PROOF OF THE LOCAL REM CONJECTURE FOR NUMBER PARTITIONING II: GROWING ENERGY SCALES

CHRISTIAN BORGS[1], JENNIFER CHAYES[1], STEPHAN MERTENS[2], CHANDRA NAIR[1]

Abstract. We continue our analysis of the number partitioning problem with n weights chosen i.i.d. from some fixed probability distribution with density ρ. In Part I of this work, we established the so-called local REM conjecture of Bauke, Franz and Mertens. Namely, we showed that, as $n \to \infty$, the suitably rescaled energy spectrum above some *fixed* scale α tends to a Poisson process with density one, and the partitions corresponding to these energies become asymptotically uncorrelated. In this part, we analyze the number partitioning problem for energy scales α_n that *grow with* n, and show that the local REM conjecture holds as long as $n^{-1/4}\alpha_n \to 0$, and fails if α_n grows like $\kappa n^{1/4}$ with $\kappa > 0$.

We also consider the SK-spin glass model, and show that it has an analogous threshold: the local REM conjecture holds for energies of order $o(n)$, and fails if the energies grow like κn with $\kappa > 0$.

1. Introduction

1.1. **Number Partitioning.** In this paper we continue the study of the energy spectrum of the number partition problem (Npp) with randomly chosen weights. We refer the reader to [BCMN05] for a detailed motivation of this study, but for completeness, we repeat the main definitions.

We consider random instances of the Npp with weights $X_1, \dots, X_n \in \mathbb{R}$ taken to be independently and identically distributed according to some density $\rho(X)$ with finite second moment (for simplicity of notation, we will choose the second moment to be one). Given these weights, one seeks a partition of these numbers into two subsets such that the sum of numbers in one subset is as close as possible to the sum of numbers in the other subset. Each of the 2^n partitions can be encoded as $\boldsymbol{\sigma} \in \{-1,+1\}^n$, where $\sigma_i = 1$ if X_i is put in one subset and $\sigma_i = -1$ if X_i is put in the other subset; in the physics literature, such partitions $\boldsymbol{\sigma}$ are identified with *Ising spin configurations*. The cost function to be minimized over all spin configurations $\boldsymbol{\sigma}$ is the *energy*

$$E(\boldsymbol{\sigma}) = \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \sigma_i X_i \right|, \tag{1.1}$$

where, as in [BCMN05], we have inserted a factor $1/\sqrt{n}$ to simplify the equations in the rest of the paper.

Note that this scaling implies that the *typical* energies are of order one, and the maximal energies are of order $\sqrt{n}$. Indeed, if $\boldsymbol{\sigma}$ is chosen uniformly at random and $X_1, \dots, X_n$ are i.i.d. with second moment one, the random variable $n^{-1/2}\sum_i \sigma_i X_i$ converges to a standard normal as $n \to \infty$, implying in particular that for a typical

Date: August 23, 2005.

configuration, $E(\boldsymbol{\sigma})$ is of order one. The maximal energy, on the other hand, is equal to $n^{-1/2}\sum_i |X_i|$. By the law of large numbers, this implies that the maximal energy is asymptotically equal to $\sqrt{n}$ times the expectation of $|X|$.

As usual, the correlation between two different partitions $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ is measured by the *overlap* between $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$, defined as

$$q(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \frac{1}{n}\sum_{k=1}^{n} \sigma_k \tilde{\sigma}_k. \tag{1.2}$$

Note that the spin configurations $\boldsymbol{\sigma}$ and $-\boldsymbol{\sigma}$ correspond to the same partition and therefore of course have the same energy. Thus there are $N = 2^{n-1}$ distinct partitions and (with probability one) also N distinct energies. The *energy spectrum* is the sorted increasing sequence $E_1, ..., E_N$ of the energy values corresponding to these N distinct partitions. Taking into account that, for each i, there are two configurations with energy E_i, we define $\boldsymbol{\sigma}^{(i)}$ to be the random variable which is equal to one of these two configurations with probability 1/2, and equal to the other with probability 1/2. Then the overlap between the configurations corresponding to i^{th} and j^{th} energies is the random variable $q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)})$.

As noted in [BCMN05], neither the distribution of the energies, nor the distribution of the overlaps changes if one replaces the density $\rho(X)$ by the symmetrized density $\frac{1}{2}(\rho(X)+\rho(-X))$. We may therefore assume without loss of generality that $\rho(X) = \rho(-X)$. Under this assumption, it is easy to see that the energies $E(\boldsymbol{\sigma})$ for the different configurations $\boldsymbol{\sigma}$ are identically distributed. Let us stress, however, that these energies are *not* independently distributed; the energies between different configurations are *correlated* random variables. Indeed, there are $N = 2^{n-1}$ energies, $E_1, \dots E_n$, constructed from only n independent variables $X_1, \dots X_n$.

Consider now a very simple model, the so-called random energy model (REM) first introduced by Derrida [Der81] in a different context. The defining property of the REM is that N energies $E(\boldsymbol{\sigma})$ are taken to be *independent*, identically distributed random variables. In the REM, the asymptotic energy spectrum for large N can be easily determined with the help of large order statistics: if the energies are ordered in increasing order and $\alpha \geq 0$ is any fixed energy scale, the suitably rescaled energy spectrum above α converges to a Poisson process. More precisely, if the distribution of $E(\boldsymbol{\sigma})$ has a non-vanishing, continuous density $g(\alpha)$ at α and E_{r+1} is the first energy above α, then the rescaled energies $(E_{r+1} - \alpha)Ng(\alpha)$, $(E_{r+2} - \alpha)Ng(\alpha)$, $\dots$ converge to a Poisson process with density one.

In spite of the correlations between the energies $E(\boldsymbol{\sigma})$ in the NPP, it had been conjectured [Mer00, BFM04] that as $n \to \infty$, the energy spectrum above any fixed energy α behaves asymptotically like the energy spectrum of the REM, in the sense that the suitably rescaled spectrum above α again becomes a Poisson process. In [BFM04] it was also conjectured that the overlaps corresponding to adjacent energies are asymptotically uncorrelated, so that the suitably normalized overlaps converges to a standard normal. These two, at first sight highly speculative, claims were collectively called the *local REM conjecture*. This conjecture was supported by detailed simulations.

In Part I of this paper [BCMN05], we proved the local REM conjecture for the NPP with a distribution ρ that has finite second moment and lies in $L^{1+\epsilon}$ for some $\epsilon > 0$. More precisely, under these conditions, we proved that for all $i \neq j$, the suitably normalized overlap between the configurations corresponding to the i^{th} and

j^{th} energy above α becomes asymptotically normal, and that the rescaled energies $(E_{r+1}-\alpha)\xi_n^{-1}$, $(E_{r+2}-\alpha)\xi_n^{-1}$, ... with rescaling factor

$$\xi_n = \sqrt{\frac{\pi}{2}}2^{-(n-1)}e^{\frac{\alpha^2}{2}} \tag{1.3}$$

converge to a Poisson process with density one. Recalling that the normalization in (1.1) corresponds to typical energies of order one, this this establishes the local REM conjecture for *typical* energies.

In [BFM04], the authors expressed the belief that the weak convergence of the rescaled energies to a Poisson process should extend to values of α that grow slowly enough with n, although computational limitations prevented them from supporting this stronger claim by simulations. At first, one might think that the local REM conjecture could hold for $\alpha = o(\sqrt{n})$. Indeed, since the maximal energy is of order $\sqrt{n}$, it is clear that the conjecture is false for $\alpha = c\sqrt{n}$ with large enough c. But if this were the only obstruction, then one might hope that the conjecture could hold up to $\alpha = o(\sqrt{n})$. As we will see in this paper, this is *not the case*; the conjecture will only hold for $\alpha = o(n^{1/4})$.

Before stating this precisely, let us be a little more careful with the scaling of the energy spectrum. Note that the scaling factor (1.3) is equal to $[Ng(\alpha)]^{-1}$, where $N = 2^{n-1}$ is the number of energies, and $g(\alpha) = \sqrt{2/\pi}e^{-\alpha^2/2}$ is the density of the absolute value of a standard normal, in accordance with the expected asymptotic density of $E(\boldsymbol{\sigma})$ according to the local limit theorem. But it is well know that, in general, the local limit theorem does not hold in the tails of the distribution. For growing α_n, the REM conjecture should therefore be stated with a scaling factor that is equal to the inverse of 2^{n-1} multiplied by the density of the energy $E(\boldsymbol{\sigma})$ at α. We call the REM conjecture with this scaling the *modified* REM conjecture. It is this modified REM conjecture that one might naively expect to hold for $\alpha = o(\sqrt{n})$.

It turns out, however, that at least for the Npp, this distinction does not make much of a difference. For $\alpha = o(n^{1/4})$, the original and the modified conjectures are equivalent, and the original REM conjecture holds, while for α growing like a constant times $n^{1/4}$, both the original and the modified REM conjectures fail. So for the Npp, the threshold for the validity of the REM conjecture is $n^{1/4}$, not $\sqrt{n}$ as one might have naively guessed.

1.2. **The SK Spin Glass.** In a follow-up paper to [BFM04], Bauke and Mertens generalized the local REM conjecture for the Npp to a large class of disordered systems, including many other models of combinatorial optimization as well as several spin glass models [BM04]. Motivated by this conjecture, Bovier and Kurkova developed an axiomatic approach to Poisson convergence, covering, in particular many types of spin glasses like the Edwards-Anderson model and the Sherrington-Kirkpatrick model.

The Sherrington-Kirkpatrick model (SK model) is defined by the energy function

$$E(\boldsymbol{\sigma}) = \frac{1}{\sqrt{n}}\sum_{i,j=1}^{n} X_{ij}\sigma_i\sigma_j. \tag{1.4}$$

As before, $\boldsymbol{\sigma}$ is a spin configuration in $\{-1,+1\}^n$, but now the random input is given in terms of n^2 random variables X_{ij} with $i,j \in \{1,\dots,n\}$, usually taken to be i.i.d. standard normals. Again $E(\boldsymbol{\sigma}) = E(-\boldsymbol{\sigma})$, leading to $N = 2^{n-1}$ *a priori* different energies $E_1 \leq E_2 \leq \cdots \leq E_N$. Note, however, that the normalization in (1.4) corresponds to typical energies of order $\sqrt{n}$ and maximal energies of order n, in accordance with the standard physics notation.

