

Electron. J. Probab. **27** (2022), article no. 44, 1–27.  
ISSN: 1083-6489 <https://doi.org/10.1214/22-EJP767>

# Local semicircle law for Curie-Weiss type ensembles

Michael Fleermann\*    Werner Kirsch†    Thomas Kriecherbauer‡

## Abstract

We derive local semicircle laws for random matrices with exchangeable entries which exhibit correlations that decay slowly in the dimension  $N$  of the matrix. To be precise, any  $\ell$ -point correlation  $\mathbb{E}[Y_1 \cdots Y_\ell]$  between distinct matrix entries  $Y_1, \dots, Y_\ell$  may decay at a rate of only  $N^{-\ell/2}$ . We call our ensembles of *(high temperature) Curie-Weiss type*, and Curie-Weiss( $\beta$ )-distributed entries directly fit within our framework in the high temperature regime  $\beta \in [0, 1]$ . Using rank-one perturbations, we show that even in the low-temperature regime  $\beta \in (1, \infty)$ , where  $\ell$ -point correlations survive in the limit, the local semicircle law still holds after rescaling the matrix entries with a constant which depends on  $\beta$  but not on  $N$ .

**Keywords:** random matrix; local semicircle law; exchangeable entries; correlated entries; Curie-Weiss entries.

**MSC2020 subject classifications:** 60B20.

Submitted to EJP on June 1, 2021, final version accepted on March 18, 2022.

## 1 Introduction

The local semicircle law is a relatively recent result that was derived to gain a more detailed understanding of the convergence of the empirical spectral distributions (ESDs) of random matrices to the semicircle distribution. Further, it was also used to establish universality results for Wigner matrices. A common formulation of this type of theorem is a uniform alignment of the Stieltjes transforms of the ESDs  $\sigma_N$  and the semicircle distribution  $\sigma$ , see [6], for example. Another formulation of the local law is as follows, cf. [34]: For any sequence of intervals  $(I_N)_N$ , whose diameter is not decaying to zero too quickly,  $\sigma_N(I_N)$  can be well approximated by  $\sigma(I_N)$  for large  $N$ . In fact, the second formulation of the local law will follow from the first, as we will show further below in Theorem A.8. And it is precisely this second formulation which lends the local law its name: Even when zooming in onto smaller and smaller intervals, the ESDs are well-approximated by the semicircle distribution (see also [20] for a translation of this convergence concept to the setting of classical probability theory).

\*FernUniversität in Hagen, Germany. E-mail: michael.fleermann@fernuni-hagen.de

†FernUniversität in Hagen, Germany. E-mail: werner.kirsch@fernuni-hagen.de

‡Universität Bayreuth, Germany. E-mail: thomas.kriecherbauer@uni-bayreuth.de

Although there were some previous results into the direction of a local law in [26] and [14], it is safe to say that on the level of strength available today, it was established by Erdős, Schlein and Yau in [13] and by Tao and Vu in [33]. Ever since, the results were strengthened (see [23] and [22], for example) and proof layouts were refined to make the theory more accessible to a broader audience. Indeed, the local laws are displayed in a pedagogical manner in the text [6] by Benaych-Georges and Knowles and the book [15] by Erdős and Yau. Both of these texts have their roots in the joint publication [11].

As the semicircle law itself, the local semicircle law was initially considered for random matrices with independent and identically distributed entries, see [13]. After the seminal work [13] the local semicircle law was established rapidly for more and more general classes of complex Hermitian or real symmetric random matrices with independent entries. For example, in [11] such a result was derived under the assumptions that the entries are centered, that the variances of the matrix entries sum up to 1 along rows and columns, and that certain upper bounds on these variances and on higher moments of the entries are satisfied. A few years later, [3] provided a local law in an even more general setting where stochasticity of the matrix of variances is not required. In this situation the limit law is generally not the semicircle. Instead, its Stieltjes transformation is determined by the quadratic vector equation (see [3] and references therein, in particular [27, 32] where the importance of the quadratic vector equation was already observed).

In a next step it is reasonable to relax the independence condition. A natural way to do this is to assume that correlations between matrix entries are decaying with a growing distance within the matrix (modulo symmetry of the matrix). With respect to local laws, the following results for such ensembles can be found in the literature: In [1], the local law was proved for random matrices with correlated Gaussian entries, where the covariance matrix is assumed to possess a certain translation invariant structure. In [2], ensembles with correlated entries were considered, where the correlation decays arbitrarily polynomially fast in the distance of the entries. This result has been improved by [12] (who reference an older preprint version of [2]), where fast polynomial decay is assumed only for entries outside of neighborhoods of a size growing slower than  $\sqrt{N}$ , and a slower correlation decay between entries within these neighborhoods. Another correlation structure was analyzed in [8], where correlation was only allowed for entries close to each other and independence was assumed otherwise. What all four mentioned publications have in common is that the local *semicircle* law is not the main object of interest, but rather the existence of *some* local limit.

Another way to relax the independence assumption is to assume exchangeability for the matrix entries. In order to discuss such classes of models we begin by reminding the reader of some results on the level of the *global law* that are relevant for our results. In [30] the random matrices are filled from an infinite sequence of exchangeable random variables. Observe that for this class of randomness there is no decay of correlations neither with respect to some spatial structure nor with respect to the dimension of the matrix. As a consequence the limiting spectral behavior has a new feature even on the level of the *global law*. In [30] a semicircle law is proved for such matrices, however as a rule with a *support that is random*. This has to do with the fact that the arithmetic mean of an (infinite) exchangeable sequence of random variables converges in distribution to a measure  $\mu$  which is *not* concentrated in a single point except for the case of independent variables. De Finetti's theorem states that the distribution of an infinite sequence of exchangeable random variables is in fact a mixture of product measures with respect to the measure  $\mu$  (see e. g. [25] for the general case and [29] for the  $\{0, 1\}$ -valued case). We call  $\mu$  the *de-Finetti measure*.

In the papers [24] and [31] the authors follow a path interpolating between indepen-

dence and exchangeability in the above sense by considering exchangeable ( $\{-1, 1\}$ -valued) random variables for each matrix but with a de-Finetti measure  $\mu_N$  which depends on the dimension  $N$  of the matrix. The measures  $\mu_N$  they consider are more and more concentrated near a single point [24] or near two points [31]. In fact, their main example is a Curie-Weiss process, described by a Gibbs measure which is well-known in statistical physics. This process marks a threshold between independence and the full exchangeability in the following sense: Depending on a parameter of the model,  $\beta$  in physics interpreted as an inverse temperature, the arithmetic mean of the Curie-Weiss ensemble converges in distribution to a Dirac measure concentrated in a single point ( $\beta \leq 1$ ) or to a measure concentrated in two points ( $\beta > 1$ ). This establishes a phase transition in the behavior of the model, in a sense from ‘near independent’ to ‘strongly correlated’. The model exhibits correlation between the random variables which is decaying for  $\beta \leq 1$  with increasing dimension  $N$  but is constant (due to exchangeability) within the random matrix of given dimension  $N$ . The papers [24] and [31] prove a *global* semicircle law in probability both for  $\beta \leq 1$  and for  $\beta > 1$ .

Before returning to local laws we mention a few more results on the global scale for related models. Limit laws for random matrices with Curie-Weiss spins were first proved in [21] where independent diagonals were filled with Curie-Weiss entries and a phase transition from the semicircle law to the limit laws of random Toeplitz matrices could be shown at the critical temperature ( $\beta = 1$ ). The above mentioned results of [24, 31] were strengthened to almost sure convergence in [19] where, in addition, they were generalized to corresponding ensembles of band matrices. Also in [19], a correlated Gaussian ensemble fitting the framework of [24] was introduced, which has its roots in the analysis of [17]. In [7], the results in [19] were further improved and expanded, in particular allowing for block matrices with Curie-Weiss spins, where the limiting spectral distribution need not be the semicircle distribution anymore. In [18], Marchenko-Pastur and semicircle laws for sample covariance matrices with Curie-Weiss entries were derived.

Let us now return to the discussion of the local laws that are the main focus of the present paper. It continues the analysis of the first author in [16], where he answered a question of the second author that was in part motivated by the desire to establish a local law for ensembles with Curie-Weiss distributed entries studied in [24], see Example 2.8. Our investigations led us to introduce matrix ensembles of *Curie-Weiss type*, see Definition 2.7. These ensembles share two important features with Curie-Weiss ensembles in the high temperature regime ( $\beta \leq 1$ ). First, there is slow decay of correlations as a function of the matrix dimension  $N$ . In fact, it follows from condition (2.2) of Definition 2.7 that the  $\ell$ -point correlation  $\mathbb{E}[Y_1 \cdots Y_\ell]$  between any  $\ell$  distinct matrix entries in the upper right half of the matrix may decay at a rate of order  $N^{-\ell/2}$ , a rate that is actually assumed in the critical case  $\beta = 1$ . Note again that there is no decay with respect to a spatial structure so that the results of [1, 2, 8, 12] do not apply to our models. Instead, and this is the second property that Curie-Weiss type ensembles share with Curie-Weiss ensembles, we require for each  $N$  that the matrix entries have a de-Finetti representation (i.e. a representation as a mixture of product measures). Just for the sake of clarity we remind the reader that de Finetti’s theorem applies to infinite sequences of random variables. For finite sequences of random variables, exchangeability is a weaker condition than assuming a de-Finetti representation as we do (cf. [4]).

A crucial aspect of our definition of Curie-Weiss type ensembles is that it also allows to treat the low temperature case ( $\beta > 1$ ) in which the correlations do not decay at all. In this case we can apply a (random) rank 1 perturbation that leads to an auxiliary matrix ensemble with decaying correlations but without the Curie-Weiss property of having spin entries  $\pm 1$ . Therefore, additional conditions (2.3) and (2.5) in combination with

(2.6) need to be introduced that are trivially satisfied in the spin case. These conditions are then used to prove certain large-deviation inequalities, see Theorems 3.2 and 3.3, that are central in the proof of the local law. In order to indicate that Curie-Weiss type ensembles might also be useful in other contexts we present in Example 2.9 a specific Gaussian matrix ensemble with correlated entries that can be shown to be of Curie-Weiss type.

We close the Introduction by describing the plan of the paper with an outline of the main idea of the proof.

Section 2 begins with a sequence of definitions that lead to our main object of study, the ensembles of Curie-Weiss type introduced in Definition 2.7. We then argue that Curie-Weiss( $\beta$ ) ensembles fall into this class for  $0 \leq \beta \leq 1$  (Example 2.8) as well as a specific family of correlated Gaussian ensembles (Example 2.9). After recalling some basic concepts that have proved to be useful in establishing local laws, we formulate with Theorem 2.12 our main result, a weak local law for Curie-Weiss type ensembles. The remainder of Section 2 is then devoted to formulate and prove a weak local law for Curie-Weiss( $\beta$ ) ensembles with  $\beta > 1$  (Theorem 2.15), where correlations do not decay with the matrix dimension  $N$  and an auxiliary ensemble is introduced by a (random) rank 1 perturbation that allows the application of Theorem 2.12.

Section 3 contains the proof of our main result Theorem 2.12. Our proof follows closely the arguments presented in [6] for the case of independent entries, in particular the proof of the weak local law in [6, Section 5]. As the strategy of proof is described lucidly in [6, Section 4] we now focus on the differences between the present paper and [6]. Crucial in the proof of the local law are the estimates contained in [6, Lemma 5.4] that correspond to our Lemma 3.1. The assumption of centeredness and independence of the matrix entries enters the proof of [6, Lemma 5.4] via the large deviation bounds provided by [6, Lemma 3.6]. It is here where we need to deviate from [6]. With Theorem 3.2 we remove the assumption of centeredness which is necessary as in our de-Finetti representation we have a mixture of independent ensembles that are generally not centered. Theorem 3.3 then generalizes from independent entries to entries of de-Finetti type and it is in the proof of this theorem where the conditions (2.2) - (2.6) of Definition 2.7 are used and their origin can be understood.

In the Appendix, we present various extensions and corollaries of Theorem 2.12. With help of the general Lemma A.1, we extend the uniformness of Theorem 2.12 in Theorem A.2. We use this result in combination with Lemma A.5 to prove Theorem A.6, which analyzes the approximation of the semicircle density by a kernel density estimate which is based on the empirical spectral distribution. Lastly, in Theorem A.7 and Theorem A.8 we analyze absolute and relative differences of interval probabilities of the empirical spectral distributions and the semicircle distribution.

## 2 Setup and main results

### 2.1 Ensembles of Curie-Weiss type

We will first explain some notation and introduce random matrices of Curie-Weiss type. The expectation operator  $\mathbb{E}$  will always denote the expectation with respect to a generic probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Euclidian spaces  $\mathbb{R}^n$  will always be equipped with Borel- $\sigma$ -algebras induced by the standard topology. The space  $\mathcal{M}_1(\mathbb{R})$  of all probability measures on  $\mathbb{R}$  will be equipped with the topology of weak convergence and the associated Borel  $\sigma$ -algebra. In addition, probability spaces with finite sample space will always be equipped with the power set as  $\sigma$ -algebra. If  $I$  is an index set and for all  $i \in I$ ,  $Z_i$  is a mathematical object, then we write  $Z_I := (Z_i)_{i \in I}$ . On the other hand, if for all  $i \in I$ ,  $M_i$  is a set, then we write  $M^I := \prod_{i \in I} M_i$  as the cartesian product. Lastly, if we write  $a = a(b)$ , where  $a$  is

an expression and  $b$  is a parameter vector, then this means that  $a$  depends on the choice of  $b$ . The following definition is based on [28] and [31].

**Definition 2.1.** Let  $I$  be a finite index set and  $Y_I$  be a family of  $\mathbb{R}$ -valued random variables on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Then the random vector  $Y_I$  is called of de-Finetti type, if there is a probability space  $(T, \mathcal{T}, \mu)$  and a measurable mapping

$$\begin{aligned} P : (T, \mathcal{T}) &\longrightarrow \mathcal{M}_1(\mathbb{R}) \\ t &\longmapsto P_t \end{aligned}$$

such that for all measurable sets  $B \subseteq \mathbb{R}^I$ , we find

$$\mathbb{P}(Y_I \in B) = \int_T P_t^{\otimes I}(B) d\mu(t), \quad (2.1)$$

where  $P_t^{\otimes I} := \otimes_{i \in I} P_t$  is the  $I$ -fold product measure on  $\mathbb{R}^I$ .

It should be noted that for a random vector  $Y_I$  to be of de-Finetti type is solely a property of the *distribution* of  $Y_I$  and not a property of the probability space on which  $Y_I$  is defined. To be more precise, it means that the push-forward distribution  $\mathbb{P}^{Y_I}$  is a mixture of product distributions  $P_t^{\otimes I}$ ,  $t \in T$ . In this context,  $\mu$  is also called *mixing distribution* or simply *mixture*. We will also call  $(T, \mathcal{T}, \mu, P)$  *mixing space*. Further properties of de-Finetti type variables are clarified in the following remark:

**Remark 2.2.** Let  $Y_I$  be of de-Finetti type as in Definition 2.1, then we observe:

1. For any subset  $J \subseteq I$ ,  $Y_J$  is of de-Finetti type with respect to the same mixing space  $(T, \mathcal{T}, \mu, P)$ .
2. For any  $t \in T$ , the coordinates of the identity map on  $(\mathbb{R}^I, P_t^{\otimes I})$  are i.i.d.  $P_t$ -distributed.
3. The random variables  $Y_I$  are *exchangeable*, that is, if  $\pi : I \rightarrow I$  is a bijection, then  $(Y_i)_{i \in I}$  and  $(Y_{\pi(i)})_{i \in I}$  have the same distribution.

