



# Point Processes, Hole Events, and Large Deviations: Random Complex Zeros and Coulomb Gases

Subhroshekhar Ghosh<sup>1</sup> · Alon Nishry<sup>2</sup>

Received: 10 February 2017 / Revised: 12 October 2017 / Accepted: 20 October 2017 /  
Published online: 7 March 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

**Abstract** We consider particle systems (also known as point processes) on the line and in the plane and are particularly interested in “hole” events, when there are no particles in a large disk (or some other domain). We survey the extensive work on hole probabilities and the related large deviation principles (LDP), which has been undertaken mostly in the last two decades. We mainly focus on the recent applications of LDP-inspired techniques to the study of hole probabilities and the determination of the most likely configurations of particles that have large holes. As an application of this approach, we illustrate how one can confirm some of the predictions of Jancovici, Lebowitz, and Manificat for large fluctuation in the number of points for the (two-dimensional)  $\beta$ -Ginibre ensembles. We also discuss some possible directions for future investigations.

**Keywords** Point processes · Particle systems · Coulomb gases · Random matrices · Random polynomials · Hole probabilities · Large deviations · Empirical measures

**Mathematics Subject Classification** Primary 60G55 · Secondary 60F10

---

Communicated by Sylvia Serfaty.

✉ Subhroshekhar Ghosh  
subhrowork@gmail.com

Alon Nishry  
alonish@tauex.tau.ac.il

<sup>1</sup> National University of Singapore, Singapore, Singapore

<sup>2</sup> Tel Aviv University, Tel Aviv, Israel

## 1 Introduction

Random point configurations, also known as point processes, have been an object of key interest in the last few decades - both in probability theory and in the statistical physics literature. The most extensive results have been obtained in Euclidean spaces of dimensions 1 and 2, although higher dimensions and other geometries have also been studied.

A point process  $\Pi$ , usually defined to live on a Polish space  $\Sigma$  equipped with a regular Borel measure  $\mu$ , is a probability distribution over the space of locally finite point configurations on  $\Sigma$ . We recall here that a Polish space is a separable and completely metrizable topological space. It is well known ([20, Chap. 9]) that, under mild conditions, the statistical behavior of a point process is described by its various *k-point intensities* ( $k = 1, 2, \dots$ ), which are roughly the joint probability densities of having particles at  $k$  specified locations in  $\Sigma$ . For almost all interesting point processes, these  $k$  point intensities are absolutely continuous with respect to  $\mu^{\otimes k}$  (referred to as the *background measure*), and the resulting Radon Nikodym derivatives are known as the  $k$  point intensity functions. Often, in Euclidean spaces or other homogeneous spaces, key point processes exhibit *invariance*, which is to say that the law of the process is invariant under the isometries of  $\Sigma$ .

The most fundamental example of a point process is the Poisson point process. The Poisson point process, defined on the space  $\Sigma$  with respect to the background measure  $\mu$ , is the unique point process on  $\Sigma$  that exhibits statistical independence of its point configurations in disjoint domains, with the particle count in a domain  $D \subset \Sigma$  obeying a Poisson distribution with mean  $\mu(D)$ . This characterizing property of spatial independence makes many important statistical properties easy to compute, which is the reason behind the popularity of the Poisson process as a probabilistic model for many real-world systems ([20, Chap. 2]). At the same time, it renders the Poisson process ineffective in modeling many natural phenomena, particularly those involving local repulsion, like electron systems.

In this context, several *natural* models have emerged that embody nontrivial spatial correlation, including local repulsion, and at the same time are amenable to analysis. These models often have their origin in statistical physics, principal among them being Coulomb systems. The Coulomb system of size  $n$  in dimension  $d$  and inverse temperature  $\beta$  is given by the joint distribution

$$p(x_1, \dots, x_n) = Z_n^{-1} \exp \left( \frac{1}{2} \beta \left[ \sum_{i \neq j} \rho(|x_i - x_j|) - n \sum_{i=1}^n V(x_i) \right] \right),$$

where  $\rho$  is the fundamental solution to the Laplacian in  $d$  dimensions (in particular, the logarithm function in 2 dimensions), and  $V$  is an external field (or confining potential). It is also of considerable interest to consider a similar system in 1D, with  $\rho$  the logarithm function; this model (or rather its infinite particle limit) is popularly known as the Dyson log gas.

In 1 and 2 dimensions, at inverse temperature  $\beta = 2$ , the Coulomb system with logarithmic interactions (a.k.a. *Dyson log gas* in 1D) is known to be a *determinantal*

point process, meaning that its correlation functions are given by certain determinants. When  $V(x) = |x|^2$ , these ensembles can also be described as the set of eigenvalues of certain random matrices - in 1 dimension it is the Gaussian Wigner matrix (GUE), and in 2 dimensions the Ginibre ensemble (having independent standard complex Gaussian entries). Both of these ensembles have well-defined weak limits that are determinantal point processes with infinitely many particles. In 1 dimension, the fundamental solution to the Laplacian is  $f(x) = |x|$ , and it is natural to consider a Coulomb system with this interaction potential. This system has been extensively studied by Aizenmann, Lebowitz, Martin, Yalcin, and others (see, e.g., [2,3,42,47] for some of the delicate results on this model).

Another important two-dimensional model is the zero set of the standard Gaussian Entire Function (henceforth abbreviated as GEF). The GEF is given by the Gaussian Taylor series

$$F(z) = \sum_{k=0}^{\infty} \frac{\xi_k}{\sqrt{k!}} z^k, \tag{1}$$

where the  $\xi_k$ -s are independent standard complex Gaussians. We mention that the truncation of this series at degree  $n$  is called the Weyl polynomial of degree  $n$ . The studies of the GEF and the Weyl polynomials have their origin in statistical physics, where they have been investigated in the context of quantum chaotic dynamics ([13]).

An object of key interest in the study of point processes is the “hole” event  $\mathcal{H}_R$ , entailing that a large disk (or interval, according to the dimension) of radius  $R$  around the origin does not contain any particles. Of course, this is a *rare event*, and  $\mathbb{P}[\mathcal{H}_R] \rightarrow 0$  as  $R \rightarrow \infty$ . The quantitative asymptotics of how this decay takes place throws light on the statistical structure of the point process, and has been studied in fine detail for many key processes. A closely related but much less understood question pertains to what causes such a large “hole” to appear. This involves understanding the typical configuration of particles outside the “hole”, and until recently such results were available only for  $\beta = 2$  Coulomb systems in 1 and 2 dimensions ([40]). Very recently, progress has been made on this front for the GEF zeros process, as well as for holes of general shapes for the Ginibre ensemble. This is based on large deviation techniques, which brings us to the third key object in this paper, namely large deviation principles (abbreviated henceforth as LDPs).

Roughly speaking, a sequence of random variables  $X_n$ , defined on a common Polish space  $\Sigma$ , is said to satisfy an LDP with rate  $a_n \uparrow \infty$ , and rate function  $I: \Sigma \rightarrow \mathbb{R}_+$ , if, for any ‘nice’ set  $\mathcal{F} \subset \Sigma$ , we have

$$\mathbb{P}[X_n \in \mathcal{F}] \approx \exp\left(-a_n \inf_{x \in \mathcal{F}} I(x)\right), \quad n \rightarrow \infty.$$

In the above display,  $\approx$  is understood in the sense that  $\frac{\log \mathbb{P}[X_n \in \mathcal{F}]}{a_n} \rightarrow -\inf_{x \in \mathcal{F}} I(x)$  as  $n \rightarrow \infty$ . For us, the most interesting case is when the random variables  $X_n$  are *empirical measures* of the points (that is, discrete counting measures of the points, normalized to be probability measures), and  $\Sigma$  is a space of probability measures on

$\mathbb{R}^d$ . Large deviation principles of this type have been extensively studied for various random matrix models (see, e.g., [8, 9, 35, 40]) in the last two decades. More recently, large deviation principles have been understood for several random polynomial models (see, e.g., [16, 32, 67, 68]).

In [1, 33] the main ingredient of the approach to the “hole configuration” is to consider the “hole” event as a “rare” event in the setting of the LDP for the relevant matrix or polynomial ensemble. This intuitively leads to the conclusion that the (limiting) intensity measure of the particles outside the hole must be the minimizer of the large deviation rate functional, under the constraint of the existence of the hole. This approach seems to be rather promising in investigating related problems for point processes. We provide more details on this approach in Sect. 6, where we study the two-dimensional  $\beta$ -Ginibre ensembles (also known as jellium or the one-component plasma).

The main thrust of this work is on a certain set of ideas that tie together point processes, large deviations, and the study of the hole event. Such focus naturally leaves out several important strands of work related to various combinations of these concepts. For instance, we mention the recent series of works studying various fine properties of the large deviation principle for Coulomb systems. In particular, these works establish rigorous connections of the LDP to the concept of *renormalized energy* ([6, 14, 43, 45, 57, 61, 62]). Another direction of recent investigations involves the study of spatial *rigidity* structures that arise in several of these natural models ([10–12, 28–30, 34, 55]). Beyond that, there is the extensive research on universality in random matrix ensembles (see, e.g., [24, 65]). We will not pursue these matters here.

## 2 Large Deviations for Empirical Measures

A sequence of random variables  $X_n$ , defined on a common Polish space  $\Sigma$ , is said to satisfy an LDP with rate  $a_n \uparrow \infty$  and a convex lower semi-continuous rate function  $I: \Sigma \rightarrow \mathbb{R}_+$  if, for any Borel measurable set  $\mathcal{F} \subset \Sigma$ , we have

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}[X_n \in \mathcal{F}] \leq - \inf_{x \in \mathcal{F}^o} I(x),$$

where  $\mathcal{F}^o$  is the interior of the set  $\mathcal{F}$  and

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}[X_n \in \mathcal{F}] \geq - \inf_{x \in \overline{\mathcal{F}}} I(x),$$

where  $\overline{\mathcal{F}}$  is the topological closure of the set  $\mathcal{F}$ .

**Definition 2.1** A rate function  $I: \Sigma \rightarrow \mathbb{R}_+$  is *good* if all its level sets  $\{x: I(x) \leq \alpha\}$  are compact subsets of  $\Sigma$ .

### 2.1 Eigenvalues of Random Matrices

Large deviations for empirical measures of random matrices have been studied by multiple authors. In this section, we will only focus on LDPs for the empirical measure of some specific families of random matrices, including Gaussian (and other unitarily invariant) Hermitian ensembles in 1D, and the (real and complex) Ginibre ensemble in 2D. We direct readers interested in a more extensive survey, including dynamical aspects related to evolution under the Dyson Brownian motion, to [35].

To describe the results, we need to introduce some notation. We will denote by  $\mathcal{M}_1(\mathbb{R})$  and  $\mathcal{M}_1(\mathbb{C})$ , respectively, the space of probability measures on  $\mathbb{R}$  and  $\mathbb{C}$ . For a finite set of points  $\Lambda := \{z_1, \dots, z_N\}$ , we define the empirical measure

$$\mathcal{E}(\Lambda) := \frac{1}{N} \sum_{i=1}^N \delta_{z_i},$$

where  $\delta_\lambda$  is the delta measure (of unit mass) at the point  $\lambda$ . The empirical measures of eigenvalues live in the space  $\mathcal{M}_1(\mathbb{R})$  (or  $\mathcal{M}_1(\mathbb{C})$  as appropriate) and, for “nice enough” ensembles, obey LDPs in the same space.

#### 2.1.1 The Ginibre Ensemble

We begin with the LDP for the Ginibre ensemble. For the original paper, we refer the reader to [9]. The (real or complex) Ginibre ensemble (of order  $n$ ) is the ensemble of eigenvalues of  $n \times n$  random matrices with i.i.d. Gaussian entries (resp., real or complex) with mean zero and variance  $n^{-1}$ . The (infinite) Ginibre ensemble is the limit, in distribution, of the finite Ginibre ensembles.

In the complex case, the (infinite) Ginibre ensemble turns out to be the 2D Coulomb gas at inverse temperature  $\beta = 2$ . It also turns out to be a determinantal point process on  $\mathbb{C}$  with kernel  $K(z, w) = \exp(z\bar{w})$  and background measure  $d\mu(z) = \pi^{-1}e^{-|z|^2}dm(z)$ , i.e. the standard Gaussian measure. Its distribution is invariant under the rigid motions of the plane, and it serves as a crucial example of an (invariant) 2D point process that is relevant to the physical literature. The finite  $n$  joint density, also known as the density of states in the physical literature, is given by

$$\varrho(z_1, \dots, z_n) = |\Delta(z_1, \dots, z_n)|^2 \exp\left(-\sum_{i=1}^n |z_i|^2\right), \tag{2}$$

where

$$\Delta(z_1, \dots, z_n) = \prod_{j < k} (z_j - z_k)$$

denotes the Vandermonde determinant.