Consider an energy scale $\alpha_n \geq 0$, and let E_{r+1} be the first energy above α_n. To obtain the REM approximation for the SK model, we observe that the random variable $E(\boldsymbol{\sigma})$ is a Gaussian with density

$$\tilde{g}(x) = \frac{1}{\sqrt{2\pi n}} e^{-x^2/2n}. \tag{1.5}$$

The REM approximation for the SK model therefore suggests that the rescaled energy spectrum $(E_{r+1} - \alpha_n)\tilde{\xi}_n^{-1}$, $(E_{r+2} - \alpha_n)\tilde{\xi}_n^{-1}$, ... with rescaling factor

$$\tilde{\xi}_n = \sqrt{2\pi n}\, 2^{-(n-1)} e^{\frac{\alpha_n^2}{2n}} \tag{1.6}$$

converges to a Poisson process with density one. Recalling that typical energies are now of order $\sqrt{n}$, the local REM conjecture for *typical energies* in the SK model thus claims that for $\alpha_n = O(\sqrt{n})$, the rescaled energy spectrum converges to a Poisson process with density one, with overlaps which again tend to zero as $n \to \infty$.

This conjecture was proved in a very nice paper by Bovier and Kurkova [BK05a]. In fact, they proved that the conjecture remains valid as long as $\alpha_n = O(n^\eta)$ with $\eta < 1$. To get some insight into still faster growing α_n, Bovier and Kurkova then considered the generalized random energy model (GREM) of Derrida [Der85]. For this model, they proved [BK05b] that the local REM conjecture holds up to $\alpha_n = \beta_0 n$, where β_0 is the inverse transition temperature of the GREM, and fails once α_n exceeds this threshold. Based on these results for the GREM, Bovier and Kurkova then suggested [BK05c] that the β_0 might be the threshold for other disordered spin systems as will.

As we will show in this paper, this is not the case, at least not for the SK model, for which we prove that the REM conjecture holds up to the threshold $\alpha_n = o(n)$, and becomes invalid as soon as $\limsup \alpha_n/n > 0$, see Theorem 2.2 below for the precise statement. Thus even the scaling with n of the threshold does not obey the naive expectation derived from the GREM. Note that for the SK model there is no difference between the original REM conjecture and the modified REM conjecture, since the density of $E(\boldsymbol{\sigma})$ is Gaussian for all energy scales.

1.3. **Organization of the Paper.** This paper is organized as follows. In the next section, we precisely state the assumptions on our model and formulate our main results, see Theorems 2.1 and 2.2. In Section 3, we then describe our main proof strategy for the NPP. Since the proof strategy for the SK model only requires minor modifications (the proof is, in fact, much easier), we defer the discussion of this model to the last subsection, Section 3.7. The next four sections contain the details of the proof: As a warmup, we start with the NPP with Gaussian noise, where our strategy is most straightforward. Next, in Section 5, we move to the NPP with a general distribution. This section contains the meat of our proof: the establishment of a large deviations estimate for the probability density of several (weakly dependent) random variables. In Section 6 we give the proof of Theorem 2.2, and in Section 7 we establish several auxiliary results needed in the rest of the paper. We conclude the paper with a section summarizing our results and discussing possible extensions, Section 8.

2. Statement of Results

2.1. **Number Partitioning.** Let $X_1, ..., X_n$ be independent random variables distributed according to the common density function $\rho(x)$. We assume that ρ has second moment one and satisfies the bound

$$\int_{-\infty}^{\infty} |\rho(x)|^{1+\epsilon}\ dx < \infty \tag{2.1}$$

for some $\epsilon > 0$. Since neither the distribution of the overlaps nor the energy spectrum changes if we replace $\rho(x)$ by $\frac{1}{2}(\rho(x)+\rho(-x))$, we will assume that $\rho(x) = \rho(-x)$. We use the symbol $\mathbb{P}_n(\cdot)$ to denote the probability with respect to the joint probability distribution of $X_1, ..., X_n$, and the symbol $\mathbb{E}_n(\cdot)$ to denote expectations with respect to $\mathbb{P}_n(\cdot)$.

As in the introduction, we represent the 2^n partitions of the integers $\{1, .., n\}$ as spin configurations $\boldsymbol{\sigma} \in \{-1, +1\}^n$ and define the energy of $\boldsymbol{\sigma}$ as in (1.1). Recalling that the distribution of $E(\boldsymbol{\sigma})$ does not depend on $\boldsymbol{\sigma}$, let $g_n(\cdot)$ be the density of $E(\boldsymbol{\sigma})$, and let ξ_n be the modified scaling factor

$$\xi_n = \frac{1}{2^{n-1} g_n(\alpha_n)}. \tag{2.2}$$

We now introduce a continuous time process $\{N_n(t) \colon t \geq 0\}$ where $N_n(t)$ is defined as the number of points in the energy spectrum that lie between α_n and $\alpha_n + t\xi_n$.

Let $E_1, \ldots, E_N$ be the increasing spectrum of the energy values corresponding to the $N = 2^{n-1}$ distinct partitions. Given $\alpha_n \geq 0$, let r_n be the random variable defined by $E_{r_n} < \alpha_n \leq E_{r_n+1}$. For $j > i > 0$, we then define the rescaled overlap Q_{ij} as the random variable

$$Q_{ij} = \frac{1}{4} \sum_{\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}} n^{1/2} q(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) \tag{2.3}$$

where the sum goes over the four pairs of configurations with $E(\boldsymbol{\sigma}) = E_{r_n+i}$ and $E(\tilde{\boldsymbol{\sigma}}) = E_{r_n+j}$. Instead of the overlap Q_{ij}, we will sometimes consider the following variant: consider two distinct configurations $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ chosen uniformly at random from all pairs of distinct configurations. We then define $Q_{n,t}$ as the rescaled overlap $n^{1/2} q(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})$ conditioned on the event that $E(\boldsymbol{\sigma})$ and $E(\tilde{\boldsymbol{\sigma}})$ fall into the energy interval $[\alpha_n, \alpha_n + t\xi_n]$. We will refer to $Q_{n,t}$ as the *overlap between two typical configurations contributing to* $N_n(t)$.

The main results of this paper are statements ii) and iii) of the following theorem. The first statement is a corollary of the proof of ii) and iii) and implies that for $\alpha_n = o(n^{1/4})$, the original and the modified REM conjecture are equivalent.

Theorem 2.1. *Let $X_1, ..., X_n \in \mathbb{R}$ be i.i.d. random variables drawn from a probability distribution with second moment one and even density ρ. If ρ obeys the assumption* (2.1) *for some $\epsilon > 0$ and has a Fourier transform that is analytic in a neighborhood of zero, then the following holds:*

i) Let $g_n(\cdot)$ be the density of $E(\boldsymbol{\sigma})$, and let $g(\alpha) = \sqrt{2/\pi} e^{-\alpha^2/2}$. If $\alpha_n = o(n^{1/4})$, then $g_n(\alpha_n) = g(\alpha_n)(1 + o(1))$.

ii) Let $\alpha_n = o(n^{1/4})$, and let $j > i > 0$ be arbitrary integers not depending on n. As $n \to \infty$, the process $N_n(t)$ converges weakly to a Poisson process with density one, and both Q_{ij} and $Q_{n,t}$ converge in distribution to standard normals.

iii) Let $\alpha_n = \kappa n^{1/4}$ for some finite $\kappa > 0$. Then $\mathbb{E}_n[N_n(t)] = t + o(1)$, but the process $N_n(t)$ does not converge to a Poisson process, and $Q_{n,t}$ does not converge to a standard normal.

In order to prove the above theorem, we will analyze the factorial moments of the process $Z_n(t)$. We will show that for $\alpha_n = o(n^{1/4})$, they converge to those of a Poisson process, and for $\alpha_n = \kappa n^{1/4}$ with $\kappa > 0$, they do not converge to the moments of a Poisson process. Together with suitable upper bounds on the moments of $Z_n(t)$, this allows us to establish non-convergence to Poisson for $\kappa > 0$, but, unfortunately, it does not allow us to establish convergence to some other distribution.

The situation is slightly more "constructive" for the overlap distribution: here we are able to determine the limiting distribution of $Q_{n,t}$ for $\alpha_n = \kappa n^{1/4}$ with $\kappa > 0$. In this regime, the distribution of $Q_{n,t}$ converges to a convex combination of two shifted Gaussians: with probability 1/2 a Gaussian with mean κ^2 and variance one, and with probability 1/2 a Gaussian with mean $-\kappa^2$ and variance one, so in particular $Q_{n,t}$ is not asymptotically normal. As we will see, this is closely connected to the failure of Poisson convergence; see Remark 4.2 in Section 4 and Remark 5.8 in Section 5 below.

2.2. **SK Spin Glass.** We consider the SK model with energies given by (1.4) and random coupling X_{ij} which are i.i.d. standard normals. Let $N = 2^{n-1}$, and let $E_1, \ldots, E_N$ and $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(N)}$ be as defined in the introduction.

Given an energy scale $\alpha_n \geq 0$ and two integers $j > i > 0$, we again introduce Q_{ij} as the random variable defined in (2.3), with r_n given by the condition that $E_{r_n} < \alpha_n \leq E_{r_n+1}$. Finally, we define $N_n(t)$ to be the number of points in the energy spectrum of (1.4) that lie between α_n and $\alpha_n + t\tilde{\xi}_n$, with $\tilde{\xi}_n$ given by (1.6). We say that the local REM conjecture holds if $N_n(t)$ converges weakly to a Poisson process with density one, and Q_{ij} converges in distribution to a standard normal for all $j > i > 0$.

Our proofs for the Npp can then be easily generalized to give the following theorem.

Theorem 2.2. *There exists a constant $\epsilon_0 > 0$ such the following statements hold for all sequences of positive real numbers α_n with $\alpha_n \leq \epsilon_0 n$:*

(i) $\mathbb{E}(N_n(t)) \to t$ as $n \to \infty$.

(ii) The local REM conjecture for the SK model holds if and only if $\alpha_n = o(n)$.

3. Proof Strategy

In this section we describe our proof strategy for Theorems 2.1 and 2.2. We explain our ideas using the example of the Npp, referring to the SK model only in the last subsection, Section 3.7.

3.1. **Factorial moments.** Consider a finite family of non-overlapping intervals $[c_1, d_1]$, $\ldots$, $[c_m, d_m]$ with $d_\ell > c_\ell \geq 0$, and let $\gamma_\ell = d_\ell - c_\ell$. Weak convergence of the process $\{N_n(t) \colon t \geq 0\}$ to a Poisson process of density one is equivalent to the statement that for each such family, the increments $N_n(d_1) - N_n(c_1)$, $\ldots$, $N_n(d_m) - N_n(c_m)$ converge to independent Poisson random variables with rates $\gamma_1, \ldots, \gamma_m$.

Let $Z_n(a,b)$ be the number of point in the energy spectrum that fall into the interval $[a,b]$. We then rewrite $N_n(d_\ell)-N_n(c_\ell)$ as $Z_n(a_n^\ell, b_n^\ell)$, where $a_n^\ell = \alpha + c_\ell\xi_n$ and $b_n^\ell = \alpha + d_\ell\xi_n$, with ξ_n defined in equation (2.2). We prove convergence of the increments to independent Poisson random variables by proving convergence of the multidimensional factorial moments, i.e., by proving the following theorem. To simplify our notation, we will henceforth drop the index n on both the symbol $\mathbb{E}_n$ and $\mathbb{P}_n$.

