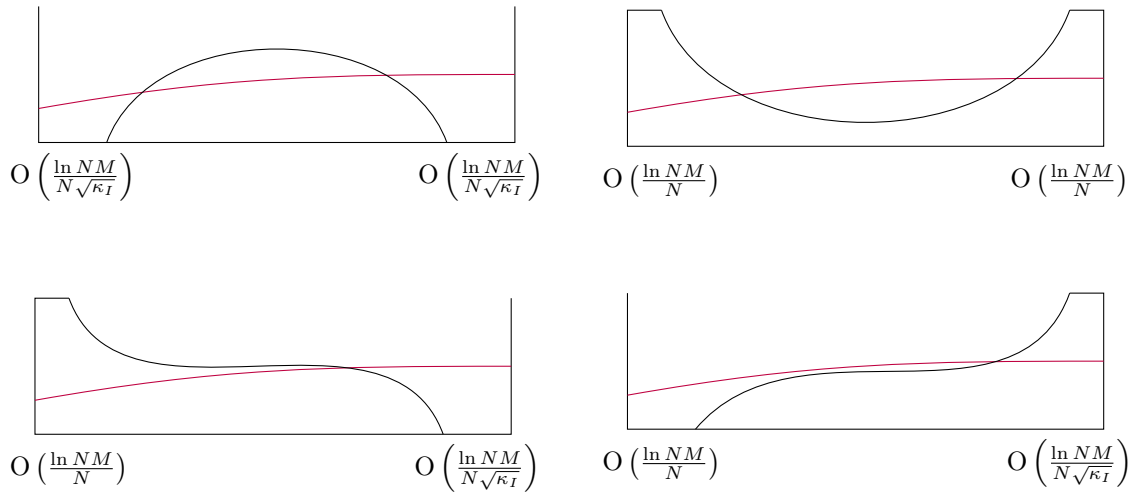
(3.22) follows from (3.27) and (3.33). And (3.23) follows from (3.22) by the same argument as in [17, Corollary 2.16] (see also the proof of Proposition 2.3).  $\square$

We collect below a few estimates on the local equilibrium measure that we prove in the appendix.

**Proposition 3.17.** *[Estimates on  $\alpha, \beta$ ] We assume that  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of “good” boundary conditions,  $I$  is as defined in (3.7) or (3.8), and  $[\alpha, \beta] \subset I$  is from Proposition 3.13. Then we have the following estimates for the local equilibrium measure:*

1. *The bulk case: Let  $\kappa_I = \text{dist}(I, \{A, B\})$ . If  $\rho_W(x) \equiv 0$  on  $[a, \alpha]$ , then  $\alpha - a = O(\ln NM/N\sqrt{\kappa_I})$ . If  $\rho_W(x) \equiv N/L\theta$  on  $[a, \alpha]$ , then  $\alpha - a = O(\ln NM/N)$ . If  $\rho_W(x) \equiv 0$  on  $[\beta, b]$ , then  $b - \beta = O(\ln NM/N\sqrt{\kappa_I})$ . If  $\rho_W(x) \equiv N/L\theta$  on  $[\beta, b]$ , then  $b - \beta = O(\ln NM/N)$ .*
2. *The left edge case: Let  $\kappa_I = \text{dist}(b, A)$ . If  $\rho_W(x) \equiv 0$  on  $[\beta, b]$ , then  $b - \beta = O(\ln NM/N\sqrt{\kappa_I})$ . If  $\rho_W(x) \equiv N/L\theta$  on  $[\beta, b]$ , then  $b - \beta = O(\ln NM/N)$ .*

Figure 1: equilibrium measure of localized discrete  $\beta$ -ensemble(bulk case).





**Proposition 3.18.** [Estimates on the local density in the bulk] We assume that the particle configuration  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of “good” boundary conditions, and  $I = [a, b]$  is a bulk interval, i.e.  $K \in \llbracket \lfloor L/2 \rfloor + 1, N_0 \rrbracket$  (where  $N_0 := \max\{i : \gamma_i \leq B - r_0 - r\}$ ),  $a = (\ell_K + \theta)/N$  and  $b = (\ell_{K+L+1} - \theta)/N$ . Then for any  $x \in I$  such that  $x - a \asymp |I|$  and  $b - x \asymp |I|$ , the following holds

$$\left| \rho_W(x) - \frac{N}{L} \rho_V(x) \right| \lesssim \frac{\ln NM}{L} \frac{N\sqrt{\kappa_I}}{L},$$

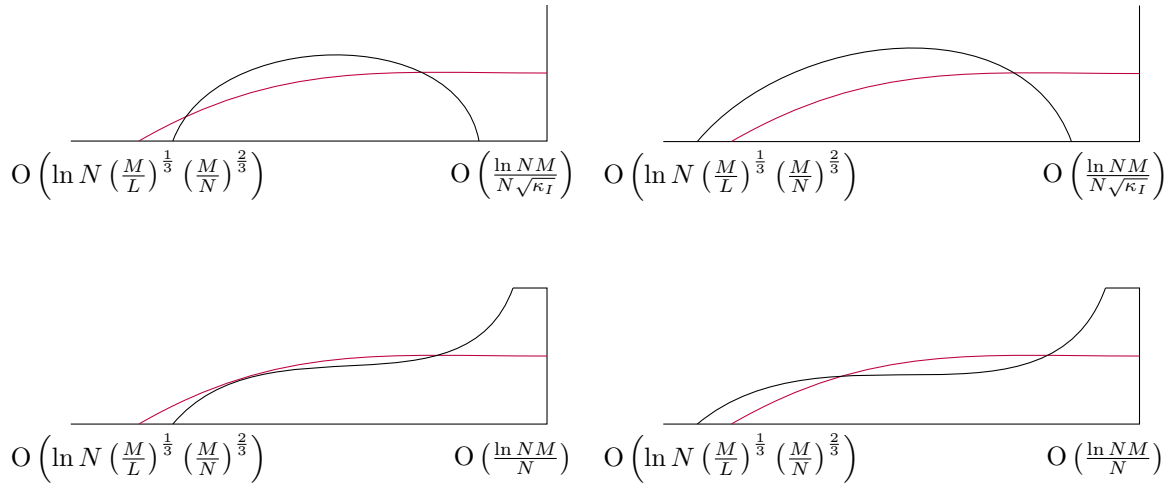
where  $\kappa_I = \text{dist}(I, \{A, B\})$ .

**Proposition 3.19.** [Estimates on  $\alpha$  when  $K = 0$ ] We assume that  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of “good” boundary conditions, and  $I$  is as defined in (3.7) or (3.8). Then for the left edge case, i.e.  $K = 0$ , the local equilibrium measure satisfies

$$|A - \alpha| = O\left(\ln N \left(\frac{M}{L}\right)^{1/3} \left(\frac{M}{N}\right)^{2/3}\right),$$

where  $\alpha$  is as in Proposition 3.13.

Figure 2: equilibrium measure of localized discrete  $\beta$ -ensemble (left edge case,  $\kappa_I \asymp (L/N)^{2/3}$ ).



**Proposition 3.20.** [Estimates on the density when  $K = 0$ ] We assume that the particle configuration  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M,L,K}$ , i.e. it is in the set of “good” boundary conditions. We consider the left edge case, i.e.  $K = 0$  and  $I = [a, b]$ , where  $a = a(N)/N$  and  $b = (\ell_{L+1} - \theta)/N$ . Let  $\kappa_I = \text{dist}(b, A)$ , then  $\kappa_I \asymp (L/N)^{2/3}$ . For any  $x \in I$  such that  $x \geq \min\{A, \alpha\}$  and  $b - x \asymp \kappa_I$ , the following holds

1. if  $\alpha \geq A$ ,

$$\left| \rho_W(x) - \frac{N}{L} \rho_V(x) \right| \lesssim \begin{cases} \frac{N}{L} \sqrt{\alpha - A}, & 0 \leq x - \alpha \leq \alpha - A, \\ \frac{N}{L} \frac{\alpha - A}{\sqrt{x - \alpha}} + \frac{\ln NM}{L} \frac{N}{L} \sqrt{x - \alpha}, & x - \alpha \geq \alpha - A, \end{cases} \quad (3.34)$$

2. if  $\alpha \leq A$ ,

$$\left| \rho_W(x) - \frac{N}{L} \rho_V(x) \right| \lesssim \begin{cases} \frac{N}{L} \ln \frac{A - \alpha}{x - \alpha} \sqrt{A - \alpha}, & \alpha - A \leq x - A \leq A - \alpha, \\ \frac{N}{L} \frac{A - \alpha}{\sqrt{x - \alpha}} + \frac{\ln NM}{L} \frac{N}{L} \sqrt{x - \alpha}, & x - A \geq A - \alpha. \end{cases} \quad (3.35)$$

*Proof of Theorem 2.6.* By our assumption, with probability at least  $1 - e^{-cM}$ , the rigidity estimates (2.29) hold on the scale  $M/N$  in the spectral domain  $\mathcal{D}_{M, r_0} \cup \mathcal{D}_*$ . It follows from Proposition 3.7, that with probability at least  $1 - e^{-cM/2}$  with respect to the marginal measure, we have  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M, L, K}$ , i.e. it is in the set of "good" boundary conditions. In the following we fix a "good" boundary condition  $(\ell_1, \ell_2, \dots, \ell_K, \ell_{K+L+1}, \ell_{K+L+2}, \dots, \ell_N) \in \mathcal{G}_{M, L, K}$  and prove that (2.30) and (2.31) hold with probability  $1 - e^{-cL \ln N}$  with respect to the conditioned measure  $\mathbb{P}_L$  (as defined in (3.9)). Since by our choice of parameters,  $M \ll L$ , this implies Theorem 2.6.

For any  $z = E + i\eta \in \mathcal{D}_{M, r_0+r}$  as defined in (1.13), we denote

$$i_0 = \operatorname{argmin}_{1 \leq i \leq N} |E - \gamma_i|,$$

where  $\gamma_i$  are classical particle locations of the equilibrium measure  $\mu$  as in (1.5). Then  $i_0 \leq N_0 + 1$ , where  $N_0 = \max\{i : \gamma_i \leq B - r_0 - r\}$ . There are bulk cases and edge cases,

1. The bulk cases: if  $i_0 > \lfloor 3L/4 \rfloor$ , we take  $K = i_0 - \lfloor L/4 \rfloor$ , then  $K \leq N_0$ . Let  $I = [a, b]$ , where  $a = (\ell_K + \theta)/N$  and  $b = (\ell_{K+L+1} - \theta)/N$ , and  $\kappa_I = \operatorname{dist}(I, \{A, B\})$ , then  $\kappa_I \gtrsim (L/M)^{2/3}$ . We take a bump function  $\chi_0 : \mathbb{R} \rightarrow \mathbb{R}$ , such that  $\chi_0(x) = 1$  on  $[\gamma_{K+\lfloor L/5 \rfloor}, \gamma_{K+\lfloor 4L/5 \rfloor}]$ , and  $\chi_0(x)$  vanishes outside  $[\gamma_{K+\lfloor L/6 \rfloor}, \gamma_{K+\lfloor 5L/6 \rfloor}]$ . Moreover,  $\chi_0'(x) = O(|I|^{-1})$  and  $\chi_0''(x) = O(|I|^{-2})$ .
2. The left edge cases: if  $i_0 \leq \lfloor 3L/4 \rfloor$ , we take  $K = 0$ . Let  $I = [a(N)/N, b]$ , where  $b = (\ell_{L+1} - \theta)/N$ , and  $\kappa_I = \operatorname{dist}(b, A)$ , then  $\kappa_I \asymp (L/N)^{2/3}$ . We take a bump function  $\chi_0 : \mathbb{R} \rightarrow \mathbb{R}$ , such that  $\chi_0(x) = 1$  on  $[\hat{a}_N, \gamma_{\lfloor 4L/5 \rfloor}]$  (where  $\hat{a}_N = \min\{a(N)/N, \hat{a}\}$ ), and  $\chi_0(x)$  vanishes for  $x \geq \gamma_{\lfloor 5L/6 \rfloor}$ . Moreover,  $\chi_0'(x) = O(\kappa_I^{-1})$  and  $\chi_0''(x) = O(\kappa_I^{-2})$ .

We decompose the difference  $G_N(z) - G_\mu(z)$ ,

$$G_N(z) - G_\mu(z) = \int \frac{\chi_0(x)}{z - x} (d\mu_N(x) - d\mu(x)) + \int \frac{1 - \chi_0(x)}{z - x} (d\mu_N(x) - d\mu(x)). \quad (3.36)$$

In the following we estimate the second term in (3.36). We first discuss the bulk case. Since  $\chi_0(x) = 1$  on  $[\gamma_{K+\lfloor L/5 \rfloor}, \gamma_{K+\lfloor 4L/5 \rfloor}]$ , we have

$$\int \frac{1 - \chi_0(x)}{z - x} (d\mu_N(x) - d\mu(x)) = \int_{x \geq \gamma_{K+\lfloor 4L/5 \rfloor}} \frac{1 - \chi_0(x)}{z - x} (d\mu_N(x) - d\mu(x)) + \int_{x \leq \gamma_{K+\lfloor L/5 \rfloor}} \frac{1 - \chi_0(x)}{z - x} (d\mu_N(x) - d\mu(x)). \quad (3.37)$$

We estimate the first term on the righthand side of (3.37). The estimate for the second term is analogous, which is similar to the argument for (C.16). Similar to the construction in Remark 3.4, we take a dyadic

partition of unity  $\chi_1, \chi_2, \dots, \chi_{n'}$  with scale  $\tilde{\eta} = L/N$ , such that  $1 = \chi_0(x) + \chi_1(x) + \chi_2(x) + \chi_3(x) + \dots + \chi_{n'}(x)$  for  $[\gamma_{K+\lfloor 4L/5 \rfloor}, \hat{b}_N]$ .  $\chi_k(y)$  is supported on interval  $I_k$ , with  $|I_k| \asymp 2^k L/N \sqrt{\kappa_I}$ . It follows

$$\left| \int_{x \geq \gamma_{K+\lfloor 4L/5 \rfloor}} \frac{1 - \chi_0(x)}{z - x} (d\mu_N(x) - d\mu(x)) \right| = \left| \sum_{k=1}^{n'} \int \frac{\chi_k(x)}{z - x} (d\mu_N(x) - d\mu(x)) \right|.$$

Moreover, from our choice of the index  $K$ ,  $\gamma_{K+\lfloor 4L/5 \rfloor} - E \asymp L/N \sqrt{\kappa_I}$  and  $\text{dist}(I_k, E) \asymp 2^k L/N \sqrt{\kappa_I}$ . Let

$$h_k(x) = \frac{\chi_k(y)|I_k|}{z - x}.$$

Then we have,  $h_k(x) = O(1)$ ,  $h'_k(x) = O(|I_k|^{-1})$  and  $h''_k(x) = O(|I_k|^{-2})$ . Moreover, either  $I_k \subset [\gamma_{K+\lfloor 4L/5 \rfloor}, B - r_0]$ , or  $I_k$  is not completely contained in  $[\gamma_{K+\lfloor 4L/5 \rfloor}, B - r_0]$ , in which case  $|I_k| \gtrsim 1$ . By Proposition 3.2 and 3.3, we have

$$\begin{aligned} \sum_{k=1}^{n'} \left| \int \frac{\chi_k(x)}{z - x} (d\mu_N(x) - d\mu(x)) \right| &= \sum_{k=1}^{n'} |I_k|^{-1} \left| \int h_k(x) (d\mu_N(x) - d\mu(x)) \right| \\ &\lesssim \sum_{k=1}^{n'} |I_k|^{-1} \frac{\ln NM}{N} \lesssim \frac{\ln NM}{N|I_1|} \lesssim \frac{\ln NM}{L} \sqrt{\kappa_I}. \end{aligned}$$

The same estimate holds for the second term in (3.37), and it follows

$$\left| \int \frac{1 - \chi_0(x)}{z - x} (d\mu_N(x) - d\mu(x)) \right| \lesssim \frac{\ln NM}{L} \sqrt{\kappa_I}. \quad (3.38)$$

For the left edge case, we have

$$\int \frac{1 - \chi_0(x)}{z - x} (d\mu_N(x) - d\mu(x)) = \int_{x \geq \lfloor 4L/5 \rfloor} \frac{1 - \chi_0(x)}{z - x} (d\mu_N(x) - d\mu(x)).$$

Exactly the same argument leads to the estimate (3.38).

For the first term in (3.36), it depends only on particles in the interval  $I$ , we can rewrite it as

$$\frac{L}{N} \int \frac{\chi_0(x)}{z - x} (d\mu_L(x) - \rho_W(x) dx) + \int \frac{\chi_0(x)}{z - x} \left( \frac{L}{N} \rho_W(x) - \rho_V(x) \right) dx. \quad (3.39)$$

In the following we prove with probability at least  $1 - e^{-\ln NL}$ , with respect to the conditioned measure  $\mathbb{P}_L$  (as defined in (3.9)) (3.39) is small.

An estimate of the first term in (3.39) follows from Proposition 3.16. We take  $g(x) = \chi_0(x)/(z - x)$  in (3.23). Then  $\|g\|_{1/2} \lesssim (\ln N)^{1/2}/\eta$ , and it follows, with probability at least  $1 - e^{-\ln NL}$ , with respect to the conditional measure (3.9),

$$\left| \frac{L}{N} \int \frac{\chi_0(x)}{z - x} (d\mu_L(x) - \rho_W(x) dx) \right| \leq \frac{C \ln N \sqrt{L}}{N \eta}. \quad (3.40)$$

For the second term in (3.39), we first discuss the bulk cases.  $\chi_0(x)$  is supported on  $[\gamma_{K+\lfloor L/6 \rfloor}, \gamma_{K+\lfloor 5L/6 \rfloor}]$ . From our choice of the index  $K$ , for any  $x \in [\gamma_{K+\lfloor L/6 \rfloor}, \gamma_{K+\lfloor 5L/6 \rfloor}]$ , we have  $b - x \asymp x - a \asymp |I|$ . Thanks to Proposition 3.18, we have

$$\left| \int \frac{\chi_0(x)}{z-x} \left( \frac{L}{N} \rho_W(x) - \rho_V(x) \right) dx \right| \lesssim \int \frac{\chi_0(x)}{|z-x|} \frac{\ln NM}{L} \sqrt{\kappa_I} dx \lesssim \frac{(\ln N)^2 M}{L} \sqrt{\kappa_I}. \quad (3.41)$$

It follows by combining (3.36), (3.38), (3.40) and (3.41), with probability at least  $1 - e^{-\ln NL}$ , with respect to the conditional measure (3.9),

$$|G_N(z) - G_\mu(z)| \leq C(\ln N)^2 \left( \frac{\sqrt{L}}{N\eta} + \frac{M}{L} \sqrt{\kappa_I} \right).$$

For the second term in (3.39), the analysis for the left edge case is more delicate. As in Proposition 3.20, there are two cases. If  $\alpha \leq A$ , by using (3.35), we have

$$\begin{aligned} \left| \int \frac{\chi_0(x)}{z-x} \left( \frac{L}{N} \rho_W(x) - \rho_V(x) \right) dx \right| &\lesssim \int_{\alpha \leq x \leq 2A-\alpha} \frac{\chi_0(x)}{|z-x|} \left| \frac{L}{N} \rho_W(x) - \rho_V(x) \right| dx \\ &+ \int_{x \geq 2A-\alpha} \frac{\chi_0(x)}{|z-x|} \frac{A-\alpha}{\sqrt{x-\alpha}} dx + \frac{\ln NM}{L} \int_{x \geq 2A-\alpha} \frac{\chi_0(x)}{|z-x|} \sqrt{x-\alpha} dx. \end{aligned} \quad (3.42)$$

Let  $\kappa_E = \text{dist}(E, \{A, B\})$ . For the first term in (3.42), we have by (3.35),

$$\int_{\alpha \leq x \leq 2A-\alpha} \frac{\chi_0(x)}{|z-x|} \left| \frac{L}{N} \rho_W(x) - \rho_V(x) \right| dx \lesssim \sqrt{\alpha - A} \int_{\alpha \leq x \leq 2A-\alpha} \frac{\chi_0(x)}{|z-x|} \ln \frac{A-\alpha}{x-\alpha} dx. \quad (3.43)$$

If  $\eta + \kappa_E \geq 2(A - \alpha)$ , then

$$\int_{\alpha \leq x \leq 2A-\alpha} \frac{\chi_0(x)}{|z-x|} \ln \frac{A-\alpha}{x-\alpha} dx \lesssim \int_{\alpha \leq x \leq 2A-\alpha} \frac{1}{\eta + \kappa_E} \ln \frac{A-\alpha}{x-\alpha} dx \lesssim \frac{\ln N(A-\alpha)}{\eta + \kappa_E}. \quad (3.44)$$

If  $\eta + \kappa_E \leq 2(A - \alpha)$ , we trivially bound

$$\int_{\alpha \leq x \leq 2A-\alpha} \frac{\chi_0(x)}{|z-x|} \ln \frac{A-\alpha}{x-\alpha} dx \lesssim \ln N. \quad (3.45)$$

By combining (3.43), (3.44) and (3.45), we have

$$\int_{\alpha \leq x \leq 2A-\alpha} \frac{\chi_0(x)}{|z-x|} \left| \frac{L}{N} \rho_W(x) - \rho_V(x) \right| dx \lesssim \ln N \left( \frac{(A-\alpha)^{3/2}}{\eta + \kappa_E} \wedge (A-\alpha)^{1/2} \right). \quad (3.46)$$

For the second term in (3.42), by a similar argument, we have

$$\int_{x \geq 2A-\alpha} \frac{\chi_0(x)}{|z-x|} \frac{A-\alpha}{\sqrt{x-\alpha}} dx \lesssim \ln N \left( \frac{A-\alpha}{\sqrt{\eta + \kappa_E}} \wedge (A-\alpha)^{1/2} \right). \quad (3.47)$$

For the last term in (3.42), we have

$$\frac{\ln NM}{L} \int_{x \geq 2A - \alpha} \frac{\chi_0(x)}{|z - x|} \sqrt{x - \alpha} dx \lesssim \frac{(\ln N)^2 M}{L} \left(\frac{L}{N}\right)^{1/3}, \quad (3.48)$$

where we used that  $x - \alpha \leq b - \alpha \leq (L/N)^{2/3}$ . By combining (3.42), (3.46), (3.47) and (3.48), we have

$$\begin{aligned} \left| \int \frac{\chi_0(x)}{z - x} \left( \frac{L}{N} \rho_W(x) - \rho_V(x) \right) dx \right| &\leq C \left( \ln N \left( \frac{A - \alpha}{\sqrt{\kappa_E + \eta}} \wedge (A - \alpha)^{1/2} \right) + \frac{(\ln N)^2 M}{L} \left(\frac{L}{N}\right)^{1/3} \right) \\ &\leq C (\ln N)^2 \left( \left(\frac{M}{L}\right)^{1/2} \left(\frac{L}{N}\right)^{1/3} \wedge \frac{M}{L} \left(\frac{L}{N}\right)^{2/3} \frac{1}{\sqrt{\kappa_E + \eta}} + \frac{M}{L} \left(\frac{L}{N}\right)^{1/3} \right). \end{aligned} \quad (3.49)$$

A similar argument implies that (3.49) holds for the case  $\alpha \geq A$ . It follows by combining (3.36), (3.38), (3.40) and (3.49), with probability at least  $1 - e^{-\ln NL}$ , with respect to the conditional measure (3.9),

$$|G_N(z) - G_\mu(z)| \leq C (\ln N)^2 \left( \frac{\sqrt{L}}{N\eta} + \left(\frac{M}{L}\right)^{1/2} \left(\frac{L}{N}\right)^{1/3} \wedge \frac{M}{L} \left(\frac{L}{N}\right)^{2/3} \frac{1}{\sqrt{\kappa_E + \eta}} + \frac{M}{L} \left(\frac{L}{N}\right)^{1/3} \right).$$

This finishes the proof of Theorem 2.6.  $\square$

## 4 Edge Universality

In this section we prove the edge universality of the discrete  $\beta$ -ensembles for any  $\beta \geq 1$ , i.e. the distributions of extreme particles are the same as those of continuous  $\beta$ -ensembles, governed by Tracy-Widom  $\beta$  distributions. Heuristically, the edge universality states that extreme particles after scaling by  $N^{2/3}$  converge in law to Tracy-Widom  $\beta$  distribution. We can view discrete  $\beta$ -ensembles as “discretized” continuous  $\beta$ -ensembles on the scale  $1/N$ . Since the discretization scale is much smaller than the scale of the fluctuations of extreme particles, the discretization does not affect the distribution of extreme particles, and the edge universality of the discrete  $\beta$ -ensembles follows. To do this, we study the conditioned discrete  $\beta$ -ensemble (4.1) of the first few particles, and compare it with its continuous analogue (4.2). We show that these two laws are the same up to negligible error. It is proven in [29] that, under mild conditions, the distribution of extreme particles of (4.2) converges to Tracy-Widom  $\beta$  distribution. The edge universality of the discrete  $\beta$ -ensembles follows.

We fix small constant  $r > 0$ , small exponents  $\mathfrak{a}, \mathfrak{b} > 0$ , and let  $L = N^{\mathfrak{b}}$ . For any  $(\ell_{L+1}, \ell_{L+2}, \dots, \ell_N) \in \mathbb{W}_N^\theta$ , we define the conditioned discrete  $\beta$ -ensemble,

$$\mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L) = \frac{1}{Z_L^{\text{dis}}} \prod_{1 \leq i < j \leq L} \frac{\Gamma(\ell_j - \ell_i + 1) \Gamma(\ell_j - \ell_i + \theta)}{\Gamma(\ell_j - \ell_i) \Gamma(\ell_j - \ell_i + 1 - \theta)} \prod_{i=1}^L \hat{w}(\ell_i; L), \quad (4.1)$$

where

$$\hat{w}(Nu; L) = e^{-NV_N(u)} e^{-\beta \Theta(N^{2/3 - \mathfrak{a}}(u - A))} \prod_{j > L} \frac{\Gamma(\ell_j - Nu + 1) \Gamma(\ell_j - Nu + \theta)}{\Gamma(\ell_j - Nu) \Gamma(\ell_j - Nu + 1 - \theta)},$$

where (4.1) is slightly different from (3.9), since we introduced an additional quadratic potential  $\Theta(u) := (u+1)^2 \mathbf{1}_{u \leq -1}$ , or  $\Theta(u) := 2(u+1)^2 \mathbf{1}_{u \leq -1}$  to prevent the particles  $\ell_i$  from deviating far in the left direction.

We define the continuous version of  $\mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L)$ , let  $\beta = 2\theta$ ,

$$d\mathbb{P}_L^{\text{cont}}(dx_1, dx_2, \dots, dx_L) = \frac{1}{Z_L^{\text{cont}}} \prod_{1 \leq i < j \leq L} |x_j - x_i|^\beta \prod_{i=1}^L \hat{w}(x_i; L) dx_i, \quad (4.2)$$

where  $\mathbb{P}_L^{\text{cont}}$  is a measure on  $x_1 < x_2 < \dots < x_L \leq \ell_{L+1} - \theta$ .

**Definition 4.1.** We define the set of “good” boundary conditions,  $\mathcal{R}_L^* = \mathcal{R}_L^*(\mathbf{a}, r)$ , as particle configurations

$$(\ell_{L+1}, \ell_{L+2}, \dots, \ell_N) \in \mathbb{W}_N^\theta,$$

such that with probability  $1 - \exp(-c(\ln N)^2)$  with respect to the conditioned discrete  $\beta$ -ensemble  $\mathbb{P}_L^{\text{dis}}$ :

1. The Stieltjes transform satisfies

$$|G_N(z) - G_\mu(z)| \leq \frac{(\ln N)^2}{N\eta},$$

uniformly for any  $z \in \mathcal{D}_r \cup \mathcal{D}_*$ .