**Lemma 2.3.** Let  $Y_I$  be of de-Finetti type with respect to the mixing space  $(T, \mathcal{T}, \mu, P)$ . Then it holds for any measurable function  $F : \mathbb{R}^I \rightarrow \mathbb{C}$ :

$$\mathbb{E}F(Y_I) = \int_T \int_{\mathbb{R}^I} F(y_I) dP_t^{\otimes I}(y_I) d\mu(t),$$

where the left-hand side of the equation is well-defined iff the right-hand side is.

*Proof.* The statement follows by standard arguments: The claim is easily verified for step functions of the form  $F = \sum_{k=1}^K \alpha_k \mathbb{1}_{A_k}$ , where  $K \in \mathbb{N}$ ,  $\alpha_k \geq 0$  and  $A_k \subseteq \mathbb{R}^I$  are measurable. The case for  $F \geq 0$  is then concluded via Beppo-Levi. The  $\mathbb{R}$ -valued case is seen by decomposing  $F = F_+ - F_-$ , and the final  $\mathbb{C}$ -valued case is then shown by decomposing  $F = \operatorname{Re} F + i \operatorname{Im} F$ .  $\square$

A prominent example of random variables of de-Finetti type is given by Curie-Weiss spins:

**Definition 2.4.** Let  $M \in \mathbb{N}$  be arbitrary and  $Y_1, \dots, Y_M$  be random variables defined on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Let  $\beta \geq 0$ , then we say that  $Y_1, \dots, Y_M$  are Curie-Weiss( $\beta, M$ )-distributed, if for all  $y_1, \dots, y_M \in \{-1, 1\}$  we have that

$$\mathbb{P}(Y_1 = y_1, \dots, Y_M = y_M) = \frac{1}{Z_{\beta, M}} \cdot e^{\frac{\beta}{2M} (\sum y_i)^2},$$

where  $Z_{\beta, M}$  is a normalization constant. The parameter  $\beta$  is called inverse temperature.

The Curie-Weiss( $\beta, M$ ) distribution is used to model the behavior of  $M$  ferromagnetic particles (spins) at the inverse temperature  $\beta$ . At low temperatures, that is, if  $\beta$  is large, all magnetic spins are likely to have the same alignment, resembling a strong magnetic effect. In contrast, at high temperatures (if  $\beta$  is small), spins can act almost independently, resembling a weak magnetic effect. At infinitely high temperature, that is, if  $\beta = 0$ , Curie-Weiss spins are simply i.i.d. Rademacher distributed random variables. For details on the Curie-Weiss model we refer to [9], [35] and [28]. The Curie-Weiss distribution is an important model in statistical mechanics. It is exactly solvable and features a phase transition at  $\beta = 1$ . The behavior of Curie-Weiss spins differs significantly in the regimes  $\beta = 0$ ,  $\beta \in (0, 1)$ ,  $\beta = 1$  and  $\beta \in (1, \infty)$ , as exemplified by the next lemma. In particular, we will see exactly at which speed  $\ell$ -point correlations between Curie-Weiss spins decay, and that for  $\beta > 1$  these correlations do not vanish at all:

**Lemma 2.5.** Fix  $\ell \in \mathbb{N}$  and let for all  $M \geq \ell$ ,  $(Y_1^{(M)}, \dots, Y_\ell^{(M)})$  be part of a Curie-Weiss( $\beta, M$ ) distributed random vector. If  $\ell$  is even, the following statements hold:

i) If  $\beta = 0$ , then  $\mathbb{E}Y_1^{(M)} \dots Y_\ell^{(M)} = 0$ .

ii) If  $\beta \in (0, 1)$ , then for some constant  $c = c(\beta, \ell) > 0$ :

$$\mathbb{E}Y_1^{(M)} \dots Y_\ell^{(M)} \sim cM^{-\ell/2} \text{ as } M \rightarrow \infty.$$

iii) If  $\beta = 1$ , then for some constant  $c = c(\ell) > 0$ :

$$\mathbb{E}Y_1^{(M)} \dots Y_\ell^{(M)} \sim cM^{-\ell/4} \text{ as } M \rightarrow \infty.$$

iv) If  $\beta \in (1, \infty)$ , then

$$\mathbb{E}Y_1^{(M)} \dots Y_\ell^{(M)} \sim c^\ell$$

as  $M \rightarrow \infty$ , where  $c = c(\beta) \in (0, 1)$  is the unique positive number such that  $\tanh(\beta c) = c$ .

If  $\ell$  is odd, then for all  $\beta \geq 0$  one has  $\mathbb{E}Y_1^{(M)} \dots Y_\ell^{(M)} = 0$ .

*Proof.* See Theorem 5.17 in [28].  $\square$

The next theorem shows that the discrete distribution of Curie-Weiss spins has a de-Finetti representation in the sense of Definition 2.1.

**Theorem 2.6.** If  $Y_1, \dots, Y_M$  are Curie-Weiss( $\beta, M$ )-distributed with  $\beta \geq 0$ , then they are of de-Finetti type with respect to the mixing space  $((-1, 1), \mathcal{B}_{(-1,1)}, \mu_M^\beta, P)$ , where

$$P : (-1, 1) \longrightarrow \mathcal{M}_1(\mathbb{R})$$

$$t \longmapsto P_t = \frac{1-t}{2}\delta_{-1} + \frac{1+t}{2}\delta_1.$$

Here,  $\mathcal{B}_{(-1,1)}$  denotes the Borel  $\sigma$ -algebra over the interval  $(-1, 1)$  and  $\mu_M^\beta$  is the Dirac measure  $\delta_0$  for  $\beta = 0$ , whereas if  $\beta > 0$ ,  $\mu_M^\beta$  is the Lebesgue-continuous probability distribution with density on  $(-1, 1)$  given by

$$t \mapsto f_M(t) := C \cdot \frac{e^{-\frac{M}{2}F_\beta(t)}}{1-t^2} \mathbf{1}_{(-1,1)}(t),$$

where  $C = C(\beta, M)$  is a normalization constant and for all  $t \in (-1, 1)$  we define

$$F_\beta(t) := \frac{1}{\beta} \left( \frac{1}{2} \ln \left( \frac{1+t}{1-t} \right) \right)^2 + \ln(1-t^2).$$

Further, if  $\beta \leq 1$ , the mixtures  $(\mu_M^\beta)_{M \in \mathbb{N}}$  satisfy the following moment decay:

$$\forall p \in 2\mathbb{N} : \int_{(-1,+1)} t^p d\mu_M^\beta(t) \leq \frac{K_{\beta,p}}{M^{\frac{p}{4}}},$$

where  $K_{\beta,p} \in \mathbb{R}_+$  is a constant that depends on  $\beta$  and  $p$  only.

*Proof.* This was shown rigorously in [28], see Theorem 5.6, Remark 5.7, Proposition 5.9 and Theorem 5.17 in their text.  $\square$

The Curie-Weiss type ensembles (sequences of random matrices) which we study in this paper are defined as follows:

**Definition 2.7.** An ensemble of real symmetric random matrices  $N \times N$  matrices  $(H_N)_N$  is called of (high temperature) Curie-Weiss type, if:

a) For all  $N \in \mathbb{N}$  it holds

$$(H_N(i,j))_{1 \leq i \leq j \leq N} = \left( \frac{1}{\sqrt{N}} X_N(i,j) \right)_{1 \leq i \leq j \leq N},$$

where  $(X_N(i,j))_{1 \leq i \leq j \leq N}$  is of de-Finetti type with respect to some mixing space  $(T_N, \mathcal{T}_N, \mu_N, P^{(N)})$ .

b) Set for all  $\ell, N \in \mathbb{N}$  and  $t \in T_N$ ,  $m_N^{(\ell)}(t) := \int_{\mathbb{R}} x^\ell dP_t^{(N)}(x)$  the  $\ell$ -th moment of  $P_t^{(N)}$ . Then it holds:

$$\forall p \in 2\mathbb{N} : \exists K_p \in \mathbb{R}_+ : \forall N \in \mathbb{N} :$$

$$\text{First moment condition: } \int_{T_N} |m_N^{(1)}(t)|^p d\mu_N(t) \leq \frac{K_p}{N^{\frac{p}{2}}} \quad (2.2)$$

$$\text{Second moment condition: } \int_{T_N} |1 - m_N^{(2)}(t)|^p d\mu_N(t) \leq \frac{K_p}{N^{\frac{p}{2}}} \quad (2.3)$$

$$\text{Central first moment condition: } \forall t \in T_N : \int_{\mathbb{R}} |y - m_N^{(1)}(t)|^p dP_t^{(N)}(y) \leq K_p(t) \quad (2.4)$$

$$\text{Central second moment condition: } \forall t \in T_N : \int_{\mathbb{R}} |y^2 - m_N^{(2)}(t)|^p dP_t^{(N)}(y) \leq K_p(t) \quad (2.5)$$

where  $K_p(t) = K_p^{(N)}(t)$  is a family of constants satisfying

$$\forall p \in 2\mathbb{N} : \forall N \in \mathbb{N} : \int_{T_N} K_p(t) d\mu_N(t) \leq K_p \quad \text{and} \quad \int_{T_N} K_p^2(t) d\mu_N(t) \leq K_p. \quad (2.6)$$

Notationally, for the remainder of this paper, we set  $[N] := \{1, \dots, N\}$  for all  $N \in \mathbb{N}$ .

**Example 2.8.** Let  $0 \leq \beta$  be arbitrary and let for each  $N \in \mathbb{N}$  the random variables  $(\tilde{X}_N(i,j))_{i,j \in [N]}$  be Curie-Weiss( $\beta, N^2$ )-distributed. Define the Curie-Weiss( $\beta$ ) ensemble  $(H_N)_N$  by setting

$$\forall N \in \mathbb{N} : \forall (i,j) \in [N]^2 : H_N(i,j) = \begin{cases} \frac{1}{\sqrt{N}} \tilde{X}_N(i,j) & \text{if } i \leq j \\ \frac{1}{\sqrt{N}} \tilde{X}_N(j,i) & \text{if } i > j. \end{cases}$$

If  $\beta \in [0, 1]$ , by Theorem 2.6,  $(H_N)_N$  is an ensemble of Curie-Weiss type with mixtures  $(\mu_N)_N := (\mu_{N^2}^\beta)_N$ . To see this, condition a) in Definition 2.7 is clear by construction

where the spaces  $(T_N, \mathcal{T}_N)$  and the map  $P^{(N)}$  are the same for all  $N$ , only the mixture  $\mu_N$  changes with  $N$ . For condition b), note that  $m_N^{(1)}(t) = t$  and  $m_N^{(2)}(t) = 1$  for all  $N \in \mathbb{N}$  and  $t \in (-1, 1)$ . So by Lemmas 2.3 and 2.5, Conditions (2.2), (2.3), (2.4), (2.5), and (2.6) are satisfied.

**Example 2.9.** In this example we shall see that specific families of correlated Gaussian ensembles easily fit within the framework of Definition 2.7. Fix some number  $\alpha \geq 1$  and let  $(X_N(i, j))_{1 \leq i \leq j \leq N}$  be of de-Finetti type with respect to the mixing space  $(T_N, \mathcal{T}_N, \mu_N, P^{(N)}) = (\mathbb{R}, \mathcal{B}, \mathcal{N}(0, N^{-\alpha}), (P_t^{(N)})_{t \in \mathbb{R}})$ , where  $P_t^{(N)} = P_t = \mathcal{N}(t, 1)$ . In other words, we have a two-step random experiment where in the first step we choose a common mean  $t \in \mathbb{R}$  according to  $\mathcal{N}(0, N^{-\alpha})$ , and in the second step we choose  $N(N+1)/2$  i.i.d. realizations of  $\mathcal{N}(t, 1)$ . Then  $m_N^{(1)}(t) = t$ ,  $m_N^{(2)}(t) = 1 + t^2$  and for  $p \in 2\mathbb{N}$ , verifying the conditions of Definition 2.7:

$$\begin{aligned}
 (2.2) \quad & \int_{\mathbb{R}} |m_N^{(1)}(t)|^p d\mu_N(t) = \int_{\mathbb{R}} t^p d\mathcal{N}(0, N^{-\alpha})(t) = \frac{(p-1)!!}{N^{\frac{\alpha p}{2}}} \\
 (2.3) \quad & \int_{\mathbb{R}} |1 - m_N^{(2)}(t)|^p d\mu_N(t) = \int_{\mathbb{R}} t^{2p} d\mathcal{N}(0, N^{-\alpha})(t) = \frac{(2p-1)!!}{N^{\alpha p}} \\
 (2.4) \quad & \int_{\mathbb{R}} |y - m_N^{(1)}(t)|^p dP_t(y) = \int_{\mathbb{R}} (y - t)^p d\mathcal{N}(t, 1)(y) = (p-1)!! \\
 (2.5) \quad & \int_{\mathbb{R}} |y^2 - m_N^{(2)}(t)|^p dP_t(y) = \int_{\mathbb{R}} (y^2 - t^2 - 1)^p d\mathcal{N}(t, 1)(y) \\
 & \leq \int_{\mathbb{R}} 4^p (y^{2p} + t^{2p} + 1) d\mathcal{N}(0, 1)(y) = 4^p (2p-1)!! + 4^p t^{2p} + 4^p \\
 (2.6) \quad & \int_{\mathbb{R}} (p-1)!! d\mathcal{N}(0, N^{-\alpha})(t) = (p-1)!! \quad \text{and} \\
 & \int_{\mathbb{R}} 4^p (2p-1)!! + 4^p t^{2p} + 4^p d\mathcal{N}(0, N^{-\alpha})(t) \leq 3 \cdot 4^p (2p-1)!!
 \end{aligned}$$

We conclude that with the choice for the constant  $K_p := 3 \cdot 4^p (2p-1)!!$ , all conditions of Definition 2.7 are satisfied. As it turns out, the distribution of  $X := (X_N(i, j))_{1 \leq i \leq j \leq N}$  can be identified as a multivariate correlated Gaussian ensemble. To this end, set  $d := N(N+1)/2$  and let  $\lambda \in \mathbb{R}^d$  be arbitrary. We calculate

$$\begin{aligned}
 \mathbb{E} \exp(i \langle \lambda | X \rangle) &= \int_{\mathbb{R}} \int_{\mathbb{R}^d} \exp(i \langle \lambda | x \rangle) dP_t^{\otimes d}(x) d\mathcal{N}(0, N^{-\alpha})(t) \\
 &= \int_{\mathbb{R}} \exp \left( it \sum \lambda_i - \frac{1}{2} \sum \lambda_i^2 \right) d\mathcal{N}(0, N^{-\alpha})(t) \\
 &= \exp \left( -\frac{1}{2} \sum \lambda_i^2 \right) \exp \left( -(\sum \lambda_i)^2 \frac{1}{2N^\alpha} \right) \\
 &= \exp \left( -\frac{1}{2} \left\langle \left( I + \frac{1}{N^\alpha} \mathcal{E} \right) \lambda | \lambda \right\rangle \right),
 \end{aligned}$$

which is the characteristic function of a  $\mathbb{R}^d$  valued Gaussian correlated random vector with variance  $1 + N^{-\alpha}$  and covariances  $N^{-\alpha}$ . In our calculation, we used Lemma 2.3 in the first step, the characteristic function of  $\mathcal{N}(t, 1)^{\otimes d}$  in the second, the characteristic function of  $\mathcal{N}(0, N^{-\alpha})$  in the third step, and in the last step  $I$  denotes the  $d \times d$  identity matrix and  $\mathcal{E}$  the  $d \times d$  matrix with entries 1.