Theorem 3.1. *Let ρ be as in Theorem 2.1, let $\alpha_n = o(n^{1/4})$ be a sequence of positive real numbers, let m be a positive integer, let $[c_1,d_1]$, $\dots$, $[c_m,d_m]$ be a family of non-overlapping, non-empty intervals, and let $(k_1,..,k_m)$ be an arbitrary m-tuple of positive integers. For $\ell=1,\dots,m$, set $a_n^\ell = \alpha_n + c_\ell\xi_n$, $b_n^\ell = \alpha_n + d_\ell\xi_n$, and $\gamma_\ell = d_\ell - c_\ell$. Under these conditions, we have*

$$\lim_{n\to\infty} \mathbb{E}[\prod_{\ell=1}^m (Z_n(a_n^\ell, b_n^\ell))_{k_\ell}] = \prod_{\ell=1}^m \gamma_\ell^{k_\ell}, \tag{3.1}$$

where, as usual, $(Z)_k = Z(Z-1)\dots(Z-k+1)$.

Theorem 3.1 establishes that $\{N_n(t)\colon t\geq 0\}$ converges to a Poisson process with density one if $\alpha_n = o(n^{1/4})$. The convergence of the overlaps is an easy corollary to the proof of this theorem. The details are exactly the same as in [BCMN05], and will not be repeated here.

The failure of Poisson convergence for faster growing α_n is easiest to explain for the case in which $X_1,\dots,X_n$ are standard normals, since this does not require us to distinguish between the original and the modified REM conjectures. Our proof is again based on the analysis of the factorial moments of $N_n(t)$.

More precisely, we will show that $E[N_n(t)]$ converges to t, while the second factorial moment, $E[(N_n(t))_2]$ does not converge to t^2. Note that this fact by itself is not enough to exclude convergence to a Poisson process since convergence of the factorial moments is, in general, only a sufficient condition for weak convergence to a Poisson process. But combined with suitable estimates on the growth of the third moment, the fact that $\mathbb{E}[(N_n(t))_2]$ does not converge to t^2 is enough. This follows from the following lemma, which is an easy consequence of a standard theorem on uniformly integrable sequences of random variables (see, e.g., Theorem 25.12 and its corollary in [Bi94]).

Lemma 3.2. *Let $Z_n\geq 0$ be a sequence of random variables such that $\mathbb{E}[Z_n^r]$ is bounded uniformly in n for some $r<\infty$. If Z_n converges weakly to a Poisson random variable with rate $\gamma>0$, then $\lim_{n\to\infty}\mathbb{E}[(Z_n)_k] = \gamma^k$ for all $k<r$.*

Combined with this lemma, the next theorem establishes the third statement of Theorem 2.1 if the weights $X_1,\dots,X_n$ are Gaussian. The proof for non-Gaussian weights will be given in Section 5.4.

Theorem 3.3. *Let $X_1,...,X_n\in\mathbb{R}$ be i.i.d. random variables with normal distribution, and let $\alpha_n = o(\sqrt{n})$.*

(i) $\mathbb{E}[N_n(t)] = t + o(1)$ for all fixed $t>0$.

(ii) Let m, a_n^ℓ, b_n^ℓ, γ_ℓ, and $k_1,..,k_m$ be as in Theorem 3.1. For $k=\sum_{\ell=1}^m k_\ell \geq 2$, we then have

$$\mathbb{E}[\prod_{\ell=1}^m (Z_n(a_n^\ell, b_n^\ell))_{k_\ell}] = \Big(\prod_{\ell=1}^m \gamma_\ell^{k_\ell}\Big) e^{\frac{k(k-1)}{4n}\alpha_n^4} e^{O(\alpha_n^6 n^{-2}) + o(1)}. \tag{3.2}$$

Theorem 3.3 will be proved in Section 4, and Theorem 3.1 will be proved in Section 5. In the remainder of this section, we will map out the general proof strategy, and in the process, establish several properties which will be used in the proofs of Theorems 3.1 and 3.3.

3.2. **Analysis of first moment.** In order to analyze the first moment we rewrite $Z_n(a_n^\ell, b_n^\ell)$ as

$$Z_n(a_n^\ell, b_n^\ell) = \sum_{\boldsymbol{\sigma}} I^{(\ell)}(\boldsymbol{\sigma}) \tag{3.3}$$

where $I^{(\ell)}(\boldsymbol{\sigma})$ is 1/2 times an indicator function of the event that the energy $E(\boldsymbol{\sigma})$ falls into the interval $[a_n^\ell, b_n^\ell]$ (the factor 1/2 compensates for the fact that the sum in (3.3) goes over all configurations $\boldsymbol{\sigma} \in \{-1,+1\}^n$, and therefore counts every distinct partition twice). Taking expectations of both sides we see that

$$\mathbb{E}[Z_n(a_n^\ell, b_n^\ell)] = \frac{1}{2} \sum_{\boldsymbol{\sigma}} \mathbb{P}\Big(E(\boldsymbol{\sigma}) \in [a_n^\ell, b_n^\ell]\Big). \tag{3.4}$$

Since we have taken $\rho(x) = \rho(-x)$, the distribution of $E(\boldsymbol{\sigma})$ is identical for all $\boldsymbol{\sigma}$, so that the probability on the right hand side is independent of $\boldsymbol{\sigma}$. In order to prove that $\mathbb{E}[Z_n(a_n^\ell, b_n^\ell)] = \gamma_\ell(1 + o(1))$, we will therefore want to prove that

$$\mathbb{P}\Big(E(\boldsymbol{\sigma}) \in [a_n^\ell, b_n^\ell]\Big) = 2^{-(n-1)}\gamma_\ell(1 + o(1)). \tag{3.5}$$

Rewriting the left hand side as

$$\mathbb{P}\Big(E(\boldsymbol{\sigma}) \in [a_n^\ell, b_n^\ell]\Big) = \int_{a_n^\ell}^{b_n^\ell} g_n(y)dy, \tag{3.6}$$

where $g_n(\cdot)$ is the density of $E(\boldsymbol{\sigma})$ with respect to the Lesbesgue measure on $\mathbb{R}_+ = \{x \in \mathbb{R}\colon x \geq 0\}$, and recalling the definition $\xi_n = [2^{n-1}g_n(\alpha_n)]^{-1}$, we see that the bound (3.5) is equivalent to the statement that $g_n(y) = g_n(\alpha_n)(1+o(1))$ whenever $y = \alpha_n + O(\xi_n)$.

This bound is easily established in the Gaussian case, where it holds as long as $\alpha_n \leq c\sqrt{n}$ for some $c < \sqrt{2 \log 2}$. Indeed, let us write $E(\boldsymbol{\sigma})$ as $|H(\boldsymbol{\sigma})|$, where

$$H(\boldsymbol{\sigma}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_i X_i. \tag{3.7}$$

If the random variables $X_1, \ldots, X_n$ are standard Gaussians, the random variables $H(\boldsymbol{\sigma})$ are standard Gaussians as well, implying that $E(\boldsymbol{\sigma})$ has density

$$g(y) = \sqrt{\frac{2}{\pi}} e^{-y^2/2} \tag{3.8}$$

with respect to the Lesbesgue measure on $\mathbb{R}_+ = \{x \in \mathbb{R}\colon x \geq 0\}$. If $\alpha_n \leq c\sqrt{n}$ for some $c < \sqrt{2 \log 2}$, then $\xi_n = o(e^{-\epsilon n})$ for some $\epsilon > 0$, implying that for $y = \alpha_n + O(\xi_n)$, we have $g(y) = g(\alpha_n)(1+O(\sqrt{n}\xi_n)) = g(\alpha_n)(1+o(1))$, as desired. This proves (3.5) and hence the bound $\mathbb{E}[Z_n(a_n^\ell, b_n^\ell)] = \gamma_\ell(1 + o(1))$ provided $\alpha_n \leq c\sqrt{n}$ for some $c < \sqrt{2 \log 2}$. Note that this already establishes the first moment bound stated in Theorem 3.3.

For more general distributions, the proof is more complicated since $g_n(\alpha_n)$ is no longer given by a simple formula like (3.8). But given our assumptions on ρ, we will be able to show that under the assumption that $\alpha_n = o(\sqrt{n})$, it can be written in the form

$$g_n(\alpha_n) = \sqrt{\frac{2}{\pi}} e^{-nG(\alpha_n n^{-1/2})}(1 + o(1)) \tag{3.9}$$

where G is an even function which is analytic in a neighborhood of zero and satisfies the bound

$$G(x) = \frac{x^2}{2} + O(x^4). \tag{3.10}$$

The proof of the representation (3.9) involves an integral representation for g_n combined with a steepest descent analysis and will be given in Section 5.2.

The bounds (3.9) and (3.10) have several immediate consequences. First, they clearly imply that $g_n(\alpha_n) = g(\alpha_n)(1 + o(1))$ if $\alpha_n = o(n^{1/4})$, proving the first statement of Theorem 2.1. Second, they imply that ξ_n decays exponentially in n if $\alpha_n = o(\sqrt{n})$. Using the bounds (3.9) and (3.10) once more, we conclude that $g_n(y) = g_n(\alpha_n)(1 + o(1) + O(\alpha_n \xi_n)) = g_n(\alpha_n)(1 + o(1))$ whenever $y = \alpha_n + O(\xi_n)$. For $\alpha_n = o(\sqrt{n})$, we therefore get convergence of the first moment, $\mathbb{E}[Z_n(a_n^\ell, b_n^\ell)] = \gamma_\ell(1 + o(1))$, implying in particular that $\mathbb{E}[Z_n(t)] = t + o(1)$ as claimed in the last statement of Theorem 2.1.

3.3. **Factorial moments as sums over pairwise distinct configurations.** Next we turn to higher factorial moments. Before studying these factorial moments, let us consider the standard moments $\mathbb{E}[\prod_{\ell=1}^m (Z_n(a_n^\ell, b_n^\ell))^{k_\ell}]$. In view of (3.3), these moments can be written as a sum over k configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$, where $k = \sum_{\ell=1}^m k_\ell$. As already observed in [BCP01], the factorial moments can be expressed in a similar way, the only difference being that the sum over configurations is now a sum over pairwise distinct configurations, i.e., configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)} \in \{-1, +1\}^n$ such that $\boldsymbol{\sigma}^{(i)} \neq \pm \boldsymbol{\sigma}^{(j)}$ for all $i \neq j$. Explicitly,

$$\begin{aligned} \mathbb{E}\Big[\prod_{\ell=1}^m (Z_n(a_n^\ell, b_n^\ell))_{k_\ell}\Big] &= \sum_{\substack{\boldsymbol{\sigma}^{(1)},\ldots,\boldsymbol{\sigma}^{(k)}: \\ \boldsymbol{\sigma}^{(i)} \neq \pm \boldsymbol{\sigma}^{(j)}}} \mathbb{E}\Big[\prod_{j=1}^k I^{(\ell(j))}(\boldsymbol{\sigma}^{(j)})\Big] \\ &= \frac{1}{2^k} \sum_{\substack{\boldsymbol{\sigma}^{(1)},\ldots,\boldsymbol{\sigma}^{(k)}: \\ \boldsymbol{\sigma}^{(i)} \neq \pm \boldsymbol{\sigma}^{(j)}}} \mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j = 1, \ldots, k\Big), \end{aligned} \tag{3.11}$$

where the sums run over pairwise distinct configurations and $\ell(j) = 1$ if $j = 1, \ldots, k_1$, $\ell(j) = 2$ if $j = k_1 + 1, \ldots, k_1 + k_2$, and so on. See [BCMN05] for the (straightforward) derivation of (3.11).