2. The extreme eigenvalue satisfies  $\ell_1/N \geq A - N^{-2/3+\alpha}$ .

If  $(\ell_{L+1}, \ell_{L+2}, \dots, \ell_N)$  is in the set of “good” boundary conditions, then there exists some  $r' > 0$ , such that

$$|\ell_k/N - \gamma_k| \leq N^{-2/3+\alpha} k^{-1/3}, \quad k \in [L+1, r'N],$$

and with high probability with respect to  $\mathbb{P}_L^{\text{dis}}$ ,

$$|\ell_k/N - \gamma_k| \leq N^{-2/3+\alpha} k^{-1/3}, \quad k \in [1, L].$$

It follows from Theorem 1.8 and Corollary 2.5, that with probability at least  $1 - \exp(-c(\ln N)^2)$  with respect to  $\mathbb{P}_N$ , we have  $(\ell_{L+1}, \ell_{L+2}, \dots, \ell_N) \in \mathcal{R}_L^*(\mathbf{a}, r)$ .

Thanks to the additional quadratic potential  $\Theta$ , similar to the proof of Proposition 3.16, we have the following estimate on the extreme particles of  $\mathbb{P}_L^{\text{cont}}$ ,

**Proposition 4.2.** For any “good” boundary condition  $(\ell_{L+1}, \ell_{L+2}, \dots, \ell_N) \in \mathcal{R}_L^*(\mathbf{a}, r)$ , there exists universal constant  $C > 0$  such that

$$\mathbb{P}_L^{\text{cont}} \left( x_1/N \leq A - C(L \ln N)^{1/2} N^{-2/3+\alpha} \right) \leq e^{-L \ln N}. \quad (4.3)$$

*Proof.* We recall that  $W$  is defined in (3.15). We denote by  $\rho_W^{\text{cont}}$  the unconstrained equilibrium measure of the potential  $W$ , which is characterized by

$$W(x) - 2\theta \int_a^b \ln |x - y| \rho_W^{\text{cont}}(y) dy - f_W = F_W(x) = \begin{cases} = 0, & \text{on } \rho_W^{\text{cont}}(x) > 0, \\ \geq 0, & \text{on } \rho_W^{\text{cont}}(x) = 0. \end{cases} \quad (4.4)$$

The same argument as in Proposition (3.16) gives,

$$\mathbb{P}_L \left( \sum_{i=1}^L \Theta(N^{2/3-\alpha}(\ell_i/N - A)) + L^2 \mathcal{D}^2[\rho_W^{\text{cont}} - \rho_L] \geq \gamma^2 L^2 \right) \leq \exp(C \ln NL - \gamma^2 L^2),$$

where  $\rho_L$  is the convolution of the empirical particle density  $1/L \sum_{i=1}^L \delta_{\ell_i/N}$  with the uniform measure on the interval  $[-N^{-p}/2, N^{-p}/2]$  for some  $p > 2$ , and the energy functional  $\mathcal{D}$  is defined in (3.21). In particular, we have

$$\mathbb{P}_L \left( \Theta(N^{2/3-\alpha}(\ell_1/N - A)) \geq CL \ln N \right) \leq \exp(-\ln NL),$$

which leads to (4.3) by rearranging.  $\square$

Following [29], we further define the following modification of  $\mathcal{R}_L^*$ . We define  $\mathcal{R}_L^\#$  which adds the condition of a level repulsion near the boundary,

$$\mathcal{R}_L^\# = \mathcal{R}_L^\#(\mathbf{a}, r) := \{(\ell_{L+1}, \dots, \ell_N) \in \mathcal{R}_L^*(\mathbf{a}/3, r) : |\ell_{L+2} - \ell_{L+1}| \geq N^{1/3-\alpha} L^{-1/3}\}.$$

The proof of edge universality of the discrete  $\beta$ -ensembles for any  $\beta \geq 1$  follows from a direct comparison between  $\mathbb{P}_L^{\text{dis}}$  and  $\mathbb{P}_L^{\text{cont}}$ . The following weak level repulsion estimates for  $\mathbb{P}_L^{\text{dis}}$  and  $\mathbb{P}_L^{\text{cont}}$  follows the same arguments as those in [29, (D.1), (D.2)]. We postpone their proofs to Appendix D.

**Proposition 4.3.** *Let  $\beta > 0$ . There are constants  $C$  such that for  $(\ell_{L+1}, \dots, \ell_N) \in \mathcal{R}_L^*(\mathbf{a}, r)$ ,  $k \in \llbracket 1, L \rrbracket$  and any  $L^{4/3} N^{-1/3} \ll s \ll \min\{L^{-2}, N^{-\alpha}\}$ , we have*

$$\mathbb{P}_L^{\text{dis}} \left( |\ell_{k+1} - \ell_k| \leq s L^{-1/3} N^{1/3} \right) \leq C s L^2, \quad (4.5)$$

$$\mathbb{P}_L^{\text{cont}} \left( |x_{k+1} - x_k| \leq s L^{-1/3} N^{1/3} \right) \leq C s L^2. \quad (4.6)$$

Moreover, thanks to the rigidity estimates for the local measure  $\mathbb{P}_L^{\text{dis}}$ , we have better estimate, for  $k \in \llbracket 1, L \rrbracket$  and any  $L^{4/3} N^{-1/3} \ll s \ll N^{-\alpha}$

$$\mathbb{P}_L^{\text{dis}} \left( |\ell_{k+1} - \ell_k| \leq s k^{-1/3} N^{1/3} \right) \leq C (s N^\alpha + \exp(-c(\ln N)^2)). \quad (4.7)$$

In the following we fix a small number  $s$  which will be chosen later. We define  $\mathcal{G}^{\text{dis}}$  to be the set of particle configurations  $(\ell_1, \ell_2, \dots, \ell_L) \in \mathbb{W}_N^\theta$ , such that

$$\ell_1/N \geq A - C(L \ln N)^{1/2} N^{-2/3+\alpha}, \quad |\ell_{k+1} - \ell_k| \geq s L^{-1/3} N^{1/3}, \quad k \in \llbracket 1, L \rrbracket.$$

We define the continuous version of  $\mathcal{G}^{\text{dis}}$  as

$$\mathcal{G}^{\text{cont}} = \{(x_1, x_2, \dots, x_L) : ([x_1]_\theta, [x_2]_\theta, \dots, [x_L]_\theta) \in \mathcal{G}^{\text{dis}}\},$$

where  $[x_k]_\theta = \inf\{\ell \geq x_k : \ell \in a(N) + k\theta + \mathbb{Z}\}$ . We remark that  $[x_k]_\theta - 1 < x_k \leq [x_k]_\theta$ . Thanks to Propositions 4.2, 4.3 and a union bound, we have

$$\mathbb{P}_L^{\text{dis}}(\mathcal{G}^{\text{dis}}) \geq 1 - C s L^3, \quad \mathbb{P}_L^{\text{cont}}(\mathcal{G}^{\text{cont}}) \geq 1 - C s L^3. \quad (4.8)$$

**Proposition 4.4.** For any “good” boundary condition  $(\ell_{L+1}, \ell_{L+2}, \dots, \ell_N) \in \mathcal{R}_L^*(\mathbf{a}, r)$ , and particle configurations  $(x_1, x_2, \dots, x_L) \in \mathcal{G}^{\text{cont}}$ , we have

$$Z_L^{\text{dis}} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L) = \exp \left\{ O \left( \frac{L^{4/3} N^{\mathbf{a}}}{s N^{1/3}} \right) \right\} Z_L^{\text{cont}} \mathbb{P}_L^{\text{cont}}(x_1, x_2, \dots, x_L), \quad (4.9)$$

where  $\ell_k = \lceil x_k \rceil_\theta = \inf \{ \ell \geq x_k : \ell \in a(N) + k\theta + \mathbb{Z} \}$ , for  $k \in \llbracket 1, L \rrbracket$ . Moreover

$$\mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L) = \left( 1 + O \left( s L^3 + \frac{L^{4/3} N^{\mathbf{a}}}{s N^{1/3}} \right) \right) \mathbb{P}_L^{\text{cont}}(x_1, x_2, \dots, x_L). \quad (4.10)$$

In the above estimates, we assume  $L^{4/3} N^{\mathbf{a}-1/3} \ll s \ll L^{-3}$ .

*Proof.* For any  $1 \leq i < j \leq L$ , since  $\ell_j - \ell_i \geq s L^{-1/3} N^{1/3} (j - i) \gg 1$ , we can estimate the interaction terms in (4.1)

$$\frac{\Gamma(\ell_j - \ell_i + 1) \Gamma(\ell_j - \ell_i + \theta)}{\Gamma(\ell_j - \ell_i) \Gamma(\ell_j - \ell_i + 1 - \theta)} = |\ell_j - \ell_i|^{2\theta} e^{O(1/(\ell_j - \ell_i))} = |x_j - x_i|^{2\theta} e^{O(1/(\ell_j - \ell_i))}.$$

Therefore we have

$$\begin{aligned} \prod_{1 \leq i < j \leq L} \frac{\Gamma(\ell_j - \ell_i + 1) \Gamma(\ell_j - \ell_i + \theta)}{\Gamma(\ell_j - \ell_i) \Gamma(\ell_j - \ell_i + 1 - \theta)} &= e^{\sum_{1 \leq i < j \leq L} O(1/(\ell_j - \ell_i))} \prod_{1 \leq i < j \leq L} |x_j - x_i|^{2\theta} \\ &= \exp \left\{ O \left( \frac{L^{4/3} \ln N}{s N^{1/3}} \right) \right\} \prod_{1 \leq i < j \leq L} |x_j - x_i|^{2\theta}. \end{aligned} \quad (4.11)$$

In the following we estimate,  $\hat{w}(\ell_k; L) / \hat{w}(x_k; L)$ . Notice that we have the following asymptotics,

$$\partial_z \ln \Gamma(z) = \ln z - \frac{1}{2z} + O \left( \frac{1}{z^2} \right).$$

Moreover, the derivative of the potential  $V_N$  can be written in terms of the equilibrium measure,

$$V'_N(u) = V'(u) + O(N^{-1/3}) = 2\theta P.V. \int_A^B \frac{\rho_V(y)}{u - y} dx + 1_{u \notin [A, B]} O \left( \sqrt{(u - A)(u - B)} \right) + O(N^{-1/3}). \quad (4.12)$$

where we used assumption 1.7. Therefore, for any  $x \in [N(A - C(L \ln N)^{1/2} N^{-2/3+\mathbf{a}}), \ell_{L+1} - s L^{-1/3} N^{1/3}]$ , we have

$$\partial_x \ln \hat{w}(x; L) = -V'_N \left( \frac{x}{N} \right) + O \left( \frac{(L \ln N)^{1/2}}{N^{1/3+\mathbf{a}}} \right) + 2\theta \sum_{j \geq L+1} \frac{1}{x - \ell_j} + O \left( \frac{1}{(x - \ell_j)^2} \right). \quad (4.13)$$

We deduce from [29, Lemma C.1] that

$$-2\theta P.V. \int_A^B \frac{\rho_V(y)}{x/N - y} + 2\theta \sum_{j \geq L+CN^{\mathbf{a}}} \frac{1}{x - \ell_j} = O \left( \frac{L^{1/3} N^{\mathbf{a}}}{N^{1/3}} \right), \quad (4.14)$$



Moreover, since  $x \leq \ell_{L+1} - sL^{-1/3}N^{1/3}$ , we have

$$2\theta \sum_{j=L+1}^{L+CN^a} \frac{1}{x - \ell_j} + 2\theta \sum_{j \geq L+1} O\left(\frac{1}{(x - \ell_j)^2}\right) = O\left(\frac{L^{1/3} \ln N}{sN^{1/3}}\right). \quad (4.15)$$

We get the following estimate by combining (4.12), (4.13), (4.14) and (4.15),

$$\partial_x \ln \hat{w}(x; L) = O\left(\frac{L^{1/3}N^a}{sN^{1/3}}\right).$$

Therefore, we deduce that

$$\frac{\hat{w}(\ell_k; L)}{\hat{w}(x_k; L)} = e^{C \sup_{x_k \leq x \leq \ell_k} |\partial_x \ln \hat{w}(x; L)| |\ell_k - x_k|} = \exp\left\{O\left(\frac{L^{1/3}N^a}{sN^{1/3}}\right)\right\}, \quad (4.16)$$

where we used that  $|\ell_k - x_k| = O(1)$ .

It follows by combining (4.11) and (4.16) together,

$$Z_L^{\text{dis}} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L) = \exp\left\{O\left(\frac{L^{4/3} \ln N}{sN^{1/3}}\right) + O\left(\frac{L^{4/3}N^a}{sN^{1/3}}\right)\right\} Z_L^{\text{cont}} \mathbb{P}_L^{\text{cont}}(x_1, x_2, \dots, x_L).$$

and (4.9) follows.

For (4.10), we need to show  $Z_L^{\text{dis}} \approx Z_L^{\text{cont}}$ . Thanks to (4.8), it follows

$$\begin{aligned} (1 + O(sL^3))Z_L^{\text{dis}} &= Z_L^{\text{dis}} \mathbb{P}_L^{\text{dis}}(\mathcal{G}^{\text{dis}}) = \sum_{(\ell_1, \ell_2, \dots, \ell_L) \in \mathcal{G}^{\text{dis}}} Z_L^{\text{dis}} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L) \\ &= \sum_{(\ell_1, \ell_2, \dots, \ell_L) \in \mathcal{G}^{\text{dis}}} \exp\left\{O\left(\frac{L^{4/3}N^a}{sN^{1/3}}\right)\right\} Z_L^{\text{cont}} \mathbb{P}_L^{\text{cont}}(\{x_1, x_2, \dots, x_L : [x_k]_\theta = \ell_k, \quad k \in [1, L]\}) \\ &= \exp\left\{O\left(\frac{L^{4/3}N^a}{sN^{1/3}}\right)\right\} Z_L^{\text{cont}} \mathbb{P}_L^{\text{cont}}(\mathcal{G}^{\text{cont}}) = (1 + O(sL^3)) \exp\left\{O\left(\frac{L^{4/3}N^a}{sN^{1/3}}\right)\right\} Z_L^{\text{dis}}. \end{aligned}$$

By rearranging the above expression, we get

$$Z_L^{\text{dis}} = \left(1 + O\left(sL^3 + \frac{L^{4/3}N^a}{sN^{1/3}}\right)\right) Z_L^{\text{cont}}, \quad (4.17)$$

and (4.10) follows from (4.9) and (4.17). □

The edge universality of discrete  $\beta$ -ensemble follows easily from Proposition 4.4.

*Proof of Theorem 1.11.* Thanks to Theorems 1.8 and (4.3) in Proposition 4.3, the event  $\mathcal{R}_L^*(\mathbf{a}, r) \mathcal{R}_L^\#(\mathbf{a}, r)$  holds with probability  $1 - O(N^{-2a/3})$ . (1.16) follows from the following statement: there exists some  $\chi > 0$ , for any boundary conditions  $(\ell_{L+1}, \ell_{L+2}, \dots, \ell_N) \in \mathcal{R}_L^*(\mathbf{a}, r) \cap \mathcal{R}_L^\#(\mathbf{a}, r)$ , it holds,

$$\left| \mathbb{E}_{\mathbb{P}_L^{\text{dis}}} \left[ O\left(\left(N^{2/3}k^{1/3}s_A^{2/3}(\ell_k/N - \gamma_k)\right)_{k \in \Lambda}\right) \right] - \mathbb{E}_{G\beta E} \left[ O\left(\left(N^{2/3}k^{1/3}(x_k/N - \tilde{\gamma}_k)\right)_{k \in \Lambda}\right) \right] \right| = O(N^{-\chi}). \quad (4.18)$$

In the following we prove (4.18). Notice that

$$\begin{aligned}
& \mathbb{E}_{\mathbb{P}_L^{\text{cont}}} \left[ O \left( \left( N^{2/3} k^{1/3} s_A^{2/3} (x_k/N - \gamma_k) \right)_{k \in \Lambda} \right) \right] \\
&= \mathbb{E}_{\mathbb{P}_L^{\text{cont}}} \left[ \mathbf{1}_{\{\mathcal{G}^{\text{cont}}\}} O \left( \left( N^{2/3} k^{1/3} s_A^{2/3} (\lceil x_k \rceil_\theta / N - \gamma_k) \right)_{k \in \Lambda} \right) \right] + O \left( (L/N)^{1/3} + sL^3 \right) \\
&= \mathbb{E}_{\mathbb{P}_L^{\text{dis}}} \left[ \mathbf{1}_{\{\mathcal{G}^{\text{dis}}\}} O \left( \left( N^{2/3} k^{1/3} s_A^{2/3} (\ell_k/N - \gamma_k) \right)_{k \in \Lambda} \right) \right] + O \left( sL^3 + \frac{L^{4/3} N^\alpha}{sN^{1/3}} \right) \\
&= \mathbb{E}_{\mathbb{P}_L^{\text{dis}}} \left[ O \left( \left( N^{2/3} k^{1/3} s_A^{2/3} (\ell_k/N - \gamma_k) \right)_{k \in \Lambda} \right) \right] + O \left( sL^3 + \frac{L^{4/3} N^\alpha}{sN^{1/3}} \right),
\end{aligned} \tag{4.19}$$

where we used (4.8) and  $|x_k - \lceil x_k \rceil_\theta| < 1$  for the second and fourth lines, and (4.10) for the third line. Recall that  $L = N^\mathfrak{b}$  and  $\mathfrak{b} < 1/13$ . We can take  $\chi = 1/6 - \alpha/2 - 13\mathfrak{b}/6$ ,  $s = N^{-1/6 + \alpha/2 - 5\mathfrak{b}/6}$ , and  $\alpha$  small enough such that  $\chi > 0$ . Then the error in (4.19) is bounded by  $O(N^{-\chi})$ .

In particular (4.19) implies that

$$\mathbb{P}_L^{\text{cont}}(x_1 \geq \gamma_1 - 2N^{-2/3+\alpha}) \geq 1/2, \quad |\mathbb{E}_{\mathbb{P}_L^{\text{cont}}}[x_k - \gamma_k]| \leq 2N^{-2/3+\alpha} k^{-1/3}, \quad k \in \llbracket 1, L \rrbracket.$$

Therefore the boundary condition  $(\ell_{L+1}, \ell_{L+2}, \dots, \ell_N)$  satisfies the assumptions in [29, Theorem 3.3]. As a result, there exists a small  $\chi > 0$ ,

$$\left| \mathbb{E}_{\mathbb{P}_L^{\text{cont}}} \left[ O \left( \left( N^{2/3} k^{1/3} s_A^{2/3} (x_k/N - \gamma_k) \right)_{k \in \Lambda} \right) \right] - \mathbb{E}_{G\beta E} \left[ O \left( \left( N^{2/3} k^{1/3} (x_k/N - \tilde{\gamma}_k) \right)_{k \in \Lambda} \right) \right] \right| = O(N^{-\chi}). \tag{4.20}$$

(4.18) follows from combining (4.19) and (4.20). □

## 5 Multi-cut Case

In this section, we indicate how the arguments used so far in the article can be carried to the multi-cut stochastic systems with fixed filling fractions, which is the full model studied in [17]. We recall the definition of the multi-cut model below.

We fix an integer  $k > 0$ , whose meaning is the number of segments in the support of the measure ( $k = 1$  corresponds to the one-cut case). For each  $N$  we take  $k$  integers  $n_1(N), \dots, n_k(N)$ , such that  $\sum_{i=1}^k n_i(N) = N$  and  $k$  disjoint intervals  $[a_1(N), b_1(N)], \dots, [a_k(N), b_k(N)]$  of the real line ordered from left to right.

We assume that  $b_i(N) + \theta \leq a_{i+1}(N)$  for  $i \in \llbracket 1, k-1 \rrbracket$ . The numbers  $a_i(N), b_i(N)$  must also satisfy the conditions  $b_i(N) - \theta n_i(N) - a_i(N) \in \mathbb{Z}_{>0}$ . Further, the number  $n_i(N)$  counts the number of the particles in the  $i$ -th interval; to make this statement precise we define the sets of indices  $I_j \subset \llbracket 1, N \rrbracket$  for  $j \in \llbracket 1, k \rrbracket$ , via

$$I_j = \llbracket 1 + \sum_{m=1}^{j-1} n_m(N), \sum_{m=1}^j n_m(N) \rrbracket.$$

We also set  $I_j^+$  and  $I_j^-$  to be the maximal and minimal elements of  $I_j$ , respectively.

**Definition 5.1.** *The state space  $\mathbb{W}_N^\theta$  consists of  $N$ -tuples  $\ell_1 < \ell_2 < \dots < \ell_N$  such that for each  $j = 1, \dots, k$ :*

1. *If  $i = I_j^-$ , then  $\ell_i - a_i(N) \in \mathbb{Z}_{>0}$ .*
2. *If  $i = I_j^+$ , then  $b_i(N) - \ell_i \in \mathbb{Z}_{>0}$ .*
3. *If  $i \in I_j$ , but  $i \neq I_j^\pm$ , then  $\ell_{i+1} - \ell_i - \theta \in \mathbb{Z}_{\geq 0}$ .*

We also take a weight function  $w(x; N) > 0$  in the interior of the intervals  $(a_i(N), b_i(N))$ , and  $w(a_i(N); N) = w(b_i(N); N) = 0$  for  $i \in \llbracket 1, k \rrbracket$ . The multi-cut discrete  $\beta$ -ensemble is given by the probability measure  $\mathbb{P}_N$  on  $\mathbb{W}_N^\theta$

$$\mathbb{P}_N(\ell_1, \dots, \ell_N) = \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} \frac{\Gamma(\ell_j - \ell_i + 1) \Gamma(\ell_j - \ell_i + \theta)}{\Gamma(\ell_j - \ell_i) \Gamma(\ell_j - \ell_i + 1 - \theta)} \prod_{i=1}^N w(\ell_i; N). \quad (5.1)$$

For the study of the multi-cut model, the Assumptions 1.2, 1.3 and 1.5 and 1.7 need to be replaced by their counterparts from [17, Section 3.2], as listed below

**Assumption 5.2.** *We require that as  $N \rightarrow \infty$ , for each  $i \in \llbracket 1, k \rrbracket$ ,*

$$a_i(N) = N\hat{a}_i + O(\ln(N)), \quad b_i(N) = N\hat{b}_i + O(\ln(N)), \quad w(a_i(N); N) = w(b_i(N); N) = 0,$$

and

$$\hat{a}_1 < \hat{b}_1 < \hat{a}_2 < \dots < \hat{a}_k < \hat{b}_k.$$

We require the weight  $w(x; N)$  in the intervals  $[a_i(N) + 1, b_i(N) - 1]$  for  $i \in \llbracket 1, k \rrbracket$ , has the form

$$w(x; N) = \exp\left(-NV_N\left(\frac{x}{N}\right)\right),$$

for a function  $V_N$  that is continuous in the intervals  $[a_i(N) + 1, b_i(N) - 1]$ , and such that

$$V_N(u) = V(u) + O\left(\frac{\ln(N)}{N}\right),$$

uniformly over  $u \in [(a_i(N) + 1)/N, (b_i(N) - 1)/N]$ . The function  $V(u)$  is differentiable and the following bound holds for a constant  $C > 0$ ,

$$|V'(u)| \leq C \left(1 + \sum_{i=1}^k \left(|\ln(u - \hat{a}_i)| + |\ln(u - \hat{b}_i)|\right)\right).$$

**Assumption 5.3.** *There exists a constant  $\mathfrak{d} > 0$  such that for any  $i \in \llbracket 1, k \rrbracket$  and  $N$  large enough, the filling fractions satisfy*

$$\mathfrak{d} \leq n_i(N)/N \leq \theta^{-1}(\hat{b}_i - \hat{a}_i) - \mathfrak{d}.$$

As in the one-cut case, we denote the empirical particle density as  $\mu_N = N^{-1} \sum_{i=1}^N \delta_{\ell_i/N}$ , and let  $\mu(dx) = \rho_V(x)dx$  be the constrained equilibrium measure of the discrete  $\beta$ -ensemble  $\mathbb{P}_N$ , which is given by the minimizer of the following functional

$$-\theta \int_{x \neq y} \ln|x-y| \rho(x) \rho(y) dx dy + \int V(x) \rho(x) dx, \quad (5.2)$$

over all the densities  $\rho(x)$ , supported on  $\cup_{i=1}^k [\hat{a}_i, \hat{b}_i]$ , with  $0 \leq \rho(x) \leq \theta^{-1}$ , and

$$\int_{\hat{a}_i}^{\hat{b}_i} \rho(x) dx = \hat{n}_i, \quad i \in \llbracket 1, k \rrbracket.$$

We remark that  $\rho_V$  depends on  $\hat{n}_i$ , and the latter depend on  $N$ , so  $\rho_V$  depends on  $N$ . However, we will hide this dependence from our notations. We denote by  $F_V$  the (non-centered) effective potential given by

$$F_V(x) := -2\theta \int_{-\infty}^{\infty} \ln|x-y| \rho_V(y) dy + V(x),$$

On each single cut  $[\hat{a}_i, \hat{b}_i]$ , there exists a constant  $f_i$  so that  $F_V$  satisfies:

1.  $F_V(x) - f_i \geq 0$ , for all  $x$  in voids in  $[\hat{a}_i, \hat{b}_i]$ , i.e. maximal closed connected intervals where  $\rho_V(x) = 0$ ;
2.  $F_V(x) - f_i \leq 0$ , for all  $x$  in saturated regions in  $[\hat{a}_i, \hat{b}_i]$ , i.e. maximal closed connected intervals where  $\rho_V(x) = \theta^{-1}$ ;
3.  $F_V(x) - f_i = 0$ , for all  $x$  in bands in  $[\hat{a}_i, \hat{b}_i]$ , i.e. maximal open connected intervals where  $0 < \rho_V(x) < \theta^{-1}$ .

We denote the classical particle locations  $\gamma_1, \gamma_2, \dots, \gamma_N$  corresponding to the constrained equilibrium measure  $\mu$ , as

$$\frac{i-1/2}{N} = \int_{-\infty}^{\gamma_i} \rho_V(x) dx, \quad i \in \llbracket 1, N \rrbracket. \quad (5.3)$$

**Assumption 5.4.** For  $i \in \llbracket 1, k \rrbracket$ , there exist open sets  $\mathcal{M}_i$ , which contains the interval  $[\hat{a}_i, \hat{b}_i]$ , such that the functions  $\psi_N^\pm(x)$  defined through

$$\frac{w(x; N)}{w(x-1; N)} = \frac{\psi_N^+(x)}{\psi_N^-(x)},$$

can be chosen in such a way that

$$\psi_N^\pm(x) = \phi^\pm\left(\frac{x}{N}\right) + \mathcal{O}\left(\frac{1}{N}\right),$$

uniformly over  $x/N$  in compact subsets of  $\mathcal{M} := \cup_{i=1}^k \mathcal{M}_i$ . All the aforementioned functions are analytic.