## 2.2 Stochastic domination, Resolvents and Stieltjes transforms

For the statement of the local law and its proof we need the concepts of stochastic domination, resolvents and Stieltjes transforms. The first time the concept of stochastic



domination was used was in [10]. We will say that a statement which depends on  $N \in \mathbb{N}$  holds *v-finally*, where  $v$  is a parameter(-vector), if the statement holds for all  $N \geq N^*(v)$ .

**Definition 2.10.** Let  $X = X^{(N)}$  be a sequence of complex-valued and  $Y = Y^{(N)}$  be a sequence of non-negative random variables, then we say that  $X$  is stochastically dominated by  $Y$ , if for all  $\epsilon, D > 0$  there is a constant  $C_{\epsilon,D} \geq 0$  such that

$$\forall N \in \mathbb{N} : \mathbb{P} \left( |X^{(N)}| > N^\epsilon Y^{(N)} \right) \leq \frac{C_{\epsilon,D}}{N^D}.$$

In this case, we write  $X \prec Y$  or  $X^{(N)} \prec Y^{(N)}$ . If both  $X$  and  $Y$  depend on a possibly  $N$ -dependent index set  $U = U^{(N)}$ , such that  $X = (X^{(N)}(u), N \in \mathbb{N}, u \in U^{(N)})$  and  $Y = (Y^{(N)}(u), N \in \mathbb{N}, u \in U^{(N)})$ , then we say that  $X$  is stochastically dominated by  $Y$  uniformly in  $u \in U$ , if for all  $\epsilon, D > 0$  we can find a  $C_{\epsilon,D} \geq 0$  such that

$$\forall N \in \mathbb{N} : \sup_{u \in U^{(N)}} \mathbb{P} \left( |X^{(N)}(u)| > N^\epsilon Y^{(N)}(u) \right) \leq \frac{C_{\epsilon,D}}{N^D}. \quad (2.7)$$

In this case, we write  $X \prec Y$  or  $X(u) \prec Y(u), u \in U$  or  $X^{(N)}(u) \prec Y^{(N)}(u), u \in U^{(N)}$ , where the first version is used if  $U$  is clear from the context. In above situation, if all  $Y(u)$  are strictly positive, then we say that  $X$  is stochastically dominated by  $Y$ , simultaneously in  $u \in U$ , if for all  $\epsilon, D > 0$  we can find a  $C_{\epsilon,D} \geq 0$ , such that

$$\forall N \in \mathbb{N} : \mathbb{P} \left( \sup_{u \in U^{(N)}} \frac{|X^{(N)}(u)|}{Y^{(N)}(u)} > N^\epsilon \right) \leq \frac{C_{\epsilon,D}}{N^D},$$

and then we write  $\sup_{u \in U} |X(u)|/Y(u) \prec 1$  or  $\sup_{u \in U^{(N)}} |X^{(N)}(u)|/Y^{(N)}(u) \prec 1$ .

**Remark 2.11.** Simultaneous stochastic domination implies uniform stochastic domination (for the other direction, see Lemma A.1). Further, in order to show  $X \prec Y$ , it suffices to show that (2.7) holds for all  $\epsilon$  small enough, that is, for all  $\epsilon \in (0, \epsilon_0]$  for some  $\epsilon_0 > 0$ . In addition, it suffices to show (2.7) for  $(\epsilon, D)$ -finally all  $N \in \mathbb{N}$ .

Stochastic domination admits several important and intuitive rules of calculation. For example,  $\prec$  is transitive and reflexive, and if  $X_1 \prec Y_1$  and  $X_2 \prec Y_1$ , then both  $X_1 + X_2 \prec Y_1 + Y_2$  and  $X_1 \cdot X_2 \prec Y_1 \cdot Y_2$ . For more rules of calculation and their proofs, see e.g. [16]. In what follows, we will follow largely the notation in [6]. In particular, we will drop the index  $N$  from many – but not all –  $N$ -dependent quantities. Let  $H = H_N$  be an ensemble of Curie-Weiss type,  $z \in \mathbb{C} \setminus \mathbb{R}$ , then we denote by  $G(z) := (H - z)^{-1}$  its resolvent at  $z$ . The resolvent  $G$  of  $H$  carries all the spectral information of  $H$  which is contained in its empirical spectral distribution

$$\sigma = \sigma_N := \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}, \quad (2.8)$$

where  $\lambda_1, \dots, \lambda_N$  are the eigenvalues of  $H$ , which are all real-valued due to the symmetry of  $H$ . The relationship between  $G$  and  $\sigma$  is given by inspecting the Stieltjes transform  $s := S_\sigma$  of  $\sigma$ . In general, the Stieltjes transform  $S_\nu$  of a probability measure  $\nu$  on  $(\mathbb{R}, \mathcal{B})$  is given by the map

$$S_\nu : \mathbb{C}_+ \longrightarrow \mathbb{C}_+ \\ z \longmapsto \int_{\mathbb{R}} \frac{1}{x - z} d\nu(x),$$

so using (2.8) we obtain

$$s(z) = S_\sigma(z) = \int_{\mathbb{R}} \frac{1}{x - z} d\sigma(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z} = \frac{1}{N} \text{tr } G(z).$$

As  $N \rightarrow \infty$  we want to analyze the weak convergence behavior of  $\sigma$  to the *semicircle distribution*  $\mu$ , which is the probability distribution on  $(\mathbb{R}, \mathcal{B})$  with Lebesgue density  $x \mapsto f_\sigma(x) := (2\pi)^{-1} \sqrt{(4 - x^2)_+}$ . We denote by  $m := S_\mu$  the Stieltjes transform of  $\mu$ . Then we obtain with [5, p. 32]:

$$\forall z \in \mathbb{C}_+ : m(z) = \frac{-z + \sqrt{z^2 - 4}}{2}.$$

### 2.3 Main results

We are now ready to state the main results of this paper. Notationally, whenever a  $z \in \mathbb{C}_+$  is considered, we set

$$\eta := \eta(z) := \operatorname{Im}(z), \quad E := E(z) := \operatorname{Re}(z) \quad \text{and} \quad \kappa := \kappa(z) := ||E| - 2|. \quad (2.9)$$

**Theorem 2.12.** Fix  $\tau \in (0, 1)$  and define the domains

$$\mathcal{D}_N(\tau) := \left[-\frac{1}{\tau}, \frac{1}{\tau}\right] + i \left[\frac{1}{N^{1-\tau}}, \frac{1}{\tau}\right] \quad \text{and} \quad \mathcal{D}_N^*(\tau) := [-2 + \tau, 2 - \tau] + i \left[\frac{1}{N^{1-\tau}}, \frac{1}{\tau}\right].$$

Let  $H$  be a Curie-Weiss type ensemble,  $G(z) = (H - z)^{-1}$  and

$$\Lambda(z) := \max_{i,j} |G_{ij}(z) - m(z)\delta_{ij}|.$$

Then it holds

$$\max(\Lambda(z), |s(z) - m(z)|) \prec \frac{\frac{1}{\sqrt{N\eta}}}{\sqrt{\kappa + \eta + \frac{1}{\sqrt{N\eta}}}}, \quad z \in \mathcal{D}_N(\tau) \quad (2.10)$$

so in particular

$$\max(\Lambda(z), |s(z) - m(z)|) \prec \frac{1}{\sqrt{N\eta}}, \quad z \in \mathcal{D}_N^*(\tau). \quad (2.11)$$

Note that each (2.10) and (2.11) are to be viewed as two separate statements in that each of the terms in the maximum is dominated by the error term on the right hand side. By properties of  $\prec$ , this is equivalent to the maximum being dominated. For corollaries and many implications of Theorem 2.12, we refer the reader to Appendix A.

**Remark 2.13.** In the literature, the statement of the form of Theorem 2.12 is called *weak local law* – see Proposition 5.1 in [6] and Theorem 7.1 in [15] – since in the study of independent entries, smaller error bounds are known to hold (except for the term  $\Lambda(z)$  in (2.11)). The authors of the current paper plan to derive such stronger results also for Curie-Weiss type ensembles. It should be noted that our error term is slightly smaller than those in the cited statements, since

$$\frac{\frac{1}{\sqrt{N\eta}}}{\sqrt{\kappa + \eta + \frac{1}{\sqrt{N\eta}}}} \leq \frac{1}{(N\eta)^{\frac{1}{4}}} \quad \text{and} \quad \frac{\frac{1}{\sqrt{N\eta}}}{\sqrt{\kappa + \eta + \frac{1}{\sqrt{N\eta}}}} \leq \frac{1}{\sqrt{N\eta\kappa}}. \quad (2.12)$$

However, the error term we use also appears naturally in the works of [6] and [15] who then chose to simplify it probably because it only serves there as an intermediate step to prove the strong local law. In the present paper, the bound in (2.10) is the final bound and we left it in this form.

**Corollary 2.14.** Let  $\beta \in [0, 1]$  and  $(H_N)_N$  be a Curie-Weiss( $\beta$ ) ensemble as in Example 2.8. Then as argued there,  $(H_N)_N$  is an ensemble of Curie-Weiss type. Therefore, the local law as in Theorem 2.12 holds for the Curie-Weiss( $\beta$ ) ensemble.

Next, we would like to analyze what can be said about the Curie-Weiss( $\beta$ ) ensemble  $(H_N)_N = (N^{-1/2}X_N)_N$  if  $\beta > 1$ . Here,  $\ell$ -point correlations  $\mathbb{E}X_N(i_1, j_1) \cdots X_N(i_\ell, j_\ell)$  – where  $X_N(i_1, j_1), \dots, X_N(i_\ell, j_\ell)$  are distinct random spins in  $X_N$  – do not vanish as  $N \rightarrow \infty$ . In [31] it was shown that the semicircle law holds in probability for the ensemble  $((1 - c(\beta)^2)^{-1/2}H_N)_N$ , where  $c(\beta) > 0$  defines the unique solution in  $(0, 1)$  of the equation  $\tanh(\beta c) = c$ . Additionally, by the work in [19] it immediately follows that the semicircle law holds almost surely for  $((1 - c(\beta)^2)^{-1/2}H_N)_N$ . Now, the question is whether the limit law also holds locally for  $((1 - c(\beta)^2)^{-1/2}H_N)_N$ .

**Theorem 2.15.** *Let  $\beta > 1$  and  $(H_N)_N$  be a Curie-Weiss( $\beta$ ) ensemble as in Example 2.8. Then for the rescaled ensemble  $((1 - c(\beta)^2)^{-1/2}H_N)_N$  the local semicircle law holds, that is,*

$$|s(z) - m(z)| \prec \frac{\frac{1}{\sqrt{N\eta}}}{\sqrt{\kappa + \eta + \frac{1}{\sqrt{N\eta}}}}, \quad z \in \mathcal{D}_N(\tau) \quad (2.13)$$

as well as

$$|s(z) - m(z)| \prec \frac{1}{\sqrt{N\eta}}, \quad z \in \mathcal{D}_N^*(\tau). \quad (2.14)$$

Theorem 2.15 is proved by showing that a rank-1 perturbation of  $(H_N)_N$  is, in fact, of Curie-Weiss type as in Definition 2.7, and by seeing that a rank-1 perturbation does not affect the local law for  $|s(z) - m(z)|$ .

Let us explain the heuristics behind the proof of Theorem 2.15, using the notation of Theorem 2.6. If  $\beta > 1$ , it is well-known that the mixing distribution  $\mu_N^\beta$  will converge weakly to the probability measure  $1/2 \cdot \delta_{-c} + 1/2 \cdot \delta_c$  where  $c = c(\beta)$  is as above. As a result, for large  $N$  the entries in the upper right triangle of  $X_N = \sqrt{N}H_N$ , where  $H_N$  is a Curie-Weiss( $\beta$ ) ensemble, are approximately either i.i.d.  $P_c$  or  $P_{-c}$  distributed (with corresponding means  $c$  resp.  $-c$  and variance  $1 - c^2$  in both cases), each with probability  $1/2$ , depending on if  $\mu_N^\beta$  drew  $c$  or  $-c$ . For each of these cases, we standardize  $X_N$  so that its upper right triangle contains i.i.d. standardized entries, resembling the Wigner case. To this end, denote by  $\mathcal{E}_N$  the  $N \times N$  matrix consisting entirely of ones, and set

$$Y_N := \frac{1}{\sqrt{1 - c^2}}(X_N - \mathbb{1}_{S_N > 0}c\mathcal{E}_N + \mathbb{1}_{S_N \leq 0}c\mathcal{E}_N),$$

where  $S_N$ , the sum of the spins of the upper right triangle of  $X_N$ , is a proxy to decide whether  $\mu_N^\beta$  tends to  $c$  or  $-c$ . Now for each realization,  $Y_N$  is just a rank 1 perturbation of  $(1 - c^2)^{-1/2}X_N$ , leaving the limiting spectral distribution of the  $N^{-1/2}$ -normalized ensemble unchanged. This was the initial approach in [31]. In our situation it is unclear whether  $Y_N$  is of de-Finetti type, so we do not know whether it is a Curie-Weiss type ensemble as in Definition 2.7. The solution is to find a different random variable to decide when to add or subtract  $c\mathcal{E}_N$ . The key idea now is that we use a very specific construction of the probability space on which our Curie-Weiss ensembles are defined. We use the product space  $\otimes_{N \in \mathbb{N}}(\mathcal{M}_N \otimes \mathcal{S}_N)$ , where  $\mathcal{M}_N := ((-1, 1), \mathcal{B}((-1, 1)), \mu_{N^2}^\beta)$  and  $\mathcal{S}_N$  equals  $\{\pm 1\}^{N \times N}$ , the latter equipped with kernels  $(P_t^{\otimes N^2})_{t \in (-1, 1)}$ . On each factor  $(\mathcal{M}_N \otimes \mathcal{S}_N)$  we define the probability measure  $\mu_{N^2}^\beta(dt) \otimes P_t^{\otimes N^2}$ , which is the product of a probability measure and a kernel. Denote by  $M_N^\beta$  resp.  $\tilde{X}_N$  the projection onto the first resp. second component of  $(\mathcal{M}_N \otimes \mathcal{S}_N)$ , then we obtain mixing variables  $M_N^\beta$  which is  $\mu_{N^2}^\beta$  distributed alongside Curie-Weiss( $\beta, N^2$ )-distributed random variables  $\tilde{X}_N$  which are utilized in Example 2.8 to produce the Curie-Weiss( $\beta$ ) ensemble  $X_N$ . Now we consider

$$Z_N := \frac{1}{\sqrt{1 - c^2}}(X_N - \mathbb{1}_{M_N^\beta > 0}c\mathcal{E}_N + \mathbb{1}_{M_N^\beta \leq 0}c\mathcal{E}_N).$$