In order to prove convergence of the higher factorial moments, we therefore would like to show that the probability

$$\mathbb{P}\left(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all j}\right)$$

asymptotically factors into the product $\prod_j \mathbb{P}(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}])$. Unfortunately, this asymptotic factorization does not hold for arbitrary families of distinct configurations $\boldsymbol{\sigma}^{(1)} \ldots \boldsymbol{\sigma}^{(k)}$. This problem is already present for α_n that are bounded as $n \to \infty$ (see [BCMN05]), and – in a milder form – it is even present for the special case of $\alpha_n = 0$ treated in [BCP01].

3.4. **Typical and atypical configurations.** As in [BCMN05] and [BCP01], it is useful to distinguish between "typical" and "atypical" sets of configurations $\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}$ when analyzing the right hand side of (3.11). To this end, we consider the matrix M formed by the row vectors $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$. Given a vector $\boldsymbol{\delta} \in \{-1, 1\}^k$, we then define $n_{\boldsymbol{\delta}}(\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)})$ as the number of times the column vector $\boldsymbol{\delta}$ appears in the matrix M,

$$n_{\boldsymbol{\delta}} = n_{\boldsymbol{\delta}}(\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}) = |\{j \leq n : (\sigma_j^{(1)}, ..., \sigma_j^{(k)}) = \boldsymbol{\delta}\}|. \tag{3.12}$$

If $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)} \in \{-1, +1\}^n$ are chosen independently and uniformly at random, then the expectation of $n_{\boldsymbol{\delta}}$ is equal to $n2^{-k}$ for all $\delta \in \{-1, +1\}^k$. By a standard martingale argument, for most configurations, the difference between $n_{\boldsymbol{\delta}}$ and $n2^{-k}$ is therefore not much larger than $\sqrt{n}$. More precisely, for any $\lambda_n \to \infty$ as $n \to \infty$, all but a vanishing fraction of the configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$ obey the condition

$$\max_{\boldsymbol{\delta}} |n_{\boldsymbol{\delta}}(\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}) - \frac{n}{2^k}| \leq \sqrt{n}\, \lambda_n n, \tag{3.13}$$

see Lemma 3.9 in [BCMN05] for a proof.

The proof of Theorems 3.1 and 3.3 now proceeds in two steps. First, we show that the contribution of the configurations that violate (3.13) is negligible as $n \to \infty$, and second we analyze the configurations satisfying (3.13). It turns out that first part is quite complicated and requires distinguishing several sub-classes of configurations, but this analysis has already been carried out in [BCMN05], resulting in bounds that are sharp enough for growing α_n as well. So the only additional work needed is a sharp analysis of the typical configurations.

The next lemma summarizes the main results from [BCMN05] needed in this paper. To state it, we define $R_{n,k}(\lambda_n)$ as

$$R_{n,k}(\lambda_n) = \frac{1}{2^k} \sideset{}{'}\sum_{\boldsymbol{\sigma}^{(1)}, \cdots, \boldsymbol{\sigma}^{(k)}} \mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j = 1, \ldots, k\Big), \tag{3.14}$$

where the sum runs over pairwise distinct configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$ that are either linearly dependent or violate the bound (3.13). We also use the notation

$$q_{\max} = \max_{i \neq j} |q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)})| \tag{3.15}$$

for the maximal off-diagonal overlap of $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$.

Lemma 3.4. *Let λ_n be a sequence of positive numbers.*

(1) *Then the number of configurations $\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}$ that violate the condition* (3.13) *is bounded by $2^{nk} 2^{k+1} e^{-\frac{1}{2}\lambda_n^2}$.*
(2) *Assume that both α_n and λ_n are of order $o(\sqrt{n})$. Then there are constants $c, C < \infty$ depending only on k, $\gamma_1, \ldots, \gamma_k$, and the sequence λ_n, such that for n sufficiently large we have*

$$R_{n,k}(\lambda_n) \leq C n^c e^{\frac{1}{2}k\alpha_n^2(1+o(1))} \big(\xi_n^{1/n_0} + e^{-\lambda_n^2/2}\big). \tag{3.16}$$

Here $n_0 = (1 + \epsilon)/\epsilon$ with ϵ as in (2.1).
(3) *Let $\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}$ be an arbitrary set of row vectors satisfying* (3.13). *Then*

$$q_{\max} \leq 2^k \frac{\lambda_n}{\sqrt{n}}, \tag{3.17}$$

and hence for $\lambda_n = o(\sqrt{n})$ *and* n *sufficiently large, we have that* $\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}$ *are linearly independent.*

Proof. Statements (1) and (3) are copied from Lemma 3.8 in [BCMN05], while statement (2) is a consequence of the bounds (3.65) and (3.67) in [BCMN05] and the fact that $2^n \xi_n = e^{\frac{1}{2}\alpha_n^2(1+o(1))} e^{O(1)}$, a bound which follows immediately from (3.9), (3.10), and the definition (2.2) of ξ_n. □

3.5. **Factorization for typical configurations.** In view of Lemma 3.4, it will be enough to analyze the expectations on the right hand side of (3.11) for configurations $\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}$ that satisfy (3.13) and are linearly independent, provided we choose λ_n in such a way that $\alpha_n = o(\lambda_n)$, $\lambda_n = o(\sqrt{n})$ and $e^{-\lambda_n^2/2}$ decays faster than any power of n.

Consider therefore a family of linearly independent configurations satisfying the condition (3.13). The main technical result of this paper is that the following approximate factorization statement

$$\begin{aligned}\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j = 1, \dots, k\Big)& \\ = \prod_{j=1}^{k} \mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}]\Big) e^{O(\alpha_n^2 q_{\max}) + o(1)}&\end{aligned} \tag{3.18}$$

is valid whenever n is large enough and $\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}$ obey the condition (3.13). (For the non-Gaussian case, we will also need that $\alpha_n = o(n^{1/4})$, while the assumption $\alpha_n = o(\sqrt{n})$ stated at the beginning of this subsection is enough in the Gaussian case.)

If $X_1, \dots, X_n$ are standard normals, the bound (3.18) is quite intuitive and not hard to prove. Indeed, let $H(\boldsymbol{\sigma})$ be the random variable defined in (3.7), and let $\kappa^{(k)}(\cdot)$ be the joint density of $H(\boldsymbol{\sigma}^{(1)})$, $\dots$, $H(\boldsymbol{\sigma}^{(k)})$. Using the notation $\mathbf{x}$ and $\mathbf{y}$ for vectors with components $x_1, \dots, x_k$ and $y_1, \dots, y_k$, respectively, we then express the joint density of $E(\boldsymbol{\sigma}^{(1)})$, $\dots$, $E(\boldsymbol{\sigma}^{(k)})$ with respect to the Lebesque measure on $\mathbb{R}_+^k$ as

$$g^{(k)}(\mathbf{y}) = \sum_{\substack{x_1, \dots x_k: \\ y_i = \pm x_i}} \kappa^{(k)}(\mathbf{x}). \tag{3.19}$$

If $X_1, \dots, X_n$ are standard Gaussians, the joint distribution of $H(\boldsymbol{\sigma}^{(1)})$, $\dots$, $H(\boldsymbol{\sigma}^{(k)})$ is Gaussian as well, with mean zero and covariance matrix

$$C_{ij} = E[H(\boldsymbol{\sigma}^{(i)}) H(\boldsymbol{\sigma}^{(j)})] = q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)}), \tag{3.20}$$

leading to the representation

$$\kappa^{(k)}(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}} \frac{1}{\sqrt{\det C}} \exp\Big(-\frac{1}{2}(\mathbf{x}, C^{-1}\mathbf{x})\Big), \tag{3.21}$$

whenever C is invertible. As usual, C^{-1} denotes the matrix inverse of C, and $(\mathbf{x}, C^{-1}\mathbf{x}) = \sum_{i,j} x_i C_{ij}^{-1} x_j$. Observing that $C_{ij}^{-1} = \delta_{ij} + O(q_{\max})$, this immediately leads to the bound (3.18); see Section 4 for details.

In the non-Gaussian case, we do not have an explicit formula for the joint density of $E(\boldsymbol{\sigma}^{(1)})$, $\dots$, $E(\boldsymbol{\sigma}^{(k)})$, thus making the proof of the approximate factorization formula (3.18) much more difficult. Basically, the proof requires a multi-dimensional local limit theorem for a case in which the arguments of the probability density under investigation grow with n. In order to prove this local limit theorem, we will use an integral representation for the probabilities on the left hand side of (3.18) and then use a saddle point analysis to prove (3.18). In contrast to [BCMN05] and [BCP01], where it was sufficient to analyze the saddle point of the integral in the original domain of integration, the case with growing α_n considered here requires a more sophisticated analysis, involving the shift to a complex saddle point in a complex space of dimension k; see Section 5 for the details.

Let us close this section by showing how to complete the proof of Theorem 3.1 once the bound (3.18) is established. To this end, let us first consider the case where α_n grows somewhat more slowly than $o(n^{1/4})$, in particular assume that $\alpha_n = o(n^{1/6})$. Choosing $\lambda_n = n^{1/6}$, we then invoke the bound (3.17) from Lemma 3.4 (3) to conclude that $q_{\max} = O(n^{-1/3})$ and $\alpha_n^2 q_{\max} = o(1)$ whenever $\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}$ satisfy the bound (3.13). For a family of configurations satisfying (3.13), the the multiplicative error term in the joint probability on the right hand side of (3.18) is therefore equal to $1 + o(1)$, leading to asymptotic factorization of the probabilities on the right hand side of (3.11). Using Lemma 3.4 (2) to bound the sum over families of configurations not satisfying (3.13), and Lemma 3.4 (1) to show that the number of families of configurations satisfying (3.13) is equal to $2^{nk}(1 + o(1))$, this gives the bound (3.1) from Theorem 3.1.

For still more quickly growing α_n, it is not enough to use the worst case bound (3.17). Instead, we would like to use that a typical family of configurations has maximal off-diagonal overlap of order $n^{-1/2}$. For a typical family of configurations, we therefore get asymptotic factorization as long as $\alpha_n = o(n^{1/4})$. The next lemma, to be proved in Section 7, implies that the error term coming from atypical configurations does not destroy the asymptotic factorization.