We recall the Stieltjes transform  $G_\mu$  of the equilibrium measure  $\mu$  as in (1.7), and the two functions  $R_\mu, Q_\mu$  as in (1.8),

$$\begin{aligned} R_\mu(z) &:= \phi^-(z)e^{-\theta G_\mu(z)} + \phi^+(z)e^{\theta G_\mu(z)}, \\ Q_\mu(z) &:= \phi^-(z)e^{-\theta G_\mu(z)} - \phi^+(z)e^{\theta G_\mu(z)}. \end{aligned} \quad (5.4)$$

In the multi-cut case, they all depend on  $N$ . Moreover, under Assumptions 5.2, 5.3 and 5.4,  $R_\mu(z)$  is analytic on  $\mathcal{M}$  [17, Proposition 5.11].

**Assumption 5.5.** *We assume there exists a function  $H(z)$  analytic in  $\mathcal{M}$  and numbers  $\{A_i, B_i\}_{i=1}^k$  such that*

- $\hat{a}_i < A_i < B_i < \hat{b}_i$  for  $i \in \llbracket 1, k \rrbracket$ ;
- $Q_\mu(z) = H(z) \prod_{i=1}^k \sqrt{(z - A_i)(z - B_i)}$ , where the branch of square root is such that  $\sqrt{(z - A_i)(z - B_i)} \sim z$  as  $z \rightarrow \infty$ ;
- $H(z) \neq 0$  for all  $z$  in a neighborhood of  $\cup_{i=1}^k [\hat{a}_i, \hat{b}_i]$ .

**Assumption 5.6.** *For any  $u$  in a small neighborhood of  $\cup_{i=1}^k [A_i, B_i]$ ,  $V_N(u)$  is analytic and the following holds*

$$V'_N(u) = V'(u) + O(N^{-1/3}).$$

Under Assumptions 5.2, 5.3, 5.4 and 5.5, when restricted on a single cut  $[\hat{a}_i, \hat{b}_i]$ , the equilibrium measure  $\rho_V|_{[\hat{a}_i, \hat{b}_i]}$  behaves as in one-cut case. Especially, it has a single band, and the square root behaviors at  $A_1, B_1, \dots, A_k, B_k$ .

To state the rigidity theorem in the multi-cut case, we need some more definitions. We fix a small parameter  $\mathfrak{a} > 0$ , and an index  $i \in \llbracket 1, k \rrbracket$ , and define the spectral domains

$$\begin{aligned} \mathcal{D}_{i,r} &:= \mathcal{D}_{i,r}^{\text{int}} \cup \mathcal{D}_i^{\text{ext}}, \\ \mathcal{D}_{i,r}^{\text{int}} &:= \{E + i\eta \in \mathcal{M}_i \cap \mathbb{C}_+ : E \in [A_i, B_i - r], \eta \sqrt{\kappa_E + \eta} \geq N^\mathfrak{a}/N\}, \\ \mathcal{D}_i^{\text{ext}} &:= \{E + i\eta \in \mathcal{M}_i \cap \mathbb{C}_+ : E \leq A_i, \eta \geq (N^\mathfrak{a}/N)^{2/3}\}, \end{aligned} \quad (5.5)$$

and

$$\begin{aligned} \mathcal{D}_* &:= \mathcal{D}_*^{\text{int}} \cup \mathcal{D}_*^{\text{ext}}, \\ \mathcal{D}_*^{\text{int}} &:= \{E + i\eta \in \mathcal{M} \cap \mathbb{C}_+ : E \in \cup_{i=1}^k [A_i, B_i], \eta \sqrt{\kappa_E + \eta} \geq N^{-(1-\mathfrak{a})/2}\}, \\ \mathcal{D}_*^{\text{ext}} &:= \{E + i\eta \in \mathcal{M} \cap \mathbb{C}_+ : E \notin \cup_{i=1}^k [A_i, B_i], \eta \geq N^{-(1-\mathfrak{a})/3}\}. \end{aligned} \quad (5.6)$$

where  $\kappa_E = \text{dist}(E, \{A_1, B_1, \dots, A_k, B_k\})$ .

As in the one-cut case, we define the dual equilibrium measure  $\mu^{\text{dual}}$ ,

$$\mu^{\text{dual}}(dx) = \rho_V^{\text{dual}}(x)dx, \quad \rho_V^{\text{dual}}(x) = \theta^{-1} \sum_{i=1}^k \mathbf{1}_{[\hat{a}_i, \hat{b}_i]} - \rho_V(x), \quad (5.7)$$

and the Stieltjes transform  $G_\mu^{\text{dual}}$  of the dual equilibrium measure  $\mu^{\text{dual}}$ . For any  $z = E + i\eta \in \mathbb{C}_+$ , let  $\kappa_E = \text{dist}(E, \{A_1, B_1, \dots, A_k, B_k\})$ ,  $\hat{\kappa}_E = \text{dist}(E, \{\hat{a}_1, \hat{b}_1, \dots, \hat{a}_k, \hat{b}_k\})$ , we define the control parameter:

$$\Theta_\mu(z) = \min \left\{ \frac{|\text{Im}[G_\mu(z)]|}{N\eta}, \frac{|\text{Im}[G_\mu^{\text{dual}}(z)]|}{N\eta} + \frac{1}{N(\eta + \hat{\kappa}_E)} \right\}. \quad (5.8)$$

Thanks to the square root behavior of the equilibrium density  $\rho_V(x)$  at  $A_1, B_1, \dots, A_k, B_k$ , the asymptotic behavior of  $\Theta_\mu$  depends on  $\eta$ ,  $\kappa_E$  and  $\hat{\kappa}_E$ ,

$$\Theta_\mu(z) \asymp \begin{cases} \sqrt{\kappa_E + \eta}/N\eta, & E \in \cup_{i=1}^k [A_i, B_i], \\ 1/N\sqrt{\kappa_E + \eta}, & E \notin \cup_{i=1}^k [A_i, B_i] \text{ on voids,} \\ 1/N\sqrt{\kappa_E + \eta} + 1/N(\hat{\kappa}_E + \eta), & E \notin \cup_{i=1}^k [A_i, B_i] \text{ on saturated regions,} \end{cases} \quad (5.9)$$

For any  $v \in \mathbb{C}_+$ ,  $0 \leq t \leq 1$ ,  $\alpha = \pm 1, \pm i$ , and large number  $K = K(N, \text{Im}[v]) \leq N$ , we introduce the deformed probability measure

$$\mathbb{P}_N^{K,t,v,\alpha} = \frac{Z_N}{Z_N^{K,t,v,\alpha}} \mathbb{P}_N e^{2K \text{Re}[\sum_{i=1}^N \ln(1 + \frac{\alpha t}{N(v - \ell_i/N)})]}. \quad (5.10)$$

As in the one-cut case, we can use the loop equation to improve the estimates of the Green's function. For the proof of the following Theorem, we need to introduce the hyperelliptic integrals to invert the loop equation, which is slightly different from that of Theorem 2.4. We will sketch the proof at the end of this section.

**Theorem 5.7.** *Let  $v = E + i\eta \in \mathcal{M} \cap \mathbb{C}_+$ ,  $\kappa_E = \text{dist}(E, \{A_1, B_1, \dots, A_k, B_k\})$ ,  $\hat{\kappa}_E = \text{dist}(E, \{\hat{a}_1, \hat{b}_1, \dots, \hat{a}_k, \hat{b}_k\})$  and  $\Theta_\mu$  as defined in (5.8). We assume Assumptions 5.2, 5.3, 5.4, 5.5, and that  $N\eta\sqrt{\kappa_E + \eta} \gg 1$ . For any  $0 \leq t \leq 1$ ,  $\alpha = \pm 1, \pm i$ , and  $K \ll (N\eta)^2 \sqrt{\kappa_E + \eta}$ , we assume that with high probability w.r.t.  $\mathbb{P}_N^{K,t,v,\alpha}$*

$$G_N(v) = G_\mu(v) + O(\varepsilon), \quad (5.11)$$

then we have

$$\mathbb{E}_{\mathbb{P}_N^{K,t,v,\alpha}} [G_N(v) - G_\mu(v)] = O\left(\frac{\varepsilon^2}{\sqrt{\kappa_E + \eta}} + \varepsilon_0\right), \quad (5.12)$$

where,

$$\varepsilon_0 = \frac{1}{\sqrt{\kappa_E + \eta}} \left( \Theta_\mu(v) + \frac{K \min\{|\text{Im}[G_\mu(v)], |\text{Im}[G_\mu^{\text{dual}}(v)]|\}}{(N\eta)^2} + \frac{\ln N}{(N\eta)^2} + \frac{\ln N}{N} + \frac{\ln NK}{(N\eta)^3} + \frac{K^2}{(N\eta)^4} \right). \quad (5.13)$$

Let  $\tilde{\varepsilon}_0 = \max\{\varepsilon_0, (\ln N)^2 K^{-1}\}$ . If there exists some  $\mathfrak{c} > 0$ , such that  $\varepsilon \leq N^{-\mathfrak{c}} \sqrt{\kappa_E + \eta}$ , then we have

$$\mathbb{P}_N (|G_N(v) - G_\mu(v)| \lesssim s\tilde{\varepsilon}_0) \geq 1 - e^{-csK\tilde{\varepsilon}_0}, \quad (5.14)$$

for any  $s \geq 1$ .

Using the large deviation estimate [17, Proposition 5.6] as input for Theorem 5.9, we have the same statement as Corollary 2.5 for the multi-cut case:

**Corollary 5.8.** *We assume Assumptions 5.2, 5.3, 5.4 and 5.5. Fix  $s \geq (\ln N)^2$ . With probability at least  $1 - e^{-cs}$  with respect to  $\mathbb{P}_N$ , it holds uniformly for any  $z = E + i\eta \in \mathcal{D}_*$  (as defined in (5.6)),*

$$|G_N(z) - G_\mu(z)| \lesssim \frac{s}{N\eta}. \quad (5.15)$$

With the estimate (5.15), and Proposition 3.3, we can localize at a single cut  $[\hat{a}_i, \hat{b}_i]$ , the effect of the other cuts is equivalent to a smooth potential of size  $O(s/N)$ . The same argument as in one-cut case leads to the following rigidity and edge universality statement.

**Theorem 5.9.** *We assume Assumptions 5.2, 5.3, 5.4, 5.5 and that  $[\hat{a}_i, A_i]$  is a void region, i.e.  $\rho_V(x) = 0$  on  $[\hat{a}_i, A_i]$ , then the following holds:*

1. *Fix  $s \geq (\ln N)^2$  and small  $r > 0$ . With probability at least  $1 - e^{-cs}$ , it holds uniformly for any  $z = E + i\eta \in \mathcal{D}_{i,r}$ ,*

$$|G_N(z) - G_\mu(z)| \leq \frac{s}{N\eta}. \quad (5.16)$$

2. *Fix a small  $c$  such that  $0 < c < \alpha/4$ . With probability at least  $1 - \exp(-c(\ln N)^2)$ , it holds uniformly for  $z = E + i(N^\alpha/N)^{2/3} \in \mathcal{M}_i$  with  $E \leq A_i - N^{2c}(N^\alpha/N)^{2/3}$ ,*

$$|G_N(z) - G_\mu(z)| \ll \frac{1}{N\eta}. \quad (5.17)$$

A similar statement holds if  $[B_i, \hat{b}_i]$  is a void region.

**Theorem 5.10.** *Fix  $\beta \geq 1$ ,  $\theta = \beta/2$ ,  $0 < \mathfrak{b} < 1/13$  and  $L = N^\mathfrak{b}$ . We assume Assumptions 5.2, 5.3, 5.4, 5.5, 5.6 and that  $[\hat{a}_i, A_i]$  is a void region,  $\rho_V(u) = (1 + O(u - A_i))_{s_{A_i}} \sqrt{u - A_i}/\pi$ . Take any fixed  $m \geq 1$  and a continuously differentiable compactly supported function  $O : \mathbb{R}^m \rightarrow \mathbb{R}$ . For any index set  $\Lambda \in [1, L]$  with  $|\Lambda| = m$ , we denote the shifted index set  $\Lambda_i = \{k + I_i^- - 1 : k \in \Lambda\}$ , then*

$$\left| \mathbb{E}_{\mathbb{P}_N} \left[ O \left( \left( N^{2/3} k^{1/3} s_{A_i}^{2/3} (\ell_k/N - \gamma_k) \right)_{k \in \Lambda_i} \right) \right] - \mathbb{E}_{G_{\beta E}} \left[ O \left( \left( N^{2/3} k^{1/3} (x_k/N - \tilde{\gamma}_k) \right)_{k \in \Lambda} \right) \right] \right| = O(N^{-\chi}), \quad (5.18)$$

where  $\gamma_k$  are classical eigenvalue locations of  $\rho_V$ , and  $\tilde{\gamma}_k$  are classical eigenvalue locations of semi-circle distribution. A similar statement holds if  $[B_i, \hat{b}_i]$  is a void region.

*Proof of Theorem 5.9.* We recall the notations  $\phi_N^+(x), \phi_N^-(x), R_N(zN), \mathcal{E}(z)$  from (2.41), (2.42) and (2.44). In the multi-cut setting,  $R_N(zN)$  is also an analytic function in  $\mathcal{M}$  as proven in [17, Theorem 4.1]. We take two sets of oriented contours:  $\mathcal{C}_1$  consists of two clockwise oriented circles, one is centered at  $v$  with radius  $\eta/2$ , and the other is centered at  $\bar{v}$  with radius  $\eta/2$ .  $\mathcal{C}_2$  consists of  $k$  counterclockwise oriented contours:  $\mathcal{C}_2^1, \mathcal{C}_2^2, \dots, \mathcal{C}_2^k$ , such that each  $\mathcal{C}_2^i \subset \mathcal{M}_i$  ( $1 \leq i \leq k$ ) encloses a small neighborhood of  $[\hat{a}_i, \hat{b}_i]$ . By dividing

(2.43) by  $2\pi i(z-v)H(z)$  (as defined in Assumption 5.5), and integrating along the union of contours  $\mathcal{C}_1 \cup \mathcal{C}_2$ , we get

$$0 = I_1 + I_2 + I_3 + I_4, \quad (5.19)$$

where  $I_1, I_2, I_3, I_4$  are defined in (2.45). We have the same estimate for  $I_1, I_3, I_4$  as in the proof of Theorem 2.4:

$$\begin{aligned} I_1 &= \theta \prod_{i=1}^k \sqrt{(v-A_i)(v-B_i)} \mathbb{E}[G_N(v) - G_\mu(v)] \\ &\quad + O\left(\varepsilon^2 + \frac{|\mathbb{E}[\operatorname{Im}[G_N(v) - G_\mu(v)]]|}{N\eta} + \Theta_\mu(v) + \frac{1}{N} + \frac{\ln N}{(N\eta)^2}\right) \\ I_3 &= O\left(\frac{\ln N}{N}\right) \\ I_4 &\lesssim \frac{K \min\{\mathbb{E}[|\operatorname{Im}[G_N(v)]|], \mathbb{E}[|\operatorname{Im}[G_N^{\text{dual}}(v)]|]\}}{(N\eta)^2} + O\left(\frac{1}{N} + \frac{K}{N^2\eta} + \frac{\ln NK}{(N\eta)^3} + \frac{K^2}{(N\eta)^4}\right). \end{aligned} \quad (5.20)$$

For the estimate of  $I_2$  in the multi-cut setting, we need to introduce certain notations from the theory of hyperelliptic integrals to invert the loop equation [17, Section 6].

Let  $P(z) = p_0 + p_1z + \dots + p_{k-2}z^{k-2}$  be a polynomial of degree  $k-2$ , and consider the map

$$\Omega : P(z) \mapsto \left( \frac{1}{2\pi i} \int_{\mathcal{C}_2^1} \frac{P(z)dz}{\prod_{i=1}^k \sqrt{(z-A_i)(z-B_i)}}, \dots, \frac{1}{2\pi i} \int_{\mathcal{C}_2^k} \frac{P(z)dz}{\prod_{i=1}^k \sqrt{(z-A_i)(z-B_i)}} \right).$$

Note that the sum of the integrals in the definition of  $\Omega$  equals zero. And it turns out that  $\Omega$  is an isomorphism between  $(k-1)$ -dimensional vector spaces for any  $k \geq 2$ . Using  $\Omega$  we can now define a more complicated map  $\Upsilon$ . Given a function  $f(z)$  defined on the contours  $\mathcal{C}_2^i$  and such that the sum of its integrals over these contours is zero, we define a function  $\Upsilon_z[f]$  through

$$\Upsilon_z[f] = f(z) + \frac{P(z)}{\prod_{i=1}^k \sqrt{(z-A_i)(z-B_i)}},$$

where  $P(z)$  is the unique polynomial of degree at most  $k-2$ , such that for each  $i \in \llbracket 1, k \rrbracket$ ,

$$\frac{1}{2\pi i} \int_{\mathcal{C}_2^i} \Upsilon_z[f] dz = 0.$$

The polynomial  $P(z)$  can be evaluated in terms of the map  $\Omega$  via

$$P = \Omega^{-1} \left( -\frac{1}{2\pi i} \int_{\mathcal{C}_2^1} f(z) dz, \dots, -\frac{1}{2\pi i} \int_{\mathcal{C}_2^k} f(z) dz \right).$$

The same as in (2.51), we have

$$I_2 = \frac{\theta}{2\pi i} \int_{\mathcal{C}_2} \frac{\mathbb{E}[G_N(z) - G_\mu(z)] \prod_{i=1}^k \sqrt{(z-A_i)(z-B_i)}}{z-v} dz + O\left(\frac{\ln N}{N}\right). \quad (5.21)$$



Since  $\mathbb{E}[G_N(z) - G_\mu(z)]$  is analytic outside the contours of integration and decays as  $1/z^2$  as  $z \rightarrow \infty$ , the integral is a polynomial  $P_N(v)$  of degree at most  $k - 2$  (to see that one uses  $(z - v)^{-1} = z^{-1} \sum_{n \geq 0} (v/z)^n$ ). We conclude,

$$I_2 = P_N(v) + O\left(\frac{\ln N}{N}\right). \quad (5.22)$$

Combining the estimates (5.19), (5.20), and (5.22) all together, we deduce that

$$\begin{aligned} \theta \prod_{i=1}^k \sqrt{(v - A_i)(v - B_i)} \mathbb{E}[G_N(v) - G_\mu(v)] + P_N(v) &= O\left(\frac{1}{N\eta} + \frac{K}{(N\eta)^2}\right) |\mathbb{E}[\text{Im}[G_N(v) - G_\mu(v)]]| \\ &+ O\left(\varepsilon^2 + \Theta_\mu(v) + \frac{K \min\{|\text{Im}[G_\mu(v)|], |\text{Im}[G_\mu^{\text{dual}}(v)|]\}}{(N\eta)^2} + \frac{\ln N}{(N\eta)^2} + \frac{\ln N}{N} + \frac{\ln NK}{(N\eta)^3} + \frac{K^2}{(N\eta)^4}\right). \end{aligned} \quad (5.23)$$

Notice that  $\prod_{i=1}^k \sqrt{(v - A_i)(v - B_i)} \asymp \sqrt{\kappa_E + \eta}$ , and recall  $\varepsilon_0$  from (5.13),

$$\begin{aligned} &\theta \mathbb{E}[G_N(v) - G_\mu(v)] + \frac{P_N(v)}{\prod_{i=1}^k \sqrt{(v - A_i)(v - B_i)}} \\ &= O\left(\frac{1}{N\eta\sqrt{\kappa_E + \eta}} + \frac{K}{(N\eta)^2\sqrt{\kappa_E + \eta}}\right) |\mathbb{E}[\text{Im}[G_N(v) - G_\mu(v)]]| + O\left(\frac{\varepsilon^2}{\sqrt{\kappa_E + \eta}} + \varepsilon_0\right). \end{aligned}$$

Now we are in a position to apply the map  $\Upsilon_v$ . Indeed, the integral of  $G_N(z)$  around  $\mathcal{C}_2^i$  is deterministic and equals  $n_i(N)/N$ . On the other hand, the integral of  $G_\mu(z)$  around  $\mathcal{C}_2^i$  equals the total mass of  $\mu(x)$  inside  $\mathcal{C}_2^i$ , which is  $\hat{n}_i = n_i(N)/N$ . Thus the integral of  $\mathbb{E}[G_N(z) - G_\mu(z)]$  around each loop  $\mathcal{C}_2^i$  vanishes. So  $\mathbb{E}[G_N(z) - G_\mu(z)]$  is given by

$$\begin{aligned} &\mathbb{E}[G_N(z) - G_\mu(z)] \\ &= \Upsilon_v \left[ O\left(\frac{1}{N\eta\sqrt{\kappa_E + \eta}} + \frac{K}{(N\eta)^2\sqrt{\kappa_E + \eta}}\right) |\mathbb{E}[\text{Im}[G_N(v) - G_\mu(v)]]| + O\left(\frac{\varepsilon^2}{\sqrt{\kappa_E + \eta}} + \varepsilon_0\right) \right] \\ &= O\left(\frac{1}{N\eta\sqrt{\kappa_E + \eta}} + \frac{K}{(N\eta)^2\sqrt{\kappa_E + \eta}}\right) |\mathbb{E}[\text{Im}[G_N(v) - G_\mu(v)]]| + O\left(\frac{\varepsilon^2}{\sqrt{\kappa_E + \eta}} + \varepsilon_0\right), \end{aligned} \quad (5.24)$$

where in the last line we used that the map  $\Upsilon_v$  depends only on the contours  $\mathcal{C}_2$ , and is Lipschitz.

By our assumption  $N\eta\sqrt{\kappa_E + \eta} \gg 1$ , and  $(N\eta)^2\sqrt{\kappa_E + \eta} \gg K$ , it follows by rearranging (5.24),

$$\mathbb{E}[G_N(z) - G_\mu(z)] = O\left(\frac{\varepsilon^2}{\sqrt{\kappa_E + \eta}} + \varepsilon_0\right).$$

This finishes the proof of (5.12). The remaining of the Theorem 5.7 can be proven in the same way as in the one-cut case. □

## 6 Examples and Generalizations

In this section, we discuss the generalizations of Theorems 1.8 and 1.11 in the following four senses:

1. In the special case when  $\theta = 1$ , thanks to the duality between particles and holes (see [1]), our results imply the rigidity and edge universality for both void and *saturated* regions.
2. Thanks to certain large deviation estimates for the largest and smallest particles, our results imply the rigidity and edge universality of stochastic systems supported on *infinite* intervals.
3. By conditioning on filling fractions, our results imply the edge universality of multi-cut stochastic systems *without fixed filling fractions*.
4. Thanks to certain large deviation estimates for the number of total particles, our results imply the edge universality of stochastic systems *without fixed number of particles*.

We demonstrate how the generalizations of Theorems 1.8 and 1.11 imply the rigidity and edge universality of certain stochastic systems, i.e., Binomial ensembles, discretized  $\beta$ -ensemble with convex potentials, Lozenge Tilings and Jack deformation of Plancherel measures.

### 6.1 $\theta = 1$ Binomial Weight

The first example is the binomial ensemble, which is a special case of the discrete  $\beta$ -ensemble defined in Section 1.1. This ensemble is also known as the *Krawtchouk orthogonal polynomial ensemble*. We demonstrate how to use the duality between the particles and holes to prove the rigidity and edge universality for both void and saturated regions.

Let  $\mathbb{P}_N$  be a discrete  $\beta$ -ensemble as defined in Section 1.1 with  $\theta = 1$ , i.e.,

$$\mathbb{P}_N(\ell_1, \ell_2, \dots, \ell_N) = \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} (\ell_i - \ell_j)^2 \prod_{i=1}^N w(\ell_i; N),$$

on  $\mathbb{W}_N$ , where  $\mathbb{W}_N$  is the set of ordered  $N$ -tuples of integers

$$0 \leq \ell_1 < \ell_2 < \dots < \ell_N \leq M. \tag{6.1}$$

We can consider its dual particle locations. More precisely, for any given particle configuration  $\ell_1, \ell_2, \dots, \ell_N$ , we denote the locations of holes,  $0 \leq x_1 < x_2 < \dots < x_{M-N+1} \leq M$ . The distribution  $\mathbb{P}_N$  induces a distribution on the configuration of the holes,

$$\mathbb{P}_N^{\text{dual}}(x_1, \dots, x_{M-N+1}) = \mathbb{P}_N(\ell_1, \ell_2, \dots, \ell_N) = \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} (\ell_i - \ell_j)^2 \prod_{i=1}^N w(\ell_i; N). \tag{6.2}$$

Up to some constant, we can rewrite the righthand side of (6.2) in terms of  $x_1, x_2, \dots, x_{M-N+1}$ ,

$$\begin{aligned} \prod_{1 \leq i < j \leq N} (\ell_i - \ell_j)^2 \prod_{i=1}^N w(\ell_i; N) &= \frac{\prod_{\substack{1 \leq i \leq N, \\ 0 \leq k \leq M, k \neq \ell_i}} |\ell_i - k|}{\prod_{\substack{1 \leq i \leq M-N+1, \\ 0 \leq k \leq M, k \neq x_i}} |x_i - k|} \prod_{1 \leq i < j \leq N} (x_i - x_j)^2 \prod_{i=1}^N w(\ell_i; N) \\ &\propto \frac{1}{\prod_{\substack{1 \leq i \leq M-N+1, \\ 0 \leq k \leq M, k \neq x_i}} (x_i - k)^2} \prod_{1 \leq i < j \leq N} (x_i - x_j)^2 \prod_{i=1}^{N-M+1} w(x_i; N)^{-1}. \end{aligned}$$

where we used that

$$\begin{aligned} \prod_{\substack{1 \leq i \leq N, \\ 0 \leq k \leq M, k \neq \ell_i}} |\ell_i - k| \prod_{\substack{1 \leq i \leq M-N+1, \\ 0 \leq k \leq M, k \neq x_i}} |x_i - k| &= \prod_{0 \leq j \neq k \leq M} |j - k| \\ \prod_{i=1}^N w(\ell_i; N) \prod_{i=1}^{N-M+1} w(x_i; N) &= \prod_{j=1}^M w(j; N), \end{aligned}$$

are both constants depending on  $M, N$ .

Therefore, there exists some constant  $Z_N^{\text{dual}}$  so that the distribution of the configuration of the holes takes the following form

$$\mathbb{P}_N^{\text{dual}}(x_1, \dots, x_{M-N+1}) = \frac{1}{Z_N^{\text{dual}}} \prod_{1 \leq i < j \leq M-N+1} (x_i - x_j)^2 \prod_{i=1}^{N-M+1} w(x_i; N)^{-1} \prod_{\substack{0 \leq k \leq M, \\ k \neq x_i}} (x_i - k)^{-2}, \quad (6.3)$$

which is again a discrete  $\beta$ -ensemble with  $\theta = 1$  and weight,

$$w^{\text{dual}}(x; N) = w(x; N)^{-1} \prod_{\substack{0 \leq k \leq M, \\ k \neq x}} (x - k)^{-2}.$$

This is called the dual ensemble as studied in [1, Section 3.2]. The saturated regions (void regions) of the equilibrium measure of  $\mathbb{P}_N$  become the void regions (saturated regions) of the equilibrium measure of  $\mathbb{P}_N^{\text{dual}}$ . In the following we use the Krawtchouk orthogonal polynomial ensemble as an example to illustrate how to use the duality between the particles and holes to prove the rigidity and edge universality for both void and saturated regions.