In what follows, let  $S = \mathbb{R}$  or  $\mathbb{C}$ , with the particular value being specified by the context. Let  $\mathcal{M}_1^{sym}(\mathbb{C})$  denote the space of probability measures on  $\mathbb{C}$  that are

symmetric about the real axis. Define the *logathmic potential* of a measure  $\mu \in \mathcal{M}_1(\mathbb{C})$ :

$$U_\mu(z) = \int \log |z - w| \, d\mu(w).$$

The *logarithmic energy* of  $\mu$  is given by

$$\Sigma(\mu) = \int U_\mu(z) \, d\mu(z) = \iint \log |z - w| \, d\mu(z) d\mu(w).$$

Note that  $\Sigma(\mu)$  can equal  $-\infty$ , e.g., in the case of measures  $\mu$  having an atomic component. Define also the functionals  $I^{[S]} : \mathcal{M}_1(\mathbb{C}) \mapsto [0, \infty]$  as

$$I^{[\mathbb{R}]}(\mu) = \begin{cases} \frac{1}{2} \int |z|^2 d\mu(z) - \frac{1}{2} \Sigma(\mu) - 3/8 & \text{if } \mu \in \mathcal{M}_1^{\text{sym}}(\mathbb{C}), \\ \infty & \text{otherwise,} \end{cases}$$

and

$$I^{[\mathbb{C}]}(\mu) = \int |z|^2 d\mu(z) - \Sigma(\mu) - 3/4.$$

We can state the LDP for the (real or complex) Ginibre ensemble (denoted resp.  $\mathcal{G}_n^{[\mathbb{R}]}$  or  $\mathcal{G}_n^{[\mathbb{C}]}$ ) as follows:

**Theorem 2.1** ([9, 37]) *The sequence of empirical measures  $\mathcal{E}(\mathcal{G}_n^{[S]})$  obey a large deviation principle in the space  $\mathcal{M}_1(\mathbb{C})$  with rate  $n^2$  and good rate function  $I^{[S]}$ .*

The rate function  $I^{[S]}$  is minimized by the uniform measure on the unit disk. Consequently, the (random) empirical measures  $\mathcal{E}(\mathcal{G}_n^{[S]})$  converge a.s. to the uniform measure on the unit disk.

### 2.1.2 One-Dimensional log-Gas

We now consider  $n$  particles on the real line, whose joint probability density is given by

$$Z_n^{-1} |\Delta(x_1, \dots, x_n)|^\beta \exp \left( -n \sum_{i=1}^n V(x_i) \right),$$

where  $\beta > 0$ , the confining potential  $V : \mathbb{R} \rightarrow \mathbb{R}$  is a continuous function such that, for some  $\beta' > 1$  satisfying  $\beta' \geq \beta$ , we have

$$\lim_{|x| \rightarrow \infty} \frac{V(x)}{\beta' \log |x|} > 1,$$

and  $Z_n = Z_n(\beta, V)$  is a normalizing constant. Important special cases include  $\beta = 1, 2$  and  $V(x) = |x|^2/2$ , which correspond to the eigenvalues of the well-known Gaussian Orthogonal Ensemble (GOE) and Gaussian Unitary Ensemble (GUE) respectively. In other words,  $\beta = 1$  and  $2$  respectively correspond to the Hermitian versions of the real and complex Ginibre ensembles (tridiagonal random matrix models for general  $\beta > 0$  are known for the quadratic confining potential, see [23]).

To state the large deviations principle for such Hermitian random matrices (see, e.g., [5, 8]), we introduce the rate function  $I_\beta^V : \mathcal{M}_1(\mathbb{R}) \rightarrow [0, \infty]$  as:

$$I_\beta^V = \begin{cases} \int V(x)d\mu(x) - \frac{\beta}{2}\Sigma(\mu) - c_\beta^V & \text{if } \int V(x)d\mu(x) < \infty, \\ \infty & \text{otherwise,} \end{cases}$$

where

$$c_\beta^V = \inf_{\nu \in \mathcal{M}_1(\mathbb{R})} \left\{ \int V(x)d\nu(x) - \frac{\beta}{2}\Sigma(\nu) \right\}.$$

With these definitions in hand, we state the LDP for the above Hermitian random matrices as follows:

**Theorem 2.2** ([8]) *The sequence of empirical measures  $\mathcal{E}(\mathbb{P}_{V,\beta}^n)$  obey a large deviation principle in the space  $\mathcal{M}_1(\mathbb{R})$  with rate  $n^2$  and good rate function  $I_\beta^V$ .*

It is known that  $I_\beta^V$  attains its minimum value in the space  $\mathcal{M}_1(\mathbb{R})$  at a unique measure  $\sigma_\beta^V$  that is compactly supported and is characterized by

$$V(x) - \beta U_{\sigma_\beta^V}(x) \begin{cases} = C_\beta^V & \text{for } \sigma_\beta^V - a.e.x, \\ > C_\beta^V & \text{for all } x \notin \text{supp}(\sigma_\beta^V), \end{cases}$$

where  $U_\mu$  denotes the logarithmic potential of the measure  $\mu$ , and  $C_\beta^V$  is some constant. An upshot of this is that the (random) empirical measures  $\mathcal{E}(\mathbb{P}_{V,\beta}^n)$  converge a.s. to the measure  $\sigma_\beta^V$ .

We mention in passing that large deviation principles are also known for  $\beta$ -Ginibre ensembles (defined analogously to the  $\beta$  ensembles in 1D by using a general  $\beta$  exponent on the Vandermonde in (2)); these correspond to the 2D Coulomb gas (for general inverse temperature  $\beta$ ). For details, we refer to [38].

### 2.2 Zeros of Random Polynomials

The theory of large deviations for empirical measures of zeros of random polynomials is of more recent origin. One of the earliest articles in this direction, namely [67], deals with the crucial case of (complex) Gaussian random polynomials, i.e., random polynomials with independent Gaussian coefficients (with mean zero and possibly decaying variances). Depending on the mode of decay of the variances, we obtain

several distinguished “standard ensembles” - Kac (constant variance of coefficients), Elliptic (coefficient of  $z^k$  has variance  $\binom{n}{k}k!$ ) and Weyl (coefficient of  $z^k$  has variance  $1/k!$ ). [67] covers all these cases, as well as more general scalings of coefficients. In fact, [67] works in the more general setting of Gaussian measures on polynomial spaces of degree  $n$  that live on the Riemann surface  $\mathbb{C}P^1$ . These Gaussian measures are determined by inner products naturally induced from a metric  $h$  and measure  $\nu$  on  $\mathbb{C}P^1$ , the only condition being that the pair  $(h, \nu)$  satisfy the so-called Bernstein–Markov property. The results obtained on the Riemann sphere  $\mathbb{C}P^1$  can be transferred (via the stereographic projection) to the complex plane. For a detailed exposition of this, we refer the reader to [15] (which also deals with the case of real Gaussian coefficients).

### 2.2.1 Weyl Polynomials

In this article, we will focus on the crucial case of the Weyl polynomials,

$$P_n(z) = \sum_{k=0}^n \frac{\xi_k}{\sqrt{k!}} z^k,$$

which are naturally related to the so-called standard planar Gaussian Entire Function (GEF, see (1)). Viewed over the complex plane, this corresponds to  $h$ , the standard Euclidean metric, and  $\nu$ , the standard Gaussian measure on  $\mathbb{C}$ .

In what follows, we will denote by  $Z_n = \{z_1, \dots, z_n\}$  the zero set of the Weyl polynomial of degree  $n$ , scaled down by  $\sqrt{n}$ . Also recall that for any measure  $\mu \in \mathcal{M}_1(\mathbb{C})$ , we denote by  $U_\mu$  and  $\Sigma(\mu)$  the logarithmic potential and the logarithmic energy of  $\mu$ , respectively. We can now state the following LDP for zeros of Weyl polynomials:

**Theorem 2.3** ([16,67]) *The sequence of empirical measures  $\mathcal{E}(Z_n)$  satisfy a large deviation principle in the space  $\mathcal{M}_1(\mathbb{C})$  with rate  $n^2$  and good rate function*

$$I^Z(\mu) = 2 \sup_{z \in \mathbb{C}} \left( U_\mu(z) - \frac{1}{2}|z|^2 \right) - \Sigma(\mu) - C.$$

The minimizer (and, consequently, the a.s. limit of the  $Z_n$ -s) of the above rate function is the uniform measure on the unit disk. The constant  $C$  is such that  $I$  evaluated at this measure is 0.

A word is in order here about the ‘unusual’ form of the rate function, and especially the appearance of the nonlinear and rather nondifferentiable sup term. The key to this lies in an expression for the joint density of (the scaled) zeros for the Weyl polynomial (w.r.t. Lebesgue measure on  $\mathbb{C}^n$ ), which can be written as

$$\rho(z_1, \dots, z_n) \propto |\Delta(z_1, \dots, z_n)|^2 \left( \frac{n}{\pi} \int_{\mathbb{C}} |Q_{Z_n}(w)|^2 e^{-n|w|^2} dm(w) \right)^{-(n+1)}, \quad (3)$$

where  $Q_{Z_n}$  is the monic polynomial



$$Q_{Z_n}(w) = (w - z_1) \dots (w - z_n),$$

and  $m$  is the Lebesgue measure on  $\mathbb{C}$ . The Vandermonde determinant leads to the logarithmic energy term in the rate function. In addition, we have

$$|Q_{Z_n}(w)|^2 e^{-n|w|^2} = \exp \left( 2n \left[ U_{\mathcal{E}(Z_n)}(w) - \frac{1}{2}|w|^2 \right] \right),$$

and we see that for large  $n$ , the main contribution to the integral in (3) is coming from the maximum over  $w \in \mathbb{C}$  of the term inside the square brackets.

### 2.2.2 Other Polynomials

LDPs are known for empirical measures of zeros of many other random polynomial and polynomial-like ensembles, in addition to the models described above. Some key examples are [25, 32, 68] and [16].

Some of these ensembles pose specific technical challenges of their own in establishing the LDP. As an example, we can consider the LDP for the Kac polynomial ensemble with exponential coefficients ([32]). The major new difficulty is that all the coefficients are now positive a.s. This restricts the possible zero sets of such polynomials, the precise nature of which was not fully understood until recently. E.g., Obrechhoff’s theorem ([54]) provides a necessary (but not sufficient) condition that the number of zeros of such a polynomial in a conical sector (around  $\mathbb{R}_+$ ) can grow at most linearly with the angle at the apex of the cone. This issue makes an impact even on the form of the LDP rate functional:

**Theorem 2.4** ([32]) *The empirical measure of zeros of Kac polynomials with exponential coefficients satisfy an LDP at rate  $n^2$  and good rate function*

$$I(\mu) = \begin{cases} U_\mu(1) - \frac{1}{2}\Sigma(\mu) & \text{if } \mu \in \mathcal{P}, \\ \infty & \text{otherwise,} \end{cases}$$

where  $\mathcal{P}$  is the set of all measures in  $\mathcal{M}_1(\mathbb{C})$  that are weak limits of empirical measures of polynomials with positive coefficients.

The approach of [32] exploits certain aspects of a potential theoretic description of the set  $\mathcal{P}$  obtained by [7]. The universality results of [16] employ comparison techniques with appropriate ensembles already known to have an LDP.

## 3 Hole Events and Hole Probabilities

Hole events and hole probabilities have classically been a key object of interest in the study of point processes (a.k.a. particle systems). An important example of this is the well-known result that hole probabilities for a determinantal point process are given by certain Fredholm determinants related to its kernel ([48, Chap. 6], [60]).

To fix ideas, let  $D_r$  denote the (open) disk (in dimension one an interval) of radius  $r$ , centered at the origin. The *hole event*, denoted  $\mathcal{H}_r$ , is the event where there are no points of the point process in  $D_r$ . The hole probability at radius  $r$  is  $\mathbb{P}[\mathcal{H}_r]$ , which clearly decays to 0 as  $r \rightarrow \infty$ . A very well-studied question in point process theory is the manner of decay of  $\mathbb{P}[\mathcal{H}_r]$  (more precisely, its logarithmic asymptotics). Typically, the logarithm of the hole probability decays like a power law, whose exponent depends upon the point process under consideration, and is thought to shed light on its ‘rigidity’. By rigidity in this setting, we envisage lattice-like behavior. In particular, the heuristic is that the faster the decay rate of the hole probability (that is, the higher the exponent discussed above), the stronger is the lattice-like behavior. E.g., as we shall see below, the exponent for the Poisson process (in 2D) is 2. On the other hand, for compactly supported i.i.d. perturbations of the lattice  $\mathbb{Z}^2$ , the hole probability is 0 for large enough hole sizes, and hence, heuristically speaking, the above exponent is  $\infty$ .

In 2D, the simplest example of a homogeneous point process, namely the Poisson process (with unit intensity) gives a decay of  $\mathbb{P}[\mathcal{H}_r] = \exp(-\text{Area}(D_r)) = \exp(-\pi r^2)$ , so the decay exponent is 2. For the 2D Coulomb gas (inverse temperature  $\beta = 2$ , a.k.a. the Ginibre ensemble), it has been shown that this exponent is 4 ([58], [39, Chap. 7]); i.e., the hole probability exhibits the decay  $r^{-4} \log \mathbb{P}[\mathcal{H}_r] \rightarrow -\frac{1}{4}$  as  $r \rightarrow \infty$ . The larger exponent of the Ginibre process already attests to a stronger global spatial correlation compared to the Poisson process (the latter being characterized by the spatial independence of its points). For the application of LDP techniques to study hole probabilities for the Ginibre ensemble, see, for example, Sect. 4.3.