Lemma 3.5. *Let $c > 0$, let k be a positive integer, and let α_n be a sequence of positive numbers such that $\alpha_n n^{-1/4} \to 0$ as $n \to \infty$. If f is a function from the set of all families of configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)} \in \{-1, +1\}^n$ into $\mathbb{R}$ such that $|f(\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)})| \le c\alpha_n^2 q_{\max}$, then*

$$2^{-nk} \sum_{\boldsymbol{\sigma}^{(1)},\ldots,\boldsymbol{\sigma}^{(k)}} e^{f(\boldsymbol{\sigma}^{(1)},\ldots,\boldsymbol{\sigma}^{(k)})} = 1 + o(1) \tag{3.22}$$

as $n \to \infty$.

We will now show that Lemmas 3.4 and 3.5 combined with (3.18) and (3.5) immediately imply Theorem 3.1. Indeed, let us choose λ_n in such a way that $\alpha_n = o(\lambda_n)$, $\lambda_n = o(\sqrt{n})$ and $e^{-\lambda_n^2/2}$ decays faster than any power of n. Invoking (3.11) and using Lemma 3.4 (2) to bound the sum over configurations which do not satisfy (3.13) and the bounds (3.18) and (3.5) to approximate the remaining terms on the right hand side of (3.11), we get the estimate

$$\mathbb{E}\Big[\prod_{\ell=1}^{m}(Z_n(a_n^\ell, b_n^\ell))_{k_\ell}\Big] = (1 + o(1))\Big(\prod_{i=1}^{k}\gamma_{\ell(i)}\Big)2^{-nk} \sum_{\boldsymbol{\sigma}^{(1)},\ldots,\boldsymbol{\sigma}^{(k)}}{}^{''} e^{O(\alpha_n^2 q_{\max})+o(1)} \tag{3.23}$$

where the sum $\sum''$ runs over families of linearly independent configurations $\boldsymbol{\sigma}^{(1)}$, ..., $\boldsymbol{\sigma}^{(k)}$ that satisfy (3.13). Using Lemma 3.4 (1) to extend the sum to a sum over all families of configurations $\boldsymbol{\sigma}^{(1)}$, ..., $\boldsymbol{\sigma}^{(k)} \in \{-1, +1\}^n$, we then refer to the statement of Lemma 3.5 to complete the proof of Theorem 3.1.

The above considerations also indicate why the asymptotic factorization of the factorial moments fails if α_n grows faster than $o(n^{1/4})$. Indeed, expressing the matrix elements of C as $C_{ij} = \delta_{ij} + \tilde{q}_{ij}$ where $\tilde{q}_{ij} = q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)})$ if $i \neq j$ and $\tilde{q}_{ij} = 0$ if $i = j$, $(\mathbf{x}, C^{-1}\mathbf{x}) = \sum_j x_j^2 - \sum_{i \neq j} x_i \tilde{q}_{ij} x_j + O(q_{\max}^2 \|\mathbf{x}\|_2^2)$, where, as usual, $\|\mathbf{x}\|_2$ denotes the ℓ_2 norm of $\mathbf{x}$. This in turn leads to the more precise estimate

$$\mathbb{E}\Big[\prod_{\ell=1}^{m} (Z_n(a_n^\ell, b_n^\ell))_{k_\ell}\Big] = \Big(\prod_{i=1}^{k} \gamma_{\ell(i)}\Big) 2^{-nk} \sum_{\boldsymbol{\sigma}^{(1)},\ldots,\boldsymbol{\sigma}^{(k)}} e^{\frac{\alpha_n^2}{2}\sum_{i \neq j} q_{ij} + O(\alpha_n^2 q_{\max}^2) + o(1)} + o(1) \tag{3.24}$$

for the case where $X_1, \ldots, X_n$ are standard normals. While the term $\frac{\alpha_n^2}{2}\sum_{i \neq j} q_{ij}$ is negligible if $\alpha_n = o(n^{1/4})$, it becomes important as soon as α_n grows like $n^{1/4}$ or faster, leading to the failure of asymptotic factorization for the factorial moments. The details are straightforward but a little tedious and are given in Section 4 for the Gaussian case, and in Section 5.4 for the non-Gaussian case.

3.6. **Integral representation.** As discussed in the preceding subsections, the convergence of the factorial moments of Z_n reduces to the proof of the approximate factorization formula (3.18). Following a strategy that was already used in [BCP01] and [BCMN05], our proof for the case where $X_1, \ldots, X_n$ are not normally distributed uses an integral representation for the probabilities on the left hand side of (3.18).

To derive this integral representation we first express the indicator function $I^{(\ell)}(\boldsymbol{\sigma})$ in terms of the function $\mathrm{rect}(x)$ defined to be 1 if $|x| \leq 1/2$ and 0 otherwise. Using Fourier inversion and the fact that the Fourier transform of the function $\mathrm{rect}(x)$ is equal to $\mathrm{sinc}(f) = \frac{\sin \pi f}{\pi f}$ this leads to the representation

$$I^{(\ell)}(\boldsymbol{\sigma}) = q_{n,\ell} \int_{-\infty}^{\infty} \mathrm{sinc}(f q_{n,\ell}) \cos(2\pi f t_n^\ell \sqrt{n}) e^{2\pi i f \sum_{s=1}^{n} \sigma_s X_s} df, \tag{3.25}$$

where $t_n^\ell = (a_n^\ell + b_n^\ell)/2$ denotes the center of the interval $[a_n^\ell, b_n^\ell]$, and $q_{n,\ell} = \gamma_\ell \xi_n \sqrt{n}$. Taking the expectation of both sides of (3.25) and exchanging the expectation and the integral on the right hand side (the justification of this exchange is given by Lemma 3.2 in [BCMN05]), we get the representation

$$\begin{aligned}\mathbb{P}\Big(E(\boldsymbol{\sigma}) \in [a_n^\ell, b_n^\ell]\Big) &= 2\mathbb{E}[I^{(\ell)}(\boldsymbol{\sigma})] \\ &= 2q_{n,\ell} \int_{-\infty}^{\infty} \mathrm{sinc}(f q_{n,\ell}) \cos(2\pi f t_n^\ell \sqrt{n}) \hat{\rho}^n(f) df\end{aligned} \tag{3.26}$$

where $\hat{\rho}(f) = \mathbb{E}[e^{2\pi i f X}]$ is the Fourier transform of ρ.

When deriving an integral representation for the terms on the right hand side of (3.11), we will have to take the expectation of a product of integrals of the form (3.25). Neglecting, for the moment, problems associated with the exchange of integrations and expectations, this is not difficult. Indeed, rewriting the product of integrals as

$$\begin{aligned}\prod_{j=1}^{k}\int_{-\infty}^{\infty}\operatorname{sinc}(f_j q_{n,\ell(j)})\cos(2\pi f_j t_n^{\ell(j)}\sqrt{n})e^{2\pi i f_j \sum_{s=1}^{n}\sigma_s X_s}df_j \\ = \iiint_{-\infty}^{\infty}\prod_{s=1}^{n}e^{2\pi i v_s X_s}\prod_{j=1}^{k}\operatorname{sinc}(f_j q_{n,\ell(j)})\cos(2\pi f_j t_n^{\ell(j)}\sqrt{n})df_j\end{aligned} \tag{3.27}$$

where

$$v_s = \sum_{j=1}^{k}\sigma_s^{(j)} f_j, \tag{3.28}$$

and taking expectations of both sides, we easily arrive at the representation

$$\begin{aligned}\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j=1,\dots,k\Big) = \prod_{\ell=1}^{m}(2q_{n,\ell})^{k_\ell} \\ \times \iiint_{-\infty}^{\infty}\prod_{s=1}^{n}\hat{\rho}(v_s)\prod_{j=1}^{k}\operatorname{sinc}(f_j q_{n,\ell(j)})\cos(2\pi f_j t_n^{\ell(j)}\sqrt{n})df_j.\end{aligned} \tag{3.29}$$

But here a little bit more care is needed to justify the exchange of expectation and integrals. Indeed, this exchange can only be justified if $\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}$ are linearly independent (see Lemma 3.5 in [BCMN05]), but luckily, this will be all we need.

3.7. **The SK Model.** As in the proof for the Npp, we define $Z_n(a,b)$ to be the number of points in the energy spectrum that fall into the interval $[a,b]$, where the energy of a configuration $\boldsymbol{\sigma}$ is now given by (1.4). Given a family of non-overlapping intervals $[c_1,d_1]$, …, $[c_m,d_m]$ with $c_\ell \geq 0$, and $d_\ell = c_\ell + \gamma_\ell > c_\ell$ for $\ell = 1,\dots,m$, we now consider the intervals $[a_n^\ell, b_n^\ell]$ with $a_n^\ell = \alpha + c_\ell\tilde{\xi}_n$ and $b_n^\ell = \alpha + d_\ell\tilde{\xi}_n$, where $\tilde{\xi}_n$ is defined in (1.6). Theorem 2.2 now follows immediately from the following two theorems.

Theorem 3.6. *Let $\alpha_n = o(n)$ be a sequence of positive real numbers, let m be a positive integer, and for $\ell = 1,\dots,m$, let a_n^ℓ and b_n^ℓ be as above. For an arbitrary m-tuple of positive integers $(k_1,..,k_m)$ we then have*

$$\lim_{n\to\infty}\mathbb{E}[\prod_{\ell=1}^{m}(Z_n(a_n^\ell,b_n^\ell))_{k_\ell}] = \prod_{\ell=1}^{m}\gamma_\ell^{k_\ell}. \tag{3.30}$$

Theorem 3.7. *There exists a constant $\epsilon_0 > 0$ such the following statements hold for all $c \geq 0$ and $\gamma > 0$, all sequences of positive real numbers α_n with $\alpha_n \leq \epsilon_0 n$, and a_n, b_n of the form $a_n = \alpha_n + c\tilde{\xi}_n$, $b_n = \alpha_n + (c+\gamma)\tilde{\xi}_n$.*

(i) $\lim_{n\to\infty}\mathbb{E}[Z_n(a_n,b_n)] = \gamma$.

(ii) $\mathbb{E}[(Z_n(a_n,b_n))^3]$ is bounded uniformly in n.

(iii) If $\limsup \frac{\alpha_n}{n} > 0$, then $\limsup_{n\to\infty}\mathbb{E}[(Z_n(a,b))_2] > \gamma^2$.