Fix two integers  $0 \leq N \leq M$ , with  $M = \lceil \mathfrak{m}N \rceil$  and  $\mathfrak{m} > 1$ , and consider the space  $\mathbb{W}_N$  of  $N$ -tuples of integers as defined in (6.1). The Krawtchouk orthogonal polynomial ensemble is the probability distribution  $\mathbb{P}_N$  on  $\mathbb{W}_N$ ,

$$\mathbb{P}_N(\ell_1, \dots, \ell_N) = \frac{1}{Z(N, M)} \prod_{1 \leq i < j \leq N} (\ell_i - \ell_j)^2 \prod_{i=1}^N \binom{M}{\ell_i}. \quad (6.4)$$

We remark that the partition function  $Z(N, M)$  is explicitly known in this case:

$$Z(N, M) = 2^{N(M-N+1)} (M!)^N \prod_{j=0}^{N-1} \frac{j!}{(M-j)!}.$$

We denote the empirical particle density  $\mu_N = N^{-1} \sum_{i=1}^N \delta_{\ell_i/N}$ . The empirical particle density  $\mu_N$  converges to the equilibrium measure, which is absolutely continuous. For  $\mathbf{m} > 2$ , the density of the measure  $\mu_{\mathbf{a}}$  is given by

$$\mu_{\mathbf{m}}(x) = \begin{cases} \frac{1}{\pi} \operatorname{arccot} \left( \frac{\mathbf{m} - 2}{2\sqrt{\mathbf{m} - 1 - (x - \mathbf{m}/2)^2}} \right), & |x - \frac{\mathbf{m}}{2}| < \sqrt{\mathbf{m} - 1}, \\ 0, & |x - \frac{\mathbf{m}}{2}| \geq \sqrt{\mathbf{m} - 1}, \end{cases}$$

and for  $1 < \mathbf{m} < 2$ ,

$$\mu_{\mathbf{m}}(x) = \begin{cases} \frac{1}{\pi} \operatorname{arccot} \left( \frac{\mathbf{m} - 2}{2\sqrt{\mathbf{m} - 1 - (x - \mathbf{m}/2)^2}} \right), & |x - \frac{\mathbf{m}}{2}| < \sqrt{\mathbf{m} - 1}, \\ 1, & \frac{\mathbf{m}}{2} \geq |x - \frac{\mathbf{m}}{2}| \geq \sqrt{\mathbf{m} - 1}, \\ 0, & |x - \frac{\mathbf{m}}{2}| \geq \frac{\mathbf{m}}{2}, \end{cases}$$

where  $\operatorname{arccot}(y)$  is the inverse cotangent function. We use the notations from Section 1.1,

$$\hat{a} = 0, \quad \hat{b} = \mathbf{m}, \quad A = \frac{\mathbf{m}}{2} - \sqrt{\mathbf{m} - 1}, \quad B = \frac{\mathbf{m}}{2} + \sqrt{\mathbf{m} - 1}.$$

The Assumptions 1.2, 1.3 and 1.5 are verified in [17, Section 2]. For the case  $\mathbf{m} \geq 2$ , since both  $[\hat{a}, A]$  and  $[B, \hat{b}]$  are void regions, Theorem 1.8 and 1.11 imply that the rigidity and edge universality hold. For the case  $1 < \mathbf{m} < 2$ , unfortunately, since both  $[\hat{a}, A]$  and  $[B, \hat{b}]$  are saturated regions, Theorem 1.8 and 1.11 do not say anything. However, in this case, the rigidity and edge universality follow by considering the dual particle locations. As in (6.3), for any given particle configuration  $\ell_1, \ell_2, \dots, \ell_N$ , the distribution on the configuration of the holes,  $0 \leq x_1 < x_2 < \dots < x_{M-N+1} \leq M$  is given by

$$\mathbb{P}_N^{\text{dual}}(x_1, \dots, x_{M-N+1}) = \frac{1}{Z^{\text{dual}}(N, M)} \prod_{1 \leq i < j \leq M-N+1} (x_i - x_j)^2 \prod_{i=1}^{N-M+1} w^{\text{dual}}(x_i; N),$$

which is again a discrete  $\beta$ -ensemble as defined in Section 1.1 with  $\theta = 1$ , and weight,

$$w(x; N)^{\text{dual}} = \binom{M}{x}^{-1} \prod_{\substack{0 \leq j \leq M, \\ j \neq x}} (x - j)^{-2}.$$

One can check that for the dual measure  $\mathbb{P}_N^{\text{dual}}$ , the Assumptions 1.2, 1.3 and 1.5 hold, and its equilibrium measure is given by

$$\mu_{\mathbf{m}}^{\text{dual}}(x) = \begin{cases} \frac{1}{\mathbf{m}-1} \left( 1 - \frac{1}{\pi} \operatorname{arccot} \left( \frac{\mathbf{m} - 2}{2\sqrt{\mathbf{m} - 1 - (x - \mathbf{m}/2)^2}} \right) \right) & |x - \frac{\mathbf{m}}{2}| < \sqrt{\mathbf{m} - 1}, \\ 0 & |x - \frac{\mathbf{m}}{2}| \geq \sqrt{\mathbf{m} - 1}. \end{cases}$$

Since  $[\hat{a}, A]$  and  $[B, \hat{b}]$  are both void regions for the dual equilibrium measure,  $\mu_{\mathbf{m}}^{\text{dual}}$ , Theorem 1.8 and 1.11 imply the rigidity and edge universality for the configuration of holes. Since the particle configuration is

determined by the hole configuration, the rigidity of the hole configuration implies the rigidity of the particle configuration.

In general for the discrete  $\beta$ -ensemble with  $\theta = 1$  and satisfying Assumptions 1.2, 1.3, 1.5 and 1.7, by considering the duality between particles and holes, we have

1. The optimal rigidity holds, i.e. any particle is close to its classical location given by the equilibrium measure.
2. The distributions of extreme particles or extreme holes after proper scaling are the same as the distributions of extreme eigenvalues of  $GUE$ .

## 6.2 Arbitrary Convex Potential on $\mathbb{R}$ With No Saturation

The second example is the discrete  $\beta$ -ensemble supported on infinite intervals. We take a real convex analytic function  $V(x)$ , i.e. such that  $V''(x) > 0$  for all  $x \in \mathbb{R}$ . Fix a constant  $\kappa > 0$  such that

$$\liminf_{x \rightarrow \infty} \frac{\kappa V(x)}{2\theta \ln |x|} > 1, \quad (6.5)$$

and consider the probability distribution given by

$$\mathbb{P}(\ell_1, \dots, \ell_N) = \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} \frac{\Gamma(\ell_j - \ell_i + 1)\Gamma(\ell_j - \ell_i + \theta)}{\Gamma(\ell_j - \ell_i)\Gamma(\ell_j - \ell_i + 1 - \theta)} \prod_{i=1}^N \exp\left(-\kappa N \cdot V\left(\frac{\ell_i}{N}\right)\right), \quad (6.6)$$

on  $N$ -tuples  $\ell_1 < \ell_2 < \dots < \ell_N$  such that  $\ell_j = \lambda_j + \theta j$  and  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$  are integers. We are in the framework of Section 1.1 except that now the particle configuration is supported on an infinite interval, i.e.,  $\mathbb{R}$ . We have  $a(N) = \hat{a} = -\infty$  and  $b(N) = \hat{b} = \infty$ . We hereafter assume that the equilibrium measure has no saturated regions. Then the constrained equilibrium measure  $\mu$  is the same as the unconstrained equilibrium and  $\mu$  has only one band. This is the case when  $\kappa$  is small enough. By taking

$$\psi_N^+(x) = \exp\{\kappa N (V((x-1)/N) - V(x/N))\}, \quad \psi_N^-(x) = 1, \quad \phi^+(z) = \exp\{-\kappa V'(z)\}, \quad \phi^-(z) = 1, \quad (6.7)$$

we also have the following Nekrasov's equation,

$$R_N(\xi) = \psi_N^-(\xi) \mathbb{E}_{\mathbb{P}_N} \left[ \prod_{i=1}^N \left(1 - \frac{\theta}{\xi - \ell_i}\right) \right] + \psi_N^+(\xi) \mathbb{E}_{\mathbb{P}_N} \left[ \prod_{i=1}^N \left(1 + \frac{\theta}{\xi - \ell_i - 1}\right) \right]. \quad (6.8)$$

where  $R_N(\xi)$  is analytic over  $\mathbb{C}$ . In the argument of the proof of Theorem 2.4, we take a contour integral which encloses a small neighborhood of  $[\hat{a}, \hat{b}]$ . For this to work, we need to localize the probability measure  $\mathbb{P}_N$  onto a finite interval. Thanks to [17, Proposition 9.3], with exponentially high probability the empirical particle density of (6.6) is supported on a finite interval, i.e. there exist constants  $L, c$  such that

$$\mathbb{P}_N(-L \leq \ell_1/N \leq \ell_N/N \leq L) \geq 1 - e^{-cN}, \quad (6.9)$$

for  $N$  large enough. As in [17], we define the new measure  $\hat{\mathbb{P}}_N$  as  $\mathbb{P}_N$  conditioned on the event that  $\{\sup_{1 \leq i \leq N} |\ell_i/N| \leq L\}$ . For this new density  $\hat{\mathbb{P}}_N$ , we take

$$\hat{a} = -L, \quad \hat{b} = L, \quad a(N) = \max\{\lambda + \theta < -NL : \lambda \in \mathbb{Z}\}, \quad b(N) = \min\{\lambda + N\theta > NL : \lambda \in \mathbb{Z}\},$$

and  $\psi_N^\pm(z)$  and  $\psi^\pm(z)$  are as given in (6.7). The Assumptions 1.2, 1.3, and 1.5 are verified in [17, Section 9.3]. However, since we no longer have the vanishing boundary condition  $w(a(N); N) = w(b(N); N) = 0$ , we don't have the precise Nekrasov's equation. The righthand side of (6.8), after replacing  $\mathbb{P}_N$  by  $\hat{\mathbb{P}}_N$

$$\psi_N^-(\xi)\mathbb{E}_{\hat{\mathbb{P}}_N} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{\xi - \ell_i} \right) \right] + \psi_N^+(\xi)\mathbb{E}_{\hat{\mathbb{P}}_N} \left[ \prod_{i=1}^N \left( 1 + \frac{\theta}{\xi - \ell_i - 1} \right) \right],$$

has two poles at  $a(N) + 1$  and  $b(N)$ . Fortunately, thanks to the large deviation estimate (6.9), the residues at  $a(N) + 1$  and  $b(N)$  are exponentially small. Therefore, we have the following approximated Nekrasov's equation

$$R_N(\xi) + O(e^{-cN}) = \psi_N^-(\xi)\mathbb{E}_{\hat{\mathbb{P}}_N} \left[ \prod_{i=1}^N \left( 1 - \frac{\theta}{\xi - \ell_i} \right) \right] + \psi_N^+(\xi)\mathbb{E}_{\hat{\mathbb{P}}_N} \left[ \prod_{i=1}^N \left( 1 + \frac{\theta}{\xi - \ell_i - 1} \right) \right],$$

where  $R_N(\xi)$  is an analytic function, and the error  $O(e^{-cN})$  is uniform provided that  $\text{dist}(\xi/N, [-L, L]) \geq 2/N$ . The proof of Theorem 2.4 can be carried through without any change. For the remaining proofs, we do not use Nekrasov's equation. Therefore, the rigidity and edge universality hold for both  $\hat{\mathbb{P}}_N$  and  $\mathbb{P}_N$ ,

**Corollary 6.1.** *We fix a constant  $\kappa > 0$ , take a real convex analytic function  $V(x)$  satisfying (6.5), and consider the probability  $\mathbb{P}_N$  as given in (6.6). Suppose that the equilibrium measure  $\mu(x) = \rho_V(x)dx$  is supported on  $[A, B]$ , has no saturated region, and therefore only one band. Then the rigidity and edge universality hold for  $\mathbb{P}_N$ :*

1. For any  $\mathfrak{a} > 0$ , with probability at least  $1 - \exp(-c(\ln N)^2)$ , we have

$$\left| \frac{\ell_i}{N} - \gamma_i \right| \leq \frac{N^{\mathfrak{a}}}{N^{2/3} \min\{i, N - i\}^{1/3}}, \quad i \in \llbracket 1, N \rrbracket,$$

where  $\gamma_i$  are the classical particle locations of the equilibrium measure  $\mu$ .

2. Fix  $\beta \geq 1$ ,  $0 < \mathfrak{b} < 1/13$  and  $L = N^{\mathfrak{b}}$ . Let  $\rho_V(u) = (1 + O(u - A))s_A\sqrt{u - A}/\pi$ . Take any fixed  $m \geq 1$  and a continuously differentiable compactly supported function  $O : \mathbb{R}^m \rightarrow \mathbb{R}$ . For any index set  $\Lambda \in \llbracket 1, L \rrbracket$  with  $|\Lambda| = m$ ,

$$\left| \mathbb{E}_{\mathbb{P}_N} \left[ O \left( \left( N^{2/3} k^{1/3} s_A^{2/3} (\ell_k/N - \gamma_k) \right)_{k \in \Lambda} \right) \right] - \mathbb{E}_{G\beta E} \left[ O \left( \left( N^{2/3} k^{1/3} (x_k/N - \tilde{\gamma}_k) \right)_{k \in \Lambda} \right) \right] \right| = O(N^{-\chi}),$$

where  $\tilde{\gamma}_i$  are the classical particle locations of the semi-circle distribution. The same statement holds for particles close to the right edge.

### 6.3 Lozenge Tilings

The third example is the uniformly random lozenge tilings of a hexagon with a hole. We demonstrate that by conditioning on the filling fractions, our results imply the edge universality of multi-cut stochastic systems without filling fractions.

Consider an  $A \times B \times C$  hexagon drawn on the regular triangular lattice. We cut a rhombic  $D \times D$  hole in the hexagon, where the bottom point of the hole is at distance  $t$  from the left side of the hexagon and

at height  $H$  (counted from the bottom of the hexagon along  $t$ -th vertical line). Let  $\mathbb{P}_N$  be the probability distribution of the horizontal lozenges (outside the hole) on  $t$ -th vertical line induced by the uniform measure on all tilings. Assuming  $t > \max(B, C)$ , which yields  $N = B + C - D - t$ , and introducing the coordinate system such that the lowest possible position for horizontal lozenge on the  $t$ -th vertical line is 1 and the highest one is  $A + B + C - t$ . The probability distribution  $\mathbb{P}_N$  is given by

$$\mathbb{P}_N(\ell_1, \ell_2, \dots, \ell_N) = \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} (\ell_i - \ell_j)^2 \prod_{i=1}^N w(\ell_i; N),$$

where

$$w(\ell_i; N) = (A + B + C + 1 - t - \ell_i)_{t-B} (\ell_i)_{t-C} (H - \ell_i)_D (H - \ell_i)_D,$$

and  $(a)_n$  is the Pochhammer symbol,  $(a)_n = a(a+1) \cdots (a+n-1)$ . This is a two-cut discrete  $\beta$ -ensemble with  $k = 2$  and  $\theta = 1$ . We also need two filling fractions  $n_1$  and  $n_2$ : we consider such tilings that there are  $n_1(N)$  horizontal lozenges (on  $t$ -th vertical line) below the hole and  $n_2(N)$  lozenges above. We use  $N$  as a large parameter, and suppose that

$$\begin{aligned} A &= \hat{A}N + O(1), & B &= \hat{B}N + O(1), & C &= \hat{C} + O(1), & t &= \hat{t}N + O(1), \\ H &= \hat{H}N + O(1), & D &= \hat{D}N + O(1), & n_1(N) &= \hat{n}_1N, & n_2(N) &= \hat{n}_2N. \end{aligned}$$

Using the notations from Section 5, we have

$$\hat{a}_1 = 0, \quad \hat{b}_1 = \hat{H}, \quad \hat{a}_2 = \hat{H} + \hat{D}, \quad \hat{b}_2 = \hat{A} + \hat{B} + \hat{C} - \hat{t}.$$

We further assume that  $\hat{t} > \max\{\hat{B}, \hat{C}\}$ , and there exists a constant  $\mathfrak{d} > 0$ ,

$$\mathfrak{d} < \hat{n}_1 < \hat{H} - \mathfrak{d}, \quad \mathfrak{d} < \hat{n}_2 < \hat{A} + \hat{B} + \hat{C} - \hat{t} - \hat{H} - \hat{D} - \mathfrak{d}.$$

For the density  $\mathbb{P}_N$ , the potential  $V(x)$  has the form,

$$\begin{aligned} V(x) &= (\hat{A} + \hat{B} + \hat{C} - \hat{t} - x) \ln(\hat{A} + \hat{B} + \hat{C} - \hat{t} - x) - (\hat{A} + \hat{C} - x) \ln(\hat{A} + \hat{C} - x) \\ &\quad + x \ln x - (\hat{t} - \hat{C} - x) \ln(\hat{t} - \hat{C} - x) + 2(\hat{H} - x) \ln(\hat{H} - x) - 2(\hat{H} + \hat{D} - x) \ln(\hat{H} + \hat{D} - x). \end{aligned}$$

We denote the equilibrium measure with filling fraction  $\hat{n} = (\hat{n}_1, \hat{n}_2)$  by  $\mu^{\hat{n}} = \rho^{\hat{n}}(x)dx$ , which is given by the minimizer of the following functional

$$- \int_{x \neq y} \ln|x-y| \rho(x) \rho(y) dx dy + \int V(x) \rho(x) dx, \quad (6.10)$$

over all the densities  $\rho(x)$ , supported on  $[\hat{a}_1, \hat{b}_1] \cup [\hat{a}_2, \hat{b}_2]$ , with  $0 \leq \rho(x) \leq 1$ , and

$$\int_{a_i}^{b_i} \rho(x) dx = n_i, \quad i = 1, 2.$$

The function  $Q_{\mu^{\hat{n}}}$  is a square root of a polynomial of degree 6 [17, Section 9.2]

$$Q_{\mu^{\hat{n}}}(z) = c^{\hat{n}} (z - C^{\hat{n}}) \sqrt{(z - A_1^{\hat{n}})(z - B_1^{\hat{n}})(z - A_2^{\hat{n}})(z - B_2^{\hat{n}})},$$

where  $c^{\hat{n}}$  is a constant, and

$$a_1 < A_1^{\hat{n}} < B_1^{\hat{n}} < b_1 < C^{\hat{n}} < a_2 < A_2^{\hat{n}} < B_2^{\hat{n}} < b_2.$$

The Assumptions 5.2, 5.3, 5.4, 5.5 and 5.6 are verified in Section [17, Section 9.2]. Theorem 5.9 and 5.10, combining with a duality argument as in Section 6.1, imply the rigidity and edge universality for lozenge tilings of a hexagon with a hole (with fixed filling fraction above and below the hole).

In the following, we consider the uniformly random lozenge tilings without fixed filling fraction. We denote the empirical particle density  $\mu_N = N^{-1} \sum_{i=1}^N \delta_{\ell_i/N}$ , and  $\mu^* = \rho_V^*(x)dx$  the equilibrium measure, which is given by the minimizer of the following functional

$$- \int_{x \neq y} \ln |x - y| \rho(x) \rho(y) dx dy + \int V(x) \rho(x) dx, \quad (6.11)$$

over all the densities  $\rho(x)$ , supported on  $[\hat{a}_1, \hat{b}_1] \cup [\hat{a}_2, \hat{b}_2]$ , with  $0 \leq \rho(x) \leq 1$ . We denote

$$\hat{n}_1 = \mu_N([\hat{a}_1, \hat{b}_1]), \quad \hat{n}_1^* = \mu^*([\hat{a}_1, \hat{b}_1]), \quad \hat{n}_2 = \mu_N([\hat{a}_2, \hat{b}_2]), \quad \hat{n}_2^* = \mu^*([\hat{a}_2, \hat{b}_2]),$$

and write  $\hat{n}^* = (\hat{n}_1^*, \hat{n}_2^*)$  and  $\hat{n} = (\hat{n}_1, \hat{n}_2)$ . We recall that  $\mu^{\hat{n}} = \rho_V^{\hat{n}}(x)dx$ , the equilibrium measure with filling fraction  $\hat{n}$  is given by the minimizer of the functional (6.10). Since the equilibrium measure  $\rho_V^{\hat{n}}$  depends on the filling fraction  $\hat{n}$ , we cannot expect the rigidity and edge universality if the filling fraction  $\hat{n}$  fluctuates too much. We make the following assumption: with probability going to one

$$\|\hat{n}^* - \hat{n}\|_{\infty} := \max\{|\hat{n}_1^* - \hat{n}_1|, |\hat{n}_2^* - \hat{n}_2|\} = o(N^{-2/3}). \quad (6.12)$$

**Remark 6.2.** *Under mild assumptions, statements like (6.12) are proven in [24] for multi-cut discrete  $\beta$ -ensembles without fixed filling fraction in a more general setting. In particular, for the tiling model studied in this section, (6.12) always holds.*

As proven in [17, Proposition 5.2], the equilibrium measure  $\mu^{\hat{n}}$  is Lipschitz with respect to the filling fraction, i.e. if  $\hat{n}$  and  $\hat{n}^*$  are close enough

$$\mathcal{D}(\mu^{\hat{n}}, \mu^{\hat{n}^*}) \lesssim \|\hat{n} - \hat{n}^*\|_{\infty}, \quad (6.13)$$

where  $\mathcal{D}(\cdot, \cdot)$  is defined in (2.18). As a consequence,  $Q_{\mu^{\hat{n}}}$  is Lipschitz in  $\hat{n}$  in a neighborhood  $B_{\varepsilon} = \{\hat{n} : \|\hat{n} - \hat{n}^*\| \leq \varepsilon\}$  of  $\hat{n}^*$ , and  $c^{\hat{n}}, C^{\hat{n}}, A_1^{\hat{n}}, B_1^{\hat{n}}, A_2^{\hat{n}}, B_2^{\hat{n}}$  are all Lipschitz with respect to the filling fractions in a neighborhood of  $\hat{n}^*$  ( $c^{\hat{n}^*} \neq 0$  and the  $A_1^{\hat{n}^*}, B_1^{\hat{n}^*}, A_2^{\hat{n}^*}, B_2^{\hat{n}^*}$  are all distinct so that the derivatives of  $Q_{\mu^{\hat{n}}}$  with respect to these parameters do not vanish). We observe here the large deviation estimates similar to Proposition 2.3 insures that  $B_{\varepsilon}^c$  has probability bounded by  $e^{-cN^2}$  for some  $c > 0$  for both continuous and discrete models.

Fix  $i \in \llbracket 1, 2 \rrbracket$ . We assume that  $[\hat{a}_i, A_i^{\hat{n}^*}]$  is a void region for the equilibrium measure  $\mu^*$ , and  $\rho_V^* = (1 + O(u - A_i^{\hat{n}^*}))s_{A_i^{\hat{n}^*}} \sqrt{u - A_i^{\hat{n}^*}}/\pi$ . In the following, we show that under assumption 6.12, the left few particles on the cut  $[\hat{a}_i, \hat{b}_i]$  have Tracy-Widom fluctuations.

We first show that if we condition on the filling fraction  $\hat{n}$ , satisfying (6.12), the left few particles on the cut  $[\hat{a}_i, \hat{b}_i]$  have Tracy-Widom fluctuations. Thanks to (6.13),  $[\hat{a}_i, A_i^{\hat{n}}]$  is also a void region for the equilibrium



measure  $\mu^{\hat{n}}$ , and  $\rho_V^{\hat{n}} = (1 + O(u - A_i^{\hat{n}}))s_{A_i^{\hat{n}}}\sqrt{u - A_i^{\hat{n}}}/\pi$ . The following is a consequence of Theorem 5.10. Take any fixed  $m \geq 1$  and a continuously differentiable compactly supported function  $O : \mathbb{R}^m \rightarrow \mathbb{R}$ . For any index set  $\Lambda \in \mathbb{Z}_{\geq 1}$  with  $|\Lambda| = m$ , we denote the shifted index set  $\Lambda_i = \{k + n_1 1_{i=2} : k \in \Lambda\}$ , then

$$\begin{aligned}
& \mathbb{E}_{G\beta E} \left[ O \left( \left( N^{2/3} (x_k/N + 2) \right)_{k \in \Lambda} \right) \right] + O(N^{-\chi}) \\
&= \mathbb{E} \left[ 1_{\hat{n} \in B_\varepsilon} \mathbb{E}_{\mathbb{P}_N} \left[ O \left( \left( N^{2/3} s_{A_i^{\hat{n}}}^{2/3} (\ell_k/N - A_i^{\hat{n}}) \right)_{k \in \Lambda_i} \right) \middle| \hat{n} \right] \right] + O(e^{-cN^2}) \\
&= \mathbb{E} \left[ \mathbb{E}_{\mathbb{P}_N} \left[ O \left( \left( N^{2/3} s_{A_i^{\hat{n}^*}}^{2/3} (\ell_k/N - A_i^{\hat{n}^*}) \right)_{k \in \Lambda_i} \right) \middle| \hat{n} \right] \right] + \mathbb{E}[1_{\hat{n} \in B_\varepsilon} O(N^{2/3} \|\hat{n} - \hat{n}^*\|_\infty)] + O(e^{-cN^2}) \\
&= \mathbb{E} \left[ \mathbb{E}_{\mathbb{P}_N} \left[ O \left( \left( N^{2/3} s_{A_i^{\hat{n}^*}}^{2/3} (\ell_k/N - A_i^{\hat{n}^*}) \right)_{k \in \Lambda^*} \right) \middle| \hat{n} \right] \right] + o(1),
\end{aligned} \tag{6.14}$$

where  $\Lambda^* = \{k + j - 1 : k \in \Lambda\}$ , and  $j$  is the index of the left most particle on the cut  $[a_i, b_i]$ , and in the third line we used that  $s_{A_i^{\hat{n}}}$  and  $A_i^{\hat{n}}$  are Lipschitz with respect to the filling fraction. Combining with a duality argument as in Section 6.1, the above argument implies, under assumption 6.12, extreme particles or holes for lozenge tilings of a hexagon with a hole without fixed filling fraction have Tracy-Widom distribution.

## 6.4 Jack Deformation of Plancherel Measure

The last example is the Jack Deformation of Plancherel Measure. We prove that for  $\beta \geq 1$ , the distribution of the lengths of the first few rows in Young diagrams under the Jack deformation of Plancherel measure, appropriately scaled, converges to the Tracy-Widom  $\beta$  distribution. This answers an open problem in [39, Section 1.6].

Young diagrams of size  $n$  is a collection of  $n$  boxes, arranged in left-justified rows, with the row lengths in non-increasing order. We denote a Young diagram of size  $n$  by listing the number of boxes in each row,  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{\ell_\lambda})$ , where  $\ell_\lambda$  is the length of  $\lambda$ ,  $\lambda_i$  is the number of boxes in the  $i$ -th row of  $\lambda$  and  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{\ell_\lambda}$ . In this section we study the following probability measure on Young diagrams of size  $n$ ,

$$\mathbb{M}^{(n)}(\lambda) = \frac{n! \theta^n}{H(\lambda, \theta) H'(\lambda, \theta)}, \tag{6.15}$$

where

$$\begin{aligned}
H(\lambda, \theta) &= \prod_{(i,j) \in \lambda} ((\lambda_i - j) + (\lambda'_j - i)\theta + 1), \\
H'(\lambda, \theta) &= \prod_{(i,j) \in \lambda} ((\lambda_i - j) + (\lambda'_j - i)\theta + \theta),
\end{aligned} \tag{6.16}$$

and  $(i, j) \in \lambda$  stands for the box in the  $i$ -th row and  $j$ -th column of the Young diagram  $\lambda$ , and  $\lambda'$  is the transposed Young diagram. In the expressions (6.16),  $\lambda_i - j$  is the arm length and  $\lambda'_j - i$  is the leg length of the box  $(i, j) \in \lambda$ .