The ensemble  $Z_N$  is, in fact, of de-Finetti type:

**Lemma 2.16.** Set  $Z_{ij} := Z_N(i, j)$  for all  $(i, j) \in I := \{1 \leq i \leq j \leq N\}$ . Then  $Z_I$  is of de-Finetti type with mixing space  $((-1, 1), \mathcal{B}_{(-1,1)}, \mu_{N^2}^\beta, \tilde{P})$ , where for all  $t \in (-1, 1)$  and with  $c = c(\beta)$  as above, we define

$$\tilde{P}_t := \begin{cases} \frac{1+t}{2} \delta_{\frac{1-c}{\sqrt{1-c^2}}} + \frac{1-t}{2} \delta_{\frac{-1-c}{\sqrt{1-c^2}}} & t > 0, \\ \frac{1+t}{2} \delta_{\frac{1+c}{\sqrt{1-c^2}}} + \frac{1-t}{2} \delta_{\frac{-1+c}{\sqrt{1-c^2}}} & t \leq 0. \end{cases}$$

Further, denoting by  $\tilde{m}^{(\ell)}(t) := \int_{\mathbb{R}} x^\ell d\tilde{P}_t$  the  $\ell$ -th moment of  $\tilde{P}_t$ , we obtain

$$\begin{aligned} \tilde{m}^{(1)}(t) &= \begin{cases} \frac{1}{\sqrt{1-c^2}}(t-c), & t > 0, \\ \frac{1}{\sqrt{1-c^2}}(t+c), & t \leq 0, \end{cases} \\ 1 - \tilde{m}^{(2)}(t) &= \begin{cases} \frac{2c}{1-c^2}(t-c), & t > 0, \\ \frac{-2c}{1-c^2}(t+c), & t \leq 0. \end{cases} \end{aligned}$$

*Proof.* For the duration of this proof, set  $X_{ij} := X_N(i, j)$  for all  $(i, j) \in I$ . We observe that  $Z_I$  takes values in

$$\left\{ \frac{\pm 1 - c}{\sqrt{1 - c^2}} \right\}^I \cup \left\{ \frac{\pm 1 + c}{\sqrt{1 - c^2}} \right\}^I.$$

Now let  $z_I$  be an arbitrary element in above set, w.l.o.g.  $z_I \in \{(\pm 1 + c)/\sqrt{1 - c^2}\}^I$ . Let  $y_I := \sqrt{1 - c^2} z_I$  and  $x_I := y_I - c \in \{\pm 1\}^I$ . Then

$$\begin{aligned} \mathbb{P}(Z_I = z_I) &= \mathbb{P}(X_I - c \mathbb{1}_{M_N^\beta > 0} + c \mathbb{1}_{M_N^\beta \leq 0} = y_I) \\ &= \int_{(-1,1)} \mathbb{P}(X_I - c \mathbb{1}_{M_N^\beta > 0} + c \mathbb{1}_{M_N^\beta \leq 0} = y_I | M_N^\beta = t) \mathbb{P}^{M_N^\beta}(dt) \\ &= \int_{(-1,0)} \mathbb{P}(X_I = x_I | M_N^\beta = t) \mathbb{P}^{M_N^\beta}(dt) = \int_{(-1,1)} \tilde{P}_t^{\otimes I}(z_I) \mu_{N^2}^\beta(dt), \end{aligned}$$

The moment calculations for  $\tilde{m}^{(1)}$  and  $1 - \tilde{m}^{(2)}$  are straightforward.  $\square$

**Lemma 2.17.** The ensemble  $N^{-1/2} Z_N$  – which is a rank one perturbation of  $((1 - c(\beta)^2)^{-1/2} H_N)_N$  – is a Curie-Weiss type ensemble as in Definition 2.7.

*Proof.* In Lemma 2.16 we have just shown that  $(Z_N(i, j))_{1 \leq i \leq j \leq N}$  is of de-Finetti type with mixture  $\mu_{N^2}^\beta$  as in Theorem 2.6, but with map  $t \mapsto \tilde{P}_t$  as in Lemma 2.16. Thus, condition a) of Definition 2.7 is satisfied. It remains to verify conditions (2.2), (2.3), (2.4), (2.5), and (2.6). Note that in our setting, only the mixing distribution  $\mu_{N^2}^\beta$  depends on  $N$ , but not the associated space  $T = T_N = (-1, 1)$ , nor the maps  $t \mapsto \tilde{P}_t$  and the moments  $\tilde{m}^{(l)}(t)$ . For the proof, we need two well-known facts about the distributions  $\mu_{N^2}^\beta$  on  $(-1, 1)$  when  $\beta > 1$ , see e.g. Lemma 6 in [31] (where  $c = c(\beta) \in (0, 1)$  such that  $\tanh(c\beta) = c$ ):

$$(CW1) \quad \exists C, \delta > 0 : \forall N \in \mathbb{N} : \mu_{N^2}^\beta([-c/2, c/2]) \leq C e^{-\delta N^2},$$

$$(CW2) \quad \forall \ell \in \mathbb{N} : \exists C_\ell > 0 : \forall N \in \mathbb{N} : \int_{c/2}^1 |t - c|^\ell \mu_{N^2}^\beta(dt) \leq \frac{C_\ell}{N^\ell}.$$

To show (2.2), we calculate for  $p \in 2\mathbb{N}$ :

$$\begin{aligned} \int_{(-1,1)} |\tilde{m}^{(1)}(t)|^p \mu_{N^2}^\beta(dt) &= 2 \int_{(0,1)} \left( \frac{t-c}{\sqrt{1-c^2}} \right)^p \mu_{N^2}^\beta(dt) \\ &\leq 2 \left( \frac{c}{\sqrt{1-c^2}} \right)^p C e^{-\delta N^2} + \frac{2C_p}{(\sqrt{1-c^2})^p N^p} \leq \frac{Const(p, c)}{N^p} \leq \frac{Const(p, c)}{N^{p/2}}, \end{aligned}$$

where in the first step, we used Lemma 2.16 and that the measure  $\mu_{N^2}^\beta$  is symmetric. In the second step, we split integration over  $(0, c/2)$  and  $(c/2, 1)$  and used (CW1) and (CW2). This shows (2.2), and (2.3) can be shown analogously, since again – by Lemma 2.16 – we basically integrate over  $|t - c|^p$  resp.  $|t + c|^p$ . Conditions (2.4) and (2.5) are satisfied since there is a compact subset of  $\mathbb{R}$  in which the support of every probability measure  $\tilde{P}_t$  is contained. Moreover, this implies that the bounds  $K_p(t)$  can be chosen independent of  $t$ , implying (2.6).  $\square$

The last ingredient for the proof of Theorem 2.15 is the following lemma:

**Lemma 2.18.** *Let  $Y$  be an Hermitian  $N \times N$  matrix,  $\mathcal{E}$  be an arbitrary Hermitian  $N \times N$  matrix of rank  $k$ . Then it holds for all  $z \in \mathbb{C}_+$ :*

$$|\operatorname{tr}[(Y - z)^{-1}] - \operatorname{tr}[(Y + \mathcal{E} - z)^{-1}]| \leq \frac{2k}{\eta}.$$

*Proof.* Unitary transformations of  $Y$  and  $Y + \mathcal{E}$  do not affect the l.h.s. of the statement, so we may assume  $\mathcal{E}(1, 1), \dots, \mathcal{E}(k, k) \neq 0$  but all other entries of  $\mathcal{E}$  vanish. We define for all  $i = 1, \dots, k$  the matrix  $\mathcal{E}_i$  such that  $\mathcal{E}_i(j, j) = \mathcal{E}(j, j)$  for all  $j \in \{1, \dots, i\}$  but all other entries of  $\mathcal{E}_i$  vanish. In particular,  $\mathcal{E}_k = \mathcal{E}$ . Further let  $\mathcal{E}_0$  be the matrix consisting entirely of zeros. Then – denoting for any  $N \times N$  matrix  $M$  and  $l \in \{1, \dots, N\}$  by  $M^{(l)}$  the  $l$ -th principal minor of  $M$  – we calculate

$$\begin{aligned} \operatorname{tr}[(Y - z)^{-1}] - \operatorname{tr}[(Y + \mathcal{E} - z)^{-1}] &= \sum_{i=0}^{k-1} (\operatorname{tr}[(Y + \mathcal{E}_i - z)^{-1}] - \operatorname{tr}[(Y + \mathcal{E}_{i+1} - z)^{-1}]) \\ &= \sum_{i=0}^{k-1} \left( \operatorname{tr}[(Y + \mathcal{E}_i - z)^{-1}] - \operatorname{tr}[(Y + \mathcal{E}_i)^{(i+1)} - z)^{-1}] \right) \\ &\quad + \sum_{i=0}^{k-1} \left( \operatorname{tr}[(Y + \mathcal{E}_{i+1})^{(i+1)} - z)^{-1}] - \operatorname{tr}[(Y + \mathcal{E}_{i+1} - z)^{-1}] \right) \end{aligned}$$

where for the second equality we used that  $(Y + \mathcal{E}_i)^{(i+1)} = (Y + \mathcal{E}_{i+1})^{(i+1)}$  for all  $i = 0, \dots, k-1$ . Taking absolute values, applying the triangle inequality and then the bound (A.1.12) in [5] yields the statement.  $\square$

*Proof of Theorem 2.15.* Let  $\beta > 1$  and  $(H_N)_N$  be Curie-Weiss( $\beta$ ) ensemble. Denote by  $s_N$  the Stieltjes transform corresponding to  $((1 - c(\beta)^2)^{-1/2} H_N)_N$  and by  $\tilde{s}_N$  the Stieltjes transformation corresponding to the ensemble  $N^{-1/2} Z_N$  as defined above, which is of Curie-Weiss type and a rank one perturbation of  $((1 - c(\beta)^2)^{-1/2} H_N)_N$  by Lemma 2.17.

$$|\tilde{s}_N(z) - s_N(z)| \prec \frac{1}{N\eta}, \quad z \in \mathbb{C}_+$$

by Lemma 2.18. The proof is concluded by using the estimates on  $|\tilde{s}_N - m|$  obtained by Theorem 2.12.  $\square$

### 3 Proof of Theorem 2.12

For the proof of Theorem 2.12, we follow the strategy used in [6] to prove their Proposition 5.1. Their proof works for independent entries, and it is a key observation that the ingredients which actually use the independence condition are exactly the so-called “large deviation bounds”, stated in Lemma 3.6 in [6].

To begin the proof, note that it suffices to show the statements in (2.10) and (2.11) for  $\Lambda(z)$ , since then by averaging, we obtain the  $\prec$  bounds for  $|s(z) - m(z)|$ , hence for

the maximum. Now we proceed along the lines of [6] and reveal the changes we made. We introduce the following notation: We write  $\Lambda_* := \max_{i \neq j} |G_{ij}|$ . Using the Schur complement formula, we obtain

$$\frac{1}{G_{ii}} = H_{ii} - z - \sum_{k,l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li}. \quad (3.1)$$

Here, if  $T \subseteq \{1, \dots, N\}$  is a subset, the sum  $\sum_{k,l}^{(T)}$  denotes the sum over all  $k, l \in \{1, \dots, N\} \setminus T$ , and  $G^{(T)}(z)$  denotes the resolvent of the matrix  $(H_{i,j})_{i,j \notin T}$  at  $z$ . We decompose the expression (3.1) as follows:

$$\frac{1}{G_{ii}} = -z - s + Y_i$$

where  $Y_i := H_{ii} + A_i - Z_i$  with  $Z_i := Z_i^{(1)} + Z_i^{(2)}$ , where

$$A_i := \frac{1}{N} \sum_k \frac{G_{ki} G_{ik}}{G_{ii}}, \quad Z_i^{(1)} := \sum_{k \neq l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li}, \quad Z_i^{(2)} := \sum_k^{(i)} \left( |H_{ik}|^2 - \frac{1}{N} \right) G_{kk}^{(i)}.$$

As it turns out in the analysis of the local law, the only problematic component of the error term  $Y_i$  is  $Z_i$ : Practically all the work the local law requires is to show the smallness of  $Z_i$ . In what follows we set

$$\phi := \mathbb{1}_{\Lambda \leq N^{-\tau/10}}.$$

The following lemma contains the main  $\prec$  estimates needed for the proof of Theorem 2.12, cf. Lemma 5.4 in [6].

**Lemma 3.1.** *In the above setting, we obtain*

$$(\phi + \mathbb{1}_{\eta \geq 1}) (\Lambda_* + |A_i| + |Z_i| + |G_{ii} - s|) \prec \sqrt{\frac{\operatorname{Im} m + |s - m|}{N\eta}}, \quad (3.2)$$

uniformly over all  $z \in \mathcal{D}_N(\tau)$  and  $i \in \{1, \dots, N\}$

Note that (3.2) consists of eight separate  $\prec$  statements. The proof of Lemma 3.1 can be conducted as in [6], but since we deal with correlated entries, we need new so called “large deviation bounds” as in Lemma 3.6 in [6] to deal with terms  $|Z_i|$ . Thus, the main work is to establish these bounds in our situation, which we carry out further below after the discussion of the remainder of the proof of Theorem 2.12 (culminating in Corollary 3.4). So assuming we have established Lemma 3.1, we would like to show

$$\Lambda(z) \prec \frac{\frac{1}{\sqrt{N\eta}}}{\sqrt{\kappa + \eta + \frac{1}{\sqrt{N\eta}}}}, \quad z \in \mathcal{D}_N(\tau). \quad (3.3)$$

This is achieved by adjusting the proof of Lemma 5.7 in [6] to our setting: We first establish the initial estimate

$$\Lambda \prec \frac{1}{\sqrt{N}}, \quad z \in \mathcal{D}_N(\tau) \cap \{z \in \mathbb{C}, \eta \geq 1\}, \quad (3.4)$$

which can be conducted as in Lemma 5.6 in [6]. Then, in a second step, we fix  $E \in [-\tau^{-1}, \tau^{-1}]$  and set  $\eta_k := 1 - kN^{-3}$  for all  $k = 0, 1, \dots, m(N) := \lfloor N^3 - N^{2+\tau} \rfloor$ . Then for all these  $k$  we find  $\eta_k \geq \eta_{m(N)} \geq 1 - (N^3 - N^{2+\tau})N^{-3} = N^{\tau-1}$ . Setting  $z_k := E + i\eta_k$  for all  $k \in \{0, 1, \dots, m(N)\}$  it suffices to show that

$$\Lambda(z_k) \prec \frac{\frac{1}{\sqrt{N\eta_k}}}{\sqrt{\kappa + \eta_k + \frac{1}{\sqrt{N\eta_k}}}}, \quad k \in \{0, 1, \dots, m(N)\}, \quad (3.5)$$

where the constants  $C_{\epsilon,D}$  do not depend on  $E$ . Then, by Lipschitzity of all terms involved, this establishes (3.3).