By now, the proof strategy for these theorems is straightforward: As before, the first moment is given as an integral over the energy density. For the SK model, this density is given by(1.5), leading to $\mathbb{E}[Z_n(a_n^\ell,b_n^\ell)] = \gamma_\ell(1+o(1))$ as long as $\tilde{\xi}_n = o(1)$. To analyze the higher factorial moments, we again use a representation of the form (3.11). Neglecting for the moment the issue of bounding the sum over atypical configurations, we will again study the factorization properties of probabilities of the form

$$\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j = 1, \ldots, k\Big). \tag{3.31}$$

For the SK model, this is even easier than for the NPP with Gaussian noise, since $E(\boldsymbol{\sigma})$ is now itself a Gaussian random variable, rather than the absolute value of a Gaussian. The joint distribution of $E(\boldsymbol{\sigma}^{(1)})$, …, $E(\boldsymbol{\sigma}^{(k)})$ therefore has density

$$\tilde{g}^{(k)}(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}} \frac{1}{\sqrt{\det C}} \exp\Big(-\frac{1}{2}(\mathbf{x}, C^{-1}\mathbf{x})\Big), \tag{3.32}$$

where C is the covariance matrix

$$C_{ij} = \mathbb{E}[E(\boldsymbol{\sigma}^{(i)})E(\boldsymbol{\sigma}^{(j)})] = n(q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)}))^2. \tag{3.33}$$

Expanding C_{ij}^{-1} as $C_{ij}^{-1} = \frac{1}{n}(\delta_{ij} + O(q_{\max}^2))$, we then get the following analogue of the factorization formula (3.18):

$$\begin{aligned} &\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j = 1, \ldots, k\Big) \\ &\qquad = \prod_{i=1}^{k} \mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}]\Big) e^{O(\alpha_n^2 n^{-1} q_{\max}^2) + o(1)}. \end{aligned} \tag{3.34}$$

In the exponent of the above expression, note that the additional factor of n^{-1} relative to the analogous expression for the NPP is simply a consequence of the different normalizations of the energies. The more significant difference is the factor of $q_{\max}^2$ rather than $q_{\max}$, a consequence of the difference between the covariance matrices for the two problems. For typical configurations with overlap $q_{\max} = O(n^{-1/2})$, equation (3.34) suggests that there will be asymptotic factorization if and only if $\alpha_n = o(n)$. That this is indeed the case is established in Section 6.

4. The Npp with Gaussian densities

In this section we analyze the factorization properties of the probabilities

$$\mathbb{P}\left(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for } j = 1, \ldots, k\right) \tag{4.1}$$

for the case where $X_1, \ldots, X_n$ are standard normals. Throughout this section, we will assume that $\alpha_n = o(\sqrt{n})$.

In a preliminary step, we show that it is possible to approximate the joint density $g^{(k)}(\mathbf{y})$ of $E(\boldsymbol{\sigma}^{(1)})$, …, $E(\boldsymbol{\sigma}^{(k)})$ by its value at $y_i = \alpha_n$. To this end we combine the representations (3.19) and (3.21) with the fact that C^{-1} is bounded uniformly in n if $q_{\max} = o(1)$. For $y_i \in [a_n^{\ell(i)}, b_n^{\ell(i)}]$, we then have

$$g^{(k)}(\mathbf{y}) = g^{(k)}(\boldsymbol{\alpha}_n)(1 + O(\alpha_n \xi_n) + O(\xi_n^2)) = g^{(k)}(\boldsymbol{\alpha}_n)(1 + o(1)), \tag{4.2}$$

implying that

$$\mathbb{P}\left(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for } j = 1, \ldots, k\right) = \Big(\prod_{\ell=1}^{m} (\xi_n \gamma_\ell)^{k_\ell}\Big) g^{(k)}(\boldsymbol{\alpha}_n)(1 + o(1)). \tag{4.3}$$

Having established this approximation, we now proceed to prove the factorization formula (3.18).

4.1. **Proof of the factorization formula** (3.18)**.** Let us express the right hand side of (4.3) as a sum over vectors $\mathbf{x} \in \mathbb{R}^k$ with $|x_i| = \alpha_n$, see (3.19). Recalling the representation (3.21), we expand $(\mathbf{x}, C^{-1}\mathbf{x})$ as $(\mathbf{x}, C^{-1}\mathbf{x}) = \|\mathbf{x}\|_2^2(1 + O(q_{\max})) = k\alpha_n^2 + O(q_{\max}\alpha_n^2)$, where, as before, $\|\mathbf{x}\|_2$ denotes the ℓ_2-norm of $\mathbf{x}$. For $q_{\max} = o(1)$, we further have $\det C = 1 + o(1)$, implying that

$$g^{(k)}(\boldsymbol{\alpha}_n) = \Big(\frac{2}{\pi}\Big)^{k/2} e^{-\frac{k}{2}\alpha_n^2 + O(q_{\max}\alpha_n^2) + o(1)} = (g(\alpha_n))^k e^{O(q_{\max}\alpha_n^2)+o(1)}. \tag{4.4}$$

Recalling the approximation $\mathbb{P}\Big(E(\boldsymbol{\sigma}) \in [a_n^\ell, b_n^\ell]\Big) = g(\alpha_n)\gamma_\ell \xi_n(1 + o(1))$ established in Section 3.2, the bounds (4.3) and (4.4) imply the approximate factorization formula (3.18). Note that the only conditions needed in this derivation were the conditions $\alpha_n = o(\sqrt{n})$ and $q_{\max} = o(1)$. Taking into account Lemma 3.4 (iii), we therefore have established that (3.18) holds whenever $\alpha_n = o(\sqrt{n})$ and $\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(k)}$ obey the condition (3.13) for some λ_n of order $o(\sqrt{n})$.

As shown in Section 3.5, the approximate factorization formula (3.18) immediately leads to Poisson convergence if $\alpha_n = o(n^{1/4})$. To establish that this convergence fails for faster growing α_n, requires a little bit more work. This is done in the next subsection.

4.2. **Proof of Theorem 3.3.** We start again from (4.3), but this time we expand the inverse of C a little further. Explicitly, expressing the matrix elements of C as $C_{ij} = \delta_{ij} + \tilde{q}_{ij}$ where $\tilde{q}_{ij} = q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)})$ if $i \neq j$ and $\tilde{q}_{ij} = 0$ if $i = j$, we clearly have

$$\begin{aligned}(\mathbf{x}, C^{-1}\mathbf{x}) &= \sum_j x_j^2 - \sum_{i\neq j} x_i \tilde{q}_{ij} x_j + O(q_{\max}^2 \|\mathbf{x}\|_2^2) \\ &= k\alpha_n^2 - \sum_{i\neq j} x_i \tilde{q}_{ij} x_j + O(q_{\max}^2 \alpha_n^2)\end{aligned} \tag{4.5}$$

whenever $q_{\max} = o(1)$ and $|x_i| = \alpha_n$ for all i. With the help of (4.3), (1.3), (3.19) and (3.21), this implies that

$$\begin{aligned}&\mathbb{P}\left(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for } j = 1, \dots, k\right) \\ &\quad = \Big(\prod_{\ell=1}^m \gamma_\ell^{k_\ell}\Big) 2^{-nk} \sum_{\substack{x_1,\dots x_k:\\ x_i = \pm\alpha_n}} \exp\Big(\frac{1}{2}\sum_{i\neq j} x_i \tilde{q}_{ij} x_j + O(q_{\max}^2\alpha_n^2) + o(1)\Big).\end{aligned} \tag{4.6}$$

Let us now choose λ_n in such a way that $\alpha_n = o(\lambda_n)$, $\lambda_n = o(\sqrt{n})$ and $e^{-\lambda_n^2/2}$ decays faster than any power of n. Combining (4.6) with the representation (3.11) and Lemma 3.4 (2), we then have

$$\begin{aligned}\mathbb{E}\Big[\prod_{\ell=1}^m (Z_n(a_n^\ell, b_n^\ell))_{k_\ell}\Big] &= \Big(\prod_{i=1}^k \gamma_{\ell(i)}\Big) 2^{-nk} \sum_{\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}}'' 2^{-k} \\ &\times \sum_{\substack{x_1,\dots x_k:\\ x_i = \pm\alpha_n}} \exp\Big(\frac{1}{2}\sum_{i\neq j} x_i\tilde{q}_{ij}x_j + O(q_{\max}^2\alpha_n^2) + o(1)\Big) + o(1)\end{aligned} \tag{4.7}$$

where the sum $\sum''$ runs over families of linearly independent configurations $\boldsymbol{\sigma}^{(1)}$, $\dots$, $\boldsymbol{\sigma}^{(k)}$ that satisfy (3.13). Next we claim that we can extend this sum to a sum over all families of configurations $\boldsymbol{\sigma}^{(1)}$, $\dots$, $\boldsymbol{\sigma}^{(k)}$ at the cost of an additional additive error $o(1)$. Indeed, by Lemma 3.4 (1) and (3), the number of configurations

$\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$ that are linearly dependent or violate the bound (3.13) is bounded by a constant times $2^{nk}e^{-\lambda_n^2/2}$. Since the terms in the above sum are all bounded by $e^{O(\alpha_n^2)} = e^{o(\lambda_n^2)}$, the extension from the sum $\sum''$ to a sum over all families of configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$ indeed only introduces an additive error $o(1)$. Observing finally that the exponent $\sum_{i\neq j} x_i\tilde{q}_{ij}x_j$ is invariant under the transformation $x_i \to |x_i| = \alpha_n$ and $\boldsymbol{\sigma}^{(i)} \to \operatorname{sign}(x_i)\boldsymbol{\sigma}^{(i)}$, we obtain the approximation (3.24), which we restate here as

$$\mathbb{E}\Big[\prod_{\ell=1}^{m}(Z_n(a_n^\ell, b_n^\ell))_{k_\ell}\Big] = \Big(\prod_{i=1}^{k}\gamma_{\ell(i)}\Big)2^{-nk}\sum_{\boldsymbol{\sigma}^{(1)},\ldots,\boldsymbol{\sigma}^{(k)}}\exp\Big(\frac{\alpha_n^2}{2}\sum_{i\neq j}\tilde{q}_{ij} + O(q_{\max}^2\alpha_n^2) + o(1)\Big) + o(1). \tag{4.8}$$

With the approximation (4.8) in hand, the second statement of Theorem 3.3 now follows immediately from the following lemma.