A key ingredient in the proof is the fact that the number of particles in  $D_r$ , for any determinantal point process, is given by a sum of independent Bernoulli random variables (see, e.g., [39, Chap. 4]). The parameters (success probabilities) of these Bernoullis are essentially the eigenvalues of the integral operator given by the kernel of the determinantal process restricted to  $D_r$ . An alternative approach for the Ginibre ensemble is to use the fact (first proved by Kostlan for the finite Ginibre ensemble) that the set of the squares of the moduli of the eigenvalues is distributed like a set of *independent* Gamma random variables (see [39, Theorem 4.7.3] also for the infinite ensemble).

In comparison, the study of hole probabilities for the zeros of the GEF introduces considerable challenges. The basic underlying reason for this is the absence of any tractable “integrable” structure in the GEF zeros process, as opposed to Poisson (spatial independence) or the Ginibre (determinantal). The study of the hole probability for the GEF has been undertaken in a series of papers, beginning with upper and lower bounds for  $r^{-4} \log \mathbb{P}[\mathcal{H}_r]$  ([64]) and culminating in the proof of the fact ([50]) that

$$r^{-4} \log \mathbb{P}[\mathcal{H}_r] \rightarrow -\frac{e^2}{4}, \quad \text{as } r \rightarrow \infty. \quad (4)$$

In fact, in [50] and subsequent works ([51, 52]), hole probability asymptotics have been understood for a wide class of Gaussian entire functions. One can recover (4) using LDP techniques and also study in more details the hole event, see Sects. 4.2 and 7.

Thus, the exponents of the Ginibre and the GEF zeros process match. This leads to the interesting question regarding the comparison of these two processes vis-a-vis their strength of correlations (or lattice-like behavior). This has spawned an interesting collection of results. On the one hand, there are comparison theorems for finite order correlation functions of the Ginibre and GEF zero ensembles ([53]). On the other hand, there are recent results showing significant differences in the properties of their (spatially) conditional distributions ([33,34]). A very interesting problem is to determine whether there is a natural invariant point process in the plane whose hole probability decays qualitatively faster than the decay rate of the Ginibre ensemble and the GEF zeros process.

We conclude this section by mentioning several other works related to hole probabilities, mostly of recent origin. An important instance is the study of hole probability asymptotics for zeros of a wide class of Gaussian analytic functions having a finite radius of convergence. This includes the well-studied hyperbolic GAFs (with general intensity  $L > 0$ ), whose domain of convergence is the unit disk. For the case  $L = 1$ , the zero set has been shown to be a determinantal point process in [56]. In the same paper ([56]), the asymptotics of the hole probability (as  $r \uparrow \infty$ ) has been worked out. In [17], the hole probability asymptotics have been worked out for general  $L$ . In the process, a surprising discovery is made to the effect that the form of the asymptotics (including its dependence on  $L$ ) depend crucially on whether  $L$  is sub-critical ( $0 < L < 1$ ), critical ( $L = 1$ ), or super-critical ( $L > 1$ ). Another interesting family of results involves gap probabilities (essentially, hole probabilities in 1D) for important families of 1D Gaussian processes, in particular connecting these asymptotics with simple properties of their spectral measures and so-called “persistence probabilities” ([4,21,22,26,27]). In [58] and [59], the author obtains fine quantitative estimates on various aspects of the hole probability and the hole event for the Ginibre ensemble and related determinantal processes associated with higher Landau levels.

## 4 Conditional Distribution on the Hole Event

In this section, we consider the following problem: What is the principal cause of a (rare) event of a hole of large radius? Having understood hole probabilities, the next natural question, therefore, is to try and understand the point process conditioned to have a hole of a large radius. This question, however, turns out to be a surprisingly difficult one - even in expectation.

### 4.1 The Ginibre Ensemble

Until recently, the only 2D point process for which this was understood was the Ginibre ensemble ([40]; see [58] for a more recent study of finer aspects and more quantitative results). We state the result as (see the appendix in [40]):

**Theorem 4.1** ([40]) *The conditional intensity  $\rho^R$  (w.r.t. Lebesgue measure on  $\mathbb{C}$ ) of the Ginibre eigenvalues on the event  $\mathcal{H}_R$  is given by*

$$\rho^R(z) = \frac{1}{\pi} e^{-|z|^2} \sum_{n=0}^{\infty} \frac{|z|^{2n}}{\Gamma(n+1, R^2)},$$

for  $|z| \geq R$ , where we have used the incomplete Gamma function

$$\Gamma(n+1, x) = \int_x^{\infty} e^{-t} t^n dt.$$

In particular, for  $r = R$  and  $R \gg 1$ , we have  $\rho(R) \sim \frac{1}{2}\pi R^2$  (see equation (2.13) in [40]). This roughly corresponds to the appearance of a delta measure at the edge of the hole under appropriate renormalization.

In [59], Shirai described the complete behavior of the conditional intensity of eigenvalues for the more general ‘‘Ginibre-type’’ ensembles. We mention here a version of Theorem 1.4 therein, adapted to our specific context and using our notation. For a Borel set  $D \subset \mathbb{C}$ , let  $\xi(D)$  denote the number of Ginibre eigenvalues in  $D$ . Let  $A(x, y)$  denote the annulus  $\{z \in \mathbb{C} : x \leq |z| < y\}$ . With this notation, we can state:

**Theorem 4.2** [[59]] *For  $b > a > 1$ , we have*

$$\lim_{R \rightarrow \infty} \frac{1}{R^2} \mathbb{E}[\xi(A(\sqrt{a}R, \sqrt{b}R)) | \mathcal{H}_R] = b - a$$

and

$$\lim_{R \rightarrow \infty} \frac{1}{R^2} \mathbb{E}[\xi(A(R, \sqrt{b}R)) | \mathcal{H}_R] = b.$$

The discontinuity in the limit at  $a = 1$  captures the delta measure at the edge of the hole. The above limit also shows clearly that asymptotically, beyond the hole, the conditional intensity converges to the equilibrium intensity.

We point out that only the specific situation of a ‘‘round’’ hole was considered in this approach - that is, the hole consisted of no particles present in the disk of radius  $R$ , as opposed to, say, a hole in the form of a particle-free square of side length  $R$ , with  $R \rightarrow \infty$ . This is crucial for obtaining the above results (as previously mentioned, the set of the squares of the moduli of the eigenvalues is distributed like a set of independent Gamma random variables).

A crucial deficiency of this approach is the dependence on the above explicit description of the radii of the Ginibre points, which (or any substitute thereof) is not available for the other point processes. Even for the Ginibre ensemble, this approach depends crucially on the radial symmetry of the hole, and thus precludes any understanding of holes of any shape other than a disk.

## 4.2 GEF Zeros

Very recently, progress has been achieved ([1, 33]) in understanding the conditional intensity around a large hole for point processes other than the Ginibre ensemble

and for noncircular holes. The new progress relies on a novel large deviation based approach. As an example, we state the following description for the (limiting) conditional intensity measure around a “round” hole for the GEF zeros process ([33]):

**Theorem 4.3** *Let  $m_{\mathbb{T}}$  be the uniform probability measure on the unit circle  $\{|z| = 1\}$ , and put*

$$\mu_0^Z = e \cdot dm_{\mathbb{T}} + \mathbb{1}_{\{|w| \geq \sqrt{e}\}} \frac{m}{\pi}.$$

*Then, as  $R \rightarrow \infty$ , the scaled zero counting measure  $[\mathcal{Z}_R]$  of the GEF, conditioned on having a hole in  $\{|z| < R\}$ , converges weakly to a limiting measure (in expectation, probability):*

$$\frac{1}{R^2} [\mathcal{Z}_R] \left( \frac{\cdot}{R} \right) \rightarrow \mu_0^Z, \quad \text{as } R \rightarrow \infty.$$

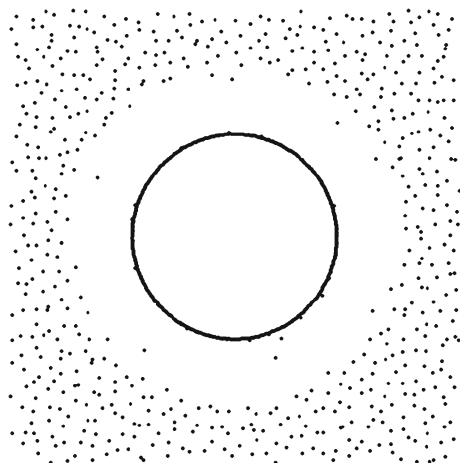
We immediately point out a key difference with the conditional intensity for the Ginibre process: the appearance of a “forbidden region” immediately beyond the hole, where the expected density of zeros vanishes as  $R \rightarrow \infty$ . To the best of our knowledge, this is the first example of such a “forbidden region”, and there is no instance, proven or conjectured, even in the physical literature, that predicts such a phenomenon. See Figs. 1 and 2 for a comparison between the distributions of the points conditioned on the hole event (with  $R = 13$ ).

The paper [33], in fact, provides quantitative estimates of the typical number of zeros in the annulus between  $R$  and  $\sqrt{e}R$ . In what follows, we denote by  $N_F(A)$  the number of zeros of the GEF in the set  $A \subset \mathbb{C}$ .

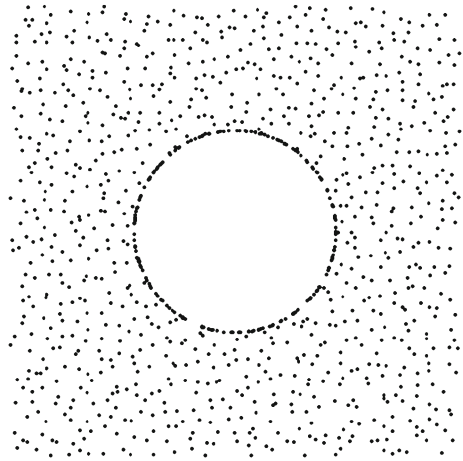
**Theorem 4.4** *Suppose  $R$  is sufficiently large,  $\epsilon \in (R^{-2}, 1)$ , that  $\gamma \in (1 + \frac{1}{2} \log \frac{1}{\epsilon} (\log R)^{-1}, 2]$ , and consider the annulus*

$$A(R(1 + \epsilon), \sqrt{e}R(1 - \epsilon)) = \{z \in \mathbb{C} : R(1 + \epsilon) \leq |z| \leq \sqrt{e}R(1 - \epsilon)\}.$$

**Fig. 1** GEF zeros, hole event



**Fig. 2** Ginibre ensemble, hole event



We have

$$\mathbb{P} [N_F(A(R(1 + \epsilon), \sqrt{\epsilon}R(1 - \epsilon))) \geq R^\gamma |\mathcal{H}_R|] \leq \exp(-C\epsilon R^{2\gamma}),$$

where  $C > 0$  is a numerical constant.

The proofs of Theorem 4.3 and Theorem 4.4 are based on a certain deviations inequality for linear statistics of the GEF zeros. We provide more details in Sect. 7.

It is an interesting problem to establish fine asymptotics, on the lines of Theorem 4.1, for the GEF zeros process. This would involve, in the best case scenario, an explicit expression for the conditional density function. At a more modest level, it can also envisage asymptotics of the conditional intensity function in various regimes, an important example of which is the rate of blowup of this function at the edge of the hole. It is also of interest to find the asymptotics of the (conditional) expected value of  $N_F(A(R(1 + \epsilon), \sqrt{\epsilon}R(1 - \epsilon)))$ , as  $R \rightarrow \infty$ .

### 4.3 Ginibre Ensemble: General Holes and Weighted Fekete Points

Noncircular holes for the Ginibre ensemble were recently studied in the paper [1]. To this end, we recall the functional

$$I(\mu) = \int |z|^2 d\mu(z) - \iint \log |z - w| d\mu(z)d\mu(w) - \frac{3}{4}, \quad \mu \in \mathcal{M}_1(\mathbb{C}).$$

We mention that it is known that the uniform measure on the unit disk  $\mu_0$  is the unique global minimizer for the above functional (and in fact,  $I(\mu_0) = 0$ ). We denote by  $D$  the open unit disk.

**Theorem 4.5** ([1]) *Let  $\mathcal{G}_n$  denote the eigenvalues of the Ginibre ensemble of order  $n$ . Let  $U \subset D$  be a subset satisfying at least one of the following conditions:*

- (Balayage condition) There exists a sequence of open sets  $U_n$  such that  $\overline{U} \subset U_n \subset D$  for all  $n$ , and the balayage measure  $\nu_n$  on  $\partial U_n$  converges weakly to the balayage measure on  $\partial U$ .
- (Exterior ball condition) There exists  $\epsilon > 0$  such that for every  $z \in \partial U$ , there exists a  $\eta \in U^{\mathbb{C}}$  such that

$$B(\eta, \epsilon) \subset U^{\mathbb{C}} \text{ and } |z - \eta| = \epsilon.$$

Then we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}[\mathcal{G}_n(\sqrt{n}U) = 0] = - \inf_{\mu \in \mathcal{M}_1(\mathbb{C}); \mu(U)=0} I(\mu).$$

Note that all convex domains satisfy the exterior ball condition.