When  $\theta = 1$ , the measure  $\mathbb{M}^{(n)}$  specializes to the well-known Plancherel measure for the symmetric groups. In general, it is called the Jack deformation of Plancherel measure, because of its connection to Jack

polynomials. We refer to [21, 39, 53, 74, 75, 86] for recent researches on the Jack measure. We prove that the distribution of the lengths of the first few rows in Young diagrams under the Jack deformation of Plancherel measure, after proper scaling, converges to the Tracy-Widom  $\beta$  distribution.

**Theorem 6.3.** *Fix  $\beta \geq 1$ ,  $\theta = \beta/2$  and integer  $m > 0$ . For any  $(t_1, t_2, \dots, t_m) \in \mathbb{R}^m$ , we have*

$$\lim_{M \rightarrow \infty} \mathbb{M}^{(M)} \left( \theta^{-5/6} M^{1/3} \left( \lambda_k / \sqrt{M} - 2\sqrt{\theta} \right) \leq t_k, 1 \leq k \leq m \right) = F_\theta(t_1, t_2, \dots, t_m), \quad (6.17)$$

where

$$F_\theta(t_1, t_2, \dots, t_m) := \lim_{N \rightarrow \infty} \mathbb{P}_{G\beta E} \left( N^{2/3} (x_k/N - 2) \leq t_k, 1 \leq k \leq m \right). \quad (6.18)$$

In the special case  $\theta = 1$ , Theorem 6.3 specializes to the well-known Baik-Deift-Johansson theorem, which was first proven in [3] for the first row, and generalized to the first few rows in [19, 58, 82]. For the cases  $\theta = 1/2$  and  $\theta = 2$ , the above theorem was proven in [4].

We can rewrite the quantities  $H(\lambda, \theta)$  and  $H'(\lambda, \theta)$  in (6.16) in a more explicit way, in terms of row lengths of  $\lambda$  (see [74, Lemma 4.1]):

$$\begin{aligned} H(\lambda, \theta) &= \prod_{1 \leq i < j \leq \ell_\lambda} \frac{\Gamma(\lambda_i - \lambda_j + (j - i - 1)\theta + 1)}{\Gamma(\lambda_i - \lambda_j + (j - i)\theta + 1)} \prod_{i=1}^{\ell_\lambda} \Gamma(\lambda_i + (\ell_\lambda - i)\theta + 1) \\ H'(\lambda, \theta) &= \prod_{1 \leq i < j \leq \ell_\lambda} \frac{\Gamma(\lambda_i - \lambda_j + (j - i)\theta)}{\Gamma(\lambda_i - \lambda_j + (j - i + 1)\theta)} \prod_{i=1}^{\ell_\lambda} \frac{\Gamma(\lambda_i + (\ell_\lambda - i)\theta + \theta)}{\Gamma(\theta)} \end{aligned} \quad (6.19)$$

For  $m \geq 1$  and any Young diagram  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{\ell_\lambda})$ , we define

$$V_m(\lambda) = \prod_{1 \leq i < j \leq m} \frac{\Gamma(\lambda_i - \lambda_j + (j - i)\theta + 1) \Gamma(\lambda_i - \lambda_j + (j - i + 1)\theta)}{\Gamma(\lambda_i - \lambda_j + (j - i - 1)\theta + 1) \Gamma(\lambda_i - \lambda_j + (j - i)\theta)}$$

and

$$W_m(\lambda) = \prod_{i=1}^m \frac{\Gamma(\theta)}{\Gamma(\lambda_i + (m - i)\theta + 1) \Gamma(\lambda_i + (m - i)\theta + \theta)},$$

where by our convention,  $\lambda_i = 0$  for  $i > \ell_\lambda$ . According to the formula (6.19), the measure  $\mathbb{M}^{(n)}$  as in (6.15) can be rewritten as

$$\mathbb{M}^{(n)}(\lambda) = \frac{n! \theta^n}{H(\lambda, \theta) H'(\lambda, \theta)} = n! \theta^n V_{\ell_\lambda}(\lambda) W_{\ell_\lambda}(\lambda). \quad (6.20)$$

The following lemma proven in [53, Lemma 6.6] will allow us to conclude that the lengths of the Young diagrams under the measure  $\mathbb{M}^{(n)}$  are of size  $O(\sqrt{n})$  with high probability.

**Lemma 6.4.** *Suppose that  $\theta > 0$ , then we have*

$$\begin{aligned} \mathbb{M}^{(n)} \left( \lambda_1 \geq 2e\sqrt{n\theta} \right) &\leq \theta^{-1} n^2 4^{-e\sqrt{n\theta}}, \\ \mathbb{M}^{(n)} \left( \ell_\lambda \geq 2e\sqrt{n/\theta} \right) &\leq \theta n^2 4^{-e\sqrt{n/\theta}}. \end{aligned}$$

Although that the number of rows of the Young diagrams under the measure  $\mathbb{M}^{(n)}$  is random, the following simple but important lemma allows us to view  $\mathbb{M}^{(n)}$  as a measure of Young diagrams with fixed number of rows.

**Lemma 6.5.** *If  $N \geq \ell_\lambda$ , then*

$$V_N(\lambda)W_N(\lambda) = V_{\ell_\lambda}(\lambda)W_{\ell_\lambda}(\lambda).$$

*Proof.* We may assume that  $N > \ell_\lambda$ . By our convention,  $\lambda_i = 0$  for  $i > \ell_\lambda$ . Hence,

$$\begin{aligned} V_N(\lambda) &= V_{\ell_\lambda}(\lambda) \prod_{\substack{1 \leq i \leq \ell_\lambda, \\ \ell_\lambda < j \leq N}} \frac{\Gamma(\lambda_i + (j-i)\theta + 1)\Gamma(\lambda_i + (j-i+1)\theta)}{\Gamma(\lambda_i + (j-i-1)\theta + 1)\Gamma(\lambda_i + (j-i)\theta)} \prod_{\ell_\lambda < i < j \leq N} \frac{\Gamma((j-i)\theta + 1)\Gamma((j-i+1)\theta)}{\Gamma((j-i-1)\theta + 1)\Gamma((j-i)\theta)} \\ &= V_{\ell_\lambda}(\lambda) \prod_{1 \leq i \leq \ell_\lambda} \frac{\Gamma(\lambda_i + (N-i)\theta + 1)\Gamma(\lambda_i + (N-i+1)\theta)}{\Gamma(\lambda_i + (\ell_\lambda-i)\theta + 1)\Gamma(\lambda_i + (\ell_\lambda-i+1)\theta)} \prod_{\ell_\lambda < i \leq N} \frac{\Gamma((N-i)\theta + 1)\Gamma((N-i+1)\theta)}{\Gamma(\theta)} \\ &= V_{\ell_\lambda}(\lambda) \frac{W_{\ell_\lambda}(\lambda)}{W_N(\lambda)}. \end{aligned}$$

This finishes the proof. □

The proof of Theorem 6.3 uses the Poissonization and de-Poissonization scheme of [3, 19, 58]. We consider the Poissonized Jack measure with rate  $M$ ,

$$\mathbb{P}^{(M)}(\lambda) = e^{-M} \sum_{n=0}^{\infty} \mathbb{M}^{(n)}(\lambda) \frac{M^n}{n!} = e^{-M} V_{\ell_\lambda}(\lambda) W_{\ell_\lambda}(\lambda) \prod_{i=1}^{\ell_\lambda} (\theta M)^{\lambda_i},$$

on the set of all partitions, where  $\mathbb{M}^{(n)}(\lambda) = 0$  if  $\sum_i \lambda_i \neq n$ , and we used (6.20). Thanks to the Poisson concentration of measure and Lemma 6.4, for  $M$  large enough, there exists  $\mathfrak{c} \geq 2e \max\{\theta^{1/2}, \theta^{-1/2}\}$  and  $c > 0$  such that

$$\mathbb{P}^{(M)}(\ell_\lambda > \mathfrak{c}\sqrt{M}) + \mathbb{P}^{(M)}(\lambda_1 > \mathfrak{c}\sqrt{M}) \leq e^{-cM^{1/2}}. \quad (6.21)$$

In the following we take  $N$  large and  $M = N^2/\mathfrak{c}^2$ , and define the measure on the Young diagrams with at most  $N$  rows

$$\tilde{\mathbb{P}}^{(M)}(\lambda) = \frac{\mathbf{1}_{\{\ell_\lambda \leq N\}}}{Z_M} e^{-M} \sum_{n=0}^{\infty} \mathbb{M}^{(n)}(\lambda) \frac{M^n}{n!}.$$

where  $Z_M$  is a normalizing constant. It follows from (6.21),  $Z_M = 1 + O(\exp(-cM^{1/2}))$ . Hence for any bounded test function  $O : \mathbb{R}^m \mapsto \mathbb{R}$ , we have

$$\mathbb{E}_{\mathbb{P}^{(M)}}[O(\lambda_1, \lambda_2, \dots, \lambda_m)] = \mathbb{E}_{\tilde{\mathbb{P}}^{(M)}}[O(\lambda_1, \lambda_2, \dots, \lambda_m)] + O\left(\exp(-cM^{1/2})\right). \quad (6.22)$$

In the following we show that the measure  $\tilde{\mathbb{P}}^{(M)}$  falls into the category of discrete  $\beta$  ensemble. Hence, the law of the first few rows of Young diagrams under  $\tilde{\mathbb{P}}^{(M)}$ , after scaling, converges to the Tracy-Widom  $\beta$

distribution. For any partition  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)$  with at most  $N$  rows, where  $\lambda_i = 0$  if  $i > \ell_\lambda$ , we define  $\ell_i = \lambda_{N-i+1} + (i-1)\theta$ , for  $1 \leq i \leq N$ . Thanks to Lemma 6.5, we can view  $\tilde{\mathbb{P}}^{(M)}$  as a measure on the particle configurations  $(\ell_1, \ell_2, \dots, \ell_N) \in \mathbb{W}_N^\theta$  with  $a(N) = -1$  and  $b(N) = \infty$  as in Definition 1.1, and

$$\begin{aligned} \tilde{\mathbb{P}}^{(M)}(\lambda) &= \frac{\mathbf{1}_{\{\ell_\lambda \leq N\}}}{Z_M} e^{-M} V_N(\lambda) W_N(\lambda) \prod_{i=1}^N (\theta M)^{\lambda_i} \\ &= \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} \frac{\Gamma(\ell_i - \ell_j + 1) \Gamma(\ell_i - \ell_j + \theta)}{\Gamma(\ell_i - \ell_j) \Gamma(\ell_i - \ell_j + 1 - \theta)} \prod_{i=1}^N \frac{(\theta M)^{\ell_i}}{\Gamma(\ell_i + 1) \Gamma(\ell_i + \theta)} =: \mathbb{P}_N(\ell_1, \ell_2, \dots, \ell_N), \end{aligned} \quad (6.23)$$

which is a discrete  $\beta$  ensemble with weight

$$w(x; N) = \frac{(\theta M)^x}{\Gamma(x+1) \Gamma(x+\theta)}.$$

In the following, we check that the measure  $\mathbb{P}_N$  in (6.23) satisfies the Assumptions 1.2, 1.3, 1.5, 1.7.

1. For Assumption 1.2, we have  $a(N) = -1$ ,  $b(N) = \infty$ ,  $\hat{a} = 0$  and  $\hat{b} = \infty$ . Although the particle configuration is supported on an infinite interval, i.e.  $[0, \infty)$ . Thanks to (6.21), we have

$$\mathbb{P}_N(\ell_N \leq (1+\theta)N) > \tilde{\mathbb{P}}^{(M)}(\lambda_1 > N) = O(\exp(-cM^{1/2})).$$

We can condition on the event that  $\ell_N \leq (1+\theta)N$  as in Section 6.2. For those particle configurations, one can check that

$$V_N(x) = \frac{1}{N} \ln(\Gamma(Nx+1) \Gamma(Nx+\theta)) - x \ln(\theta M),$$

and

$$V(x) = 2x \ln x - 2x - x \ln(\theta/c^2), \quad V'(x) = 2 \ln x - \ln(\theta/c^2).$$

2. For Assumption 1.3, we have

$$\psi_N^+(x) = \theta/c^2, \quad \psi_N^-(x) = \frac{x(x+\theta-1)}{N^2}, \quad \phi^+(x) = \theta/c^2, \quad \phi^-(x) = x^2.$$

3. For Assumption 1.5, we look for an equilibrium measure  $\mu(dx) = \rho_V(x)dx$  so that there exists  $0 < A < B$ , such that it has a unique band  $[A, B]$  and  $[0, A]$  is a saturated region. Therefore, by taking derivative on both sides of (1.4),

$$2\theta \int_0^B \frac{\rho_V(y)dy}{x-y} = V'(x),$$

for any  $x \in [A, B]$ . We notice that the total mass of  $\mu$  on  $[A, B]$  is  $1 - A/\theta$ . We can solve for  $\rho_V$  using the inverse Hilbert transform formula (A.1),

$$A = \theta - \frac{2\sqrt{\theta}}{c}, \quad B = \theta + \frac{2\sqrt{\theta}}{c}, \quad \rho_V(x) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq A, \\ \frac{2}{\pi\theta} \arctan \sqrt{\frac{B-x}{x-A}}, & A \leq x \leq B. \end{cases} \quad (6.24)$$

$\rho_V(x)dx$  is indeed our unique equilibrium measure as soon as  $0 < A < B$ , that is  $\theta \geq (2/\mathfrak{c})^2$ . The Stieltjes transform of the equilibrium measure  $\mu$  is given by

$$G_\mu(z) = \frac{1}{\theta} \ln \left( \frac{(z^2 - \theta z) - z\sqrt{(z-A)(z-B)}}{2\theta/\mathfrak{c}^2} \right).$$

where the branch of square root is such that  $\sqrt{(z-A)(z-B)} \sim z$  as  $z \rightarrow \infty$ . The quantities  $R_\mu(z)$  and  $Q_\mu(z)$  are given by

$$R_\mu(z) = z^2 - \theta z, \quad Q_\mu(z) = H(z)\sqrt{(z-A)(z-B)}, \quad H(z) = z.$$

Although  $H(z)$  vanishes at  $z = 0$ ,  $R_\mu(z)$ ,  $\psi_N^-(zN)$ ,  $\phi^-(z)$  and  $e^{\theta G_\mu(z)}$  all vanish at  $z = 0$ , which cancels the zero of  $H(z)$ . One can check that the Proof of Theorem 2.4 can be carried through without any change.

4. For Assumption 1.7, we have that  $V(x)$  is analytic in a neighborhood of  $[A, B]$ , and

$$V'_N(x) = V'(x) + O(1/N).$$

Therefore, it follows from Theorem 1.11 and (6.22), the distribution of the lengths of the first few rows in Young diagrams under the Poissonized Jack measure converges to the Tracy-Widom  $\beta$  distribution.

**Theorem 6.6.** *Fix  $\beta \geq 1$ ,  $\theta = \beta/2$ , and an index set  $\Lambda \in \mathbb{Z}_{\geq 1}$  with  $|\Lambda| = m$ . For any continuously differentiable compactly supported function  $O : \mathbb{R}^m \rightarrow \mathbb{R}$ , there exists a small  $\chi > 0$ , such that*

$$\mathbb{E}_{\mathbb{P}(M)} \left[ O \left( \theta^{-5/6} M^{1/3} \left( \lambda_k / \sqrt{M} - 2\sqrt{\theta} \right), k \in \Lambda \right) \right] - \mathbb{E}_{G\beta E} \left[ O \left( N^{2/3} (x_k/N - 2), k \in \Lambda \right) \right] = O(M^{-\chi}), \quad (6.25)$$

where  $N = \mathfrak{c}\sqrt{M}$ , provided  $M$  is large enough and  $\mathfrak{c}$  is chosen greater than  $2e \max\{\theta^{1/2}, \theta^{-1/2}\}$ .

*Proof.* From (6.24), the equilibrium measure of  $\mathbb{P}_N$  is

$$\rho_V(x) = \frac{\mathfrak{c}^{1/2}}{\theta^{5/4}\pi} (1 + O(B-x)) \sqrt{B-x}, \quad x \rightarrow B-,$$

where  $B = \theta + 2\sqrt{\theta}/\mathfrak{c}$ . Theorem 1.11 implies that

$$\mathbb{E}_{\mathbb{P}_N} \left[ O \left( \mathfrak{c}^{1/3} \theta^{-5/6} N^{2/3} (\ell_{N-k+1}/N - B), k \in \Lambda \right) \right] - \mathbb{E}_{G\beta E} \left[ O \left( N^{2/3} (x_k/N - 2), k \in \Lambda \right) \right] = O(M^{-\chi}).$$

Since  $N = \mathfrak{c}\sqrt{M}$  and  $\ell_{N-k+1} = \lambda_k + (N-k+1)\theta$ , and  $O$  is continuous differentiable, we have

$$O \left( \mathfrak{c}^{1/3} \theta^{-5/6} N^{2/3} (\ell_{N-k+1}/N - B), k \in \Lambda \right) = O \left( \theta^{-5/6} M^{1/3} \left( \lambda_k / \sqrt{M} - 2\sqrt{\theta} \right), k \in \Lambda \right) + O(N^{-1/3}).$$

Hence,

$$\mathbb{E}_{\mathbb{P}(M)} \left[ O \left( \theta^{-5/6} M^{1/3} \left( \lambda_k / \sqrt{M} - 2\sqrt{\theta} \right), k \in \Lambda \right) \right] - \mathbb{E}_{G\beta E} \left[ O \left( N^{2/3} (x_k/N - 2), k \in \Lambda \right) \right] = O(M^{-\chi}). \quad (6.26)$$

(6.25) follows from combining (6.26) and (6.22).  $\square$

In the following we prove Theorem 6.3 using the de-Poissonization argument as in [3, 19, 58].

*Proof of Theorem 6.3.* We denote  $F_n$  the distribution function of  $\lambda_1, \lambda_2, \dots, \lambda_m$  under the Jack deformation of Plancherel measure  $\mathbb{M}^{(n)}$ ,

$$F_n(x_1, x_2, \dots, x_m) = \mathbb{M}^{(n)}(\lambda_k \leq x_k, 1 \leq k \leq m).$$

We also denote the distribution function of the Poissonized Jack deformation of the Plancherel measure  $\mathbb{P}^{(M)}$ ,

$$F(M; x_1, x_2, \dots, x_m) = e^{-M} \sum_{n=0}^{\infty} \frac{M^n}{n!} F_n(x_1, x_2, \dots, x_m).$$

The measures  $\mathbb{M}^{(n)}$  can be obtained as distribution at time  $n$  of a certain random growth process of Young diagrams [21, Section 1], i.e. there exist  $q_\theta(\mu, \lambda) \geq 0$ , such that

$$\mathbb{M}^{(n-1)}(\mu) = \sum_{\lambda: \lambda \searrow \mu} q_\theta(\mu, \lambda) \mathbb{M}^{(n)}(\lambda),$$

where  $\lambda \searrow \mu$  means that  $\mu$  can be obtained from  $\lambda$  by removing one box. This implies that

$$F_n(x_1, x_2, \dots, x_m) \leq F_{n-1}(x_1, x_2, \dots, x_m).$$

Moreover, by construction,  $F_n(x_1, x_2, \dots, x_m)$  and  $F(M; x_1, x_2, \dots, x_m)$  are monotone in  $x_k$  for any  $1 \leq k \leq m$ . Hence, let  $M_\pm = M \pm 4\sqrt{M \ln M}$ , [58, Lemma 3.8] (see also [18, Lemma 4.7]) implies that

$$F(M_+; x_1, x_2, \dots, x_m) - \frac{C}{M^2} \leq F_M(x_1, x_2, \dots, x_m) \leq F(M_-; x_1, x_2, \dots, x_m) + \frac{C}{M^2}.$$

Especially, for any  $(t_1, t_2, \dots, t_m) \in \mathbb{R}^m$ , by taking  $x_k = 2\sqrt{\theta M} + \theta^{5/6} M^{1/6} t_k$ , we have

$$\begin{aligned} & F(M_+; 2\sqrt{\theta M} + \theta^{5/6} M^{1/6} t_k, 1 \leq k \leq m) - \frac{C}{M^2} \leq F_M(2\sqrt{\theta M} + \theta^{5/6} M^{1/6} t_k, 1 \leq k \leq m) \\ & \leq F(M_-; 2\sqrt{\theta M} + \theta^{5/6} M^{1/6} t_k, 1 \leq k \leq m) + \frac{C}{M^2}. \end{aligned} \quad (6.27)$$

For any  $1 \leq k \leq m$ ,  $2\sqrt{\theta M_\pm} + \theta^{5/6} M_\pm^{1/6} t_k = 2\sqrt{\theta M} + \theta^{5/6} M^{1/6} t_k + O(\sqrt{\ln M})$ , so for any  $\varepsilon > 0$  and  $M$  sufficiently large,

$$2\sqrt{\theta M_+} + \theta^{5/6} M_+^{1/6} (t_k - \varepsilon) \leq 2\sqrt{\theta M} + \theta^{5/6} M^{1/6} t_k \leq 2\sqrt{\theta M_-} + \theta^{5/6} M_-^{1/6} (t_k + \varepsilon). \quad (6.28)$$

Thanks to the monotonicity property of  $F(M_\pm, x_1, x_2, \dots, x_m)$ , (6.28) implies that  $F(M_+; 2\sqrt{\theta M_+} + \theta^{5/6} M_+^{1/6} (t_k - \varepsilon), 1 \leq k \leq m) \leq F(M_+; 2\sqrt{\theta M} + \theta^{5/6} M^{1/6} t_k, 1 \leq k \leq m)$  and  $F(M_-; 2\sqrt{\theta M_-} + \theta^{5/6} M_-^{1/6} t_k, 1 \leq k \leq m) \leq F(M_-; 2\sqrt{\theta M_-} + \theta^{5/6} M_-^{1/6} (t_k + \varepsilon), 1 \leq k \leq m)$ . Hence, (6.27) leads to

$$\begin{aligned} & F(M_+; 2\sqrt{\theta M_+} + \theta^{5/6} M_+^{1/6} (t_k - \varepsilon), 1 \leq k \leq m) - \frac{C}{M^2} \leq F_M(2\sqrt{\theta M} + \theta^{5/6} M^{1/6} t_k, 1 \leq k \leq m) \\ & \leq F(M_-; 2\sqrt{\theta M_-} + \theta^{5/6} M_-^{1/6} (t_k + \varepsilon), 1 \leq k \leq m) + \frac{C}{M^2}. \end{aligned} \quad (6.29)$$

From this and Theorem 6.6, we conclude that

$$F_\theta(t_k - \varepsilon, 1 \leq k \leq m) - o(1) \leq F_M(2\sqrt{\theta M} + \theta^{5/6}M^{1/6}t_k, 1 \leq k \leq m) \leq F_\theta(t_k + \varepsilon, 1 \leq k \leq m) + o(1)$$

as  $M \rightarrow \infty$ , where  $F_\theta$  is as defined in (6.18). Since  $\varepsilon > 0$  is arbitrary and  $F_\theta$  is continuous, this finishes the proof of Theorem 6.3.  $\square$

## A Hilbert Transform on a finite interval $[a, b]$ .

In this section, we collect some useful facts of the Hilbert transform on a finite interval  $[a, b]$ . A comprehensive study of this topic can be found in [103, Chapter 4.3].

It follows by directly computation:

$$P.V. \int_a^b \frac{1}{\sqrt{(y-a)(b-y)}} \frac{dy}{x-y} = \begin{cases} 0 & x \in [a, b], \\ \frac{\pi}{\sqrt{(x-a)(x-b)}} & x \notin [a, b]. \end{cases}$$

and

$$\int_a^b \frac{\sqrt{(y-a)(b-y)} dy}{x-y} = \begin{cases} \pi \left( x - \frac{a+b}{2} \right) & x \in [a, b], \\ \pi \left( x - \frac{a+b}{2} \right) - \pi \sqrt{(x-a)(x-b)} & x \notin [a, b], \end{cases}$$

where the branch of  $\sqrt{(x-a)(x-b)}$  is chosen such that  $\sqrt{(x-a)(x-b)} \sim x$ , as  $x \rightarrow \infty$ .

Similar to Hilbert transform, there is an inverse Hilbert transform formula for Hilbert transform on finite intervals: if  $g$  is the Hilbert transform of  $f$ ,

$$P.V. \int_a^b \frac{f(y)}{x-y} dy = g(x), \quad x \in [a, b],$$

then  $f$  is given by the inverse Hilbert transform,

$$f(x) = \frac{1}{\pi^2 \sqrt{(x-a)(b-x)}} \int_a^b \frac{\sqrt{(y-a)(b-y)}}{y-x} g(y) dy + \frac{C}{\pi \sqrt{(x-a)(b-x)}}, \quad x \in (a, b), \quad (\text{A.1})$$

where

$$C = \int_a^b f(x) dx.$$

## B Rigidity and Proof of Proposition 3.2 and 3.3

In this section we prove Proposition 3.2 and 3.3. The proofs follow a similar argument to [46, Lemma B.1], which expresses an integral in terms of the Stieltjes transform using the Helffer-Sjöstrand functional calculus.

*Proof of Proposition 3.2.* Let  $S(x + iy) = G_N(x + iy) - G_\mu(x + iy)$ . We take a cutoff function  $\chi : \mathbb{R} \rightarrow \mathbb{R}$ , such that  $\chi(y) = 1$  on  $[-\eta, \eta]$ ,  $\chi(y)$  vanishes outside  $[-2\eta, 2\eta]$ , and  $\chi'(y) = O(1/\eta)$ . By the Helffer-Sjöstrand formula we have,

$$\begin{aligned} \left| \int_{-\infty}^{\infty} f(x)(d\mu_N(x) - d\mu(x)) \right| &\leq C \int_{\mathbb{C}_+} (|f(x)| + |y||f'(x)|)|\chi'(y)||S(x + iy)|dx dy \\ &+ C \left| \int_{\mathbb{C}_+} y\chi(y)f''(x) \operatorname{Im}[S(x + iy)]dx dy \right|. \end{aligned} \quad (\text{B.1})$$

The first term in (B.1) is nonzero, only if  $x \in I$  and  $y \in [\eta, 2\eta]$ , which implies  $x + iy \in \mathcal{D}_{M, r_0}$ . Thanks to the rigidity estimate on  $\mathcal{D}_{M, r_0}$ , we have  $|S(x + iy)| \leq \tilde{\eta}/y$ . The first term is easily bounded by

$$\int_{\mathbb{C}_+} (|f(x)| + |y||f'(x)|)|\chi'(y)||S(x + iy)|dx dy \lesssim \tilde{\eta}. \quad (\text{B.2})$$

For the second term in (B.1), we define  $\eta_0 := \tilde{\eta}/\sqrt{\kappa_I}$  in the bulk case; or  $\eta_0 := \tilde{\eta}^{2/3}$  in the edge case. From our assumptions, we have  $\eta \geq \eta_0$ . We split the second term into two integrals  $\{x + iy \in \mathbb{C}_+ : x \in I, y \geq \eta_0\} \cup \{x + iy \in \mathbb{C}_+ : x \in I, y \leq \eta_0\}$ . For the integral over the first region we integrate by part,

$$\begin{aligned} \int_{x \in I} \int_{y \geq \eta_0} y\chi(y)f''(x) \operatorname{Im}[S(x + iy)]dx dy &= - \int_{x \in I} f'(x)\eta_0 \operatorname{Re}[S(x + \eta_0)]dx \\ - \int_{x \in I} \int_{y \geq \eta_0} f'(x)\partial_y(y\chi(y)) \operatorname{Re}[S(x + iy)]dx dy. \end{aligned} \quad (\text{B.3})$$

By the definition of  $\eta_0$ , the rigidity estimates hold for  $y \geq \eta_0$ . Hence, both terms are easily estimated by  $C\tilde{\eta} \ln(N)$ .