To show (3.5), pick  $\epsilon \in (0, \tau/16)$ ,  $D > 0$  and set  $\delta_k := (N\eta_k)^{-1/2}$ . Further, define the sets

$$\Xi_k := \left\{ \Lambda(z_k) \leq N^{3\epsilon} \frac{\delta_k}{\sqrt{\kappa + \eta_k + \delta_k}} \right\} \quad \text{and} \quad \Omega_k := \left\{ |s(z_k) - m(z_k)| \leq N^\epsilon \frac{\delta_k}{\sqrt{\kappa + \eta_k + \delta_k}} \right\}$$

Note that our sets  $\Xi_k$  deviate from the respective set in exposition in [6] to accomodate our choice of the error term, cf. Remark 2.13. However, the proof goes through as in [6] using the bootstrapping technique explained there: For each  $k \geq 1$ , it can be shown that  $\Xi_k \cap \Omega_k$  has high probability when conditioned on  $\Xi_{k-1} \cap \Omega_{k-1}$ . Here, we repeatedly use that the indicator  $\phi$  is active so that we can employ Lemma 3.1. In each step, only a negligible amount of probability is lost from the initial estimate (3.4), so that independently of our initial choice of  $E$ , we eventually observe

$$\sup_{k \in \{0,1,\dots,m(N)\}} \mathbb{P}(\Xi_k^c) \leq N^3(1+N^3) \frac{C_{\epsilon,D}}{N^D},$$

which establishes (3.5) and thus finishes the proof (modulo the large deviation estimates), since we may choose  $D$  arbitrarily large.

It is left to establish the large deviation bounds as in Lemma 3.6 in [6] to deal with terms  $|Z_i|$  in Lemma 3.1. We present the two-step approach developed in [16]. In the first step, our Theorem 3.2 generalizes Lemmas D.1, D.2 and D.3 in [6] to independent random variables with a common expectation  $t \in \mathbb{C}$  which may differ from zero. Notationally, for the remainder of this paper, sums over “ $i \neq j \in [N]$ ” are over all  $i$  and  $j$  in  $\{1, \dots, N\}$  with  $i \neq j$ .

**Theorem 3.2.** *Let  $N \in \mathbb{N}$  be arbitrary,  $(a_{i,j})_{i,j \in [N]}$  and  $(b_i)_{i \in [N]}$  be deterministic complex numbers,  $(Y_i)_{i \in [N]}$  and  $(Z_i)_{i \in [N]}$  be complex-valued random variables with common expectation  $m^{(1)} \in \mathbb{C}$ , so that the whole family  $\mathcal{W} := \{Y_i \mid i \in [N]\} \cup \{Z_i \mid i \in [N]\}$  is independent. Further, we assume that for all  $p \geq 2$  there exists a  $\mu_p \in \mathbb{R}_+$  such that  $\|W - m^{(1)}\|_p \leq \mu_p$  for all  $W \in \mathcal{W}$ . Then we obtain for all  $p \geq 2$ :*

$$\begin{aligned} \text{i)} \quad & \left\| \sum_{i \in [N]} b_i Y_i \right\|_p \leq \left( A_p \mu_p + \sqrt{N} |m^{(1)}| \right) \sqrt{\sum_{i \in [N]} |b_i|^2}, \\ \text{ii)} \quad & \left\| \sum_{i,j \in [N]} a_{i,j} Y_i Z_j \right\|_p \leq \left( A_p^2 \mu_p^2 + 2A_p \mu_p \sqrt{N} |m^{(1)}| + N |m^{(1)}|^2 \right) \sqrt{\sum_{i,j \in [N]} |a_{i,j}|^2}, \\ \text{iii)} \quad & \left\| \sum_{i \neq j \in [N]} a_{i,j} Y_i Y_j \right\|_p \leq \left( 4A_p^2 \mu_p^2 + 2A_p \mu_p \sqrt{N} |m^{(1)}| + N |m^{(1)}|^2 \right) \sqrt{\sum_{i \neq j \in [N]} |a_{i,j}|^2}. \end{aligned}$$

where  $A_p \in \mathbb{R}_+$  is a constant which depends only on  $p$ .

*Proof.* We show statement *iii)* first. Surely,  $(Y_i - m^{(1)})_i$  are centered and uniformly  $\|\cdot\|_p$ -bounded by  $\mu_p$  for all  $p \geq 2$ . For  $p \geq 2$  we find:

$$\begin{aligned} \left\| \sum_{i \neq j \in [N]} a_{i,j} Y_i Y_j \right\|_p & \leq \left\| \sum_{i \neq j \in [N]} a_{i,j} (Y_i - m^{(1)}) (Y_j - m^{(1)}) \right\|_p + \left\| \sum_{i \neq j \in [N]} a_{i,j} m^{(1)} (Y_j - m^{(1)}) \right\|_p \\ & + \left\| \sum_{i \neq j \in [N]} a_{i,j} m^{(1)} (Y_i - m^{(1)}) \right\|_p + \left\| \sum_{i \neq j \in [N]} a_{i,j} (m^{(1)})^2 \right\|_p =: T_1 + T_2 + T_3 + T_4. \end{aligned}$$

We will now proceed to analyze the four terms separately. To bound  $T_1$ , we have by Lemma D.3 in [6] that

$$T_1 \leq 4A_p^2\mu_p^2 \sqrt{\sum_{i \neq j \in [N]} |a_{i,j}|^2}.$$

For  $T_2$  (and analogously for  $T_3$ ) we obtain through Lemma D.1 in [6] that

$$\begin{aligned} T_2 &= |m^{(1)}| \left\| \sum_{j \in [N]} \left( \sum_{i \in [N] \setminus \{j\}} a_{i,j} \right) (Y_j - m^{(1)}) \right\|_p \\ &\leq |m^{(1)}| A_p \mu_p \sqrt{\sum_{j \in [N]} \left| \sum_{i \in [N] \setminus \{j\}} a_{i,j} \right|^2} \leq \sqrt{N} |m^{(1)}| A_p \mu_p \sqrt{\sum_{i \neq j \in [N]} |a_{i,j}|^2}, \end{aligned}$$

where we used that the Cauchy-Schwarz inequality. Lastly, we obtain

$$\begin{aligned} T_4 &= \left| \sum_{i \neq j \in [N]} a_{i,j} (m^{(1)})^2 \right| = |m^{(1)}|^2 \left| \sum_{i \neq j \in [N]} a_{i,j} \right| \\ &\leq |m^{(1)}|^2 \sqrt{\sum_{i \neq j \in [N]} |a_{i,j}|^2} \cdot \sqrt{N^2} = N |m^{(1)}|^2 \sqrt{\sum_{i \neq j \in [N]} |a_{i,j}|^2}. \end{aligned}$$

This shows that *iii*) holds. Now *ii*) is shown analogously to *iii*), with the difference that sums over  $i$  and  $j$  are always over  $[N]$  without further restrictions such as  $i \neq j$ . In addition, instead of using Lemma D.3 in [6] to bound  $T_1$ , we then use Lemma D.2 in [6] (where constants are smaller, thus we can replace  $4A_p^2\mu_p^2$  by  $A_p^2\mu_p^2$ ).

To show that *i*) holds, we calculate for  $p \geq 2$ :

$$\begin{aligned} \left\| \sum_{i \in [N]} b_i Y_i \right\|_p &= \left\| \sum_{i \in [N]} b_i ((Y_i - m^{(1)}) + m^{(1)}) \right\|_p \leq \left\| \sum_{i \in [N]} b_i (Y_i - m^{(1)}) \right\|_p + \left\| \sum_{i \in [N]} b_i m^{(1)} \right\|_p \\ &\leq A_p \mu_p \sqrt{\sum_{i \in [N]} |b_i|^2} + |m^{(1)}| \left| \sum_{i \in [N]} b_i \right| \leq (A_p \mu_p + |m^{(1)}| \sqrt{N}) \sqrt{\sum_{i \in [N]} |b_i|^2}, \end{aligned}$$

where in the third step we used Lemma D.1 in [6], and in the fourth step we used the Cauchy-Schwarz inequality.  $\square$

We proceed to show the main large deviations result in relation to the stochastic order relation  $\prec$ .

**Theorem 3.3.** *Let for all  $N \in \mathbb{N}$ ,  $Y = Y^{(N)}$  and  $W = W^{(N)}$  be  $N$ -dependent objects that satisfy the following for all  $N \in \mathbb{N}$ :*

- $W = W^{(N)}$  is a finite index set.
- $Y_W = (Y_i)_{i \in W} = (Y_i^{(N)})_{i \in W^{(N)}} = Y_{W^{(N)}}^{(N)}$  is a tuple of random variables of de-Finetti type with respect to some mixing space  $(T_N, \mathcal{T}_N, \mu_N, P^{(N)})$ .

Further, denote for all subsets  $K \subseteq W$  by  $\mathcal{F}_W(\mathbb{R}^K)$  the set of tuples  $C = (C_i)_{i \in W}$ , where for each  $i \in W$ ,  $C_i : \mathbb{R}^K \rightarrow \mathbb{C}$  is a complex-valued measurable function. Analogously, define for all subsets  $K \subseteq W$  by  $\mathcal{F}_{W \times W}(\mathbb{R}^K)$  the set of tuples  $C = (C_{i,j})_{i,j \in W}$ , where for all  $i, j \in W$ ,  $C_{i,j} : \mathbb{R}^K \rightarrow \mathbb{C}$  is a complex-valued measurable function. Then if the mapping  $P^{(N)} : T_N \rightarrow \mathcal{M}_1(\mathbb{R})$  satisfies the first moment condition (2.2) and the central first moment condition (2.4) as well as (2.6), we obtain the following large deviation bounds:



- i)  $\sum_{i \in I} B_i[Y_K]Y_i \prec \sqrt{\sum_{i \in I} |B_i[Y_K]|^2}$ , uniformly over all pairwise disjoint subsets  $I, K \subseteq W$  with  $\#I \leq N$ , and  $B \in \mathcal{F}_W(\mathbb{R}^K)$ .
- ii)  $\sum_{i \in I, j \in J} Y_i A_{i,j}[Y_K]Y_j \prec \sqrt{\sum_{i \in I, j \in J} |A_{i,j}[Y_K]|^2}$ , uniformly over all pairwise disjoint subsets  $I, J, K \subseteq W$  with  $\#I = \#J \leq N$ , and  $A \in \mathcal{F}_{W \times W}(\mathbb{R}^K)$ .
- iii)  $\sum_{i,j \in I, i \neq j} Y_i A_{i,j}[Y_K]Y_j \prec \sqrt{\sum_{i,j \in I, i \neq j} |A_{i,j}[Y_K]|^2}$ , uniformly over all pairwise disjoint subsets  $I, K \subseteq W$  with  $\#I \leq N$ , and  $A \in \mathcal{F}_{W \times W}(\mathbb{R}^K)$ .

Further, if the mapping  $P^{(N)} : T_N \rightarrow \mathcal{M}_1(\mathbb{R})$  satisfies the second moment condition (2.3) and the central second moment condition (2.5) as well as (2.6), the same bounds as in i), ii) and iii) hold after replacing  $Y_i$  and  $Y_j$  on the l.h.s. by  $1 - Y_i^2$  and  $1 - Y_j^2$ , respectively.

*Proof.* We prove iii) first: Let  $\epsilon, D > 0$  be arbitrary and choose  $p \in 2\mathbb{N}$  with  $p \geq 2$  so large that  $p\epsilon > D$ . Now, we pick an  $N \in \mathbb{N}$ , then choose pairwise disjoint subsets  $I, K \subseteq W^{(N)}$  with  $\#I \leq N$  and  $A \in \mathcal{F}_{W \times W}(\mathbb{R}^K)$  arbitrarily. To avoid division by zero, we define the set:

$$\mathcal{A}_3 := \left\{ y_K \in \mathbb{R}^K \mid \sum_{i,j \in I, i \neq j} |A_{i,j}[y_K]|^2 > 0 \right\}.$$

Then we calculate (explanations below, sums over “ $i \neq j$ ” are over all  $i, j \in I$  with  $i \neq j$ ):

$$\begin{aligned} & \mathbb{P} \left( \left| \sum_{i \neq j} Y_i A_{i,j}[Y_K]Y_j \right| > N^\epsilon \left( \sum_{i \neq j} |A_{i,j}[Y_K]|^2 \right)^{\frac{1}{2}} \right) \\ &= \mathbb{P} \left( \left| \frac{\sum_{i \neq j} Y_i A_{i,j}[Y_K]Y_j}{\left( \sum_{i \neq j} |A_{i,j}[Y_K]|^2 \right)^{\frac{1}{2}}} \right|^p \mathbf{1}_{\mathcal{A}_3}(Y_K) > N^{p\epsilon} \right) \leq \frac{1}{N^{p\epsilon}} \mathbb{E} \left| \frac{\sum_{i \neq j} Y_i A_{i,j}[Y_K]Y_j}{\left( \sum_{i \neq j} |A_{i,j}[Y_K]|^2 \right)^{\frac{1}{2}}} \right|^p \mathbf{1}_{\mathcal{A}_3}(Y_K) \\ &= \frac{1}{N^{p\epsilon}} \int_{T(N)} \int_{\mathbb{R}^K} \int_{\mathbb{R}^I} \left| \frac{\sum_{i \neq j} y_i A_{i,j}[y_K]y_j}{\left( \sum_{i \neq j} |A_{i,j}[y_K]|^2 \right)^{\frac{1}{2}}} \right|^p dP_t^{\otimes I}(y_I) \mathbf{1}_{\mathcal{A}_3}(y_K) dP_t^{\otimes K}(y_K) d\mu_N(t) \\ &\leq \frac{1}{N^{p\epsilon}} \int_{T(N)} \int_{\mathbb{R}^K} \left[ 4A_p^2 K_p^{2/p}(t) + 2A_p K_p^{1/p}(t) \sqrt{N} |m_N^{(1)}(t)| + N |m_N^{(1)}(t)|^2 \right]^p dP_t^{\otimes K}(y_K) d\mu_N(t) \\ &\leq \frac{1}{N^{p\epsilon}} \int_{T(N)} 4^p \left[ 4^p A_p^{2p} K_p^{2p}(t) + 2^p A_p^p K_p(t) N^{\frac{p}{2}} |m_N^{(1)}(t)|^p + N^p |m_N^{(1)}(t)|^{2p} \right] d\mu_N(t) \\ &\leq \frac{1}{N^{p\epsilon}} \left( 4^{2p} A_p^{2p} K_p + 8^p A_p^p \sqrt{K_p K_{2p}} + K_{2p} \right) \leq \frac{1}{ND} \cdot \text{const}(p(\epsilon, D)), \end{aligned}$$

where the first step follows from the fact that for the event in the probability to hold not all  $A_{i,j}[Y_K]$  may vanish, in the third step we used Lemma 2.3, in the fourth step we used part iii) of Theorem 3.2 (notice that the  $\mathbb{R}$ -valued coordinates  $(y_i)_{i \in I}$  are independent under  $P_t^{\otimes I}$  and have expectation  $m_N^{(1)}(t) \in \mathbb{R}$ , and also  $(\int_{\mathbb{R}} |y_i - m_N^{(1)}(t)|^p dP_t(y_i))^{1/p} \leq K_p^{1/p}(t)$  by (2.4), which makes Theorem 3.2 applicable. Further,  $\#I \leq N$ , in the fifth step we used that for  $a, b, c \geq 0$  and  $p \in \mathbb{N}$  we have  $(a + b + c)^p \leq 4^p(a^p + b^p + c^p)$ . In the sixth step, we used (2.2), (2.6), and the Cauchy-Schwarz inequality. Lastly, note that  $\text{const}(p(\epsilon, D)) := 4^{2p} A_p^{2p} K_p + 8^p A_p^p \sqrt{K_p K_{2p}} + K_{2p}$  denotes a constant which depends only on  $p$ , which in turn depends only on the choices of  $\epsilon$  and  $D$ . In particular, this constant

does not depend on the choice of  $N \in \mathbb{N}$ , the sets  $I$  and  $K$  or the function tuple  $A$ . This shows *iii*). To show *ii*), we can proceed analogously to the proof of part *iii*), using part *ii*) of Theorem 3.2 instead of part *iii*). We will show *i*) in the setting of the last statement, that is, we will replace  $Y_i$  by  $1 - Y_i^2$ : Let  $\epsilon, D > 0$  be arbitrary and choose  $p \in 2\mathbb{N}$  with  $p \geq 2$  so large that  $p\epsilon > D$ . Now, we pick an  $N \in \mathbb{N}$ , then choose pairwise disjoint subsets  $I, K \subseteq W^{(N)}$  with  $\#I \leq N$  and  $B \in \mathcal{F}_W(\mathbb{R}^K)$  arbitrarily. To avoid division by zero, we define the set

$$\mathcal{A}_1 := \left\{ y_K \in \mathbb{R}^K : \sum_{i \in I} |B_i[y_K]|^2 > 0 \right\}.$$