In the case of the exterior ball condition, the proof relies on the use of *weighted Fekete points*. Let  $E \subset \mathbb{C}$  be a (nice) closed subset, and put

$$\delta_n(E) = \sup_{z_1, \dots, z_n \in E} \left\{ \prod_{j < k} |z_j - z_k| \exp\left(-\frac{1}{2}|z_j|^2\right) \exp\left(-\frac{1}{2}|z_k|^2\right) \right\}^{\frac{2}{n(n-1)}}.$$

A set  $\mathcal{F}_n = \{z_1^*, \dots, z_n^*\} \subset E$  is said to be an  $n$ -th weighted Fekete set if the points in  $\mathcal{F}_n$  attain the supremum  $\delta_n(E)$  (such a set always exists, but is not necessarily unique). It is known (see [63]) that the sequence  $\{\delta_n(E)\}$  is decreasing, and furthermore,

$$\lim_{n \rightarrow \infty} \log \delta_n(E) = - \inf_{\mu \in (E^{\mathbb{C}})^{\mathbb{C}}; \mu(U)=0} \left[ \int |z|^2 d\mu(z) - \iint \log |z - w| d\mu(z) \right],$$

where the infimum is over all probability measures  $\mu$  such that  $\mu(E^{\mathbb{C}}) = 0$ . Heuristically, the Fekete points provide the most likely configuration of particles, conditioned on having no particles in the set  $E^{\mathbb{C}}$ .

For the infinite Ginibre ensemble, we have:

**Theorem 4.6** [1] *Let  $\mathcal{G}_\infty$  denote the infinite Ginibre ensemble, and let  $U \subset D$  be an open set satisfying either the balayage condition or the exterior ball condition as in the statement of Theorem 4.5. Then we have the hole probability asymptotics*

$$\lim_{r \rightarrow \infty} \frac{1}{r^4} \log \mathbb{P}[\mathcal{G}_\infty(rU) = 0] = - \inf_{\mu \in \mathcal{M}_1(\mathbb{C}); \mu(U)=0} I(\mu).$$

### 5 Large Fluctuations in the Number of Points and the Jancovici–Lebowitz–Manificat Law

Closely related to the hole event are the phenomena of “deficiency” and “overcrowding” in the number of particles, which entail that the number of particles in  $D_r$  is

very far from its typical value of about  $r^2$  particles (with the standard normalization for the Ginibre ensemble and the GEF zeros). This has been extensively studied both for the Ginibre ensemble and the GEF zeros ([40,41,49]), with the discovery that the fluctuations in both cases obey the Jancovici–Lebowitz–Manificat law (in short, the JLM law), that was first introduced in the context of large charge fluctuations for the 2D Coulomb gas ([40], see Conjecture 5.1 for the statement).

We start with a special case of the JLM law. Denote by  $n(R)$  the number of particles of the Ginibre ensemble inside the disk  $\{|z| < R\}$ . Here we consider the event  $\{n(R) = pR^2\}$ , where  $p \geq 0$  ( $p \neq 1$ ) is fixed. Shirai ([58]) proved

$$\lim_{R \rightarrow \infty} R^{-4} \log \mathbb{P} \left[ n(R) = pR^2 \right] = -\frac{1}{4} \left| 2p^2 \log p - (p - 1)(3p - 1) \right|. \tag{5}$$

In [33], the authors derive the corresponding result for the GEF zeros; for a complete statement, we direct the reader to [33].

### 5.1 The Jancovici–Lebowitz–Manificat Law

Compared with (5), one can certainly consider a wider range of fluctuations in the number of particles and also examine other ensembles. We find it rather surprising that the asymptotic decay of the probability of large fluctuations is described by a common ‘law’, both for the Ginibre ensemble and the GEF zeros process (this law also appears in other ensembles, such as certain randomly perturbed lattices, see [49]).

#### 5.1.1 Finite $\beta$ -Ginibre Ensemble

The paper [40] by Jancovici, Lebowitz, and Manificat considers large charge fluctuations for a one-component Coulomb system of particles of one sign embedded into a uniform background of the opposite sign. This system is mathematically equivalent to the finite two-dimensional  $\beta$ -Ginibre ensemble, which consists of  $N$  particles in the complex plane  $\mathbb{C}$ , whose joint probability density, with respect to the Lebesgue measure on  $\mathbb{C}^N$ , is given by

$$p(z_1, \dots, z_N) = (Z_N^\beta)^{-1} \prod_{j < k} |z_j - z_k|^\beta \exp \left( -\frac{\beta}{2} \sum_{j=1}^N |z_j|^2 \right). \tag{6}$$

Here  $\beta > 0$  is the inverse temperature, and  $Z_N^\beta$  is the normalizing constant (also known as the partition function).

For  $N$  large, the particles tend to be asymptotically uniformly distributed inside the disk of radius  $\sqrt{N}$  centered at the origin. Let us denote by  $n(R)$  the number of particles in the disk  $D(0, R) = \{|z| \leq R\}$ . For  $N$  large compared with  $R^2$ , we have that  $n(R)$  is typically about  $R^2$ .



Now, fix the parameters  $a > \frac{1}{2}$  and  $b \neq 0$  (where  $b \geq -1$  if  $a = 2$ , and  $b > 0$  if  $a > 2$ ), and consider the very rare event

$$\mathcal{N}_{a,b}^\beta(R) = \left\{ n(R) = \lfloor R^2 + bR^a \rfloor \right\}.$$

Based on macroscopic electrostatic considerations, the paper [40] argues that after taking the limit  $N$  to infinity, the following asymptotic probabilities for large fluctuations in  $n(R)$  are observed.

**Conjecture 5.1** (The JLM law) *With the parameters  $a, b$  as above, and for  $R \rightarrow \infty$ , we have*

$$\mathbb{P} \left[ \mathcal{N}_{a,b}^\beta(R) \right] = \exp \left( -\beta \cdot \psi(\beta; a, b, \log R) \cdot R^{\varphi(a)} (1 + o(1)) \right),$$

where

$$\varphi(a) = \begin{cases} 2a - 1, & 1/2 < a \leq 1, \\ 3a - 2, & 1 \leq a \leq 2, \\ 2a, & a \geq 2, \end{cases}$$

$$\psi(\beta; a, b, \log R) = \begin{cases} c_\beta b^2, & 1/2 < a < 1, \\ \frac{1}{6} |b|^3, & 1 < a < 2, \\ \frac{1}{2} (a - 2) b^2 \log R, & a > 2, \end{cases}$$

and  $c_\beta$  is some constant depending on  $\beta$ .

*Remark 5.1* See [40] for the precise expression for  $\psi(\beta; a, b, \gamma)$  in the case  $a = 2$  (cf. (5)). It seems that no such expression is known for  $a = 1$  (even when  $\beta = 2$ ).

The constant  $c_\beta$  is derived from the (conjectured) central limit theorem (CLT) for  $n(R)$ . Recently the CLT for smooth linear statistics was proved in [44] (in this case the dependence on  $\beta$  is explicit).

In the case of the (infinite) Ginibre ensemble ( $\beta = 2$ ), the arguments of [40] are essentially mathematically rigorous (for proofs in the case  $a = 2$ , see the aforementioned [58]). For other values of  $\beta$ , it is not even known if a limiting object for the  $\beta$ -Ginibre ensemble exists when the number of particles goes to infinity.

### 5.1.2 Fluctuations for the GEF Zeros Process

Nazarov, Sodin, and Volberg ([49]) confirmed that some of the predictions of [40] hold also for large fluctuations in the number of zeros of the GEF. More precisely, they proved the following result:

**Theorem 5.1** ([49]) *For every  $a \geq 1/2$  and every  $\epsilon > 0$ , we have*

$$\exp(-R^{\varphi(a)+\epsilon}) \leq \mathbb{P} \left( |n_F(R) - R^2| \geq R^a \right) \leq \exp(-R^{\varphi(a)-\epsilon}) \tag{7}$$

for all sufficiently large  $R > R_0(\epsilon, a)$ .

*Remark 5.2* Partial results in this direction were already obtained in [41,64].

Using the results of [33] together with the approach of [41], it is possible to establish finer asymptotics for fluctuations in the GEF zeros process that are analogous to the JLM law, in a restricted range of exponents. As an example we mention:

**Theorem 5.2** *For fixed  $b \neq 0, a \in (4/3, 2)$ , we have, for the GEF zeros process the asymptotics*

$$\mathbb{P} \left[ n_F(R) = \lfloor R^2 + bR^a \rfloor \right] = \exp \left( -\frac{2|b|^3}{3} R^{3a-2} (1 + o(1)) \right), \quad R \rightarrow \infty.$$

The lower bound in the above asymptotics can, in fact, be shown to hold for  $a \in (1, 2)$ , and it is plausible that the results hold in this range.

### 5.2 Deficiency and Overcrowding: conditional distribution

Denote by  $n_F(R)$  the number of zeros of the GEF inside the disk  $\{|z| < R\}$ . Recall that the zero counting measure  $[\mathcal{Z}_R]$  denotes the GEF zero counting measure, conditioned on the hole event in  $\{n_F(R) = 0\}$ . We now denote by  $[\mathcal{Z}_R^p]$  the GEF zero counting measure, conditioned on the event  $\{n_F(R) = \lfloor pR^2 \rfloor\}$ , with  $p \geq 0, p \neq 1$ .

**Notation:** If  $p = 0$ , we set  $q = e$ . Otherwise, for  $0 < p < e$ , let  $q = q(p)$  be the nontrivial solution of the equation  $p(\log p - 1) = q(\log q - 1)$ .

In [33], we find the limiting conditional measure for these conditional counting measures. More precisely, the scaled conditional counting measure converges weakly (say in expectation) to a limiting Radon measure on  $\mathbb{C}$ :

$$\frac{1}{R^2} [\mathcal{Z}_R^p] \left( \frac{\cdot}{R} \right) \rightarrow \mu_p^Z \quad \text{as } R \rightarrow \infty,$$

where

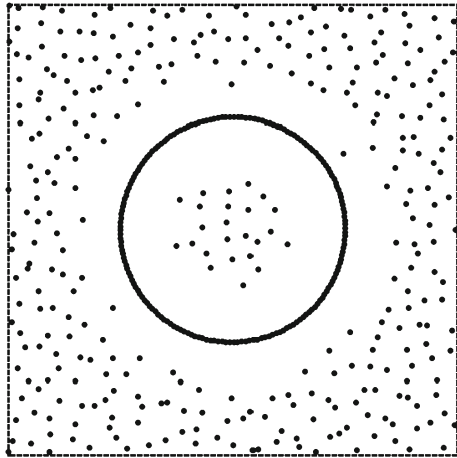
$$\mu_p^Z = \begin{cases} \left[ \mathbb{1}_{\{0 \leq |w| \leq \sqrt{p}\}} + \mathbb{1}_{\{\sqrt{q} \leq |w|\}} \right] \frac{m}{\pi} + (q - p) m_{\mathbb{T}}, & p \in [0, 1), \\ \left[ \mathbb{1}_{\{0 \leq |w| \leq \sqrt{q}\}} + \mathbb{1}_{\{\sqrt{p} \leq |w|\}} \right] \frac{m}{\pi} + (p - q) m_{\mathbb{T}}, & p \in (1, e), \\ \mathbb{1}_{\{\sqrt{p} \leq |w|\}} \frac{m}{\pi} + p m_{\mathbb{T}}, & p \geq e. \end{cases}$$

Using the determinantal structure of the Ginibre ensemble, it is not difficult to prove that a similar result holds for the (conditional) ‘‘eigenvalue’’ counting measure, with the limiting measure  $\mu_p^Z$ , replaced by

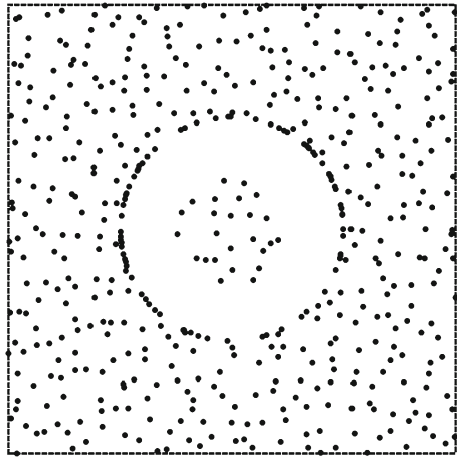
$$\mu_p^G = \begin{cases} \left[ \mathbb{1}_{\{0 \leq |w| \leq \sqrt{p}\}} + \mathbb{1}_{\{1 \leq |w|\}} \right] \frac{m}{\pi} + (1 - p) m_{\mathbb{T}}, & p \in [0, 1), \\ \left[ \mathbb{1}_{\{|w| \leq 1\}} + \mathbb{1}_{\{\sqrt{p} \leq |w|\}} \right] \frac{m}{\pi} + (p - 1) m_{\mathbb{T}}, & p > 1. \end{cases}$$

Figures 3 and 4 illustrate the case of a deficiency  $p = \frac{1}{2}$ , while Figs. 5 and 6 illustrate overcrowding for  $p = 2$ .