Lemma 4.1. *Let $C < \infty$, let $\beta_n = o(n)$, let $\mathbb{E}_k$ denote expectation with respect to the uniform measure on all families of configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$, and let R denote a function on families of configurations such that $|R(\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)})| \leq Cq_{\max}^2$. Then*

$$\mathbb{E}_k\Big[\exp\Big(\beta_n\Big(\sum_{i\neq j}\tilde{q}_{ij} + R\Big)\Big)\Big] = \exp\Big(\frac{k(k-1)}{n}\beta_n^2 + O(\beta_n^3 n^{-2}) + o(1)\Big). \tag{4.9}$$

Proof. The complete proof of the lemma will be given in Section 7.2. Here we only show that the leading term behaves as claimed, namely

$$\mathbb{E}_k\Big[\exp\Big(\beta_n\sum_{i\neq j}\tilde{q}_{ij}\Big)\Big] = \exp\Big(\frac{k(k-1)}{n}\beta_n^2 + O(\beta_n^3 n^{-2})\Big). \tag{4.10}$$

To this end, we observe that

$$\sum_{i\neq j}\tilde{q}_{ij} = \sum_{i,j} q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)}) - k = \frac{1}{n}\sum_{s=1}^{n}\Big(\sum_{i=1}^{k}\sigma_s^{(i)}\Big)^2 - k. \tag{4.11}$$

As a consequence, we have

$$\begin{aligned}\mathbb{E}_k\Big[e^{\beta_n\sum_{i\neq j}\tilde{q}_{ij}}\Big] &= e^{-k\beta_n}\mathbb{E}_k\Big[\prod_{s=1}^{n}\exp\Big(\frac{\beta_n}{n}\Big(\sum_{i=1}^{k}\sigma_s^{(i)}\Big)^2\Big)\Big] \\ &= e^{-k\beta_n}\tilde{\mathbb{E}}_k\Big[\exp\Big(\frac{\beta_n}{n}\Big(\sum_{i=1}^{k}\delta_i\Big)^2\Big)\Big]^n,\end{aligned} \tag{4.12}$$

where $\tilde{\mathbb{E}}_k$ denotes expectation with respect to the uniform measure on all configurations $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k) \in \{-1, +1\}^k$. Next we expand the expectation on the right hand side into a power series in β_n/n (recall that we assumed $\beta_n = o(n)$). Using that the expectation of $(\sum_i \delta_i)^2$ is equal to k, while the expectation of $(\sum_i \delta_i)^4$ is equal to $3k^2 - 2k$, this gives

$$\tilde{\mathbb{E}}_k\Big[\exp\Big(\frac{\beta_n}{n}\Big(\sum_{i=1}^{k}\delta_i\Big)^2\Big)\Big] = \exp\Big(\frac{\beta_n}{n}k + \frac{\beta_n^2}{n^2}k(k-1) + O(\beta_n^3 n^{-3})\Big) \tag{4.13}$$

and hence

$$\mathbb{E}_k\Big[e^{\beta_n \sum_{i\neq j}\tilde q_{ij}}\Big] = \exp\Big(\frac{k(k-1)}{n}\beta_n^2 + O(\beta_n^3 n^{-2})\Big). \tag{4.14}$$

□

Remark 4.2. For the special case where α_n grows like $n^{1/4}$, Theorem 3.3 implies that the density of the process $N_n(t)$ converges to one, while the process itself does not converge to Poisson. Another interesting consequence of our proof is the following: Consider the rescaled overlap $Q_{n,t}$ of two typical configurations contributing to $N_n(t)$ (see Section 2.1 for the precise definition of $Q_{n,t}$). Then $Q_{n,t}$ converges in distribution to the superposition of two Gaussians with mean $\pm\kappa^2$, where $\kappa = \lim_{n\to\infty}\alpha_n n^{-1/4}$,

$$\lim_{n\to\infty}\mathbb{P}(Q_{n,t}\geq y) = \frac{1}{2}\frac{1}{\sqrt{2\pi}}\int_y^\infty \frac{e^{-\frac{1}{2}(x-\kappa^2)^2}+e^{-\frac{1}{2}(x+\kappa^2)^2}}{2}dx. \tag{4.15}$$

This follows again from (4.6); in fact, now we only need this formula for $k=2$, where the sum over $\mathbf{x}$ just gives a factor $\cosh(\alpha_n^2 q(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}))$. This factor is responsible for the shift of $\pm\kappa^2$ in the limiting distribution of $Q_{n,t} = n^{1/2}q(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$.

5. The Npp with general distribution

In this section we prove the first moment estimate (3.5) and the factorization formula (3.18) for arbitrary distributions ρ obeying the assumptions of Theorem 2.1, see Propositions 5.2 and 5.6 below. As discussed in Section 3, this immediately gives Theorem 3.1. We also show in Section 5.4 how the proof of non-convergence can be generalized from the Gaussian case to the general distributions considered in Theorem 2.1.

5.1. **Properties of the Fourier transform.** Throughout this section, we will use several properties of the Fourier transform in $\hat\rho$ which we summarize in this subsection.

(i) For any $n\geq n_o$, where n_0 is the solution of $\frac{1}{1+\epsilon}+\frac{1}{n_o}=1$ with ϵ as in (2.1), we have
$$\int_{-\infty}^{\infty}|\hat\rho(f)|^n \leq \int_{-\infty}^{\infty}|\hat\rho(f)|^{n_0} = C_0 < \infty. \tag{5.1}$$
(ii) There exists $\mu_0>0$ such that $\hat\rho(f)$ is analytic in the region $|\Im\mathrm{m} f|<\mu_0$.
(iii) For any $\mu_1>0$ there exists $c_1>0$ and $\mu_2>0$ and such that
$$|\hat\rho(f)|\leq e^{-c_1} \tag{5.2}$$
whenever $|\Im\mathrm{m} f|\leq\mu_2$ and $|\Re\mathrm{e} f|\geq\mu_1$.

These properties easily follow from our assumptions on the density ρ. The bound (5.1) is a direct consequence of (2.1). Analyticity of $\hat\rho$ in a neighborhood of the origin implies existence of exponential moments and this in turn implies analyticity in a strip about the real line. To see that the bound (5.2) is true, we first observe that it obviously holds when f is real. From the fact that $d\hat\rho(z)/dz$ is bounded uniformly in z as long as the imaginary part of z is small enough, we can choose μ_2 in such a way that the bound (5.2) extends to $|\Im\mathrm{m} f|\leq\mu_2$ (with c_1 replaced by a slightly smaller constant, which we again call c_1).

5.2. **First Moment.** In this section we establish the representation (3.9) for the density $g_n(\cdot)$, see Proposition 5.2 below. As explained in Section 3.2, this representation immediately give the first moment estimate (3.5) for $\alpha_n = o(\sqrt{n})$.

We will start from the integral representation

$$g_n(\alpha) = 2\sqrt{n}\int_{-\infty}^{\infty} \cos(2\pi f\alpha\sqrt{n})\hat{\rho}^n(f)df, \tag{5.3}$$

This formula can easily be derived by first expressing the density of $H(\boldsymbol{\sigma})$ in terms of the n-fold convolution of ρ with itself, using Fourier inversion to express this density as an integral over $\hat{\rho}^n$, and then summing over the two possible choices for the sign of $H(\boldsymbol{\sigma})$ given $E(\boldsymbol{\sigma})$ (alternatively, the formula can be derived from (3.26) by sending γ_ℓ to 0).

In order to determine the asymptotic behavior of $g_n(\alpha)$, one might want to expand $\hat{\rho}^n(f)$ about its maximum, i.e. about $f = 0$. While this strategy works well for bounded α_n, it needs to be modified in the case of growing α_n. Here we will use the method of steepest descent, a method which was first introduced into the analysis of density functions by a paper of Daniels [Dan54], though of course the ideas go back to Laplace.

Before explaining this further, let us first note that, asymptotically, the integral in (5.3) can be restricted to a bounded interval about zero. Indeed, let $\mu_1 > 0$ be an arbitrary constant, let μ_2 be as in (5.2), and let n_0 be as in (5.1). For $n \geq n_0$ and $|f| \geq \mu_1$, we then have $|\hat{\rho}^n(f)| \leq |\hat{\rho}^{n_0}(f)|e^{-c_1(n-n_0)}$. As a consequence, the contribution to the integral (5.3) from the region $|f| \geq \mu_1$ is bounded by a constant times $\sqrt{n}e^{-c_1 n}$, giving the estimate

$$g_n(\alpha) = 2\sqrt{n}\int_{-\mu_1}^{\mu_1} \cos(2\pi f\alpha\sqrt{n})\hat{\rho}^n(f)df + O(\sqrt{n}e^{-c_1 n}). \tag{5.4}$$

Next we observe that both the range of integration and the function $f \mapsto \hat{\rho}(f)$ are invariant under the change $f \to -f$, implying that we can rewrite the integral as

$$2\sqrt{n}\int_{-\mu_1}^{\mu_1} e^{2\pi i f\alpha_n\sqrt{n}}\hat{\rho}^n(f)df = 2\sqrt{n}\int_{-\mu_1}^{\mu_1} e^{-n(F(f)-2\pi i\frac{\alpha_n}{\sqrt{n}}f)}df \tag{5.5}$$

where $F(f)$ is defined by

$$\hat{\rho}(f) = e^{-F(f)}. \tag{5.6}$$

As defined earlier, let μ_0 be such that $\hat{\rho}(f)$ is analytic in the region $|\Im\mathrm{m} f| < \mu_0$. Further let $0 < \eta \leq \mu_0$, and let $\mathcal{C}$ be the path in the complex plane obtained by concatenating the three line segments that join the point $-\mu_1$ to the point $-\mu_1 + i\eta$, the point $-\mu_1 + i\eta$ to the point $\mu_1 + i\eta$, and the point $\mu_1 + i\eta$ to the point μ_1, respectively. Let $\mathcal{G}$ be the region bounded by $\mathcal{C}$ and the line segment from $-\mu_1$ to μ_1. Then the function $f \mapsto \hat{\rho}(f)$ and hence the integrand in (5.5) is analytic in $\mathcal{G}$, implying that the integral on the right hand side of (5.5) is equal to the integral over $\mathcal{C}$. Our next lemma states that the contribution of the first and third line segment to this integral is negligible, effectively allowing us to "shift the path of integration" into a parallel line segment in the complex plane.

Lemma 5.1. *Given $\mu_1 > 0$, there are constants $c_1 > 0$ and $\mu_2 > 0$ such that*

$$g_n(\alpha_n) = 2\sqrt{n}\int_{-\mu_1+i\eta}^{\mu_1+i\eta} e^{-n(F(f)-2\pi i\frac{\alpha_n}{\sqrt{n}}f)}df + O(\sqrt{n}e^{-c_1 n}) \tag{5.7}$$

provided $\alpha_n \geq 0$ and $0 \leq \eta \leq \mu_2$.

Proof. Taking into account (5.4), we need to analyze the integral on the right of (5.5) which in turn is equal to the integral over the path $\mathcal{C}$ defined above, provided we choose $\mu_2 \leq \mu_0$.

Consider the integral

$$\int_{-\mu_1}^{-\mu_1+i\eta} e^{-n(F(f)-2\pi i\frac{\alpha_n}{\sqrt{n}}f)}df. \tag{5.8}$$

For f in the line segment joining $-\mu_1$ to $-\mu_1+i\eta$, the imaginary part of f is non-negative, implying that $|e^{2\pi i n\frac{\alpha_n}{\sqrt{n}}f}| \leq 1$. Due to property (5.2) of the Fourier transform $\hat{\rho}$, we furthermore have that $|\hat{\rho}(f)| = |e^{-F(f)}| \leq e^{-c_1}$ on this line segment. As a consequence, the above integral is bounded by $\eta e^{-c_1 n} = O(e^{-c_1 n})$. In a similar way, the integral over the third line segment is bounded by $O(e^{-c_1 n})$. Taking into account the multiplicative factor of $\sqrt{n}$ in (5.4) and the fact that the error term in (5.4) is bounded by $O(\sqrt{n}e^{-c_1 n})$, this proves (5.7). □

Note that the above lemma holds for arbitrary $\eta \geq 0$. As usual when applying the method of steepest descent, the value of η will be chosen in such a way that the asymptotic analysis of the integral on the right becomes as simple as possible. Here this amounts to requiring that at $f = i\eta$, the first derivative of the integrand is zero. In other words, we will choose η as the solution of the equation

$$F'(i\eta) = 2\pi i\frac{\alpha_n}{\sqrt{n}}, \tag{5.9}$$

where $F'(z) = \frac{dF(z)}{dz}$ denotes the complex derivative of F. This leads to the following proposition.