Now we calculate, with step-by-step explanations found below, and all sums over  $i$  are for  $i \in I$ :

$$\begin{aligned} & \mathbb{P} \left( \left| \sum_i (1 - Y_i^2) B_i[Y_K] \right| > N^\epsilon \left( \sum_i |B_i[Y_K]|^2 \right)^{\frac{1}{2}} \right) \\ &= \mathbb{P} \left( \left| \frac{\sum_i (1 - Y_i^2) B_i[Y_K]}{(\sum_i |B_i[Y_K]|^2)^{\frac{1}{2}}} \right|^p \mathbf{1}_{\mathcal{A}_1}(Y_K) > N^{p\epsilon} \right) \\ &\leq \frac{1}{N^{p\epsilon}} \mathbb{E} \left| \frac{\sum_i (1 - Y_i^2) B_i[Y_K]}{(\sum_i |B_i[Y_K]|^2)^{\frac{1}{2}}} \right|^p \mathbf{1}_{\mathcal{A}_1}(Y_K) \\ &= \frac{1}{N^{p\epsilon}} \int_{T^{(N)}} \int_{\mathbb{R}^K} \int_{\mathbb{R}^I} \left| \frac{\sum_i (1 - y_i^2) B_i[y_K]}{(\sum_i |B_i[y_K]|^2)^{\frac{1}{2}}} \right|^p \mathrm{d}P_t^{\otimes I}(y_I) \mathbf{1}_{\mathcal{A}_1}(y_K) \mathrm{d}P_t^{\otimes K}(y_K) \mathrm{d}\mu_N(t) \\ &\leq \frac{1}{N^{p\epsilon}} \int_{T^{(N)}} \int_{\mathbb{R}^K} \left[ A_p K_p^{1/p}(t) + \sqrt{N} |1 - m_N^{(2)}(t)| \right]^p \mathrm{d}P_t^{\otimes K}(y_K) \mathrm{d}\mu_N(t) \\ &\leq \frac{1}{N^{p\epsilon}} \int_{T^{(N)}} 2^p \left[ A_p^p K_p(t) + N^{\frac{p}{2}} |1 - m_N^{(2)}(t)|^p \right] \mathrm{d}\mu_N(t) \\ &\leq \frac{1}{N^{p\epsilon}} (2^p A_p^p K_p + K_p) \leq \frac{1}{N^D} \cdot \text{const}(p(\epsilon, D)), \end{aligned}$$

where the first step follows from the fact that for the event in the probability to hold not all  $B_i[Y_K]$  may vanish, in the third step we used Lemma 2.3, in the fourth step we used part *i*) of Theorem 3.2 (notice that the  $\mathbb{R}$ -valued coordinates  $(1 - y_i^2)_{i \in I}$  are independent under  $P_t^{\otimes I}$  and have expectation  $1 - m_N^{(2)}(t) \in \mathbb{R}$ , and also for all  $t \in T^{(N)}$ ):

$$\left( \int_{\mathbb{R}} |1 - y_i^2 - (1 - m_N^{(2)}(t))|^p \mathrm{d}P_t(y_i) \right)^{1/p} \leq K_p^{1/p}(t)$$

by (2.5), which makes Theorem 3.2 applicable. Further,  $\#I \leq N$ , in the fifth step we used that for  $a, b \geq 0$  and  $p \in \mathbb{N}$  we have  $(a + b)^p \leq 2^p(a^p + b^p)$ . In the sixth step, we used (2.3) and (2.6). Lastly, note that  $\text{const}(p(\epsilon, D)) := 2^p A_p^p K_p + K_p$  denotes a constant which depends only on  $p$ , which in turn depends only on the choices of  $\epsilon$  and  $D$ . In particular, this constant does not depend on the choice of  $N \in \mathbb{N}$ , the sets  $I$  and  $K$  or the function tuple  $B$ . This shows *i*).  $\square$

The next corollary verifies all applications of large deviation bounds needed for the main estimates, Lemma 3.1. In particular, with the next corollary at hand, Lemma 3.1 can be proven as in [6].

**Corollary 3.4.** *In the above setting, we obtain uniformly over all  $z \in \mathbb{C}_+$  and  $i \neq j \in$*

$\{1, \dots, N\}$ :

$$i) \quad Z_i^{(1)} = \sum_{k \neq l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li} \prec \left( \frac{1}{N^2} \sum_{k,l}^{(i)} |G_{kl}^{(i)}|^2 \right)^{\frac{1}{2}} \quad (3.6)$$

$$ii) \quad Z_i^{(2)} = \sum_k^{(i)} \left( |H_{ik}|^2 - \frac{1}{N} \right) G_{kk}^{(i)} \prec \left( \frac{1}{N^2} \sum_k^{(i)} |G_{kk}^{(i)}|^2 \right)^{\frac{1}{2}} \quad (3.7)$$

$$iii) \quad \sum_{k,l}^{(ij)} H_{ik} G_{kl}^{(ij)} H_{lj} \prec \left( \frac{1}{N^2} \sum_{k,l}^{(ij)} |G_{kl}^{(ij)}|^2 \right)^{\frac{1}{2}} \quad (3.8)$$

*Proof.* We prove *i)* first: Note that for all  $N \in \mathbb{N}$  and  $i \in [N]$ , the  $N - 1$  entries  $\sqrt{N}H_{ik} = X_{ik}$ ,  $k \neq i$ , are distinct entries from the family  $(X_N(a, b))_{1 \leq a \leq b \leq N}$ , which is of de-Finetti type with mixture  $\mu_N$  satisfying the first moment condition (2.2) and the central first moment condition (2.4). Further, for any  $z \in \mathbb{C}_+$  and  $k \neq l \in \{1, \dots, N\} \setminus \{i\}$  we have that  $(H^{(i)} - z)^{-1}(i, j)$  is a complex function of variables in  $(X_N(i, j))_{1 \leq i \leq j \leq N}$  disjoint from those in  $(X_N(i, j))_{1 \leq a \leq b \leq N}$ . Therefore, the statement follows with Theorem 3.3. Statement *iii)* is shown analogously, and for statement *ii)* as well, using the last statement in Theorem 3.3.  $\square$

## A Implications of Theorem 2.12

Theorem 2.12 is a statement about the supremum of certain probabilities. It can be strengthened by taking the supremum inside the probability, which is possible due to the Lipschitz continuity of all quantities involved. This will imply that  $\prec$  does not only hold uniformly for  $z \in \mathcal{D}_N(\tau)$ , but also simultaneously for these  $z$  (cf. Definition 2.10).

We formulate a general theorem, which is of help when lifting uniform  $\prec$ -statements to simultaneous ones. To this end, in addition to the domains  $\mathcal{D}_N^*(\tau)$  and  $\mathcal{D}_N(\tau)$ , we define the encompassing domains

$$\forall \tau \in (0, 1) : \forall N \in \mathbb{N} : \mathcal{C}_N(\tau) := \left[ -\frac{1}{\tau}, \frac{1}{\tau} \right] + i \left[ \frac{1}{N}, \frac{1}{\tau} \right].$$

For any sequence of regions  $\mathcal{G}_N \subseteq \mathcal{C}_N(\tau)$  and fixed  $L \in \mathbb{N}$ , define the subsets

$$\mathcal{G}_N^L := \mathcal{G}_N \cap \frac{1}{N^L}(\mathbb{Z} + i\mathbb{Z}).$$

For example, we might consider the regions  $\mathcal{G}_N^4$  for  $\mathcal{G}_N := \mathcal{D}_N(\tau)$ . We notice that  $\mathcal{G}_N^4$  forms a  $\frac{2}{N^4}$ -net in  $\mathcal{G}_N$ , which means that any  $z \in \mathcal{G}_N$  is  $\frac{2}{N^4}$ -close to some  $z' \in \mathcal{G}_N^4$ . The following theorem generalizes Remark 2.7 in [6].

**Lemma A.1.** Suppose we are given stochastic domination of the form

$$F_i^{(N)}(z) \prec \Psi^{(N)}(z), \quad i \in I_N, z \in \mathcal{G}_N^L,$$

where for all  $N \in \mathbb{N}$ :

- $\mathcal{G}_N \subseteq \mathcal{C}_N(\tau)$  is a non-empty subset with a geometry such that  $\mathcal{G}_N^L$  forms a  $\frac{2}{N^L}$ -net in  $\mathcal{G}_N$ .
- $(F_i^{(N)})_{i \in I_N}$  is a family of complex-valued functions on  $\mathcal{C}_N(\tau)$ , where  $\#I_N \leq C_1 N^{d_1}$  and for all  $i \in I_N$ ,  $F_i^{(N)}$  is  $C_2 N^{d_2}$ -Lipschitz-continuous on  $\mathcal{C}_N(\tau)$ ,

- $\Psi^{(N)}$  is an  $\mathbb{R}_+$ -valued function on  $\mathcal{C}_N(\tau)$ , which is  $C_3 N^{d_3}$ -Lipschitz-continuous and bounded from below by  $\frac{1}{C_4 N^{d_4}}$ ,

where  $C_1, \dots, C_4 > 0$ ,  $d_1, \dots, d_4 > 0$  are  $N$ -independent constants and  $L > \max(d_2 + d_4, d_3 + d_4)$ . Then we obtain the simultaneous statement:

$$\sup_{z \in \mathcal{G}_N} \max_{i \in I_N} \frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} \prec 1. \quad (\text{A.1})$$

*Proof.* The following statements hold trivially for all  $N \in \mathbb{N}$ :

$$i) \# \mathcal{G}_N^L \leq \# \mathcal{D}_N^L \leq \frac{3}{\tau} N^L \cdot \frac{2}{\tau} N^L =: C_5 N^{2L}, \quad ii) \forall z \in \mathcal{G}_N : \exists z' \in \mathcal{G}_N^L : |z - z'| \leq \frac{2}{N^L}.$$

Step 1: (A.1) holds if  $\mathcal{G}_N$  is replaced by  $\mathcal{G}_N^L$ .

This is easily done by the following calculation for  $\epsilon, D > 0$  arbitrary:

$$\mathbb{P} \left( \sup_{z \in \mathcal{G}_N^L} \max_{i \in I_N} \frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^\epsilon \right) \leq \sum_{z \in \mathcal{G}_N^L} \sum_{i \in I_N} \mathbb{P} \left( \frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^\epsilon \right) \leq C_5 N^{2L} C_1 N^{d_1} \frac{C_{\epsilon, D}}{N^D}$$

This concludes the first step by shifting  $D \rightsquigarrow D + 2L + d_1$  and absorbing  $C_1 \cdot C_5$  into  $C_{\epsilon, D+2L+d_1}$ .

Step 2: Extension from  $\mathcal{G}_N^L$  to  $\mathcal{G}_N$ .

Now, Lipschitz-continuity comes into play: For an arbitrary  $\epsilon > 0$ , suppose

$$\exists z \in \mathcal{G}_N, \exists i \in I_N : |F_i^{(N)}(z)| > \Psi^{(N)}(z) N^\epsilon.$$

Then there exists a  $z' \in \mathcal{G}_N^L$  with  $|z - z'| \leq \frac{2}{N^L}$ , and then due to Lipschitz-continuity of  $F_i^{(N)}$  and  $\Psi^{(N)}$ :

$$|F_i^{(N)}(z')| > \Psi^{(N)}(z') N^\epsilon - \frac{2}{N^L} \cdot C_2 N^{d_2} - \frac{2}{N^L} \cdot C_3 N^{d_3+\epsilon}.$$

It follows, using the lower bound on  $\Psi^{(N)}$ :

$$\frac{|F_i^{(N)}(z')|}{\Psi^{(N)}(z')} > N^\epsilon - 2 \frac{C_2 N^{d_2} + C_3 N^{d_3+\epsilon}}{N^L \Psi^{(N)}(z')} \geq N^\epsilon - 2 C_4 N^{d_4} \frac{C_2 N^{d_2} + C_3 N^{d_3+\epsilon}}{N^L}.$$

We may assume w.l.o.g. that  $\epsilon \in (0, L - d_3 - d_4)$  (see Remark 2.11). Then

$$\exists N(\epsilon) \in \mathbb{N} : \forall N \geq N(\epsilon) : N^\epsilon - 2 C_4 N^{d_4} \frac{C_2 N^{d_2} + C_3 N^{d_3+\epsilon}}{N^L} > N^{\frac{\epsilon}{2}}.$$

We have shown that for all  $N \geq N(\epsilon)$ :

$$\left[ \exists z \in \mathcal{G}_N, \exists i \in I_N : \frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^\epsilon \right] \Rightarrow \left[ \exists z' \in \mathcal{G}_N^L, \exists i \in I_N : \frac{|F_i^{(N)}(z')|}{\Psi^{(N)}(z')} > N^{\frac{\epsilon}{2}} \right].$$

Therefore, if  $D > 0$  is arbitrary, we obtain for all  $N \geq N(\epsilon)$ :

$$\mathbb{P} \left( \sup_{z \in \mathcal{G}_N} \max_{i \in I_N} \frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^\epsilon \right) \leq \mathbb{P} \left( \sup_{z \in \mathcal{G}_N^L} \max_{i \in I_N} \frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^{\frac{\epsilon}{2}} \right) \leq \frac{C_{\frac{\epsilon}{2}, D}}{N^D},$$

where we used Step 1 for the last inequality. This concludes the proof by choosing constants as  $(\epsilon, D) \mapsto C_{\frac{\epsilon}{2}, D}$  and with Remark 2.11.  $\square$

We will now show that Theorem 2.12 actually holds simultaneously.