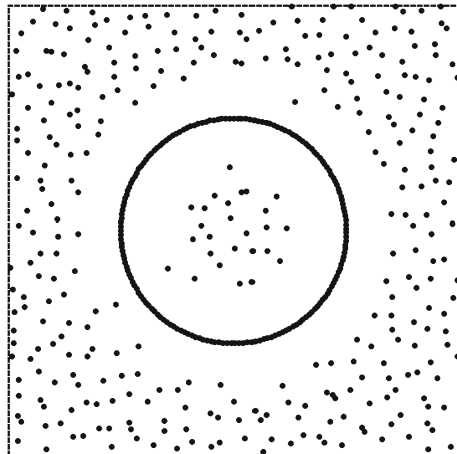
**Fig. 3** GEF, deficiency  $p = \frac{1}{2}$



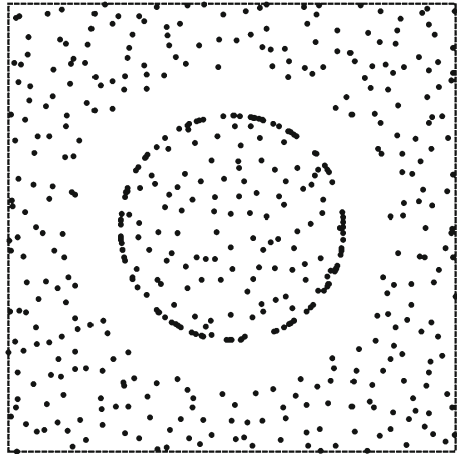
**Fig. 4** Ginibre, deficiency  $p = \frac{1}{2}$



**Fig. 5** GEF, overcrowding  $p = 2$



**Fig. 6** Ginibre, overcrowding  
 $p = 2$



## 6 Analysis of an Illustrative Model: 2D $\beta$ -Ensembles

In this section, we will provide proof sketches for the convergence of the conditional distributions of the particles and of the JLM law for the finite  $\beta$ -Ginibre ensembles. In the next section, we will briefly describe our proof from [33] for the zeros of the GEF. We recall that the two-dimensional  $\beta$ -Ginibre ensemble consists of  $N$  particles, whose joint probability density, with respect to the Lebesgue measure on  $\mathbb{C}^N$ , is given by (6).

Since, at the moment, a limiting object for the finite  $\beta$ -Ginibre ensemble is not known to exist, we choose a different limiting procedure than the one in the paper [40] (see Sect. 5.1). We fix a scaling parameter  $\alpha \geq 1$  and consider the asymptotics in terms of the large parameter  $R = \sqrt{\frac{N}{\alpha}}$ . Heuristically, with the parameter  $\alpha$ , the finite system resembles the (hypothetical) infinite system up to distances less than  $\sqrt{\alpha}R$  from the origin. In this setting, there are typically about  $R^2$  particles in the disk  $D(0, R) = \{|z| \leq R\}$ . We again denote the number of particles in this disk by  $n(R)$ .

We will illustrate below the proof of some of the predictions in Conjecture 5.1 above (using the different scaling procedure above). The proof in the case of overcrowding is similar. For  $a \geq 2$ , one has to choose the value of the scaling parameter  $\alpha$  depending on  $a$ ,  $b$ , and also  $R$  (if  $a > 2$ ).

**Theorem 6.1** *For fixed  $b > 0$ ,  $a \in (4/3, 2)$ , we have,*

$$\mathbb{P} \left[ n(R) \leq R^2 - bR^a \right] = \exp \left( -\frac{\beta b^3}{3} R^{3a-2} (1 + o(1)) \right), \quad R \rightarrow \infty.$$

Our proof of Theorem 6.1 proceeds via large deviation type estimates, and for  $a \leq 4/3$ , the error in our estimates (see Proposition 6.1) overwhelms the leading term. In the range  $a \in (1, \frac{4}{3}]$ , establishing the JLM law for the  $\beta$ -Ginibre ensembles with  $\beta \neq 2$  is an interesting open problem.

In addition to predicting the asymptotic decay of the rare events, the paper [40] describes the limiting conditional distribution of the particles. In order to simplify the presentation, we will consider just the events

$$F_p(R) = \left\{ n(R) \leq pR^2 \right\},$$

with  $p \in [0, 1)$  a fixed parameter (one can also consider overcrowding, and  $p$  can depend on  $R$ ).

In order to describe the limiting distribution, we introduce *linear statistics*, that is, the random variables

$$n(\varphi; R) = \sum_{j=1}^N \varphi\left(\frac{z_j}{R}\right),$$

where  $\varphi$  is a smooth (say  $C^2$ ) test function with compact support.

The following result describes the conditional limiting distribution, where  $m$  is the Lebesgue measure on  $\mathbb{C}$  and  $m_{\mathbb{T}}$  is the uniform probability measure on the unit circle  $\{|z| = 1\}$ .

**Theorem 6.2** *As  $R \rightarrow \infty$ , we have*

$$\mathbb{E}_{F_p(R)} [n(\varphi; R)] = R^2 \int_{\mathbb{C}} \varphi(w) d\mu_p^\alpha(w) + o(R^2),$$

where the limiting conditional measure is given by

$$\mu_p^\alpha = \left( \mathbb{1}_{\{|z| \leq \sqrt{p}\}} + \mathbb{1}_{\{1 \leq |z| \leq \sqrt{\alpha}\}} \right) \frac{m}{\pi} + (1 - p)m_{\mathbb{T}}.$$

*Remark 6.1* We write  $\mathbb{P}_F$  (resp.  $\mathbb{E}_F$ ) for the conditional probability (resp. expectation) on the event  $F$ .

*Remark 6.2* Notice that  $\mu_p^\alpha$  is *not* a probability measure. Later, it will also be convenient to work with the normalized probability measure

$$\bar{\mu}_p^\alpha = \frac{1}{\alpha} \mu_p^\alpha.$$

### 6.1 Deviation Inequality for Linear Statistics

Theorem 6.2 follows from the following deviation inequality:

**Proposition 6.1** *For  $R, \lambda > 0$ , we have*

$$\mathbb{P}_{F_p(R)} \left[ \left| n(\varphi; R) - R^2 \int_{\mathbb{C}} \varphi(w) d\mu_p^\alpha(w) \right| \geq \lambda \right] \leq \exp \left( -\frac{C\beta}{\mathfrak{D}(\varphi)} \lambda^2 + C_\varphi R^2 \log R \right),$$

where

$$\mathfrak{D}(\varphi) = \|\nabla\varphi\|_{L^2(m)}^2 = \int_{\mathbb{C}} (\varphi_x^2 + \varphi_y^2) \, dm.$$

*Remark 6.3* The proposition is nontrivial only for  $\lambda \geq CR\sqrt{\log R}$ ; hence we may assume this holds below.

The proof (motivated by the LDP approach in [9, 37]) is based on the approximation of the joint density of the particles (at the exponential scale) by a functional acting on probability measures. The (strictly convex, lower semi-continuous) function is given by

$$I_\alpha(v) = \int_{\mathbb{C}} \frac{|z|^2}{\alpha} \, dv(z) - \Sigma(v). \tag{8}$$

The global minimizer of the functional (also known as the *equilibrium measure*) is given by

$$\bar{\mu}_{\text{eq}}^\alpha = \frac{1}{\alpha} \mathbb{1}_{\{|z| \leq \sqrt{\alpha}\}}(w) \cdot \frac{m}{\pi},$$

that is, the uniform probability measure on the disk  $D(0, \sqrt{\alpha})$ .

Consider now the set of measures

$$\mathcal{F}_p = \mathcal{F}_p(\alpha) = \left\{ \mu \in \mathcal{M}_1(\mathbb{C}) : \mu(D) \leq \frac{p}{\alpha} \right\}, \quad p \in [0, 1),$$

where  $\mathcal{M}_1(\mathbb{C})$  is the set of probability measures on the complex plane  $\mathbb{C}$  and  $D$  is the unit disk  $D = D(0, 1) = \{|z| < 1\}$ . In Sect. 6.5, we show that the measure that minimizes the functional  $I_\alpha$  over the closed set  $\mathcal{F}_p$  is  $\bar{\mu}_p^\alpha$ .

We also consider measures that are ‘far’ from the minimizing measure  $\bar{\mu}_p^\alpha$ . For a test function  $\varphi$ , we put

$$\mathcal{L}_{\varphi,\lambda} = \left\{ \mu \in \mathcal{M}_1(\mathbb{C}) : \left| \int_{\mathbb{C}} \varphi(w) \, d\mu(w) - \int_{\mathbb{C}} \varphi(w) \, d\bar{\mu}_p^\alpha(w) \right| \geq \lambda \right\}.$$

A key tool required for the proof of Proposition 6.1 is the next claim, which can be regarded as an effective formulation of the fact that  $I_\alpha$  is strictly convex.

**Claim 1** *For any compactly supported measure  $\nu \in \mathcal{F}_p \cap \mathcal{L}_{\varphi,\lambda}$ , we have*

$$I_\alpha(\nu) \geq I_\alpha(\bar{\mu}_p^\alpha(w)) + \frac{2\pi}{\mathfrak{D}(\varphi)} \lambda^2.$$

*Proof* See the similar [33, Claim 11]. □

### 6.2 Approximation of the Joint Density

We start with an asymptotic estimate for the normalizing constant  $Z_N^\beta$  (see [43, Corollary 1.5]).

**Proposition 6.2** *We have*

$$\log Z_N^\beta = -\frac{1}{2}\beta N^2 \cdot I_\alpha(\bar{\mu}_{\text{eq}}^\alpha) + O(N \log N),$$

where the error term depends on  $\beta$ .

It is technically convenient to restrict the consideration to particles in a finite box. For example, one can easily prove that

$$\mathbb{P} \left[ \max_{j \in \{1, \dots, N\}} |z_j| \geq R^3 \right] \leq \exp \left( -\frac{\beta}{4} R^6 \right), \quad \forall R \geq R_0(\beta).$$

Hereafter, we assume that  $\max_j |z_j| \leq R^3$ .

#### 6.2.1 Smoothed Empirical Measure

Given  $\underline{w} = (w_1, \dots, w_N) \in \mathbb{C}^N$ , let  $\mu_{\underline{w}} = \frac{1}{N} \sum_{j=1}^N \delta_{w_j}$  be the empirical probability measure of the points. For a parameter  $t = t(R)$  (which is chosen to be  $R^{-C_1}$  for sufficiently large constant  $C_1 > 0$ ), we consider the smoothed empirical measure

$$\mu_{\underline{w}}^t = \mu_{\underline{w}} \star m_{|z|=t},$$

where  $m_{|z|=t}$  is the uniform probability measure on  $|z| = t$ . It is not difficult to verify the following (cf. [33, Claim 4]):

**Claim 2** *If  $C_1$  is chosen to be sufficiently large, then*

$$\frac{1}{N^2} \sum_{j \neq k} \log |w_j - w_k| \leq \Sigma(\mu_{\underline{w}}^t) + \frac{C \log \frac{1}{t}}{N} \leq \Sigma(\mu_{\underline{w}}^t) + \frac{C \log R}{N}$$

and

$$\frac{1}{N} \sum_{j=1}^N |w_j|^2 = \int_{\mathbb{C}} |w|^2 d\mu_{\underline{w}}^t(w) + O(R^{-2}).$$

Instead of working with the original particles  $z_j$ , it is more convenient to work with a scaled version. We set

$$L = R - t \quad \text{and} \quad \underline{w} = \frac{1}{L} \underline{z}.$$

With this scaling, our assumptions that  $n(R) \leq pR^2$  and  $|n(\varphi; R) - R^2 \int_{\mathbb{C}} \varphi(w) d\mu_p^\alpha(w)| \geq \lambda$  imply that

$$\mu_{\underline{w}}^t(D) \leq \frac{P}{\alpha}$$

and

$$\left| \int_{\mathbb{C}} \varphi(z) d\mu_{\underline{w}}^t(z) - \int_{\mathbb{C}} \varphi(w) d\bar{\mu}_p^\alpha(w) \right| \geq \frac{\lambda}{2N},$$

where we used the fact that  $N = \alpha R^2$ , the (Hölder) continuity of  $\varphi$ , and our assumption that  $\lambda \geq CR\sqrt{\log R}$ .