On the region  $\{x + iy \in \mathbb{C}_+ : x \in I, y \leq \eta_0\}$ , let  $\kappa_x := \operatorname{dist}(x, \{A, B\})$ . In the bulk case, by our assumption, we have  $\kappa_x \asymp \kappa_I$ , and

$$|\operatorname{Im}[G_\mu(x + iy)]| \asymp \sqrt{\kappa_x + y} \lesssim \sqrt{\kappa_I + \eta_0} \lesssim \sqrt{\kappa_I}. \quad (\text{B.4})$$

In the edge case, if  $x \in [A, B]$ , we have

$$|\operatorname{Im}[G_\mu(x + iy)]| \asymp \sqrt{\kappa_x + y} \lesssim \sqrt{\eta + \eta_0} \lesssim \eta^{1/2} + \tilde{\eta}^{1/3}. \quad (\text{B.5})$$

If  $x \notin [A, B]$ , we have

$$|\operatorname{Im}[G_\mu(x + iy)]| \asymp y/\sqrt{\kappa_x + y} \lesssim y^{1/2} \lesssim \tilde{\eta}^{1/3}. \quad (\text{B.6})$$

It follows for the bulk case,

$$\begin{aligned} |\operatorname{Im}[S(x + iy)]| &\lesssim |\operatorname{Im}[G_N(x + iy)]| + |\operatorname{Im}[G_\mu(x + iy)]| \\ &\lesssim \frac{\eta_0}{y} |\operatorname{Im}[G_N(x + i\eta_0)]| + |\operatorname{Im}[G_\mu(x + iy)]| \\ &\lesssim \frac{\eta_0}{y} |\operatorname{Im}[S(x + i\eta_0)]| + \frac{\eta_0}{y} |\operatorname{Im}[G_\mu(x + i\eta_0)]| + |\operatorname{Im}[G_\mu(x + iy)]|, \end{aligned} \quad (\text{B.7})$$



where we used that  $y|\operatorname{Im}[G_N(x+iy)]|$  is monotonic increasing as functions of  $y$ . It follows by combining (B.4) and (B.7)

$$|\operatorname{Im}[S(x+iy)]| \lesssim \frac{\tilde{\eta}}{y} + \frac{\eta_0\sqrt{\kappa_I}}{y} + \sqrt{\kappa_I} \lesssim \frac{\tilde{\eta}}{y}.$$

Similarly for the edge case, It follows by combining (B.5), (B.6) and (B.7)

$$|\operatorname{Im}[S(x+iy)]| \lesssim \frac{\tilde{\eta}}{y} + \frac{\eta_0\eta^{1/2}}{y}.$$

Hence we have

$$\left| \int_{x \in I} \int_{y \leq \eta_0} y \chi(y) f''(x) \operatorname{Im}[S(x+iy)] dx dy \right| \lesssim \begin{cases} \frac{\eta_0 \tilde{\eta}}{\eta} \lesssim \tilde{\eta}, & \text{bulk case,} \\ \frac{\eta_0 \tilde{\eta}}{\eta} + \frac{\eta_0^2}{\eta^{1/2}} \lesssim \tilde{\eta}, & \text{edge case.} \end{cases}$$

This finishes the proof of (3.2).

In the following, we prove that uniformly for any  $E \in [a(N)/N, B - r_0 - r/4]$ :

$$|\#\{i : \ell_i/N \leq E\} - \#\{i : \gamma_i/N \leq E\}| \lesssim N\tilde{\eta} \ln N, \quad (\text{B.8})$$

and (3.3) follows. For any  $E \in [a(N)/N, B - r_0 - r/4]$  with  $\kappa_E = \operatorname{dist}(E, \{A, B\})$ . We take two test functions  $f_1(x)$  and  $f_2(x)$  such that

1. if  $E \leq A - \tilde{\eta}^{2/3}$ , we take  $f_1(x) = 1$  on  $x \leq E - \tilde{\eta}^{2/3}$  and vanishes on  $x \geq E$ ;  $f_2(x) = 1$  on  $x \leq E$  and vanishes on  $x \geq E + \tilde{\eta}^{2/3}$ .
2. if  $A - \tilde{\eta}^{2/3} \leq E \leq A + 2\tilde{\eta}^{2/3}$ , we take  $f_1(x) = 1$  on  $x \leq A - 2\tilde{\eta}^{2/3}$  and vanishes on  $x \geq A - \tilde{\eta}^{2/3}$ ;  $f_2(x) = 1$  on  $x \leq A + 2\tilde{\eta}^{2/3}$  and vanishes on  $x \geq A + 4\tilde{\eta}^{2/3}$ .
3. if  $A + 2\tilde{\eta}^{2/3} \leq E \leq B - r_0 - r/4$ , we take  $f_1(x) = 1$  on  $x \leq E - \tilde{\eta}/\sqrt{\kappa_E/2}$  and vanishes on  $x \geq E$ ;  $f_2(x) = 1$  on  $x \leq E$  and vanishes on  $x \geq E + \tilde{\eta}/\sqrt{\kappa_E/2}$ .

Let  $I$  be the interval such that  $f'(x) \neq 0$ , we further assume that  $f'(x) = O(1/|I|)$  and  $f''(x) = O(1/|I|^2)$ . One can check that  $I$  satisfies the conditions in Proposition 3.2. The same argument as for (3.2) gives that

$$\left| \int f_i(x) (d\mu_N(x) - d\mu(x)) \right| \lesssim \tilde{\eta} \ln N, \quad i = 1, 2. \quad (\text{B.9})$$

By our choices of the functions  $f_1$  and  $f_2$ , they obey  $f_1 \leq 1_{a(N)/N \leq x \leq E} \leq f_2$ . Therefore by (B.9),

$$\begin{aligned} \frac{\#\{i : \ell_i/N \leq E\}}{N} &\leq \int f_2(x) d\mu_N(x) = \int_{x \leq E} f_2(x) d\mu(x) + \int_{x \geq E} f_2(x) d\mu(x) + O(\tilde{\eta} \ln N) \\ &= \frac{\#\{i : \gamma_i/N \leq E\}}{N} + \int_{x \geq E} f_2(x) d\mu(x) + O(\tilde{\eta} \ln N). \end{aligned}$$

For  $x \in [A, B - r]$ ,  $\rho_V(x) \asymp \sqrt{\operatorname{dist}(x, \{A, B\})}$ . We have  $\int_{x \geq E} f_2(x) d\mu(x) = O(\tilde{\eta})$ . Hence,  $\#\{i : \ell_i/N \leq E\} \leq \#\{i : \gamma_i \leq E\} + O(N\tilde{\eta} \ln N)$ . The lower bound follows the same argument.  $\square$

*Proof of Proposition 3.3.* Let  $S(x + iy) = G_N(x + iy) - G_\mu(x + iy)$ . Without loss of generality, we assume  $\mathcal{M}$  contains a distance  $\tau$  neighborhood of  $[\hat{a}, \hat{b}]$ , where  $\tau \gtrsim 1$ . We take a cutoff function  $\chi : \mathbb{R} \rightarrow \mathbb{R}$ , such that  $\chi(y) = 1$  on  $[-\tau/2, \tau/2]$ ,  $\chi(y)$  vanishes outside  $[-\tau, \tau]$ , and  $\chi'(y) = O(1)$ . By the Helffer-Sjöstrand formula we have,

$$\begin{aligned} \left| \int_{-\infty}^{\infty} f(x) (d\mu_N(x) - d\mu(x)) \right| &\leq C \int_{\mathbb{C}_+} (|f(x)| + |y||f'(x)|) |\chi'(y)| |S(x + iy)| dx dy \\ &+ C \left| \int_{\mathbb{C}_+} y \chi(y) f''(x) \operatorname{Im}[S(x + iy)] dx dy \right|. \end{aligned} \quad (\text{B.10})$$

The first term in (B.10) is nonzero, only if  $y \in [\tau/2, \tau]$ , which implies  $x + iy \in \mathcal{D}_*$ . Thanks to the rigidity estimate on  $\mathcal{D}_*$ , we have  $|S(x + iy)| \leq \tilde{\eta}/y$ . The first term is easily bounded by

$$\int_{\mathbb{C}_+} (|f(x)| + |y||f'(x)|) |\chi'(y)| |S(x + iy)| dx dy \lesssim \tilde{\eta}. \quad (\text{B.11})$$

For the second term in (B.10), we denote  $\kappa_x = \operatorname{dist}(x, \{A, B\})$ , and we split the integral domain of the second term into three domains:  $\mathcal{D}' \cup \mathcal{D}'' \cup \mathcal{D}'''$ , where

$$\begin{aligned} \mathcal{D}' &= \mathcal{D}_*, \\ \mathcal{D}'' &= \{x + iy \in \mathbb{C}_+ : x \in [A, B], y \leq N^{-(1-\alpha)/2} / \sqrt{\kappa_x + y}\}, \\ \mathcal{D}''' &= \{x + iy \in \mathbb{C}_+ : x \notin [A, B], y \leq N^{-(1-\alpha)/3}\}. \end{aligned}$$

On  $\mathcal{D}'$ , we have  $|S(x + iy)| \leq \tilde{\eta}/y$ , and

$$\left| \int_{\mathcal{D}'} y \chi(y) f''(x) \operatorname{Im}[S(x + iy)] dx dy \right| \leq C \int_{\mathcal{D}'} \tilde{\eta} dx dy \lesssim \tilde{\eta}.$$

On  $\mathcal{D}''$ , we have  $x \in [A, B]$ , and

$$\min\{|\operatorname{Im}[G_\mu(x + iy)]|, |\operatorname{Im}[G_\mu^{\text{dual}}(x + iy)]|\} \asymp \sqrt{\kappa_x + y}. \quad (\text{B.12})$$

We denote  $\eta_x = N^{-(1-\alpha)/2} / \sqrt{\kappa_x + \eta_x}$ , it follows,

$$\begin{aligned} |\operatorname{Im}[S(x + iy)]| &\lesssim |\operatorname{Im}[G_N(x + iy)]| + |\operatorname{Im}[G_\mu(x + iy)]| \\ &\lesssim \frac{\eta_x}{y} |\operatorname{Im}[G_N(x + i\eta_x)]| + |\operatorname{Im}[G_\mu(x + iy)]| \\ &\lesssim \frac{\eta_x}{y} |\operatorname{Im}[S(x + i\eta_x)]| + \frac{\eta_x}{y} |\operatorname{Im}[G_\mu(x + i\eta_x)]| + |\operatorname{Im}[G_\mu(x + iy)]|, \end{aligned} \quad (\text{B.13})$$

where we used that  $y |\operatorname{Im}[G_N(x + iy)]|$  is monotonic increasing as a function of  $y$ . Thanks to the relation (2.11), there is another way to estimate  $|\operatorname{Im}[S(x + iy)]|$ ,

$$\begin{aligned} |\operatorname{Im}[S(x + iy)]| &\lesssim |\operatorname{Im}[G_N^{\text{dual}}(x + iy)]| + |\operatorname{Im}[G_\mu^{\text{dual}}(x + iy)]| + \frac{1}{Ny} \\ &\lesssim \frac{\eta_x}{y} |\operatorname{Im}[G_N^{\text{dual}}(x + i\eta_x)]| + |\operatorname{Im}[G_\mu^{\text{dual}}(x + iy)]| + \frac{1}{Ny} \\ &\lesssim \frac{\eta_x}{y} |\operatorname{Im}[S(x + i\eta_x)]| + \frac{\eta_x}{y} |\operatorname{Im}[G_\mu^{\text{dual}}(x + i\eta_x)]| + |\operatorname{Im}[G_\mu^{\text{dual}}(x + iy)]| + \frac{1}{Ny}. \end{aligned} \quad (\text{B.14})$$

Therefore, it follows from combining (B.12), (B.13) and (B.14),

$$|\operatorname{Im}[S(x+iy)]| \lesssim \frac{\tilde{\eta}}{y} + \frac{\eta_x \sqrt{\kappa_x + y}}{y} + \frac{1}{Ny} \lesssim \frac{\tilde{\eta}}{y} + \frac{N^{-(1-\alpha)/2}}{y} + \frac{1}{Ny},$$

where we used that  $|\operatorname{Im}[S(x+i\eta_x)]| \lesssim \tilde{\eta}/\eta_x$ . The integral over  $\mathcal{D}'$  is bounded

$$\left| \int_{\mathcal{D}'} y\chi(y)f''(x) \operatorname{Im}[S(x+iy)] dx dy \right| \leq C \left( \int_{\mathcal{D}'} \tilde{\eta} + N^{-(1-\alpha)/2} dx dy \right) \lesssim \tilde{\eta} + N^{-(1-\alpha)} \lesssim \tilde{\eta}.$$

On  $\mathcal{D}''$ , we denote  $\eta_x = N^{-(1-\alpha)/3}$ . For  $x \notin [A, B]$ , we have

$$\min\{|\operatorname{Im}[G_\mu(x+iy)]|, |\operatorname{Im}[G_\mu^{\text{dual}}(x+iy)]|\} \asymp \frac{y}{\sqrt{\kappa_x + y}}. \quad (\text{B.15})$$

The same argument as above, for  $x+iy \in \mathcal{D}''$ , we have

$$|\operatorname{Im}[S(x+iy)]| \lesssim \frac{\tilde{\eta}}{y} + \frac{\eta_x}{\sqrt{\kappa_x + y}} + \frac{1}{Ny} \lesssim \frac{\tilde{\eta}}{y} + \frac{N^{-(1-\alpha)/3}}{\sqrt{\kappa_x + y}} + \frac{1}{Ny},$$

and

$$\left| \int_{\mathcal{D}''} y\chi(y)f''(x) \operatorname{Im}[S(x+iy)] dx dy \right| \leq C \left( \int_{\mathcal{D}''} \tilde{\eta} + \frac{yN^{-(1-\alpha)/3}}{\sqrt{\kappa_x + y}} dx dy \right) \lesssim \tilde{\eta} + N^{-(1-\alpha)} \lesssim \tilde{\eta}.$$

This finishes the proof of (3.4). □

## C Proof of Propositions 3.17, 3.18, 3.19 and 3.20

*Rigidity and Proof of Proposition 3.17.* Let  $J := [a, \alpha]$ . We prove the bulk case: If  $\rho_W(x) \equiv 0$  on  $[a, \alpha]$ , then  $\alpha - a = O(\ln NM/N\sqrt{\kappa_I})$ . If  $\rho_W(x) \equiv N/L\theta$  on  $[a, \alpha]$ , then  $\alpha - a = O(\ln NM/N)$ . The rest statements follow from the same argument.

If  $\rho_W(x) \equiv 0$  on  $J$ , and  $\alpha - a \gtrsim \ln NM/N\sqrt{\kappa_I}$ , we can take a bump function  $\chi(x)$  supported on  $J$ , with  $\chi'(x) \asymp 1/|J|$  and  $\chi''(x) \asymp 1/|J|^2$ . It follows from Proposition 3.16, with high probability with respect to  $\mathbb{P}_L$

$$\left| \int \chi(x)(d\mu_L(x) - \rho_W(x)dx) \right| = \int \chi(x)d\mu_L(x) \leq C\sqrt{\frac{\ln N}{L}}. \quad (\text{C.1})$$

Moreover, it follows from our definition  $\mathcal{G}_{M,L,K}$  of the set of "good" boundary conditions, and proposition 3.2, with high probability with respect to  $\mathbb{P}_L$ ,

$$\left| \int \chi(x)(d\mu_N(x) - \rho_V(x)dx) \right| \lesssim \frac{\ln NM}{N}. \quad (\text{C.2})$$

Since  $\chi(x)$  is supported on  $J \subset I$ ,

$$\int \chi(x) d\mu_N(x) = \int_I \chi(x) d\mu_N(x) = \frac{L}{N} \int \chi(x) d\mu_L(x).$$

By comparing (C.1) and (C.2), it follows

$$\int \chi(x) \rho_V(x) dx \leq \frac{C \ln NM}{N}, \quad (\text{C.3})$$

provided that  $M^2 \gg L$ . On  $[a, \alpha]$ ,  $\rho_V(x) \asymp \sqrt{\kappa_I}$ . We can take the bump function  $\chi(x)$ , such that  $\int \chi(x) \rho_V(x) dx \asymp (\alpha - a) \sqrt{\kappa_I}$ . It follows that  $\alpha - a \lesssim \ln NM / N \sqrt{\kappa_I}$ .

If  $\rho_W(x) \equiv N/L\theta$  on  $[a, \alpha]$ , and  $\alpha - a \gtrsim \ln NM / N$ , we can take a bump function  $\chi(x)$  supported on  $J$ , with  $\chi'(x) \asymp 1/|J|$  and  $\chi''(x) \asymp 1/|J|^2$ . It follows from Proposition 3.16, with high probability with respect to  $\mathbb{P}_L$

$$\left| \int \chi(x) (d\mu_L(x) - \rho_W(x) dx) \right| \leq C \sqrt{\frac{\ln N}{L}}. \quad (\text{C.4})$$

Moreover, it follows from our definition  $\mathcal{G}_{M,L,K}$  of the set of "good" boundary conditions, and proposition 3.2, with high probability with respect to  $\mathbb{P}_L$ ,

$$\left| \int \chi(x) (d\mu_N(x) - \rho_V(x) dx) \right| = \left| \int \chi(x) \left( \frac{L}{N} d\mu_L(x) - \rho_V(x) dx \right) \right| \lesssim \frac{\ln NM}{N}. \quad (\text{C.5})$$

By taking difference of (C.4) and (C.5), it follows

$$\frac{L}{N} \int \chi(x) \rho_W(x) dx - \int \chi(x) \rho_V(x) dx = \int \chi(x) (\theta^{-1} - \rho_V(x)) dx \leq \frac{C \ln NM}{N}, \quad (\text{C.6})$$

provided  $M^2 \gg L$ . On  $J$ , there exists a constant  $c > 0$  such that  $\theta^{-1} - \rho_V(x) > c$ . We can take the bump function  $\chi(x)$ , such that  $\int \chi(x) (\theta^{-1} - \rho_V(x)) dx \asymp (\alpha - a)$ . It follows that  $\alpha - a \asymp \ln NM / N$ . This finishes the proof.  $\square$

*Proof of Proposition 3.18.* From Proposition 3.6,  $\kappa_I \gtrsim (L/N)^{2/3}$ ,  $|I| \asymp L/N \sqrt{\kappa_I}$  and  $\rho_V(x) \asymp \sqrt{\kappa_I}$ . We recall  $\alpha, \beta$  from Proposition 3.13 and the relevant estimates from Proposition 3.17. We take derivative of both sides of (3.15) at  $x \in [\alpha, \beta]$ ,

$$W'(x) = \frac{2\theta N}{L} \left( P.V. \int_{\alpha}^{\beta} \frac{\rho_V(y)}{x-y} dy + \int_a^{\alpha} \frac{\rho_V(y)}{x-y} dy + \int_{\beta}^b \frac{\rho_V(y)}{x-y} dy + \int_{I^c} \frac{d\mu(y) - d\mu_N(y)}{x-y} \right), \quad (\text{C.7})$$

where we have  $F_V(x) = 0$ , since  $x \in [\alpha, \beta] \subset [A, B]$ . It follows by taking derivative of both sides of (3.18) at  $x \in [\alpha, \beta]$

$$W'(x) = 2\theta \left( P.V. \int_{\alpha}^{\beta} \frac{\rho_W(y)}{x-y} dy + \int_a^{\alpha} \frac{\rho_W(y)}{x-y} dy + \int_{\beta}^b \frac{\rho_W(y)}{x-y} dy \right), \quad (\text{C.8})$$

where we have  $F_W(x) = 0$  on  $[\alpha, \beta]$ . By taking difference of (C.7) and (C.8), we have

$$\begin{aligned} P.V. \int_{\alpha}^{\beta} \frac{\rho_W(y) - \frac{N}{L}\rho_V(y)}{x-y} dy &= \frac{N}{L} \int_{I^c} \frac{d\mu(y) - d\mu_N(y)}{x-y} \\ &- \int_a^{\alpha} \frac{\rho_W(y) - \frac{N}{L}\rho_V(y)}{x-y} dy - \int_{\beta}^b \frac{\rho_W(y) - \frac{N}{L}\rho_V(y)}{x-y} dy. \end{aligned}$$

Thanks to the inverse Hilbert transform A.1, we get

$$\begin{aligned} \rho_W(x) - \frac{N}{L}\rho_V(x) &= \frac{1}{\pi\sqrt{(x-\alpha)(\beta-x)}} \left( \frac{N}{L} \int_{I^c} h(y, x) (d\mu(y) - d\mu_N(y)) - \right. \\ &\left. - \int_a^{\alpha} h(y, x) \left( \rho_W(y) - \frac{N}{L}\rho_V(y) \right) dy - \int_{\beta}^b h(y, x) \left( \rho_W(y) - \frac{N}{L}\rho_V(y) \right) dy \right), \end{aligned} \quad (\text{C.9})$$

for  $x \in [\alpha, \beta]$ , where

$$h(y, x) = \frac{\sqrt{(y-\alpha)(y-\beta)}}{y-x} - 1,$$

and we choose the branch such that  $\sqrt{(y-\alpha)(y-\beta)} \sim y$  as  $y \rightarrow \infty$ .

For the first term in (C.9), it breaks into two terms,

$$\int_b^{\infty} h(y, x) (d\mu(y) - d\mu_N(y)) + \int_{-\infty}^a h(y, x) (d\mu(y) - d\mu_N(y)). \quad (\text{C.10})$$

We estimate the first term in (C.10), and the second part can be estimated similarly. We recall the construction from Remark 3.4. Notice that  $\text{dist}(b, \{A, B\}) \asymp \kappa_I \gtrsim (L/N)^{2/3}$ , we construct bump functions  $\chi_1, \chi_2, \dots, \chi_{n'}$  with scale  $\tilde{\eta} = M/N$ .  $\chi_k(y)$  is supported on interval  $I_k$ , with  $|I_k| \asymp 2^k M/N \sqrt{\kappa_I}$  and  $\text{dist}(I_k, \beta) \asymp 2^k M/N \sqrt{\kappa_I}$ .

$$\begin{aligned} \int_{y \geq b} h(y, x) (d\mu(y) - d\mu_N(y)) &= \int_{b_0}^{b_1} (1 - \chi_1(y)) h(y, x) (d\mu(y) - d\mu_N(y)) \\ &+ \sum_{k=1}^{n'} \int_{y \geq b} \chi_k(y) h(y, x) (d\mu(y) - d\mu_N(y)). \end{aligned} \quad (\text{C.11})$$

Notice  $1 - \chi_1$  is supported on  $[b_0, b_1]$  where  $b_0 = b$  and  $b_1 = b + O(M/N\sqrt{\kappa_I})$ . For any  $y \in [b_0, b_1]$ , since  $b - \beta \lesssim M/N\sqrt{\kappa_I}$ ,  $y - \beta \lesssim M/N\sqrt{\kappa_I}$ . Also, by our assumption  $b - x \asymp |I| \asymp L/N\sqrt{\kappa_I}$ , we have  $y - x \asymp y - \alpha \asymp |I| \asymp L/N\sqrt{\kappa_I}$ . It follows that  $h(y, x) \lesssim (M/L)^{1/2}$ . Moreover, the total mass of  $\mu$  and  $\mu_N$  on this interval is bounded by  $O(\ln NM/N)$ . Therefore,

$$\left| \int_{b_0}^{b_1} h(y, x) (1 - \chi_1(y)) (d\mu(y) - d\mu_N(y)) \right| \lesssim \frac{\ln NM}{N} \left( \frac{M}{L} \right)^{1/2} = \left( \frac{M}{L} \right)^{3/2} \frac{\ln NL}{N}. \quad (\text{C.12})$$

For  $k \geq 1$ , there are two cases,  $|I_k| \leq |I|$  or  $|I_k| \geq |I|$ . If  $|I_k| \leq |I|$ , since by our assumption  $b - x \asymp |I| \asymp L/N\sqrt{\kappa I}$ , we have  $y - x \asymp y - \alpha \asymp |I|$ ,  $y - \beta \asymp |I_k|$ , and  $I_k \subset [b, B - r_0]$ . As a function of  $y$ ,  $h(y, x)$  and its derivatives satisfy

$$|h(y, x)| \lesssim |I_k|^{1/2}|I|^{-1/2}, \quad |\partial_y h(y, x)| \lesssim |I_k|^{-1/2}|I|^{-1/2}, \quad |\partial_y^2 h(y, x)| \lesssim |I_k|^{-3/2}|I|^{-1/2}.$$

Let  $h_k(y, x) = (|I|/|I_k|)^{1/2}h(y, x)$ , then as a function of  $y$ ,  $h_k(y, x) = O(1)$ ,  $\partial_y h_k(y, x) = O(|I_k|^{-1})$  and  $\partial_y^2 h_k(y, x) = O(|I_k|^{-2})$ . By Proposition 3.2, we have

$$\begin{aligned} \sum_{k:|I_k| \leq |I|} \left| \int h(y, x) \chi_k(y) (d\mu(y) - d\mu_N(y)) \right| &= \sum_{k:|I_k| \leq |I|} \left( \frac{|I_k|}{|I|} \right)^{1/2} \left| \int h_k(y, x) (d\mu(y) - d\mu_N(y)) \right| \\ &\lesssim \sum_{k:|I_k| \leq |I|} \left( \frac{|I_k|}{|I|} \right)^{1/2} \frac{\ln NM}{N} \lesssim \frac{\ln NM}{N}. \end{aligned} \quad (\text{C.13})$$

For  $|I_k| \geq |I|$ , we have  $y - x \asymp y - \alpha \asymp y - \beta \asymp |I_k|$ , and

$$\begin{aligned} &\sum_{k:|I_k| \geq |I|} \int h(y, x) \chi_k(y) (d\mu(y) - d\mu_N(y)) \\ &= \sum_{k:|I_k| \geq |I|} \int (h(y, x) - 1) \chi_k(y) (d\mu(y) - d\mu_N(y)) + \sum_{k:|I_k| \geq |I|} \int \chi_k(y) (d\mu(y) - d\mu_N(y)) \\ &= \sum_{k:|I_k| \geq |I|} \int (h(y, x) - 1) \chi_k(y) (d\mu(y) - d\mu_N(y)) + O\left(\frac{\ln NM}{N}\right). \end{aligned} \quad (\text{C.14})$$

As a function of  $y$ ,  $h(y, x) - 1$  and its derivatives satisfy,

$$|h(y, x) - 1| \lesssim |I|/|I_k|, \quad |\partial_y (h(y, x) - 1)| \lesssim |I|/|I_k|^2, \quad |\partial_y^2 (h(y, x) - 1)| \lesssim |I|/|I_k|^3.$$