**Theorem A.2** (Simultaneous Local Law for Curie-Weiss-Type Ensembles). *In the setting of the local law for Curie-Weiss type ensembles (Theorem 2.12) we obtain*

$$\sup_{z \in \mathcal{D}_N(\tau)} \frac{\max(\Lambda(z), |s(z) - m(z)|)}{\frac{\frac{1}{\sqrt{N\eta}}}{\sqrt{\kappa + \eta + \frac{1}{\sqrt{N\eta}}}}} \prec 1 \quad (\text{A.2})$$

as well as

$$\sup_{z \in \mathcal{D}_N^*(\tau)} \frac{\max(\Lambda(z), |s(z) - m(z)|)}{\frac{1}{\sqrt{N\eta}}} \prec 1. \quad (\text{A.3})$$

*Proof.* Elementary calculations show that on the encompassing domains  $\mathcal{C}_N(\tau)$ ,  $|s(z) - m(z)|$  is  $2N^2$ -Lipschitz and  $\Lambda(z)$  is  $N^2$ -Lipschitz, hence  $F^{(N)}(z) := \max(\Lambda(z), |s_N(z) - s(z)|)$  is  $2N^2$ -Lipschitz. Further, on  $\mathcal{C}_N(\tau)$  the error terms

$$\Psi_1^{(N)}(z) := \frac{\frac{1}{\sqrt{N\eta}}}{\sqrt{\kappa + \eta + \frac{1}{\sqrt{N\eta}}}} \quad \text{and} \quad \Psi_2^{(N)}(z) := \frac{1}{\sqrt{N\eta}}$$

are  $3N/\tau$  resp.  $N/2$ -Lipschitz and lower bounded by  $\tau/(2\sqrt{N})$  resp.  $\sqrt{\tau/N}$ . Further, by Theorem 2.12 we know that  $F^{(N)}(z) \prec \Psi^{(N)}(z)$ ,  $z \in \mathcal{D}_N^4(\tau)$ . Therefore, the statement follows directly with Lemma A.1.  $\square$

**Corollary A.3.** *In the situation of Theorem 2.12, we find*

$$\sup_{z \in \mathcal{D}_N(\tau)} |s(z) - m(z)| \prec \frac{1}{N^{\frac{\tau}{4}}} \quad \text{and} \quad \sup_{z \in \mathcal{D}_N^*(\tau)} |s(z) - m(z)| \prec \frac{1}{N^{\frac{\tau}{2}}}$$

*Proof.* Since for any  $z \in \mathcal{D}_N(\tau)$  we find  $1/(N\eta)^{\frac{1}{4}} \leq 1/(N/N^{1-\tau})^{\frac{1}{4}} = N^{-\frac{\tau}{4}}$ , it follows

$$\sup_{z \in \mathcal{D}_N(\tau)} \frac{|s(z) - m(z)|}{\frac{1}{N^{\frac{\tau}{4}}}} \leq \sup_{z \in \mathcal{D}_N(\tau)} \frac{|s(z) - m(z)|}{\frac{\frac{1}{\sqrt{N\eta}}}{\sqrt{\kappa + \eta + \frac{1}{\sqrt{N\eta}}}}} \prec 1$$

by Theorem A.2. Multiplying both sides by  $1/N^{\tau/4}$  concludes the proof for the first statement, and the second statement follows analogously.  $\square$

Theorem A.2 immediately yields Corollary A.3, which allows us to conclude that with high probability,  $s$  converges uniformly to  $m$  on a growing domain  $\mathcal{D}_N(\tau)$  that approaches the real axis. Before venturing further into further corollaries, we recall how Stieltjes transforms can be used to analyze weak convergence, and why it is important for the imaginary part to reach the real axis.

For any probability measure  $\nu$  on  $(\mathbb{R}, \mathcal{B})$ , there is a close relationship between  $S_\nu$  and  $\nu$ , which is observed by analyzing the function (where  $\eta > 0$  is fixed)

$$\mathbb{R} \ni E \mapsto \frac{1}{\pi} \operatorname{Im} S_\nu(E + i\eta) = \int_{\mathbb{R}} \frac{1}{\pi} \frac{\eta}{(x - E)^2 + \eta^2} \nu(dx) = (P_\eta * \nu)(E), \quad (\text{A.4})$$

where  $*$  is the convolution and for any  $\eta > 0$ ,  $P_\eta : \mathbb{R} \rightarrow \mathbb{R}$  is the Cauchy kernel, that is,  $\forall x \in \mathbb{R} : P_\eta(x) := \frac{1}{\pi} \frac{\eta}{x^2 + \eta^2}$ , which is the Lebesgue density function of the Cauchy probability distribution with scale parameter  $\eta$ . Denoting the Lebesgue measure on  $(\mathbb{R}, \mathcal{B})$  by  $\mathbb{X}$ , we find  $(P_\eta * \nu)\mathbb{X} = (P_\eta\mathbb{X}) * \nu$ , that is, the function in (A.4) is a well-defined  $\mathbb{X}$ -density for the convolution  $(P_\eta\mathbb{X}) * \nu$ . Further, it can be verified that *i)*  $P_\eta\mathbb{X} \searrow \delta_0$  weakly as  $\eta \searrow 0$ , *ii)* the convolution is continuous with respect to weak convergence (if  $\nu_n \rightarrow \nu$  weakly and  $\nu'_n \rightarrow \nu'$  weakly, then  $\nu_n * \nu'_n \rightarrow \nu * \nu'$  weakly) and *iii)* the Dirac measure  $\delta_0$  is the neutral element of convolution. We conclude that  $(P_\eta * \nu)\mathbb{X} \rightarrow \delta_0 * \nu = \nu$  weakly as  $\eta \searrow 0$ , which proves the following well-known lemma:

**Lemma A.4.** Let  $\nu$  be a probability measure on  $(\mathbb{R}, \mathcal{B})$ . Then for any interval  $I \subseteq \mathbb{R}$  with  $\nu(\partial I) = 0$ , we find:

$$\nu(I) = \lim_{\eta \searrow 0} [(P_\eta * \nu)\mathbb{K}](I) = \lim_{\eta \searrow 0} \frac{1}{\pi} \int_I \operatorname{Im} S_\nu(E + i\eta) \mathbb{K}(dE).$$

Thus, any finite measure  $\nu$  on  $(\mathbb{R}, \mathcal{B})$  is uniquely determined by  $S_\nu$ .

Let  $\sigma_N$  be the ESDs of a sequence of Hermitian  $N \times N$  matrices  $X_N$ . Assume that  $\sigma_N$  converges weakly almost surely to the semicircle distribution  $\sigma$ , that is, convergence takes place on a measurable set  $A$  with  $\mathbb{P}(A) = 1$ . By the discussion preceding Lemma A.4, we find on  $A$  that the following commutative diagram holds, where all arrows indicate weak convergence:

$$\begin{array}{ccc} (P_\eta * \sigma_N)\mathbb{K} & \xrightarrow{N \rightarrow \infty} & (P_\eta * \sigma)\mathbb{K} \\ \eta \searrow 0 \downarrow & \searrow N \rightarrow \infty, \eta \searrow 0 & \downarrow \eta \searrow 0 \\ \delta_0 * \sigma_N = \sigma_N & \xrightarrow{N \rightarrow \infty} & \sigma \end{array}$$

In particular, the diagonal arrow indicates weak convergence  $(P_{\eta_N} * \sigma_N)\mathbb{K} \rightarrow \sigma$  as  $N \rightarrow \infty$  for any sequence  $\eta_N \searrow 0$ . But this does not tell us if also densities align, that is, if also  $P_\eta * \sigma_N \rightarrow f_\sigma$  in some sense, for example uniformly over a specified compact interval. If  $\eta = \eta_N$  drops too quickly to zero as  $N \rightarrow \infty$ , then  $P_{\eta_N} * \sigma_N$  will have steep peaks at each eigenvalue, thus will not approximate the density of the semicircle distribution uniformly. To illustrate this effect, we simulate an ESD of a  $100 \times 100$  random matrix  $X_{100}$ , where  $(\sqrt{100}X_{100}(i, j))_{1 \leq i \leq j \leq 100}$  are independent Rademacher distributed random variables. The density estimates at *bandwidths*  $\eta_1 := N^{-1/2} = 1/10$  and  $\eta_2 := N^{-1} = 1/100$  are shown in Figure 1.

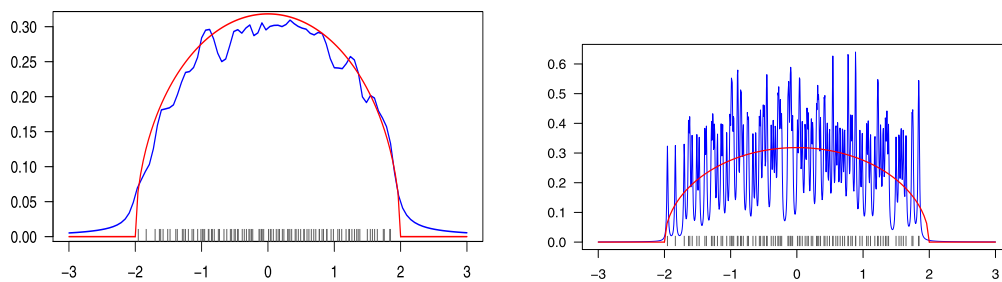


Figure 1: Red lines:  $f_\sigma$ . Blue lines:  $\frac{1}{\pi} \operatorname{Im} S_{\sigma_{100}}(\cdot + i\eta) = P_\eta * \sigma_{100}$ . Grey bars: eigenvalue locations. Left figure:  $\eta = \eta_1$ . Right figure:  $\eta = \eta_2$ .

As we see in Figure 1, we already obtain a decent approximation by the semicircle density when  $\eta = \eta_1$ , despite the low  $N = 100$ . But after reducing the scale from  $\eta_1$  to  $\eta_2$ , we observe that we do not obtain a useful approximation by the semicircle density anymore. Indeed, the scale  $N^{-1}$  is too fast to obtain uniform convergence of the estimated density to the target density, whereas a scale of  $N^{-(1-\tau)}$  for any  $\tau \in (0, 1)$  is sufficient, see our Theorem A.6, which explains Figure 1 in that it shows that we do have uniform convergence of the densities.

Before we turn to Theorem A.6, we establish that as  $\eta \searrow 0$ , the function  $E \mapsto \frac{1}{\pi} \operatorname{Im} m(E + i\eta)$ , that is  $P_\eta * \sigma$ , converges uniformly to  $f_\sigma$  over any compact interval and with a speed of  $O(\sqrt{\eta})$ .

**Lemma A.5.** *Let  $C \geq 2$  be arbitrary, then we obtain for any  $\eta \in (0, C]$ :*

$$\sup_{E \in [-C, C]} \left| \frac{1}{\pi} \operatorname{Im}(m(E + i\eta)) - f_\sigma(E) \right| \leq \sqrt{C\eta}.$$

*Proof.* Elementary calculations show that if  $(a + ib)^2 = c + id$ , where  $a, c, d \in \mathbb{R}$  and  $b > 0$ , then

$$b = \sqrt{\frac{-c + \sqrt{c^2 + d^2}}{2}}. \quad (\text{A.5})$$

With  $C \geq 2$  and  $z = E + i\eta$ , where  $E \in [-C, C]$  and  $\eta > 0$ , we find that  $z^2 - 4 = E^2 - \eta^2 - 4 + i2E\eta$ , hence with (A.5):

$$\begin{aligned} \frac{1}{\pi} \operatorname{Im} m(z) &= -\frac{\operatorname{Im}(z)}{2\pi} + \frac{\operatorname{Im}(\sqrt{z^2 - 4})}{2\pi} \\ &= \frac{1}{2\pi} \left( -\eta + \sqrt{\frac{4 + \eta^2 - E^2 + \sqrt{(E^2 - \eta^2 - 4)^2 + 4E^2\eta^2}}{2}} \right). \end{aligned}$$

Assuming at first that  $E \in [-2, 2]$ , we find

$$\left| \frac{1}{\pi} \operatorname{Im} m(z) - f_\sigma(E) \right| \leq \frac{\eta}{2\pi} + \frac{1}{2\pi} \left( \sqrt{\frac{4 + \eta^2 - E^2 + \sqrt{(E^2 - \eta^2 - 4)^2 + 4E^2\eta^2}}{2}} - \sqrt{4 - E^2} \right).$$

Using that  $\sqrt{\cdot}$  is uniformly continuous with modulus of continuity  $\sqrt{\cdot}$ , it suffices to analyze the difference of the arguments, which will then yield the desired upper bound. Now assuming that  $E \in [-C, C] \setminus [-2, 2]$  we find

$$\left| \frac{1}{\pi} \operatorname{Im} m(z) - f_\sigma(E) \right| = \left| \frac{1}{2\pi} \left( -\eta + \sqrt{\frac{4 + \eta^2 - E^2 + \sqrt{(E^2 - \eta^2 - 4)^2 + 4E^2\eta^2}}{2}} \right) \right|.$$

Considering the cases  $\eta^2 \leq E^2 - 4$  and  $\eta^2 > E^2 - 4$  separately and using  $|E| > 2$ , we obtain

$$4 + \eta^2 - E^2 + \sqrt{(E^2 - \eta^2 - 4)^2 + 4E^2\eta^2} \leq 2\eta^2 + 2C\eta,$$

which yields the desired upper bound.  $\square$

**Theorem A.6.** *In the situation of Theorem 2.12, define the scale  $\eta_N := 1/N^{1-\tau}$  for all  $N \in \mathbb{N}$  and assume  $\tau < 2/3$ . Then*

$$\sup_{E \in [-\tau^{-1}, \tau^{-1}]} \left| \frac{1}{\pi} \operatorname{Im}(s(E + i\eta_N)) - f_\sigma(E) \right| \prec \frac{1}{N^{\frac{\tau}{4}}}.$$

*Proof.* Due to Corollary A.3,

$$\sup_{E \in [-\tau^{-1}, \tau^{-1}]} \left| \frac{1}{\pi} \operatorname{Im}(s(E + i\eta_N)) - \frac{1}{\pi} \operatorname{Im}(m(E + i\eta_N)) \right| \prec \frac{1}{N^{\frac{\tau}{4}}}.$$

The statement follows with Lemma A.5, which gives

$$\sup_{E \in [-\tau^{-1}, \tau^{-1}]} \left| \frac{1}{\pi} \operatorname{Im}(m(E + i\eta_N)) - f_\sigma(E) \right| \prec \sqrt{\eta_N} = \frac{1}{N^{\frac{1}{2} - \frac{\tau}{2}}}. \quad \square$$

Theorem A.6 states in particular that at the scale  $\eta_N = N^{-(1-\tau)}$  ( $\tau \in (0, 1)$  fixed), we find uniform convergence in probability of  $P_{\eta_N} * \sigma_N$  to  $f_\sigma$  on the interval  $[-\tau^{-1}, \tau^{-1}]$ , where we have strong control on the probability estimates. In his publication [26], Khorunzhy showed for the Wigner case that for arbitrary but fixed  $E \in (-2, 2)$  and for slower scales  $\eta_N = N^{-(1-\tau)}$  ( $\tau \in (3/4, 1)$  fixed),  $P_{\eta_N} * \sigma_N(E) \rightarrow f_\sigma(E)$  in probability. Moreover, he showed that this does *not* hold in general for scales that decay too quickly, such as the scale  $\eta_N = N^{-1}$ , see his Remark 4 on page 149 in above mentioned publication. See also Figure 1 on page 22 for a visualization of these findings.

We have seen that Theorem 2.12 and Theorem A.2 guarantee closeness of the Stieltjes transforms of the ESDs and of the semicircle distribution. Theorem A.6 shows that this implies that  $f_\sigma$  can be approximated well by a kernel density estimate  $P_{\eta_N} * \sigma_N$ .