### 6.2.2 Upper Bound for the Joint Density

Define the following sets in  $\mathbb{C}^N$ :

$$N_p = N_p(t) = \left\{ \underline{w} \in \mathbb{C}^N : \mu_{\underline{w}}^t(D) \leq \frac{P}{\alpha} \right\},$$

$$L_{\varphi,\lambda} = L_{\varphi,\lambda}(t) = \left\{ \underline{w} \in \mathbb{C}^N : \left| \int_{\mathbb{C}} \varphi(z) d\mu_{\underline{w}}^t(z) - \int_{\mathbb{C}} \varphi(w) d\bar{\mu}_p^\alpha(w) \right| \geq \frac{\lambda}{2N} \right\}.$$

By the arguments outlined above, we have

$$\mathbb{P} \left[ F_p(R) \cap \left\{ \left| n(\varphi; R) - R^2 \int_{\mathbb{C}} \varphi(w) d\mu_p^\alpha(w) \right| \geq \lambda \right\} \right] \leq \exp \left( \frac{1}{2} \beta N^2 \left[ I_\alpha(\bar{\mu}_{\text{eq}}^\alpha) - \inf_{\underline{w} \in N_p \cap L_{\varphi,\lambda}} I^*(\underline{w}) \right] + O(N \log N) \right),$$

where

$$I^*(\underline{w}) = \frac{L^2}{N^2} \sum_{j=1}^N |w_j|^2 - \frac{1}{N^2} \sum_{j \neq k} \log |w_j - w_k|.$$

Claim 2 gives the following bound:

$$I^*(\underline{w}) \geq I_{\alpha'}(\mu_{\underline{w}}^t) - \frac{C \log R}{N},$$

where  $I_\alpha$  is the limiting functional defined in (8) and  $\alpha' = \frac{N}{L^2} = \alpha(1 + O(R^{-2}))$ . As one can show that

$$\inf_{\mu \in \mathcal{F}_p \cap \mathcal{L}_{\varphi,\lambda}} I_{\alpha'}(\mu) = \inf_{\mu \in \mathcal{F}_p \cap \mathcal{L}_{\varphi,\lambda}} I_\alpha(\mu) + O(R^{-2}),$$



we conclude that

$$\begin{aligned} & \mathbb{P} \left[ F_p(R) \cap \left\{ \left| n(\varphi; R) - R^2 \int_{\mathbb{C}} \varphi(w) d\mu_p^\alpha(w) \right| \geq \lambda \right\} \right] \\ & \leq \exp \left( -\frac{1}{2} \beta N^2 \inf_{\mu \in \mathcal{F}_p \cap \mathcal{L}_{\varphi, \tilde{\lambda}}} \left[ I_\alpha(\mu) - I_\alpha(\bar{\mu}_{\text{eq}}^\alpha) \right] + O(N \log N) \right), \end{aligned} \tag{9}$$

where  $\tilde{\lambda} = \frac{\lambda}{2N}$ .

### 6.3 Proof of Proposition 6.1

Using Claim 1, we obtain

$$\inf_{\mu \in \mathcal{F}_p \cap \mathcal{L}_{\varphi, \tilde{\lambda}}} \left[ I_\alpha(\mu) - I_\alpha(\bar{\mu}_p^\alpha) \right] \geq \frac{2\pi}{\mathfrak{D}(\varphi)} \cdot \left( \frac{\lambda}{2N} \right)^2 \geq \frac{C\lambda^2}{\mathfrak{D}(\varphi)N^2}. \tag{10}$$

In Sect. 6.4.1, we show the lower bound estimate

$$\mathbb{P} [F_p(R)] \geq \exp \left( -\frac{1}{2} \beta N^2 \cdot \inf_{\mu \in \mathcal{F}_p} \left[ I_\alpha(\mu) - I_\alpha(\bar{\mu}_{\text{eq}}^\alpha) \right] + O(N \log N) \right). \tag{11}$$

To prove Proposition 6.1, we combine (9), (10), (11), and the fact that  $\bar{\mu}_p^\alpha$  minimizes  $I_\alpha$  over  $\mathcal{F}_p$ .

### 6.4 Outline of the Proof of Theorem 6.1

The proof of the upper bound of Theorem 6.1 follows the same lines as the proof of Proposition 6.1 (ignoring the condition on linear statistics). From Claim 3, we find (cf. [58])

$$\begin{aligned} \mathbb{P} [F_p(R)] & \leq \exp \left( -\frac{1}{2} \beta N^2 \cdot \frac{1}{4\alpha^2} \left| 2p^2 \log p - (p-1)(3p-1) \right| + O(N \log N) \right) \\ & \leq \exp \left( -\frac{1}{8} \beta R^4 \left| 2p^2 \log p - (p-1)(3p-1) \right| + O(R^2 \log R) \right). \end{aligned}$$

To obtain the upper bound, we take  $p = 1 - bR^{a-2}$  (because of the error term, the result is not trivial for  $\frac{4}{3} < a < 2$ ).

*Remark 6.4* Notice that for  $p$  near 1, we have

$$2p^2 \log p - (p-1)(3p-1) = \frac{2}{3}(p-1)^3(1 + O(p-1)).$$

### 6.4.1 Lower Bound

The proof of the lower bound is similar to the proof of the lower bound for the hole probability in [1], and we will only sketch the idea. We again choose  $t = R^{-C_1}$  for some sufficiently large  $C_1 > 0$  and scale the points as follows:

$$\underline{w} = \frac{1}{R} \underline{z}.$$

One can construct a set of points  $\underline{w}^0$  with the following properties:

- There are at most  $pR^2$  points inside  $D$  (the open unit disk).
- Point-to-point separation:  $|w_j^0 - w_k^0| \geq Ct$  for all  $j \neq k$ .
- Separation from boundary:  $\bigcup_{w_j^0 \in D} D(w_j^0, t) \subset D$ .
- The points approximate the minimizing measure

$$\left| U_{\mu_{\underline{w}^0}^t}(z) - U_{\bar{\mu}_p^\alpha}(z) \right| \leq \frac{C \log N}{N}, \quad \forall z \in \mathbb{C}. \tag{12}$$

*Remark 6.5* One can construct such a ‘good’ set of points directly, using the radial symmetry of the problem. Another possibility is to use Fekete points for (essentially) the measure  $\bar{\mu}_p^\alpha$  (that is, weighted Fekete points with respect to the weight given by  $U_{\bar{\mu}_p^\alpha}$ ). A small difficulty with the second approach is that  $p$  can depend on  $N$  (the number of points).

Next we define a set of ‘good configurations’ of points:

$$\mathcal{G}_N = \left\{ \underline{w} \in \mathbb{C}^N : |w_j - w_j^0| < \frac{t}{2} \quad \forall j \in \{1, \dots, N\} \right\}.$$

By the separation of the points, and using (12), we have for  $\underline{w} \in \mathcal{G}_N$ :

$$\frac{1}{N^2} \sum_{j \neq k} \log |w_j - w_k| = \Sigma(\bar{\mu}_p^\alpha) + O\left(\frac{\log N}{N}\right),$$

and

$$\frac{1}{N} \sum_{j=1}^N |w_j|^2 = \int_{\mathbb{C}} |w|^2 d\bar{\mu}_p^\alpha(w) + O\left(\frac{\log N}{N}\right).$$

Also notice that the volume (in  $\mathbb{C}^N$ ) of  $\mathcal{G}_N$  is at least  $\exp(-CN \log N)$ . The lower bound is obtained by integrating the (scaled) joint density over the set  $\mathcal{G}_N$ , using the above estimates, and Claim 3.

### 6.5 Minimizing Measures

The following lemma gives a characterization of the (constrained) minimizers of the functional  $I_\alpha$  (cf. [33, Lemma 10]).

**Lemma 6.1** *Let  $\mathcal{C} \subset \mathcal{M}_1(\mathbb{C})$  be a closed and convex set of probability measures. The probability measure  $\mu_0 \in \mathcal{C}$  is the (unique) minimizer of  $I_\alpha$  over  $\mathcal{C}$  if and only if, for all  $\mu \in \mathcal{C}$ ,*

$$\int \frac{|z|^\alpha}{2\alpha} d\mu_0(z) - \Sigma(\mu_0) \leq \int \frac{|z|^\alpha}{2\alpha} d\mu(z) - \int U_{\mu_0}(z) d\mu(z).$$

*Proof* Let  $\mu \in \mathcal{C}$  be a probability measure such that  $\mu \neq \mu_0$ . Without loss of generality, we assume  $\mu$  has compact support and finite logarithmic energy. It is known ([63, Lemma I.1.8]) that  $-\Sigma(\mu - \mu_0) > 0$ .

For  $t \in (0, 1)$ , consider the measure  $\mu_t = (1 - t)\mu_0 + t\mu \in \mathcal{C}$ , and expand  $I_\alpha(\mu_t)$  to get

$$I_\alpha(\mu_t) = I_\alpha(\mu_0) + 2t \left( \int \frac{|z|^\alpha}{2\alpha} d\mu(z) - \int U_{\mu_0}(z) d\mu(z) - \int \frac{|z|^\alpha}{2\alpha} d\mu_0(z) + \Sigma(\mu_0) \right) + t^2 [-\Sigma(\mu - \mu_0)].$$

Therefore, if the linear term in  $t$  is non-negative, then  $I_\alpha(\mu_t) > I_\alpha(\mu_0)$ , and this implies (from the convexity of  $I_\alpha$ ) that  $I_\alpha(\mu) > I_\alpha(\mu_0)$ . The other direction is clear. □

We now wish to find the minimizing measure of the functional  $I_\alpha$  over the set

$$\mathcal{F}_p = \mathcal{F}_p(\alpha) = \left\{ \mu \in \mathcal{M}_1(\mathbb{C}) : \mu(D) \leq \frac{p}{\alpha} \right\},$$

which is a closed and convex subset of  $\mathcal{M}_1(\mathbb{C})$  (recall that  $D$  denotes the open unit disk).

We argue that the following probability measure is the minimizer:

$$\bar{\mu}_p^\alpha = \frac{1}{\alpha} \left( \mathbb{1}_{\{|z| \leq \sqrt{p}\}} + \mathbb{1}_{\{1 \leq |z| \leq \sqrt{\alpha}\}} \right) \frac{m}{\pi} + \frac{1-p}{\alpha} m_{\mathbb{T}}.$$

A simple calculation gives the values of the logarithmic potential on the support

$$U_{\bar{\mu}_p^\alpha}(z) = \frac{|z|^2}{2\alpha} + \frac{\log \alpha - 1}{2} + \begin{cases} c_1, & |z| \leq \sqrt{p}, \\ 0, & 1 \leq |z| \leq \sqrt{\alpha}, \end{cases}$$

where  $c_1 = \frac{p(\log p - 1) + 1}{2\alpha} > 0$ . It is also not difficult to see that

$$U_{\bar{\mu}_p^\alpha}(z) < \frac{|z|^2}{2\alpha} + \frac{\log \alpha - 1}{2} + \begin{cases} c_1, & \sqrt{p} < |z| < 1, \\ 0, & \sqrt{\alpha} < |z|. \end{cases}$$

The properties above give

$$\int \frac{|z|^2}{2\alpha} d\bar{\mu}_p^\alpha(z) - \Sigma(\bar{\mu}_p^\alpha) = \int \left[ \frac{|z|^2}{2\alpha} - U_{\bar{\mu}_p^\alpha}(z) \right] d\bar{\mu}_p^\alpha(z) = -\frac{\log \alpha - 1}{2} - c_1 \frac{p}{\alpha}.$$

On the other hand, for any  $\mu \in \mathcal{M}_1(\mathbb{C})$ , such that  $\mu(D) \leq \frac{p}{\alpha}$ , we have

$$\int \left[ \frac{|z|^2}{2\alpha} - U_{\bar{\mu}_p^\alpha}(z) \right] d\mu(z) \geq -\frac{\log \alpha - 1}{2} - \int_D c_1 d\mu \geq -\frac{\log \alpha - 1}{2} - c_1 \frac{p}{\alpha}.$$

Finally, it is straightforward to evaluate the functional  $I_\alpha$  for  $\bar{\mu}_p^\alpha$  (cf. [58]).

**Claim 3** We have for  $p \in [0, 1)$ ,

$$I_\alpha(\bar{\mu}_p^\alpha) = I_\alpha(\bar{\mu}_{\text{eq}}^\alpha) + \frac{1}{4\alpha^2} \left| 2p^2 \log p - (p - 1)(3p - 1) \right|.$$

*Remark 6.6* The above result also holds in the range  $p > 1$ .

### 7 Analysis of the GEF Zeros Process

In order to prove Theorem 4.3 and Theorem 4.4, we need to understand the behavior of linear statistics of the GEF zeros, conditioned on the hole event in  $\{|z| < R\}$ . We recall that linear statistics are the random variables

$$n_F(\varphi; R) = \sum_{z: F(z)=0} \varphi\left(\frac{z}{R}\right),$$

where  $\varphi$  is a smooth (say  $C^2$ ) test function with compact support, and we sum over all the zeros of the GEF (to prove Theorem 4.4, the test function  $\varphi$  has to depend on  $R$ , but here we ignore this technicality).