Let  $h_k(y, x) = (|I_k|/|I|)(h(y, x) - 1)$ , then as a function of  $y$ ,  $h_k(y, x) = O(1)$ ,  $\partial_y h_k(y, x) = O(|I_k|^{-1})$  and  $\partial_y^2 h_k(y, x) = O(|I_k|^{-2})$ . Moreover, either  $I_k \subset [b, B - r_0]$ , or  $I_k$  is not completely contained in  $[b, B - r_0]$ , in which case  $|I_k| \gtrsim 1$ . By Proposition 3.2 and 3.3, we have

$$\begin{aligned} \sum_{k:|I_k| \geq |I|} \left| \int (h(y, x) - 1) \chi_k(y) (d\mu(y) - d\mu_N(y)) \right| &= \sum_{k:|I_k| \geq |I|} \frac{|I|}{|I_k|} \left| \int h_k(y, x) (d\mu(y) - d\mu_N(y)) \right| \\ &\lesssim \sum_{k:|I_k| \geq |I|} \frac{|I|}{|I_k|} \frac{\ln NM}{N} \lesssim \frac{\ln NM}{N}. \end{aligned} \quad (\text{C.15})$$

It follows by combining (C.11), (C.12), (C.13) (C.14) and (C.15),

$$\left| \int_{y \geq b} h(y, x) (d\mu(y) - d\mu_N(y)) \right| \lesssim \frac{\ln NM}{N}. \quad (\text{C.16})$$

Similarly, for the second term in (C.10),

$$\left| \int_{y \leq a} h(y, x) (d\mu(y) - d\mu_N(y)) \right| \lesssim \frac{\ln NM}{N}. \quad (\text{C.17})$$

For the second term in (C.9), there are two cases. If  $\rho_W(y) \equiv 0$  on  $[a, \alpha]$ , then  $\rho_V(y) \asymp \sqrt{\kappa_I}$ ,  $\alpha - a = O(\ln NM/N\sqrt{\kappa_I})$ ,  $x - y \asymp \beta - y \asymp L/N\sqrt{\kappa_I}$ ,  $h(y, x) \asymp (N\sqrt{\kappa_I}/L)^{1/2}\sqrt{y - \alpha}$  and

$$\left| \int_a^\alpha h(y, x) \left( \rho_W(y) - \frac{N}{L} \rho_V(y) \right) dy \right| \lesssim \left( \frac{N\sqrt{\kappa_I}}{L} \right)^{3/2} \int_a^\alpha \sqrt{y - \alpha} dy \lesssim \left( \frac{\ln NM}{L} \right)^{3/2}.$$

If  $\rho_W(y) \equiv N/L\theta$  on  $[a, \alpha]$ , then  $\alpha - a = O(\ln NM/N)$ ,  $\rho_W(y) - N\rho_V(y)/L \asymp N/L$ ,  $h(y, x) \asymp (N\sqrt{\kappa_I}/L)^{1/2}\sqrt{y - \alpha}$  and

$$\left| \int_a^\alpha h(y, x) \left( \rho_W(y) - \frac{N}{L} \rho_V(y) \right) dy \right| \lesssim \left( \frac{N}{L} \right)^{3/2} \kappa_I^{1/4} \int_a^\alpha \sqrt{y - \alpha} dy \lesssim \left( \frac{\ln NM}{L} \right)^{3/2} \kappa_I^{1/4}.$$

In both cases, the second term in (C.9) is bounded,

$$\left| \int_a^\alpha h(y, x) \left( \rho_W(y) - \frac{N}{L} \rho_V(y) \right) dy \right| \lesssim \left( \frac{\ln NM}{L} \right)^{3/2}. \quad (\text{C.18})$$

Similarly, for the third term in (C.9), we have

$$\left| \int_\beta^b h(y, x) \left( \rho_W(y) - \frac{N}{L} \rho_V(y) \right) dy \right| \lesssim \left( \frac{\ln NM}{L} \right)^{3/2}. \quad (\text{C.19})$$

By combining the estimates (C.9), (C.16), (C.17), (C.18) and (C.19) all together, for  $x \in [a, b]$  such that  $b - x \asymp x - a \asymp |I| \asymp L/N\sqrt{\kappa_I}$ , we have

$$\left| \rho_W(x) - \frac{N}{L} \rho_V(x) \right| \lesssim \frac{\pi}{\sqrt{(x - \alpha)(\beta - x)}} \left( \frac{\ln NM}{L} \right) \lesssim \left( \frac{\ln NM}{L} \right) \frac{N\sqrt{\kappa_I}}{L}.$$

This finishes the proof.  $\square$

*Proof of Proposition 3.19.* We prove  $|a - \alpha| = O(\ln N(M/L)^{1/3}(M/N)^{2/3})$  for the left edge case. The proof for the right edge case is exactly the same.

We recall that for the left edge case, we have  $I = [a, b]$ , where  $a = a(N)/N$  and  $b = (\ell_{L+1} - \theta)/N$ . We denote  $\kappa_I = \text{dist}(b, A)$ , then  $\kappa_I \asymp (L/N)^{2/3}$  and  $\rho_V(x) \asymp \sqrt{[x - A]_+}$ . We take derivative of both sides of (3.15) at  $x \in [\alpha, \beta]$ ,

$$W'(x) = \frac{2\theta N}{L} \left( F'_V(x) + P.V. \int_\alpha^\beta \frac{\rho_V(y)}{x - y} dy + \int_A^\alpha \frac{\rho_V(y)}{x - y} dy + \int_\beta^b \frac{\rho_V(y)}{x - y} dy + \int_{I^c} \frac{d\mu(y) - d\mu_N(y)}{x - y} \right). \quad (\text{C.20})$$

It follows by taking derivative of both sides of (3.18) at  $x \in [\alpha, \beta]$

$$W'(x) = 2\theta \left( P.V. \int_\alpha^\beta \frac{\rho_W(y)}{x - y} dy + \int_\beta^b \frac{\rho_W(y)}{x - y} dy \right), \quad (\text{C.21})$$

where we used that  $F_W(x) = 0$  on  $[\alpha, \beta]$ .

By taking difference of (C.20) and (C.21), we have

$$\begin{aligned} P.V. \int_{\alpha}^{\beta} \frac{\rho_W(y) - \frac{N}{L}\rho_V(y)}{x-y} dy &= \frac{N}{L} \int_{I^c} \frac{d\mu(y) - d\mu_N(y)}{x-y} \\ &- \int_{\beta}^b \frac{\rho_W(y) - \frac{N}{L}\rho_V(y)}{x-y} dy + \frac{N}{L} \int_A^{\alpha} \frac{\rho_V(y)}{x-y} dy + \frac{N}{L} F'_V(x). \end{aligned}$$

Thanks to the inverse Hilbert transform A.1, we get

$$\begin{aligned} \rho_W(x) - \frac{N}{L}\rho_V(x) &= \frac{1}{\pi\sqrt{(x-\alpha)(\beta-x)}} \left( \frac{N}{L} \int_b^{\infty} h(y,x) (d\mu(y) - d\mu_N(y)) \right. \\ &- \int_{\beta}^b h(y,x) \left( \rho_W(y) - \frac{N}{L}\rho_V(y) \right) dy + \frac{N}{L} \int_A^{\alpha} h(y,x) \rho_V(y) dy + \frac{N}{\pi L} P.V. \int_{\alpha}^{\beta} \frac{\sqrt{(y-\alpha)(\beta-y)}}{y-x} F'_V(y) dy \Big), \end{aligned} \quad (\text{C.22})$$

for  $x \in [\alpha, \beta]$ , where

$$h(y,x) = \frac{\sqrt{(y-\alpha)(y-\beta)}}{y-x}.$$

and we choose the branch such that  $\sqrt{(y-\alpha)(y-\beta)} \sim y$  as  $y \rightarrow \infty$ . At  $x = \alpha$ ,  $\rho_W(x) - N\rho_V(x)/L$  is a finite number. Therefore, by plugging  $x = \alpha$  in (C.22), we get

$$\begin{aligned} 0 &= \frac{N}{L} \int_b^{\infty} \sqrt{\frac{y-\beta}{y-\alpha}} (d\mu(y) - d\mu_N(y)) - \int_{\beta}^b \sqrt{\frac{y-\beta}{y-\alpha}} \left( \rho_W(y) - \frac{N}{L}\rho_V(y) \right) dy \\ &+ \frac{N}{L} \int_A^{\alpha} \sqrt{\frac{y-\beta}{y-\alpha}} \rho_V(y) dy + \frac{N}{\pi L} \int_{\alpha}^{\beta} \sqrt{\frac{\beta-y}{y-\alpha}} F'_V(y) dy. \end{aligned} \quad (\text{C.23})$$

We first estimate the first term in (C.23). For  $y \geq b$ , the behaviors of the function  $(y-\beta)^{1/2}/(y-\alpha)^{1/2}$  and its first and second derivatives is the same as those of the function  $h(y,x)$ . Therefore, by the same argument as for (C.16), we have

$$\left| \frac{N}{L} \int_b^{\infty} \sqrt{\frac{y-\beta}{y-\alpha}} (d\mu(y) - d\mu_N(y)) \right| \lesssim \frac{\ln NM}{L}. \quad (\text{C.24})$$

For the second term in (C.23), by the same argument as for (C.19), we have

$$\left| \int_{\beta}^b \sqrt{\frac{y-\beta}{y-\alpha}} \left( \rho_W(y) - \frac{N}{L}\rho_V(y) \right) dy \right| \lesssim \left( \frac{\ln NM}{L} \right)^{3/2}. \quad (\text{C.25})$$

For the last two terms in (C.23), exact one of them is nonzero. More precisely, if  $A \leq \alpha$ , the last term in (C.23) vanishes. Otherwise the third term in (C.23) vanishes. If  $A \leq \alpha$ , we have the following estimate for



the third term in (C.23),

$$\frac{N}{L} \int_A^\alpha \sqrt{\frac{y-\beta}{y-\alpha}} \rho_V(y) dy \asymp \frac{N}{L} \int_A^\alpha \left(\frac{L}{N}\right)^{1/3} \sqrt{\frac{y-A}{\alpha-y}} dy \asymp \left(\frac{N}{L}\right)^{2/3} (\alpha-A), \quad (\text{C.26})$$

where we used  $\beta-y \asymp \kappa_I \asymp (L/N)^{2/3}$  and  $\rho_V(y) \asymp \sqrt{[y-A]_+}$ . If  $\alpha \leq A$ , we have the following estimate for the last term in (C.23),

$$\frac{N}{\pi L} \int_\alpha^A \sqrt{\frac{y-\beta}{y-\alpha}} F'_V(y) dy \asymp \frac{N}{L} \int_\alpha^A \left(\frac{L}{N}\right)^{1/3} \sqrt{\frac{A-y}{y-\alpha}} dy \asymp \left(\frac{N}{L}\right)^{2/3} (A-\alpha), \quad (\text{C.27})$$

where we used  $\beta-y \asymp \kappa_I \asymp (L/N)^{2/3}$  and  $F'_V(y) \asymp \sqrt{[A-y]_+}$ .

By combining (C.23), (C.24), (C.25), (C.26), and (C.27) all together, we get

$$\left(\frac{N}{L}\right)^{2/3} |\alpha-A| \lesssim \frac{\ln NM}{L},$$

and the statement follows by rearranging.  $\square$

*Proof of Proposition 3.20.* We prove the estimate for the left edge case. The proof for the right edge case is exactly the same. It follows from taking difference of (C.22) and (C.23),

$$\begin{aligned} \rho_W(x) - \frac{N}{L} \rho_V(x) &= \frac{1}{\pi \sqrt{(x-\alpha)(\beta-x)}} \left( \frac{N}{L} \int_b^\infty \left( h(y,x) - \sqrt{\frac{y-\beta}{y-\alpha}} \right) (d\mu(y) - d\mu_N(y)) \right. \\ &- \int_\beta^b \left( h(y,x) - \sqrt{\frac{y-\beta}{y-\alpha}} \right) \left( \rho_W(y) - \frac{N}{L} \rho_V(y) \right) dy + \frac{N}{L} \int_A^\alpha \left( h(y,x) - \sqrt{\frac{y-\beta}{y-\alpha}} \right) \rho_V(y) dy \\ &+ \frac{N}{\pi L} P.V. \int_\alpha^\beta \left( \frac{\sqrt{(y-\alpha)(\beta-y)}}{y-x} - \sqrt{\frac{y-\beta}{y-\alpha}} \right) F'_V(y) dy \Big) \\ &= \frac{1}{\pi} \sqrt{\frac{x-\alpha}{\beta-x}} \left( \frac{N}{L} \int_b^\infty g(y,x) (d\mu(y) - d\mu_N(y)) - \int_\beta^b g(y,x) \left( \rho_W(y) - \frac{N}{L} \rho_V(y) \right) dy \right. \\ &\left. + \frac{N}{L} \int_A^\alpha g(y,x) \rho_V(y) dy + \frac{N}{\pi L} P.V. \int_\alpha^\beta \frac{\sqrt{\beta-y}}{(y-x)\sqrt{y-\alpha}} F'_V(y) dy \right), \end{aligned} \quad (\text{C.28})$$

where

$$g(y,x) = \frac{\sqrt{y-\beta}}{(y-x)\sqrt{y-\alpha}}.$$

We first estimate the first term on the righthand side of (C.28), which is similar to the argument for (C.16). We recall the construction from Remark 3.4. Notice that  $\text{dist}(b,A) = \kappa_I \gtrsim (L/N)^{2/3}$ , we construct

bump functions  $\chi_1, \chi_2, \dots, \chi_{n'}$  with scale  $\tilde{\eta} = M/N$ , from Remark 3.4.  $\chi_k(y)$  is supported on interval  $I_k$ , with  $|I_k| \asymp 2^k M/N \sqrt{\kappa_I}$  and  $\text{dist}(I_k, b) \asymp 2^k M/N \sqrt{\kappa_I}$ .

$$\begin{aligned} \int_{y \geq b} g(y, x) (d\mu(y) - d\mu_N(y)) &= \int_{b_0}^{b_1} (1 - \chi_1(y)) g(y, x) (d\mu(y) - d\mu_N(y)) \\ &+ \sum_{k=1}^{n'} \int_{y \geq b} \chi_k(y) g(y, x) (d\mu(y) - d\mu_N(y)). \end{aligned} \quad (\text{C.29})$$

Notice  $1 - \chi_1$  is supported on  $[b_0, b_1]$  where  $b_0 = b$  and  $b_1 = b + O(M/N \sqrt{\kappa_I})$ . For any  $y \in [b_0, b_1]$ , since  $b - \beta \lesssim M/N \sqrt{\kappa_I}$ ,  $y - \beta \lesssim M/N \sqrt{\kappa_I}$ . Also, by our assumption  $b - x \asymp \kappa_I \asymp (L/N)^{2/3}$ . It follows by combining Proposition 3.19 that  $y - x \asymp y - \alpha \asymp \kappa_I \asymp (L/N)^{2/3}$ . All these estimates together imply  $g(y, x) \lesssim (M/L)^{1/2}$  for  $y \in [b_0, b_1]$ . Moreover, the total mass of  $\mu$  and  $\mu_N$  on this interval is bounded by  $O(\ln NM/N)$ . Therefore,

$$\left| \int_{b_0}^{b_1} g(y, x) (1 - \chi_1(y)) (d\mu(y) - d\mu_N(y)) \right| \lesssim \ln N \left( \frac{M}{N} \right)^{1/3} \left( \frac{M}{L} \right)^{7/6}. \quad (\text{C.30})$$

For  $k \geq 1$ , there are two cases,  $|I_k| \leq \kappa_I$  or  $|I_k| \geq \kappa_I$ . In the first case that  $|I_k| \leq \kappa_I$ , we  $I_k \subset [b, B - r_0]$ . Moreover, if  $y \in I_k$  with  $|I_k| \leq \kappa_I$ , by our assumption  $b - x \asymp \kappa_I \asymp L/N \sqrt{\kappa_I}$ , and Proposition 3.19, we have  $y - x \asymp y - \alpha \asymp \kappa_I \asymp (L/N)^{2/3}$ , and  $y - \beta \asymp |I_k|$ . As a function of  $y$ ,  $g(y, x)$  and its derivatives satisfy

$$|g(y, x)| \lesssim |I_k|^{1/2} \kappa_I^{-3/2}, \quad |\partial_y g(y, x)| \lesssim |I_k|^{-1/2} \kappa_I^{-3/2}, \quad |\partial_y^2 g(y, x)| \lesssim |I_k|^{-3/2} \kappa_I^{-3/2}.$$

Let  $g_k(y, x) = (\kappa_I^{3/2}/|I_k|^{1/2})g(y, x)$ , then as a function of  $y$ ,  $g_k(y, x) = O(1)$ ,  $\partial_y g_k(y, x) = O(|I_k|^{-1})$  and  $\partial_y^2 g_k(y, x) = O(|I_k|^{-2})$ . By Proposition 3.2, we have

$$\begin{aligned} \sum_{k: |I_k| \leq \kappa_I} \left| \int g(y, x) \chi_k(y) (d\mu(y) - d\mu_N(y)) \right| &= \sum_{k: |I_k| \leq \kappa_I} \frac{|I_k|^{1/2}}{\kappa_I^{3/2}} \left| \int g_k(y, x) (d\mu(y) - d\mu_N(y)) \right| \\ &\lesssim \sum_{k: |I_k| \leq \kappa_I} \frac{|I_k|^{1/2} \ln NM}{\kappa_I^{3/2} N} \lesssim \frac{\ln NM}{N \kappa_I} \asymp \frac{\ln NM}{N} \left( \frac{N}{L} \right)^{2/3}. \end{aligned} \quad (\text{C.31})$$

For  $|I_k| \geq \kappa_I$ , we have  $y - x \asymp y - \alpha \asymp y - \beta \asymp |I_k|$ , and

$$|g(y, x)| \lesssim |I_k|^{-1}, \quad |\partial_y g(y, x)| \lesssim |I_k|^{-2}, \quad |\partial_y^2 g(y, x)| \lesssim |I_k|^{-3}.$$

Let  $g_k(y, x) = |I_k|g(y, x)$ , then as a function of  $y$ ,  $g_k(y, x) = O(1)$ ,  $\partial_y g_k(y, x) = O(|I_k|^{-1})$  and  $\partial_y^2 g_k(y, x) = O(|I_k|^{-2})$ . Moreover, either  $I_k \subset [b, B - r_0]$ , or  $I_k$  is not completely contained in  $[b, B - r_0]$ , in which case  $|I_k| \gtrsim 1$ . By Proposition 3.2 and 3.3, we have

$$\begin{aligned} \sum_{k: |I_k| \geq \kappa_I} \left| \int g(y, x) \chi_k(y) (d\mu(y) - d\mu_N(y)) \right| &= \sum_{k: |I_k| \geq \kappa_I} \frac{1}{|I_k|} \left| \int g_k(y, x) (d\mu(y) - d\mu_N(y)) \right| \\ &\lesssim \sum_{k: |I_k| \geq \kappa_I} \frac{\ln NM}{N |I_k|} \lesssim \frac{\ln NM}{N \kappa_I} \asymp \frac{\ln NM}{N} \left( \frac{N}{L} \right)^{2/3}. \end{aligned} \quad (\text{C.32})$$

It follows by combining (C.30), (C.31) and (C.32), and noticing  $\beta - x \asymp \kappa_I \asymp (L/N)^{2/3}$ , we have

$$\left| \frac{1}{\pi} \sqrt{\frac{x-\alpha}{\beta-x}} \frac{N}{L} \int_b^\infty g(y,x) (d\mu(y) - d\mu_N(y)) \right| \lesssim \frac{\ln NM}{L} \frac{N}{L} \sqrt{x-\alpha}. \quad (\text{C.33})$$

For the second term on the righthand side of (C.28), we have  $g(y,x) \asymp \sqrt{y-\beta}/\kappa_I^{3/2}$ . It follows from the same argument as for (C.19),

$$\left| \frac{1}{\pi} \sqrt{\frac{x-\alpha}{\beta-x}} \int_\beta^b g(y,x) \left( \rho_W(y) - \frac{N}{L} \rho_V(y) \right) dy \right| \lesssim \left( \frac{\ln NM}{L} \right)^{3/2} \frac{N}{L} \sqrt{x-\alpha}. \quad (\text{C.34})$$

For the last two terms on the righthand side of (C.28), exact one of them is nonzero. More precisely, if  $A \leq \alpha$ , the last term vanishes; if  $\alpha \leq A$ , the third term vanishes. We first exam the first case. If  $A \leq \alpha$ , we can estimate the third term in (C.28),

$$\frac{1}{\pi} \sqrt{\frac{x-\alpha}{\beta-x}} \frac{N}{L} \int_A^\alpha g(y,x) \rho_V(y) dy \lesssim \frac{N}{L} \sqrt{x-\alpha} \int_A^\alpha \frac{\sqrt{\alpha-A}}{(x-y)\sqrt{\alpha-y}} dy, \quad (\text{C.35})$$

where we used  $\beta - x, \beta - y \asymp (L/N)^{2/3}$ , and  $\rho_V(y) \asymp \sqrt{[y-A]_+} \leq \sqrt{\alpha-A}$ . For the integral in (C.35), there are two cases,

$$\int_A^\alpha \frac{1}{(x-y)\sqrt{\alpha-y}} dy \lesssim \begin{cases} \frac{1}{\sqrt{x-\alpha}}, & 0 \leq x-\alpha \leq \alpha-A, \\ \frac{\sqrt{\alpha-A}}{x-\alpha}, & x-\alpha \geq \alpha-A. \end{cases} \quad (\text{C.36})$$

(3.34) follows from (C.28), (C.33), (C.34), (C.35) and (C.36).

If  $\alpha \leq A$ , we estimate the last term in (C.28). There are two cases, if  $x \geq A$ , the last term on the righthand side of (C.28) is a normal integral,

$$\frac{1}{\pi} \sqrt{\frac{x-\alpha}{\beta-x}} \frac{N}{L} \int_\alpha^A \frac{\sqrt{\beta-y}}{(y-x)\sqrt{y-\alpha}} F'_V(y) dy \lesssim \frac{N}{L} \sqrt{x-\alpha} \int_\alpha^A \frac{\sqrt{A-\alpha}}{(x-y)\sqrt{y-\alpha}} dy,$$

where we used  $\beta - x, \beta - y \asymp (L/N)^{2/3}$ ,  $F'_V(y) \asymp \sqrt{[A-y]_+} \leq \sqrt{A-\alpha}$ . It follows (C.36),

$$\frac{1}{\pi} \sqrt{\frac{x-\alpha}{\beta-x}} \frac{N}{L} \int_\alpha^A \frac{\sqrt{\beta-y}}{(y-x)\sqrt{y-\alpha}} F'_V(y) dy \lesssim \begin{cases} \frac{N}{L} \sqrt{A-\alpha}, & 0 \leq x-A \leq A-\alpha, \\ \frac{N}{L} \frac{A-\alpha}{\sqrt{x-\alpha}}, & x-A \geq A-\alpha. \end{cases} \quad (\text{C.37})$$

If  $\alpha \leq x \leq A$ , we denote

$$F'_V(y) = -\sqrt{A-y} S_A(y), \quad t(y) = \sqrt{\frac{\beta-y}{y-\alpha}} F'_V(y).$$

where  $S_A(y)$  is an analytic function as defined in the discussion below (1.12). Then we have

$$P.V. \int_\alpha^A \frac{\sqrt{\beta-y}}{(y-x)\sqrt{y-\alpha}} F'_V(y) dy = \int_\alpha^A \frac{t(y) - t(x)}{y-x} dy + t(x) P.V. \int_\alpha^A \frac{1}{y-x} dy. \quad (\text{C.38})$$

For the second term in (C.38),

$$t(x)P.V. \int_{\alpha}^A \frac{1}{y-x} dy \asymp \left(\frac{L}{N}\right)^{1/3} \sqrt{\frac{A-x}{x-\alpha}} \ln \frac{A-\alpha}{\text{dist}(x, \{\alpha, A\})} \lesssim \left(\frac{L}{N}\right)^{1/3} \sqrt{\frac{A-\alpha}{x-\alpha}} \ln \frac{A-\alpha}{x-\alpha}, \quad (\text{C.39})$$

where we used  $\beta - x \asymp (L/N)^{2/3}$ . For the first term in (C.38), we rewrite it as the sum of two terms

$$-\sqrt{\frac{(\beta-x)(A-x)}{x-\alpha}} \int_{\alpha}^A \frac{S_A(y) - S_A(x)}{y-x} dy - S_A(x) \int_{\alpha}^A \frac{\sqrt{\frac{(\beta-y)(A-y)}{y-\alpha}} - \sqrt{\frac{(\beta-x)(A-x)}{x-\alpha}}}{y-x} dy. \quad (\text{C.40})$$

For the first term in (C.40),

$$-\sqrt{\frac{(\beta-x)(A-x)}{x-\alpha}} \int_{\alpha}^A \frac{S_A(y) - S_A(x)}{y-x} dy \lesssim \left(\frac{L}{N}\right)^{1/3} \sqrt{\frac{A-x}{x-\alpha}} \lesssim \left(\frac{L}{N}\right)^{1/3} \sqrt{\frac{A-\alpha}{x-\alpha}}, \quad (\text{C.41})$$

where we used that  $S_A$  is an analytic function, so it is Lipschitz. For the second term in (C.40), the integrand is bounded by

$$\left(\frac{N}{L}\right)^{1/3} \left( \sqrt{\frac{A-y}{y-\alpha}} + \sqrt{\frac{A-x}{x-\alpha}} \right)^{-1} + \left(\frac{L}{N}\right)^{1/3} \frac{A-\alpha}{(x-\alpha)\sqrt{(A-y)(y-\alpha)} + (y-\alpha)\sqrt{(A-x)(x-\alpha)}}, \quad (\text{C.42})$$

where we used  $\beta - \alpha \asymp \beta - x \asymp \beta - y \asymp (L/N)^{2/3}$ . For the first term in (C.42),

$$\left(\frac{N}{L}\right)^{1/3} \int_{\alpha}^A \left( \sqrt{\frac{A-y}{y-\alpha}} + \sqrt{\frac{A-x}{x-\alpha}} \right)^{-1} dy \lesssim \left(\frac{N}{L}\right)^{1/3} \sqrt{(A-\alpha)(x-\alpha)} \lesssim \left(\frac{L}{N}\right)^{1/3} \sqrt{\frac{A-\alpha}{x-\alpha}}, \quad (\text{C.43})$$

where in the last inequality, we used  $x - \alpha \leq A - \alpha \lesssim (L/N)^{2/3}$  from Proposition 3.19. For the second term in (C.42), if  $x \geq (\alpha + A)/2$ , then  $x - \alpha \asymp A - \alpha$ , and

$$\begin{aligned} & \left(\frac{L}{N}\right)^{1/3} \int_{\alpha}^A \frac{A-\alpha}{(x-\alpha)\sqrt{(A-y)(y-\alpha)} + (y-\alpha)\sqrt{(A-x)(x-\alpha)}} dy \\ & \leq \left(\frac{L}{N}\right)^{1/3} \int_{\alpha}^A \frac{A-\alpha}{(x-\alpha)\sqrt{(A-y)(y-\alpha)}} dy \lesssim \left(\frac{L}{N}\right)^{1/3}. \end{aligned} \quad (\text{C.44})$$

If  $x \leq (\alpha + A)/2$ , then  $A - x \asymp A - \alpha$ , and

$$\begin{aligned} & \left(\frac{L}{N}\right)^{1/3} \int_{\alpha}^A \frac{A-\alpha}{(x-\alpha)\sqrt{(A-y)(y-\alpha)} + (y-\alpha)\sqrt{(A-x)(x-\alpha)}} dy \\ & \lesssim \frac{A-\alpha}{\sqrt{x-\alpha}} \left(\frac{L}{N}\right)^{1/3} \int_{\alpha}^A \frac{dy}{\sqrt{y-\alpha} (\sqrt{(x-\alpha)(A-y)} + \sqrt{(y-\alpha)(A-\alpha)})} \\ & \lesssim \frac{A-\alpha}{\sqrt{x-\alpha}} \left(\frac{L}{N}\right)^{1/3} \left( \int_{\alpha}^{(\alpha+A)/2} \frac{dy}{\sqrt{y-\alpha}\sqrt{A-\alpha}(\sqrt{x-\alpha} + \sqrt{y-\alpha})} + \int_{(\alpha+A)/2}^A \frac{dy}{(A-\alpha)^{3/2}} \right) \\ & \lesssim \sqrt{\frac{A-\alpha}{x-\alpha}} \left(\frac{L}{N}\right)^{1/3} \ln \frac{A-\alpha}{x-\alpha}. \end{aligned} \quad (\text{C.45})$$

It follows from combining the above estimates all together, for  $\alpha \leq x \leq A$ ,

$$\frac{1}{\pi} \sqrt{\frac{x-\alpha}{\beta-x}} \frac{N}{L} \int_{\alpha}^A \frac{\sqrt{\beta-y}}{(y-x)\sqrt{y-\alpha}} F'_V(y) dy \lesssim \frac{N}{L} \sqrt{A-\alpha} \ln \frac{A-\alpha}{x-\alpha}. \quad (\text{C.46})$$

(3.34) follows from (C.28), (C.33), (C.34), (C.37) and (C.46). This finishes the proof.  $\square$

## D Level Repulsion and Proof of Proposition 4.3

*Proof of Proposition 4.3.* The continuous case, i.e. (4.6), is the the same statement as in [29, Lemma D.1]. We present the proof for (4.5) when  $k = L$ , which is a discrete analogue of the proof of [29, Lemma D.1]. The general case when  $k \in [1, L]$  follows by further conditioning on  $\ell_{k+1}, \ell_{k+1}, \dots, \ell_L$ .