Next, we state a semicircle law on small scales, which is a probabilistic evaluation of how well the semicircle distribution predicts the fraction of eigenvalues in given intervals  $I \subseteq \mathbb{R}$ . Interestingly, a variant of the following theorem (see Theorem A.8 below) even constitutes the local law *per se* in [34]. Notationally, if  $A \subseteq \mathbb{R}$  is a subset, denote by  $\mathcal{I}(A)$  the set of all intervals  $I \subseteq A$ .

**Theorem A.7** (Semicircle Law on Small Scales). *In the setting of the local law for Curie-Weiss type ensembles (Theorem 2.12), we obtain the two statements*

$$\sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \prec \frac{1}{N^{\frac{1}{4}}} \quad \text{and} \quad \sup_{I \in \mathcal{I}([-2+\tau, 2-\tau])} |\sigma_N(I) - \sigma(I)| \prec \frac{1}{N^{\frac{1}{2}}}.$$

*Proof.* The proof can be carried out analogously to the proof of Theorem 2.8 in [6].  $\square$

Due to Theorem A.7, for any  $\epsilon \in (0, 1/4)$  and  $D > 0$  we find a constant  $C_{\epsilon, D} \geq 0$  such that

$$\forall N \in \mathbb{N} : \mathbb{P} \left( \sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \leq \frac{N^\epsilon}{N^{\frac{1}{4}}} \right) > 1 - \frac{C_{\epsilon, D}}{N^D}, \quad (\text{A.6})$$

This tells us that when predicting interval probabilities of  $\sigma_N$  by those of  $\sigma$ , the absolute error will be bounded by  $N^{-(1/4-\epsilon)}$ . Note that for small intervals this is not a good statement: Then the error bound of  $N^{-(1/4-\epsilon)}$  is useless, since both  $\sigma_N(I)$  and  $\sigma(I)$  are small anyway. The natural way to remedy this would be to consider the relative deviation  $\sigma_N(I)/\sigma(I)$ . This yields the following theorem, which for Tao and Vu actually constitutes “The Local Semicircle Law” (instead of a statement as Theorem 2.12 involving Stieltjes transforms), see their Theorem 7 in [34, p. 7].

**Theorem A.8** (Interval-Type Local Semicircle Laws). *In the setting of Theorem 2.12, we obtain*

i) For all  $\tau \in (0, 1/4)$ :

$$\sup_{\substack{I \in \mathcal{I}(\mathbb{R}) \\ |I| \geq \frac{1}{N^{1/4-\tau}}}} \frac{|\sigma_N(I) - \sigma(I)|}{|I|} \prec \frac{1}{N^\tau}.$$

ii) For all  $\tau \in (0, 1/2)$ :

$$\sup_{\substack{I \in \mathcal{I}([-2+\tau, 2-\tau]) \\ |I| \geq \frac{1}{N^{1/2-\tau}}}} \left| \frac{\sigma_N(I)}{\sigma(I)} - 1 \right| \prec \frac{1}{N^\tau}.$$

*Proof.* From Theorem A.7 it follows immediately that

$$\sup_{\substack{I \in \mathcal{I}(\mathbb{R}) \\ |I| \geq \frac{1}{N^{1/4-\tau}}}} \frac{|\sigma_N(I) - \sigma(I)|}{|I|} \leq \sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| N^{\frac{1}{4}-\tau} \prec \frac{1}{N^\tau},$$



which proves statement i), and ii) can be shown analogously by using the second statement of Theorem A.7.  $\square$

## References

- [1] Oskari H. Ajanki, László Erdős, and Torben Krüger. “Local spectral statistics of Gaussian matrices with correlated entries”. In: *J. Stat. Phys.* 163.2 (2016), pp. 280–302. ISSN: 0022-4715. DOI: 10.1007/s10955-016-1479-y. URL: <https://doi.org/10.1007/s10955-016-1479-y>.
- [2] Oskari H. Ajanki, László Erdős, and Torben Krüger. “Stability of the matrix Dyson equation and random matrices with correlations”. In: *Probab. Theory Related Fields* 173.1-2 (2019), pp. 293–373. ISSN: 0178-8051. DOI: 10.1007/s00440-018-0835-z. URL: <https://doi.org/10.1007/s00440-018-0835-z>.
- [3] Oskari H. Ajanki, László Erdős, and Torben Krüger. “Universality for general Wigner-type matrices”. In: *Probab. Theory Related Fields* 169.3-4 (2017), pp. 667–727. ISSN: 0178-8051. DOI: 10.1007/s00440-016-0740-2. URL: <https://doi.org/10.1007/s00440-016-0740-2>.
- [4] David J. Aldous. “Exchangeability and related topics”. In: *École d’été de probabilités de Saint-Flour, XIII—1983*. Vol. 1117. Lecture Notes in Math. Springer, Berlin, 1985, pp. 1–198. DOI: 10.1007/BFb0099421. URL: <https://doi.org/10.1007/BFb0099421>.
- [5] Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*. Second. Springer Series in Statistics. Springer, New York, 2010, pp. xvi+551. ISBN: 978-1-4419-0660-1. DOI: 10.1007/978-1-4419-0661-8. URL: <https://doi.org/10.1007/978-1-4419-0661-8>.
- [6] Florent Benaych-Georges and Antti Knowles. “Lectures on the Local Semicircle Law for Wigner Matrices”. Sept. 2018. URL: <http://www.unige.ch/~knowles/SCL.pdf>. MR3792624
- [7] Riccardo Catalano, Michael Fleermann, and Werner Kirsch. “Random Band and Block Matrices with Correlated Entries”. Feb. 2022. URL: <https://arxiv.org/abs/2202.04707>.
- [8] Ziliang Che. “Universality of random matrices with correlated entries”. In: *Electron. J. Probab.* 22 (2017), Paper No. 30, 38. DOI: 10.1214/17-EJP46. URL: <https://doi.org/10.1214/17-EJP46>. MR3629874
- [9] Richard S. Ellis. *Entropy, large deviations, and statistical mechanics*. Classics in Mathematics. Reprint of the 1985 original. Springer-Verlag, Berlin, 2006, pp. xiv+364. ISBN: 978-3-540-29059-9; 3-540-29059-1. DOI: 10.1007/3-540-29060-5. URL: <https://doi.org/10.1007/3-540-29060-5>. MR2189669
- [10] László Erdős, Antti Knowles, and Horng-Tzer Yau. “Averaging fluctuations in resolvents of random band matrices”. In: *Ann. Henri Poincaré* 14.8 (2013), pp. 1837–1926. ISSN: 1424-0637. DOI: 10.1007/s00023-013-0235-y. URL: <https://doi.org/10.1007/s00023-013-0235-y>.
- [11] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. “The local semicircle law for a general class of random matrices”. In: *Electron. J. Probab.* 18 (2013), no. 59, 58. DOI: 10.1214/EJPv18-2473. URL: <https://doi.org/10.1214/EJPv18-2473>. MR3068390
- [12] László Erdős, Torben Krüger, and Dominik Schröder. “Random matrices with slow correlation decay”. In: *Forum Math. Sigma* 7 (2019), Paper No. e8, 89. DOI: 10.1017/fms.2019.2. URL: <https://doi.org/10.1017/fms.2019.2>.
- [13] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. “Local semicircle law and complete delocalization for Wigner random matrices”. In: *Comm. Math. Phys.* 287.2 (2009), pp. 641–655. ISSN: 0010-3616. DOI: 10.1007/s00220-008-0636-9. URL: <https://doi.org/10.1007/s00220-008-0636-9>.
- [14] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. “Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices”. In: *Ann. Probab.* 37.3 (2009), pp. 815–852. ISSN: 0091-1798. DOI: 10.1214/08-AOP421. URL: <https://doi.org/10.1214/08-AOP421>. MR2537522
- [15] László Erdős and Horng-Tzer Yau. *A dynamical approach to random matrix theory*. Vol. 28. Courant Lecture Notes in Mathematics. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017, pp. ix+226. ISBN: 978-1-4704-3648-3. MR3699468

- [16] Michael Fleermann. “Global and Local Semicircle Laws for Random Matrices with Correlated Entries”. PhD thesis. FernUniversität in Hagen, Germany, 2019. DOI: 10.18445/20190612-122137-0.
- [17] Michael Fleermann. “The empirical spectral distribution of symmetric random matrices with correlated entries. An asymptotic analysis employing the method of moments.” MA thesis. University of Münster, Germany, 2015.
- [18] Michael Fleermann and Johannes Heiny. “High-dimensional sample covariance matrices with Curie-Weiss entries”. In: *ALEA Lat. Am. J. Probab. Math. Stat.* 17.2 (2020), pp. 857–876. DOI: 10.30757/alea.v17-33. URL: <https://doi.org/10.30757/alea.v17-33>. MR4169603
- [19] Michael Fleermann, Werner Kirsch, and Thomas Kriecherbauer. “The almost sure semicircle law for random band matrices with dependent entries”. In: *Stochastic Process. Appl.* 131 (2021), pp. 172–200. ISSN: 0304-4149. DOI: 10.1016/j.spa.2020.09.004. URL: <https://doi.org/10.1016/j.spa.2020.09.004>.
- [20] Michael Fleermann, Werner Kirsch, and Gabor Toth. “Interval type local limit theorems for lattice type random variables and distributions”. In: *Stochastics* 0.0 (2021), pp. 1–12. DOI: 10.1080/17442508.2021.1980570. eprint: <https://doi.org/10.1080/17442508.2021.1980570>. URL: <https://doi.org/10.1080/17442508.2021.1980570>.
- [21] Olga Friesen and Matthias Löwe. “A phase transition for the limiting spectral density of random matrices”. In: *Electron. J. Probab.* 18 (2013), no. 17, 17. DOI: 10.1214/EJP.v18-2118. URL: <https://doi.org/10.1214/EJP.v18-2118>. MR3035745
- [22] F. Götze, A. Naumov, and A. Tikhomirov. “Local semicircle law under fourth moment condition”. In: *J. Theoret. Probab.* 33.3 (2020), pp. 1327–1362. ISSN: 0894-9840. DOI: 10.1007/s10959-019-00907-y. URL: <https://doi.org/10.1007/s10959-019-00907-y>.
- [23] Friedrich Götze, Alexey Naumov, Alexander Tikhomirov, and Dmitry Timushev. “On the local semicircular law for Wigner ensembles”. In: *Bernoulli* 24.3 (2018), pp. 2358–2400. ISSN: 1350-7265. DOI: 10.3150/17-BEJ931. URL: <https://doi.org/10.3150/17-BEJ931>. MR3757532
- [24] Winfried Hochstättler, Werner Kirsch, and Simone Warzel. “Semicircle law for a matrix ensemble with dependent entries”. In: *J. Theoret. Probab.* 29.3 (2016), pp. 1047–1068. ISSN: 0894-9840. DOI: 10.1007/s10959-015-0602-3. URL: <https://doi.org/10.1007/s10959-015-0602-3>.
- [25] Olav Kallenberg. *Foundations of modern probability*. Vol. 99. Probability Theory and Stochastic Modelling. Third edition [of 1464694]. Springer, Cham, [2021] ©2021, pp. xii+946. ISBN: 978-3-030-61871-1; 978-3-030-61870-4. DOI: 10.1007/978-3-030-61871-1. URL: <https://doi.org/10.1007/978-3-030-61871-1>. MR4226142
- [26] A. Khorunzhy. “On smoothed density of states for Wigner random matrices”. In: *Random Oper. Stochastic Equations* 5.2 (1997), pp. 147–162. ISSN: 0926-6364. DOI: 10.1515/rose.1997.5.2.147. URL: <https://doi.org/10.1515/rose.1997.5.2.147>. MR1454341
- [27] A. M. Khorunzhy and L. A. Pastur. “On the eigenvalue distribution of the deformed Wigner ensemble of random matrices”. In: *Spectral operator theory and related topics*. Vol. 19. Adv. Soviet Math. Amer. Math. Soc., Providence, RI, 1994, pp. 97–127. DOI: 10.1007/978-3-0348-8545-4\_23. URL: [https://doi.org/10.1007/978-3-0348-8545-4\\_23](https://doi.org/10.1007/978-3-0348-8545-4_23).
- [28] Werner Kirsch. “A Survey on the Method of Moments”. Oct. 2015. URL: <https://www.fernuni-hagen.de/stochastik/downloads/momente.pdf>.
- [29] Werner Kirsch. “An elementary proof of de Finetti’s theorem”. In: *Statist. Probab. Lett.* 151 (2019), pp. 84–88. ISSN: 0167-7152. DOI: 10.1016/j.spl.2019.03.014. URL: <https://doi.org/10.1016/j.spl.2019.03.014>. MR3938896
- [30] Werner Kirsch and Thomas Kriecherbauer. “Random matrices with exchangeable entries”. In: *Rev. Math. Phys.* 32.7 (2020), pp. 2050022, 38. ISSN: 0129-055X. DOI: 10.1142/S0129055X20500221. URL: <https://doi.org/10.1142/S0129055X20500221>. MR4127501
- [31] Werner Kirsch and Thomas Kriecherbauer. “Semicircle law for generalized Curie-Weiss matrix ensembles at subcritical temperature”. In: *J. Theoret. Probab.* 31.4 (2018), pp. 2446–2458. ISSN: 0894-9840. DOI: 10.1007/s10959-017-0768-y. URL: <https://doi.org/10.1007/s10959-017-0768-y>.

- [32] Dimitri Shlyakhtenko. “Random Gaussian band matrices and freeness with amalgamation”. In: *Internat. Math. Res. Notices* 20 (1996), pp. 1013–1025. ISSN: 1073-7928. DOI: 10.1155/S1073792896000633. URL: <https://doi.org/10.1155/S1073792896000633>.
- [33] Terence Tao and Van Vu. “Random matrices: universality of local eigenvalue statistics”. In: *Acta Math.* 206.1 (2011), pp. 127–204. ISSN: 0001-5962. DOI: 10.1007/s11511-011-0061-3. URL: <https://doi.org/10.1007/s11511-011-0061-3>.
- [34] Terence Tao and Van Vu. “The Universality Phenomenon for Wigner Ensembles”. Feb. 2012. URL: <https://arxiv.org/pdf/1202.0068>. MR3288230
- [35] Colin Thompson. *Mathematical Statistical Mechanics*. Princeton University Press, 2015. MR0548873

**Acknowledgments.** We are very grateful to two unknown referees who helped to improve the manuscript by their questions and remarks. In particular, Example 2.9 was constructed in response to one of their questions.

---

# Electronic Journal of Probability

## Electronic Communications in Probability

---

### Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS<sup>1</sup>)
- Easy interface (EJMS<sup>2</sup>)

### Economical model of EJP-ECP

- Non profit, sponsored by IMS<sup>3</sup>, BS<sup>4</sup>, ProjectEuclid<sup>5</sup>
- Purely electronic

### Help keep the journal free and vigorous

- Donate to the IMS open access fund<sup>6</sup> (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

---

<sup>1</sup>LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

<sup>2</sup>EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

<sup>3</sup>IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

<sup>4</sup>BS: Bernoulli Society <http://www.bernoulli-society.org/>

<sup>5</sup>Project Euclid: <https://projecteuclid.org/>

<sup>6</sup>IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>