The analysis presented in the previous section requires some modifications in order to obtain the conditional intensity for the GEF zeros process. The main idea is to approximate the (scaled) GEF with the Weyl polynomials

$$P_{\alpha,R}(z) = \sum_{k=0}^N \frac{\xi_k}{\sqrt{k!}} (Rz)^k,$$

where  $\alpha > 1$  is a large parameter (eventually depending on  $R$ ), and  $N = N(\alpha, R) = \lfloor \alpha R^2 \rfloor$  is the degree of  $P_{\alpha,R}$ .

Roughly speaking, with this choice of the parameters, the scaled GEF  $F(Rz)$  (defined in (1)) and the polynomial  $P_{\alpha,R}$  have a very similar behavior inside a disk of radius  $\sqrt{\beta}$ , as long as  $\beta \ll \alpha$ . Therefore, by taking  $\alpha$  large, we can obtain an understanding of the conditional intensity of the Gaussian zeros (under conditioning by  $\mathcal{H}_R$ ) by analyzing the same problem for the polynomials.

*Remark 7.1* It turns out that in order to carry out this approximation scheme, we need to let  $\alpha \rightarrow \infty$  logarithmically with  $R$ . A drawback of this is that we cannot use ‘off-the-shelf’ large deviation principles for empirical measures of random polynomial zeros such as [67].

### 7.1 Joint Probability Density and the Limiting Functional

The joint density of the zeros of  $P_{\alpha,R}$  with respect to the Lebesgue measure on  $\mathbb{C}^N$  is given by (cf. (3))

$$\begin{aligned} &\rho_R(z_1, \dots, z_N) \\ &= (C_{N,R,\alpha})^{-1} |\Delta(z_1, \dots, z_N)|^2 \left( \frac{R^2}{\pi} \int_{\mathbb{C}} |Q_N(w)|^2 e^{-R^2|w|^2} dm(w) \right)^{-(N+1)} \end{aligned} \tag{13}$$

where  $Q_N$  is the polynomial

$$Q_N(w) = (w - z_1) \dots (w - z_N)$$

and  $C_{N,R,\alpha}$  is a normalizing constant. As was briefly outlined in 2.2.1, at the exponential scale, one can approximate this density with the limiting functional

$$I_\alpha^Z(\mu) = 2 \sup_{z \in \mathbb{C}} \left( U_\mu(z) - \frac{1}{2\alpha} |z|^2 \right) - \Sigma(\mu) - C_\alpha,$$

where we used  $\frac{N}{R^2} \cong \alpha$ .

*Remark 7.2* The global minimizer of the functional  $I_\alpha^Z$  is the uniform probability on the disk  $\{|z| \leq \sqrt{\alpha}\}$ , which we denoted by  $\bar{\mu}_{\text{eq}}^\alpha$ .

#### 7.1.1 Deviation Inequality

The analysis is done in a similar way to the case of the  $\beta$ -Ginibre ensembles (Sects. 6.1 and 6.2).

Recall that  $N_F(R)$  is the number of zeros of the GEF inside the disk  $\{|z| < R\}$ . To fix ideas, consider the following event:

$$F_p^Z(R) = \left\{ N_F(R) \leq pR^2 \right\}, \quad p \in [0, 1),$$

and introduce the following sets of measures:

$$\mathcal{F}_p = \left\{ \mu \in \mathcal{M}_1(\mathbb{C}) : \mu(D) \leq \frac{p}{\alpha} \right\}, \quad p \in [0, 1)$$

and

$$\mathcal{L}_{\varphi,\lambda}^Z = \left\{ \mu \in \mathcal{M}_1(\mathbb{C}) : \left| \int_{\mathbb{C}} \varphi(w) \, d\mu(w) - \int_{\mathbb{C}} \varphi(w) \, d\bar{\nu}_p^\alpha(w) \right| \geq \lambda \right\},$$

where  $D$  is the open unit disk and  $\bar{\nu}_p^\alpha$  is the restriction of the limiting measure  $\mu_p^Z$  (introduced in Sect. 5.2) to the disk  $\{|z| \leq \sqrt{\alpha}\}$ , and normalized to be a probability measure.

In a similar fashion to (9), one can show

$$\begin{aligned} & \mathbb{P} \left[ F_p^Z(R) \cap \left\{ \left| n_F(\varphi; R) - R^2 \int_{\mathbb{C}} \varphi(w) \, d\mu_p^Z(w) \right| \geq \lambda \right\} \right] \\ & \leq \exp \left( -N^2 \inf_{\mu \in \mathcal{F}_p \cap \mathcal{L}_{\varphi,\tilde{\lambda}}^Z} \left[ I_\alpha^Z(\mu) - I_\alpha^Z(\bar{\mu}_{\text{eq}}^\alpha) \right] + O(N \log N) \right), \end{aligned}$$

where  $\tilde{\lambda} = \frac{\lambda}{2N}$ . Then, combining Claim 11 from [33] (corresponding to Claim 1 for the Ginibre ensemble), together with a lower bound estimate for the probability of  $F_p^Z(R)$ , we obtain the deviation inequality

$$\mathbb{P}_{F_p^Z(R)} \left[ \left| n_F(\varphi; R) - R^2 \int_{\mathbb{C}} \varphi(w) \, d\mu_p^Z(w) \right| \geq \lambda \right] \leq \exp \left( -\frac{C}{\mathfrak{D}(\varphi)} \lambda^2 + C_\varphi R^2 \log^2 R \right).$$

*Remark 7.3* The actual proof is technically more involved, in large part because the choice of the value of  $N$  has to be random.

### 7.1.2 Lower Bound for $\mathbb{P} \left[ F_p^Z(R) \right]$

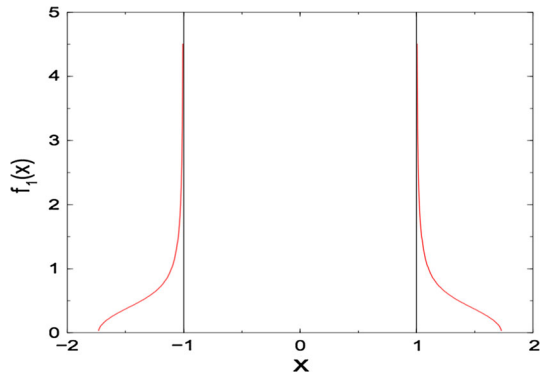
Because of the circular symmetry of the problem, one can use analytic techniques to obtain the lower bound for the probability of the event  $F_p^Z(R)$ , which are not available in the case of the  $\beta$ -Ginibre ensembles.

The main idea is to use Rouché’s theorem. More precisely, recalling that the GEF is given by the Gaussian Taylor series

$$F(z) = \sum_{k=0}^{\infty} \frac{\xi_k}{\sqrt{k!}} z^k,$$

we explicitly construct an event where the term  $\left| \xi_{k_0} \frac{z^{k_0}}{\sqrt{k_0!}} \right|$ ,  $k_0 = \lfloor pR^2 \rfloor$ , dominates the sum over all the other terms (on the circle  $\{|z| = R\}$ ). This simple but effective method originally appeared in the paper [64] and was later used in many other problems of this type.

**Fig. 7** Conditional intensity profile in 1D



### 7.1.3 The Minimizing Measures

In order to find the limiting conditional measures for the GEF zeros process, one has to identify the (probability) measure  $\bar{\nu}_p^\alpha$  which minimizes the functional  $I_\alpha^Z$  over the set  $\mathcal{F}_p$ . The interested reader can find the details in [33, Section 5].

## 8 Conditional Intensities in 1D

The conditional intensity around a hole has been studied in 1D (where it is usually called a ‘gap’), in the context of the *GUE* (Gaussian Unitary Ensemble) point process in [46]. In this section, we briefly describe the approach in [46] and compare and contrast the results therein with the situation we already discussed in 2D.

In 1D, under the natural scaling (necessary for obtaining an LDP from the GUE), the “droplet” (that is, the minimizer of the LDP rate functional) assumes the form of the famous *semicircle distribution*. With the normalization of [46], this density is given by

$$f(x) = \frac{1}{\pi} \sqrt{2 - x^2}, \quad |x| \leq \sqrt{2}.$$

Under this scaling, the “hole” assumes the form of an interval  $(\zeta_1, \zeta_2)$ . In the important case where  $(\zeta_1, \zeta_2)$  is an interval symmetric about the origin, denoted by  $(-w, w)$ , the (scaled) conditional intensity has the form

$$f_w(x) = \frac{1}{\pi} \sqrt{\frac{L^2 - x^2}{x^2 - w^2}} |x|, \quad x \in [-L, -w] \cup [w, L],$$

where  $L = \sqrt{w^2 + 2}$ . In particular, we note that there is no singular component, in contrast with both of the models we considered in 2D. Figure 7 depicts the density for  $w = 1$ .

In [46], the authors approach this problem by obtaining a singular integral equation for  $f_w$ , which is deduced essentially from a variational perturbation of the LDP rate

functional around the minimizing measure. The authors then illustrate two approaches to solving this singular integral equation - one of them being a Riemann–Hilbert approach and the other being via an application of Tricomi’s theorem ([66]).

In [46], the authors also study the particle distribution for *atypical indices* for the GUE. The *index*  $N_+$  of a configuration of  $N$  particles on  $\mathbb{R}$  is the number of particles on  $\mathbb{R}_+$ . By symmetry, the typical value of  $N_+/N = 1/2$ . Using similar variational techniques (as discussed above) on the LDP rate functional for the GUE ensemble, in [46] the authors obtain the asymptotics of the probability of an atypical index  $N_+/N = c \neq 1/2$ , as well as the typical particle profile given such an atypical index. This can be compared with [6], where a similar problem has been studied for the Ginibre ensemble using free boundary techniques.

## 9 Simulation of the Hole Event and Numerical Aspects

Numerical methods to effectively simulate the distribution of zeros (or eigenvalues) conditioned on a large hole is a challenging problem, because of the rarity of the hole event and the strong correlation among the particles. E.g., for the Ginibre ensemble, the asymptotics of the hole probabilities can be understood via the statistical independence of their absolute values, but this approach is not useful for simulations, because it carries no information about the correlations between the particles.

In the present paper, Fig. 1 is obtained by simulating the GEF (and then the zeros thereof) under the conditions (on the coefficients) that produce the tight lower bound for the hole probability (as alluded to in Sect. 7.1.2). Figure 2 is obtained by manually moving the eigenvalues of a Ginibre matrix (that are inside the disk) to the disk’s boundary. In [46], a modified Metropolis Hastings algorithm was studied for simulating such conditional distributions (conditioned on hole, overcrowding, or deficiency events) for the GUE process in 1D.

A 2D analogue of such an algorithm, for the Ginibre ensemble, would consist of the following: We start with a “legitimate” particle configuration, namely one that satisfies the constraint of having a hole. E.g., a reasonable initial configuration would be equi-spaced points on the boundary of the hole. Given a “legitimate” configuration  $(\lambda_1, \dots, \lambda_N)$ , we generate a new one  $(\lambda'_1, \dots, \lambda'_N)$  by perturbing a (randomly picked) particle by a small Gaussian noise, conditioned to avoid the “hole”. Ideally, we then replace the current configuration with the new one with probability

$$\min \left( \frac{f_N(\lambda'_1, \dots, \lambda'_N)}{f_N(\lambda_1, \dots, \lambda_N)}, 1 \right),$$

where  $f_N$  is the probability density function of the Ginibre ensemble of size  $N$ . However, the generation of new configurations that avoid the hole introduces an inherent asymmetry, which has to be taken into account in the acceptance probability for this approach. More precisely, one has to add the ratio of the probability to move from the new configuration to the old one, over the probability to move from the old configuration to the new one (which are not the same).



**Fig. 8** Simulation of a circular hole by a modified Metropolis Hastings algorithm



Figure 8 presents the result of 10,000 iterations of the above algorithm, in the case of a circular hole. The initial configuration consists of 2,000 points uniformly distributed in the annulus outside of the excluded disk and inside the support of the equilibrium measure (indicated by the outer circle). It is more difficult to implement this algorithm efficiently for the GEF zero process, mainly because the finite particle density  $f_N$  is considerably more complicated for the zeros, involving interactions of all orders up to  $N$ .

A different but related question is to numerically solve constrained optimization problems on the space of probability measures (on a Euclidean space), like the ones arising in the hole problem for the Ginibre ensemble and the GEF zeros. In the setting of the Ginibre ensemble, the goal is to minimize a weighted logarithmic energy

$$I(\mu) = \int V(x) d\mu(x) - \Sigma(\mu)$$

of a probability measure (or, more generally, a finite Borel measure)  $\mu$ , subject to constraints on its support. This can be directly related to LDP rate functionals - for example, the LDP rate functional for the Ginibre process is a logarithmic energy with a quadratic weight  $V$ .

To our knowledge, very little is known about this problem in dimensions greater than one. In 1D, a similar numerical problem has been addressed by [18] in the weighted case (using an approach involving iterated balayage), and by [19,36] in the unweighted case. Even in the 1D situation, there are various assumptions on the Riesz measure corresponding to the weight  $V$ , which would be of interest to relax.