We set  $\ell_+ := \ell_{L+1} - \theta$  and  $\ell_- := \ell_+ - a$  with  $a := N^{a+1/3} L^{-1/3}$ . By our assumption,  $(\ell_{L+1}, \ell_{L+2}, \dots, \ell_N) \in \mathcal{R}_L^*(\mathbf{a}, r)$  is a "good" boundary condition as in Definition 4.1.

We decompose the configurational space according to the number of the particles in  $[\ell_-, \ell_+]$ , which we denote by  $n$ . For any large integer  $D > 0$  (we allow that  $D = \infty$ ), we define the partition function

$$Z_D := \sum_{n=0}^L \sum_{\substack{\ell_1 < \ell_2 < \dots < \ell_{L-n} < \ell_-, \\ \ell_- \leq \ell_{L-n+1} < \dots < \ell_L \leq \ell_+ - a/2D}} Z_L^{\text{dis}} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L).$$

We remark that  $Z_{\infty} = Z_L^{\text{dis}}$ . In the following we prove that if  $a \gg DL$  and  $D \gg L^2$ , then there exists some constant  $C$  such that

$$\frac{Z_D}{Z_{\infty}} \geq \left(1 - \frac{C}{D}\right)^{CL^2}. \quad (\text{D.1})$$

Then it follows by taking  $D = \lfloor N^a/2s \rfloor$ , that if  $s \ll L^{-2}$ , we have

$$\mathbb{P}_L^{\text{dis}}(|\ell_{L+1} - \ell_L| \geq sN^{1/3}L^{-1/3}) \geq \mathbb{P}_L^{\text{dis}}(|\ell_{L+1} - \ell_L| \geq a/2D) = \frac{Z_D}{Z_{\infty}} \geq (1 - CsL^2),$$

and this finishes the proof.

For the proof of (D.1), similarly to the proof of [29, Lemma D.1], we fix  $n$  and perform a discrete change of variable which sends  $\ell_{L-n+1}, \ell_{L-n+2}, \dots, \ell_L \in [\ell_-, \ell_+ - a/2D]$  to  $\tilde{\ell}_{L-n+1}, \tilde{\ell}_{L-n+2}, \dots, \tilde{\ell}_L \in [\ell_-, \ell_+]$ . Let  $\ell_-^n = \inf\{\ell \in a(N) + (L-n+1)\theta + \mathbb{Z}_{\geq 0} : \ell \geq \ell_-\}$ , then we have

$$\begin{aligned} y_1 &:= \ell_{L-n+1} - \ell_-^n, & y_2 &:= \ell_{L-n+2} - \ell_{L-n+1} - \theta, & \dots, & & y_n &:= \ell_L - \ell_{L-1} - \theta, \\ \tilde{y}_1 &:= \tilde{\ell}_{L-n+1} - \ell_-^n, & \tilde{y}_2 &:= \tilde{\ell}_{L-n+2} - \tilde{\ell}_{L-n+1} - \theta, & \dots, & & y_n &:= \tilde{\ell}_L - \tilde{\ell}_{L-1} - \theta, \end{aligned}$$

where  $y_1, y_2, \dots, y_n, \tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n \in \mathbb{Z}_{\geq 0}$ . We define the transition kernel

$$\mathbb{K}(\{\tilde{\ell}_i\}_{L-n+1}^L, \{\ell_i\}_{L-n+1}^L) = \prod_{i=1}^n \mathbb{K}_0(\tilde{y}_i, y_i),$$

where

$$\mathbb{K}_0(\tilde{y}, y) = \begin{cases} \delta_{\tilde{y}, y} & \text{if } y = m, \text{ where } 0 \leq m \leq D-1, \\ \frac{m}{D-1} \delta_{\tilde{y}-k, y} + \frac{D-1-m}{D-1} \delta_{\tilde{y}-(k-1), y} & \text{if } \tilde{y} = kD + m, \text{ where } 1 \leq k \text{ and } 0 \leq m \leq D-1. \end{cases}$$

One can easily check:

**Claim D.1.** *If  $\tilde{\ell}_{L-n+1}, \tilde{\ell}_{L-n+2}, \dots, \tilde{\ell}_L \in [\ell_-, \ell_+]$  and  $\mathbb{K}(\{\tilde{\ell}_i\}_{L-n+1}^L, \{\ell_i\}_{L-n+1}^L) \neq 0$ , then there exists some large constants  $C$ , such that*

1.  $\ell_{L-n+1}, \ell_{L-n+2}, \dots, \ell_L \in [\ell_-, \ell_+ - a/2D]$ , provided that  $a \geq CDL$ ;
2.  $\sum_{\tilde{\ell}_{L-n+1}, \tilde{\ell}_{L-n+2}, \dots, \tilde{\ell}_L \in [\ell_-, \ell_+]} \mathbb{K}(\{\tilde{\ell}_i\}_{L-n+1}^L, \{\ell_i\}_{L-n+1}^L) \leq (1 + C/D)^n$ ;
3. For  $L-n+1 \leq i \leq L$  and  $j > L$ , we have  $\ell_j - \tilde{\ell}_i \leq \ell_j - \ell_i$ ;
4. For  $L-n+1 \leq i \leq L$ ,  $|\tilde{\ell}_i - \ell_i| \leq C|\ell_i - \ell^n|/D \leq Ca/D$ ;
5. For  $L-n+1 \leq i < j \leq L$ ,  $|\tilde{\ell}_j - \tilde{\ell}_i| \leq (1 + C/D)|\ell_j - \ell_i|$ ;
6. For  $1 \leq i < L-n+1 \leq j \leq L$ ,  $|\tilde{\ell}_j - \ell_i| \leq (1 + C/D)|\ell_j - \ell_i|$ .

Using (3), (4), (5) and (6) in Claim D.1, following the same argument as in the proof of [29, Lemma D.1], we can compare the densities and find that

$$\begin{aligned} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_{L-n}, \tilde{\ell}_{L-n+1}, \dots, \tilde{\ell}_L) &\leq (1 + C/D)^{C(n^2 + n(L-n))} e^{CnN^{2a}/D} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L) \\ &\leq (1 + C/D)^{CL^2} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L), \end{aligned} \quad (\text{D.2})$$

provided that  $D \gg N^{2a}$ .

For simplification of notations, we denote  $\{\ell_i\} = \{\ell_{L-n+1}, \ell_{L-n+2}, \dots, \ell_L\}$ , and  $\{\tilde{\ell}_i\} = \{\tilde{\ell}_{L-n+1}, \tilde{\ell}_{L-n+2}, \dots, \tilde{\ell}_L\}$ . All the discussions above together lead to the following estimate for  $Z_\infty$ ,

$$\begin{aligned} &\sum_{\substack{\ell_1 < \ell_2 < \dots < \ell_{L-n} < \ell_-, \\ \ell_- \leq \tilde{\ell}_{L-n+1} < \dots < \tilde{\ell}_L \leq \ell_+}} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_{L-n}, \tilde{\ell}_{L-n+1}, \dots, \tilde{\ell}_L) \\ &= \sum_{\substack{\ell_1 < \ell_2 < \dots < \ell_{L-n} < \ell_-, \\ \ell_- \leq \tilde{\ell}_{L-n+1} < \dots < \tilde{\ell}_L \leq \ell_+}} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_{L-n}, \tilde{\ell}_{L-n+1}, \dots, \tilde{\ell}_L) \sum_{\{\ell_i\}} \mathbb{K}(\{\tilde{\ell}_i\}, \{\ell_i\}) \\ &\leq (1 + C/D)^{CL^2} \sum_{\substack{\ell_1 < \ell_2 < \dots < \ell_{L-n} < \ell_-, \\ \ell_- \leq \tilde{\ell}_{L-n+1} < \dots < \tilde{\ell}_L \leq \ell_+}} \sum_{\{\ell_i\}} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L) \mathbb{K}(\{\tilde{\ell}_i\}, \{\ell_i\}) \\ &\leq (1 + C/D)^{CL^2} \sum_{\substack{\ell_1 < \ell_2 < \dots < \ell_{L-n} < \ell_-, \\ \ell_- \leq \tilde{\ell}_{L-n+1} < \dots < \tilde{\ell}_L \leq \ell_+ - a/2D}} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L) \sum_{\ell_- \leq \tilde{\ell}_{L-n+1} < \dots < \tilde{\ell}_L \leq \ell_+} \mathbb{K}(\{\tilde{\ell}_i\}, \{\ell_i\}) \\ &\leq (1 + C/D)^{CL^2 + n} \sum_{\substack{\ell_1 < \ell_2 < \dots < \ell_{L-n} < \ell_-, \\ \ell_- \leq \tilde{\ell}_{L-n+1} < \dots < \tilde{\ell}_L \leq \ell_+ - a/2D}} \mathbb{P}_L^{\text{dis}}(\ell_1, \ell_2, \dots, \ell_L), \end{aligned} \quad (\text{D.3})$$

where in the first equality we used that  $\mathbb{K}$  is a transition kernel, in the third line we used (D.2), and in the last line we used (2) in Claim D.1. (D.1) follows by summing over  $n$  from 0 to  $L$  in (D.3), provided that

$a \gg DL$ ,  $D \gg L^2$ ,  $D \gg N^{2\alpha}$  and  $s \ll L^{-2}$ . With a little algebra, they are satisfied if  $L^{4/3}N^{-1/3} \ll s \ll \min\{L^{-2}, N^{-\alpha}\}$ .

The estimate (4.3) follows the previous proof combining with the rigidity estimates assumption, i.e. with probability  $1 - \exp(-c(\ln N)^2)$ , with respect to  $\mathbb{P}_L^{\text{dis}}$ , the number of particles on the interval  $[\ell_-, \ell_+]$  is bounded by  $CN^\alpha$ .

□

## References

- [1] J. Baik, T. Kriecherbauer, K. D. T.-R. McLaughlin, and P. D. Miller. Uniform asymptotics for polynomials orthogonal with respect to a general class of discrete weights and universality results for associated ensembles: announcement of results. *Int. Math. Res. Not.*, (15):821–858, 2003.
- [2] J. Baik, T. Kriecherbauer, K. T.-R. McLaughlin, and P. D. Miller. *Discrete orthogonal polynomials*, volume 164 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2007. Asymptotics and applications.
- [3] Jinho Baik, Percy Deift, and Kurt Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.
- [4] Jinho Baik and Eric M. Rains. The asymptotics of monotone subsequences of involutions. *Duke Math. J.*, 109(2):205–281, 2001.
- [5] Roland Bauerschmidt, Bourgade Paul, Miika Nikula, and Yau Horng-Tzer. Local density for two-dimensional one-component plasma. *preprint, arXiv: 1510.02074*.
- [6] F. Bekerman, A. Figalli, and A. Guionnet. Transport maps for  $\beta$ -matrix models and universality. *Comm. Math. Phys.*, 338(2):589–619, 2015.
- [7] F. Bekerman and A. Lodhia. Mesoscopic central limit theorem for general  $\beta$ -ensembles. *arXiv:1605.05206*, 2016.
- [8] Florent Bekerman. Transport maps for  $\beta$ -matrix models in the multi-cut regime. *preprint, arXiv: 1605.05206*, 2016.
- [9] G. Ben Arous and A. Guionnet. Large deviations for Wigner’s law and Voiculescu’s non-commutative entropy. *Probab. Theory Related Fields*, 108(4):517–542, 1997.
- [10] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, 2012. Anniversary edition [of MR1324786], With a foreword by Steve Lalley and a brief biography of Billingsley by Steve Koppes.
- [11] Pavel Bleher and Alexander Its. Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and universality in the matrix model. *Ann. of Math. (2)*, 150(1):185–266, 1999.
- [12] Alexei Borodin. Schur dynamics of the Schur processes. *Adv. Math.*, 228(4):2268–2291, 2011.
- [13] Alexei Borodin and Ivan Corwin. Discrete time  $q$ -TASEPs. *Int. Math. Res. Not. IMRN*, (2):499–537, 2015.

- [14] Alexei Borodin, Ivan Corwin, and Daniel Remenik. Log-gamma polymer free energy fluctuations via a Fredholm determinant identity. *Comm. Math. Phys.*, 324(1):215–232, 2013.
- [15] Alexei Borodin, Ivan Corwin, and Tomohiro Sasamoto. From duality to determinants for  $q$ -TASEP and ASEP. *Ann. Probab.*, 42(6):2314–2382, 2014.
- [16] Alexei Borodin and Patrik L. Ferrari. Anisotropic growth of random surfaces in  $2 + 1$  dimensions. *Comm. Math. Phys.*, 325(2):603–684, 2014.
- [17] Alexei Borodin, Vadim Gorin, and Alice Guionnet. Gaussian asymptotics of discrete  $\beta$ -ensembles. *Publications mathématiques de l’IHÉS*, pages 1–78, 2016.
- [18] Alexei Borodin, Andrei Okounkov, and Grigori Olshanski. Asymptotics of Plancherel measures for symmetric groups. *J. Amer. Math. Soc.*, 13(3):481–515, 2000.
- [19] Alexei Borodin, Andrei Okounkov, and Grigori Olshanski. Asymptotics of Plancherel measures for symmetric groups. *J. Amer. Math. Soc.*, 13(3):481–515, 2000.
- [20] Alexei Borodin and Grigori Olshanski.  $z$ -measures on partitions, Robinson-Schensted-Knuth correspondence, and  $\beta = 2$  random matrix ensembles. In *Random matrix models and their applications*, volume 40 of *Math. Sci. Res. Inst. Publ.*, pages 71–94. Cambridge Univ. Press, Cambridge, 2001.
- [21] Alexei Borodin and Grigori Olshanski.  $Z$ -measures on partitions and their scaling limits. *European J. Combin.*, 26(6):795–834, 2005.
- [22] Alexei Borodin and Grigori Olshanski. Asymptotics of Plancherel-type random partitions. *J. Algebra*, 313(1):40–60, 2007.
- [23] Alexei Borodin and Leonid Petrov. Integrable probability: from representation theory to Macdonald processes. *Probab. Surv.*, 11:1–58, 2014.
- [24] G. Borot, V. Gorin, and A. Guionnet. Fluctuations for multi-cut discrete  $\beta$ -ensembles and application to random tilings. *in preparation*, 2017.
- [25] G. Borot and A. Guionnet. Asymptotic Expansion of  $\beta$  Matrix Models in the multi-cut Regime. *arXiv:1303.1045*, 2013.
- [26] G. Borot and A. Guionnet. Asymptotic Expansion of  $\beta$  Matrix Models in the one-cut Regime. *Comm. Math. Phys.*, 317(2):447–483, 2013.
- [27] Gaëtan Borot, Alice Guionnet, and Karol K. Kozłowski. Large- $N$  asymptotic expansion for mean field models with Coulomb gas interaction. *Int. Math. Res. Not. IMRN*, (20):10451–10524, 2015.
- [28] Paul Bourgade, László Erdős, and Horng-Tzer Yau. Bulk universality of general  $\beta$ -ensembles with non-convex potential. *J. Math. Phys.*, 53(9):095221, 19, 2012.
- [29] Paul Bourgade, László Erdős, and Horng-Tzer Yau. Edge universality of beta ensembles. *Comm. Math. Phys.*, 332(1):261–353, 2014.
- [30] Paul Bourgade, László Erdős, and Horng-Tzer Yau. Universality of general  $\beta$ -ensembles. *Duke Math. J.*, 163(6):1127–1190, 2014.
- [31] Jonathan Breuer and Maurice Duits. Central limit theorems for biorthogonal ensembles and asymptotics of recurrence coefficients. *J. Amer. Math. Soc.*, 30(1):27–66, 2017.



- [32] A. Bufetov and V. Gorin. Fluctuations of particle systems determined by schur generating functions. *arXiv:1604.01110*, 2016.
- [33] Henry Cohn, Michael Larsen, and James Propp. The shape of a typical boxed plane partition. *New York J. Math.*, 4:137–165, 1998.
- [34] P. Deift and D. Gioev. *Random matrix theory: invariant ensembles and universality*, volume 18 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York, 2009.
- [35] P. Deift, T. Kriecherbauer, K. T.-R. McLaughlin, S. Venakides, and X. Zhou. Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.*, 52(11):1335–1425, 1999.
- [36] P.A. Deift and D. Gioev. Universality at the edge of the spectrum for unitary, orthogonal and symplectic ensembles of random matrices. *Comm. Pure Appl. Math.*, 60:867–910, 2007.
- [37] P.A. Deift and D. Gioev. Universality in random matrix theory for orthogonal and symplectic ensembles. *Int. Math. Res. Pap.*, pages ArtID rpm007, 116pp., 2007.
- [38] Percy A. Deift, Alexander R. Its, and Xin Zhou. A Riemann-Hilbert approach to asymptotic problems arising in the theory of random matrix models, and also in the theory of integrable statistical mechanics. *Ann. of Math. (2)*, 146(1):149–235, 1997.
- [39] Maciej Dołęga and Valentin Féray. Gaussian fluctuations of Young diagrams and structure constants of Jack characters. *Duke Math. J.*, 165(7):1193–1282, 2016.
- [40] P. D. Dragnev and E. B. Saff. Constrained energy problems with applications to orthogonal polynomials of a discrete variable. *J. Anal. Math.*, 72:223–259, 1997.
- [41] Ioana Dumitriu and Alan Edelman. Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830–5847, 2002.
- [42] Rick Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010.
- [43] Freeman J. Dyson. Correlations between eigenvalues of a random matrix. *Comm. Math. Phys.*, 19:235–250, 1970.
- [44] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.*
- [45] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.
- [46] László Erdős, José A. Ramírez, Benjamin Schlein, and Horng-Tzer Yau. Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation. *Electron. J. Probab.*, 15:no. 18, 526–603, 2010.
- [47] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287(2):641–655, 2009.
- [48] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

- [49] László Erdős, Horng-Tzer Yau, and Jun Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.
- [50] László Erdős, Horng-Tzer Yau, and Jun Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.*, 229(3):1435–1515, 2012.
- [51] D. Féral. On large deviations for the spectral measure of discrete Coulomb gas. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 19–49. Springer, Berlin, 2008.
- [52] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.
- [53] Jason Fulman. Stein’s method, Jack measure, and the Metropolis algorithm. *J. Combin. Theory Ser. A*, 108(2):275–296, 2004.
- [54] V. E. Gorin. Nonintersecting paths and the Hahn orthogonal polynomial ensemble. *Funktsional. Anal. i Prilozhen.*, 42(3):23–44, 96, 2008.
- [55] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.
- [56] Kurt Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.
- [57] Kurt Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math. (2)*, 153(1):259–296, 2001.
- [58] Kurt Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math. (2)*, 153(1):259–296, 2001.
- [59] Kurt Johansson. Non-intersecting paths, random tilings and random matrices. *Probab. Theory Related Fields*, 123(2):225–280, 2002.
- [60] Kurt Johansson. The arctic circle boundary and the Airy process. *Ann. Probab.*, 33(1):1–30, 2005.
- [61] Antti Knowles and Jun Yin. Eigenvector distribution of Wigner matrices. *Probab. Theory Related Fields*, 155(3-4):543–582, 2013.
- [62] Wolfgang König, Neil O’Connell, and Sébastien Roch. Non-colliding random walks, tandem queues, and discrete orthogonal polynomial ensembles. *Electron. J. Probab.*, 7:no. 5, 24, 2002.
- [63] T. Kriecherbauer and M. Shcherbina. Fluctuations of eigenvalues of matrix models and their applications. *arXiv:1003.6121*, 2010.
- [64] Manjunath Krishnapur, Brian Rider, and Bálint Virág. Universality of the stochastic Airy operator. *Comm. Pure Appl. Math.*, 69(1):145–199, 2016.
- [65] A. B. J. Kuijlaars and K. T-R McLaughlin. Generic behavior of the density of states in random matrix theory and equilibrium problems in the presence of real analytic external fields. *Comm. Pure Appl. Math.*, 53(6):736–785, 2000.
- [66] Benjamin Landon and Horng-Tzer Yau. Convergence of local statistics of Dyson Brownian motion. *to appear in Comm. Math. Phys.*, 2014.
- [67] Thomas Leblé. Local microscopic behavior for 2d coulomb gases. *to appear in Probab. Theory Related Fields*, 2015.

- [68] Ji Oon Lee and Kevin Schnelli. Local law and tracy-widom limit for sparse random matrices. *To appear in Annales Henri Poincaré*, 2016.
- [69] Ji Oon Lee, Kevin Schnelli, Ben Stetler, and Horng-Tzer Yau. Bulk universality for deformed Wigner matrices. *Ann. Probab.*, 44(3):2349–2425, 2016.
- [70] Ji Oon Lee and Jun Yin. A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.*, 163(1):117–173, 2014.
- [71] E. Levin and D. S. Lubinsky. Universality limits in the bulk for varying measures. *Adv. Math.*, 219(3):743–779, 2008.
- [72] Yiting Li. Rigidity of eigenvalues for  $\beta$  ensemble in multi-cut regime. *preprint, arXiv: 1611.06603*.
- [73] W. Lindenbergh. Eine Klasse von  $p$ -Gruppen, deren Automorphismengruppen  $p$ -Gruppen sind. *Period. Math. Hungar.*, 5:171–183, 1974.
- [74] Sho Matsumoto. Jack deformations of Plancherel measures and traceless Gaussian random matrices. *Electron. J. Combin.*, 15(1):Research Paper 149, 18, 2008.
- [75] Sho Matsumoto. Jucys-Murphy elements, orthogonal matrix integrals, and Jack measures. *Ramanujan J.*, 26(1):69–107, 2011.
- [76] M. L. Mehta. A note on correlations between eigenvalues of a random matrix. *Comm. Math. Phys.*, 20:245–250, 1971.
- [77] M. L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.
- [78] M. L. Mehta and M. Gaudin. On the density of eigenvalues of a random matrix. *Nuclear Phys.*, 18:420–427, 1960.
- [79] N. Nekrasov. Bps/cft correspondence: non-perturbative dyson-schwinger equations and qq-characters. *arXiv:1512.05388*.
- [80] N. Nekrasov and V Pestun. Seiberg-witten geometry of four dimensional  $n = 2$  quiver gauge theories. *arXiv:1211.2240*.
- [81] N. Nekrasov, V Pestun, and S. Shatashvili. Quantum geometry and quiver gauge theories. *High Energy Physics - Theory*, pages 1–83, 2013.
- [82] Andrei Okounkov. Random matrices and random permutations. *Internat. Math. Res. Notices*, (20):1043–1095, 2000.
- [83] Andrei Okounkov. The uses of random partitions. In *XIVth International Congress on Mathematical Physics*, pages 379–403. World Sci. Publ., Hackensack, NJ, 2005.
- [84] Andrei Okounkov and Nikolai Reshetikhin. Correlation function of Schur process with application to local geometry of a random 3-dimensional Young diagram. *J. Amer. Math. Soc.*, 16(3):581–603, 2003.
- [85] G. I. Olshanski. Probability measures on dual objects to compact symmetric spaces, and hypergeometric identities. *Funktsional. Anal. i Prilozhen.*, 37(4):49–73, 95, 2003.
- [86] Grigori Olshanski. Plancherel averages: remarks on a paper by Stanley. *Electron. J. Combin.*, 17(1):Research Paper 43, 16, 2010.

- [87] L. Pastur and M. Shcherbina. Universality of the local eigenvalue statistics for a class of unitary invariant matrix ensembles. *J. Stat. Phys.*, 86:109–147, 1997.
- [88] L. Pastur and M. Shcherbina. Bulk universality and related properties of Hermitian matrix models. *J. Stat. Phys.*, 130(2):205–250, 2008.
- [89] Leonid Petrov. Asymptotics of random lozenge tilings via Gelfand-Tsetlin schemes. *Probab. Theory Related Fields*, 160(3-4):429–487, 2014.
- [90] Leonid Petrov. Asymptotics of uniformly random lozenge tilings of polygons. Gaussian free field. *Ann. Probab.*, 43(1):1–43, 2015.
- [91] J. Ramirez, B. Rider, and B. Virag. Beta ensembles, stochastic airy spectrum, and a diffusion. *J. Amer. Math. Soc.*, 24:919–944, 2011.
- [92] Timo Seppäläinen. Scaling for a one-dimensional directed polymer with boundary conditions. *Ann. Probab.*, 40(1):19–73, 2012.
- [93] M. Shcherbina. On universality for orthogona ensembles of random matrices and their applications. *Commun. Math. Phys.*, 285:957–974, 2009.
- [94] M. Shcherbina. Fluctuations of linear eigenvalue statistics of  $\beta$  matrix models in the multi-cut regime. *arxiv.org/abs/1205.7062*,, 2012.
- [95] M. Shcherbina. Change of variables as a method to study general  $\beta$ -models: bulk universality. *arXiv:1310.7835*, 2013.
- [96] Alexander Soshnikov. Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.*, 207(3):697–733, 1999.
- [97] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. pages 583–602, 1972.
- [98] Terence Tao and Van Vu. Random matrices: universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.*, 298(2):549–572, 2010.
- [99] Terence Tao and Van Vu. Random matrices: sharp concentration of eigenvalues. *Random Matrices Theory Appl.*, 2(3):1350007, 31, 2013.
- [100] Craig A. Tracy and Harold Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.
- [101] Craig A. Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177(3):727–754, 1996.
- [102] Craig A. Tracy and Harold Widom. Asymptotics in ASEP with step initial condition. *Comm. Math. Phys.*, 290(1):129–154, 2009.
- [103] F. G. Tricomi. *Integral equations*. Dover Publications, Inc., New York, 1985. Reprint of the 1957 original.
- [104] Benedek Valkó and Bálint Virág. Continuum limits of random matrices and the Brownian carousel. *Invent. Math.*, 177(3):463–508, 2009.