It remains a nontrivial and highly interesting question to devise efficient numerical techniques to simulate the particle configurations for the hole (and, in the same vein, for overcrowding and deficiency) events. In the case of the hole event for the Ginibre ensemble, one can use weighted Leja points (see [63, Chapter V]) to approximate the

most likely eigenvalue configurations (given by the weighted Fekete points, mentioned in Sect. 4.3). Finding a similar method for the GEF zeros process seems to be an interesting problem.

**Acknowledgements** We thank the authors of the paper [46] for allowing us to use the picture in Fig. 7. We thank Diego Ayala for allowing us to use the picture in Figure 8. We thank the anonymous referee for numerous helpful suggestions. The work of S.G. was supported in part by the ARO Grant W911NF-14-1-0094, the NSF Grant DMS-1148711 and the NUS Grant R-146-000-250-133.

## References

1. Adhikari, K., Reddy, N.: *Hole probabilities for finite and infinite Ginibre ensembles*, International Mathematics Research Notices, rnw207, <https://doi.org/10.1093/imrn/rnw207>
2. Aizenman, M., Goldstein, S., Lebowitz, J.: Conditional equilibrium and the equivalence of micro-canonical and grandcanonical ensembles in the thermodynamic limit. *Commun. Math. Phys.* **62**(3), 279–302 (1978)
3. Aizenman, M., Martin, P.: Structure of Gibbs states of one-dimensional Coulomb systems. *Comm. Math. Phys.* **78**(1), 99–116 (1980)
4. Antezana, J., Buckley, J., Marzo, J., Olsen, J.F.: Gap probabilities for the cardinal sine. *J. Math. Anal. Appl.* **396**(2), 466–472 (2012)
5. Anderson, G., Guionnet, A., Zeitouni, O.: *An Introduction to Random Matrices*, vol. 118. Cambridge University Press, Cambridge (2010)
6. Armstrong, S.N., Serfaty, S., Zeitouni, O.: Remarks on a constrained optimization problem for the Ginibre ensemble. *Potential Anal.* **41**(3), 945–958 (2014)
7. Bergweiler, W., Eremenko, A.: Distribution of zeros of polynomials with positive coefficients. *Ann. Acad. Sci. Fenn. Math.* **40**(1), 375–383 (2015)
8. Ben Arous, G., Guionnet, A.: Large deviations for Wigner’s law and Voiculescu’s non-commutative entropy. *Probab. Theory Relat. Fields* **108**(4), 517–542 (1997)
9. Arous, G.B., Zeitouni, O.: Large deviations from the circular law. *ESAIM Probab. Stat.* **2**, 123–134 (1998)
10. Bufetov, A.: Rigidity of determinantal point processes with the Airy, the Bessel and the Gamma kernel. *Bull. Math. Sci.* **6**(1), 163–172 (2016)
11. Bufetov, A., Dabrowski, Y., Qiu, Y.: Linear rigidity of stationary stochastic processes. *Ergod. Theory Dyn. Syst.* 1–15. <https://doi.org/10.1017/etds.2016.140>
12. Bufetov, A., Qiu, Y.: Determinantal point processes associated with Hilbert spaces of holomorphic functions. *Commun. Math. Phys.* **351**(1), 1–44 (2017)
13. Bogomolny, E., Bohigas, O., Leboeuf, P.: Quantum chaotic dynamics and random polynomials. *J. Stat. Phys.* **85**(5-6), 639–679 (1996)
14. Borodin, A., Serfaty, S.: Renormalized energy concentration in random matrices. *Commun. Math. Phys.* **320**(1), 199–244 (2013)
15. Butez, R.: Large Deviations for the Empirical Measure of Random Polynomials: Revisit of the Zeitouni–Zelditch Theorem. arXiv preprint [arXiv:1509.09136](https://arxiv.org/abs/1509.09136) (2015)
16. Butez, R., Zeitouni, O.: Universal large deviations for Kac polynomials. *Electron. Commun. Probab.* **22**(6), 10 (2017)
17. Buckley, J., Nishry, A., Peled, R., Sodin, M.: Hole probability for zeroes of Gaussian Taylor series with finite radii of convergence. *Probab. Theory Relat. Fields* (2017). <https://doi.org/10.1007/s00440-017-0782-0>
18. Chesnokov, A., Deckers, K., Van Barel, M.: A numerical solution of the constrained weighted energy problem. *J. Comput. Appl. Math.* **235**(4), 950–965 (2010)
19. Coroian, D., Dragnev, P.: Constrained Leja points and the numerical solution of the constrained energy problem. *J. Comput. Appl. Math.* **131**(1), 427–444 (2001)
20. Daley, D., Vere-Jones, D.: *An introduction to the theory of point processes*, vol. I and II. Springer, Berlin (2007)
21. Dembo, A., Mukherjee, S.: No zero-crossings for random polynomials and the heat equation. *Ann. Probab.* **43**(1), 85–118 (2015)

22. Dembo, A., Mukherjee, S.: Persistence of Gaussian processes: non-summable correlations. *Probab. Theory Relat. Fields* (2016). <https://doi.org/10.1007/s00440-016-0746-9>
23. Dumitriu, I., Edelman, A.: Matrix models for beta ensembles. *J. Math. Phys.* **43**(11), 5830–5847 (2002)
24. Erdos, L.: Universality for random matrices and log-gases. *Lecture Notes for the conference Current Developments in Mathematics, 2012*. arXiv preprint [arXiv:1212.0839](https://arxiv.org/abs/1212.0839) (2012)
25. Feng, R., Zelditch, S.: Large deviations for zeros of  $P(\varphi)_2$  random polynomials. *J. Stat. Phys.* **143**(4), 619–635 (2011)
26. Feldheim, N.D., Feldheim, O.N.: Long gaps between sign-changes of gaussian stationary processes. *Int. Math. Res. Not.* **2015**(11), 3021–3034 (2015)
27. Feldheim, N.D., Feldheim, O.N., Nitzan, S.: Persistence of Gaussian stationary processes: a spectral perspective. arXiv preprint [arXiv:1709.00204](https://arxiv.org/abs/1709.00204) (math.PR)
28. Ghosh, S.: Determinantal processes and completeness of random exponentials: the critical case. *Probab. Theory Relat. Fields* **163**(3–4), 643–665 (2015)
29. Ghosh, S.: Palm measures and rigidity phenomena in point processes. *Electron. Commun. Probab.* **21**(85), 14 (2016)
30. Ghosh, S., Krishnapur, M.: Rigidity hierarchy in random point fields: random polynomials and determinantal processes. arXiv preprint [arXiv:1510.08814](https://arxiv.org/abs/1510.08814) (2015)
31. Ghosh, S., Lebowitz, J.L.: Number rigidity in superhomogeneous random point fields. *J. Stat. Phys.* **166**(3–4), 1016–1027 (2017)
32. Ghosh, S., Zeitouni, O.: Large deviations for zeros of random polynomials with iid exponential coefficients. *Int. Math. Res. Not.* **2016**(5), 1308–1347 (2016)
33. Ghosh, S., Nishry, A.: Gaussian complex zeros on the hole event: the emergence of a forbidden region. arXiv preprint [arXiv:1609.00084](https://arxiv.org/abs/1609.00084) (2016)
34. Ghosh, S., Peres, Y.: Rigidity and Tolerance in point processes: Gaussian zeroes and Ginibre eigenvalues. *Duke Math. J.* **166**(10), 1789–1858 (2017)
35. Guionnet, A.: Large deviations and stochastic calculus for large random matrices. *Probab. Surv.* **1**, 72–172 (2004)
36. Helsen, S., Van Barel, M.: A numerical solution of the constrained energy problem. *J. Comput. Appl. Math.* **189**(1), 442–452 (2006)
37. Petz, D., Hiai, F.: Logarithmic energy as an entropy functional. *Contemp. Math.* **217**, 205–221 (1998)
38. Hiai, F., Petz, D.: *The Semicircle Law, Free Random Variables, and Entropy*. American Mathematical Society, Providence (2000)
39. Hough, J.B., Krishnapur, M., Peres, Y., Virág, B.: Zeros of Gaussian analytic functions and determinantal point processes, vol. 51. American Mathematical Society, Providence (2009)
40. Jancovici, B., Lebowitz, J.L., Manificat, G.: Large charge fluctuations in classical Coulomb systems. *J. Stat. Phys.* **72**(3–4), 773–787 (1993)
41. Krishnapur, M.: Overcrowding estimates for zeroes of planar and hyperbolic Gaussian analytic functions. *J. Stat. Phys.* **124**(6), 1399–1423 (2006)
42. Kunz, H.: The one-dimensional classical electron gas. *Ann. Phys.* **85**, 303–335 (1974)
43. Leblé, T., Serfaty, S.: Large deviation principle for empirical fields of Log and Riesz gases. arXiv preprint [arXiv:1502.02970](https://arxiv.org/abs/1502.02970) (2015). To appear in *Inventiones Math*
44. Leblé, T., Serfaty, S.: Fluctuations of Two-Dimensional Coulomb Gases. arXiv preprint [arXiv:1609.08088](https://arxiv.org/abs/1609.08088) (2016)
45. Leblé, T., Serfaty, S., Zeitouni, O.: (with an appendix by W. Wu), Large deviations for the 2D two-component plasma. *Commun. Math. Phys.* **350**(1), 301–360 (2017)
46. Majumdar, S.N., Nadal, C., Scardicchio, A., Vivo, P.: How many eigenvalues of a Gaussian random matrix are positive? *Phys. Rev. E* **83**(4), 041105 (2011)
47. Martin, Ph, Yalcin, T.: The charge fluctuations in classical Coulomb systems. *J. Stat. Phys.* **22**(4), 435–463 (1980)
48. Mehta, M.: *Random Matrices*. Academic Press, New York (1962)
49. Nazarov, F., Sodin, M., Volberg, A.: The JancoviciLebowitzManificat law for large fluctuations of random complex zeroes. *Commun. Math. Phys.* **284**(3), 833–865 (2008)
50. Nishry, A.: Asymptotics of the hole probability for zeros of random entire functions. *Int. Math. Res. Not.* **15**, 2925–2946 (2010)
51. Nishry, A.: The hole probability for Gaussian entire functions. *Israel J. Math.* **186**(1), 197–220 (2011)
52. Nishry, A.: Hole probability for entire functions represented by Gaussian Taylor series. *J. d'Analyse Mathématique* **118**(2), 493–507 (2012)

53. Nazarov, F., Sodin, M.: Correlation functions for random complex zeroes: strong clustering and local universality. *Commun. Math. Phys.* **310**(1), 75–98 (2012)
54. Obrechhoff, N.: Sur un probl'eme de Laguerre. *C. R. Acad. Sci. (Paris)* **177**, 102–104 (1923)
55. Osada, H.: Interacting Brownian motions in infinite dimensions with logarithmic interaction potentials. *Ann. Probab.* **41**(1), 1–49 (2013)
56. Peres, Y., Virág, B.: Zeros of the iid Gaussian power series: a conformally invariant determinantal process. *Acta Mathematica* **194**(1), 1–35 (2005)
57. Rougerie, N., Serfaty, S.: Higher dimensional coulomb gases and renormalized energy functionals. *Commun. Pure Appl. Math.* **69**(3), 519–605 (2016)
58. Shirai, T.: Large deviations for the fermion point process associated with the exponential kernel. *J. Stat. Phys.* **123**(3), 615–629 (2006)
59. Shirai, T.: Ginibre-type point processes and their asymptotic behavior. *J. Math. Soc. Jpn.* **67**(2), 763–787 (2015)
60. Shirai, T., Takahashi, Y.: Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes. *J. Funct. Anal.* **205**(2), 414–463 (2003)
61. Sandier, E., Serfaty, S.: 2D Coulomb gases and the renormalized energy. *Ann. Probab.* **43**(4), 2026–2083 (2015)
62. Sandier, E., Serfaty, S.: 1D log gases and the renormalized energy: crystallization at vanishing temperature. *Probab. Theory Relat. Fields* **162**(3–4), 795–846 (2015)
63. Saff, E.B., Totik, V.: *Logarithmic Potentials with External Fields*, vol. 316. Springer, Berlin (2013)
64. Sodin, M., Tsirelson, B.: Random complex zeroes, III. Decay of the hole probability. *Israel J. Math.* **147**(1), 371–379 (2005)
65. Tao, T., Vu, V.: Random matrices: The universality phenomenon for Wigner ensembles. *Mod. Aspects Random Matrix Theory* **72**, 121–172 (2012)
66. Tricomi, F.G.: *Integral Equations*, vol. 5. Courier Corporation, North Chelmsford (1957)
67. Zeitouni, O., Zelditch, S.: Large deviations of empirical measures of zeros of random polynomials. *Int. Math. Res. Not.* **2010**(20), 3935–3992 (2010)
68. Zelditch, S.: Large deviations of empirical measures of zeros on Riemann surfaces. *Int. Math. Res. Not.* **2013**(3), 592–664 (2013)