Proposition 5.2. *There is an even function $G(x)$ which is real analytic in a neighborhood of zero such that*

$$g_n(\alpha_n) = \sqrt{\frac{2}{\pi}}e^{-nG(\alpha_n n^{-1/2})}(1+o(1)) \tag{5.10}$$

whenever $\alpha_n = o(\sqrt{n})$. In a neighborhood of zero, $G(x)$ can be expanded as

$$G(x) = \frac{x^2}{2} + O(x^4). \tag{5.11}$$

For $\alpha_n = o(\sqrt{n})$ and a_n^ℓ, b_n^ℓ and γ_ℓ as in Theorem 3.1 we therefore have

$$\mathbb{P}\Big(E(\boldsymbol{\sigma}) \in [a_n^\ell, b_n^\ell]\Big) = 2^{-(n-1)}\gamma_\ell(1+o(1)). \tag{5.12}$$

Proof. We first argue that the equation (5.9) has a unique solution which can be expressed as an analytic function of $x = \alpha_n/\sqrt{n}$. Consider thus the equation

$$F'(i\eta) = 2\pi i x. \tag{5.13}$$

Since $\hat{\rho}$ is an even function which is analytic in a neighborhood of zero with $\hat{\rho}(0) = 1$, the function F is even and analytic in a neighborhood of zero as well. Taking into account that the second moment of ρ is one, we now expand $F(f)$ as

$$F(f) = \frac{1}{2}(2\pi f)^2 + O(f^4) \tag{5.14}$$

and $F'(f)$ as

$$F'(f) = (2\pi)^2 f + O(f^3). \tag{5.15}$$

By the implicit function theorem, the equation (5.13) has a unique solution $\eta(x)$ in a neighborhood of zero, and $\eta(x)$ is an analytic function of x. Taking into account that F' is odd, we furthermore have that $\eta(x)$ is an odd function which is real analytic in a neighborhood of zero, and expanding $\eta(x)$ as $\eta(x) = \frac{1}{2\pi}x + O(x^3)$ we see that $\eta(x) \geq 0$ if $x \geq 0$ is sufficiently small.

Having established the existence and uniqueness of $\eta(x)$ for small enough x, we are now ready to analyze the integral in (5.7) for $\eta = \eta(x)$ and $x = \alpha_n/\sqrt{n} = o(1)$. To this end, we rewrite the integral as

$$\int_{-\mu_1}^{\mu_1} e^{-n\left(F(i\eta(x)+f)+2\pi x(\eta(x)-if)\right)} df = e^{-nG(x)} \int_{-\mu_1}^{\mu_1} e^{-n\tilde{F}(f)} df \tag{5.16}$$

where

$$G(x) = F(i\eta(x)) + 2\pi x\eta(x), \tag{5.17}$$

and

$$\tilde{F}(f) = F(i\eta(x) + f) - F(i\eta(x)) - 2\pi ixf. \tag{5.18}$$

Next we would like to show that the integral on the right hand side of (5.16) can be restricted to $|f| \leq \log n/\sqrt{n}$. To this end, we expand $\tilde{F}(f)$ about $f = 0$. Taking into account the fact that the first derivative at $f = 0$ vanishes by the definition of $\eta(x)$, we get

$$\tilde{F}(f) = \frac{1}{2}F^{(2)}(i\eta(x))f^2 + \frac{1}{6}F^{(3)}(i\eta(x))f^3 + O(f^4). \tag{5.19}$$

Using again that F is an even function of its argument, we conclude that $F^{(2)}(i\eta(x))$ is real, while $F^{(3)}(i\eta(x))$ is purely imaginary, so that

$$\Re e\, \tilde{F}(f) = \frac{1}{2}F^{(2)}(i\eta(x))f^2 + O(f^4). \tag{5.20}$$

Since $\eta(x) = O(x) = o(1)$ and $F^{(2)}(i\eta(x)) = 4\pi^2 + O(\eta(x)^2)$, we have that $\Re e\, \tilde{F}(f) \geq f^2$ provided μ_1 is sufficiently small and n is sufficiently large. As a consequence, we get that

$$\int_{-\mu_1}^{\mu_1} e^{-n\tilde{F}(f)} df = \int_{-n^{-1/2}\log n}^{n^{-1/2}\log n} e^{-n\tilde{F}(f)} df + O\Big(\frac{1}{\sqrt{n}} e^{-\log^2 n}\Big). \tag{5.21}$$

For $|f| \leq \log n/\sqrt{n}$, we now expand

$$\begin{aligned} e^{-n\tilde{F}(f)} &= \exp\Big(-\frac{n}{2}F^{(2)}(i\eta(x))f^2 + O(nf^3)\Big) \\ &= \exp\Big(-\frac{n}{2}F^{(2)}(i\eta(x))f^2\Big)\Big(1 + O(nf^3)\Big), \end{aligned} \tag{5.22}$$

leading to the approximation

$$\int_{-n^{-1/2}\log n}^{n^{-1/2}\log n} e^{-n\tilde{F}(f)} df = \sqrt{\frac{2\pi}{nF^{(2)}(i\eta(x))}}\big(1 + O(n^{-1/2})\big) = \frac{1}{\sqrt{2\pi n}}\big(1 + o(1)\big). \tag{5.23}$$

Combined with (5.7), (5.16) and (5.21) this proves (5.10).

Next, we show that $G(x)$ is an even function which is real analytic in a neighborhood of zero and obeys the bound (5.11). But this is almost obvious by now. Indeed, combining the fact that F is an even function which is real analytic in a neighborhood of zero with the fact that $\eta(x)$ is an odd function which is real analytic in a neighborhood of zero, we see that $G(x)$ is an even function which is real analytic in a neighborhood of zero. With the help of (5.14), the bound (5.11) finally follows by inserting the expansion $\eta(x) = \frac{1}{2\pi}x + O(x^3)$ into the definition (5.17) of G.

As we already argued in Section 3.2, the bound (5.12) finally follows immediately from the remaining statements of the proposition. □

5.3. **Higher Moments.** In order to establish convergence of the higher factorial moments, we start from the integral representation (3.29). Recalling the definition (3.12), let

$$n_{\min} = n_{\min}(\boldsymbol{\sigma}^{(1)}, ..., \boldsymbol{\sigma}^{(u)}) = \min\{n_{\boldsymbol{\delta}} \colon \boldsymbol{\delta} \in \{-1,+1\}^u\}. \tag{5.24}$$

In a first step, we show that under the condition that $n_{\min} \geq n_0$ where n_0 is the solution of $\frac{1}{1+\epsilon} + \frac{1}{n_o} = 1$ with ϵ as in (2.1), the integral in (3.29) is well approximated by an integral over a bounded domain, with the product of the sinc factors replaced by one and the various midpoints t_n^ℓ replaced by α_n.

Lemma 5.3. *Given $\mu_1 > 0$ there exists a constant $c_1 > 0$ such that for $\alpha_n = o(\sqrt{n})$ and $n_{\min} = n_{\min}(\boldsymbol{\sigma}^{(1)}, .., \boldsymbol{\sigma}^{(k)}) \geq n_0$ we have*

$$\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for } j = 1, \dots, k\Big) = \prod_{\ell=1}^{m} (2q_{n,\ell})^{k_\ell}$$
$$\times \Bigg(\iiint_{-\mu_1}^{\mu_1} \prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}} \prod_{j=1}^{k} \cos(2\pi f_j \alpha_n \sqrt{n}) df_j + O(e^{-c_1 n_{\min}}) + O(\sqrt{n}\xi_n) \Bigg). \tag{5.25}$$

Here the product over $\boldsymbol{\delta}$ runs over all $\boldsymbol{\delta} \in \{-1,+1\}^k$ and $\mathbf{f}\cdot\boldsymbol{\delta}$ stands for the scalar product $\sum_j f_j \delta_j$.

Proof. Consider the integral on the right hand side of (3.29). In a first step, we will consider the contribution of the region where $|f_j| > \mu_1$ for at least one j and show that it is negligible. To this end, it is clearly enough to bound the contribution of the region where $\sum_{j=1}^k |f_j| > \mu_1$. But if $\sum_{j=1}^k |f_j| > \mu_1$, then there is a $\boldsymbol{\delta} \in \{-1,+1\}$ such that $|\sum_{j=1}^k \delta_j f_j| > \mu_1$. Since $n_{\boldsymbol{\delta}} \geq 1$ for all $\boldsymbol{\delta} \in \{-1,+1\}^k$, we conclude that there exists an $s \in \{1, \dots, n\}$ such that $|v_s| > \mu_1$.

Thus consider the event that one of the $|v_s|$'s, say $|v_{t_1}|$, is larger than μ_1. Let $\boldsymbol{\delta}_1 = \{\sigma_{t_1}^{(1)}, .., \sigma_{t_1}^{(k)}\}$, and let $\boldsymbol{\delta}_2, ..., \boldsymbol{\delta}_k$ be vectors such that the rank of the matrix Δ formed by $\boldsymbol{\delta}_1, .., \boldsymbol{\delta}_k$ is k. Let $\{v_{t_2}, ..., v_{t_k}\}$ be defined by

$$v_{t_i} = \sum_{j=1}^{k} \delta_i^j f_j. \tag{5.26}$$

Since Δ has rank k, we can change the variables of integration from f_j to v_{t_j}. Let the Jacobian of this transformation be J_k, i.e., let $J_k = |\det\Delta|^{-1}$ where $\det\Delta$ is the determinant of Δ. Since Δ has entries ±1 and is non-singular, we conclude that $|\det\Delta| \geq 1$, implying that $|J_k| \leq 1$. We therefore may bound the integral over the region where where $|v_{t_1}| > \mu_1$ by

$$\begin{aligned}
&\left| \iiint_{-\infty}^{\infty} \int_{|v_{t_1}|>\mu_1} J_k \prod_{s=1}^{n} \hat{\rho}(v_s) \prod_{j=1}^{k} \operatorname{sinc}(f_j q_{n,\ell(j)}) \cos(2\pi f_j t_n^{\ell(j)} \sqrt{n}) dv_{t_j} \right| \\
&\leq \iiint_{-\infty}^{\infty} \prod_{j=2}^{k} |\hat{\rho}(v_{t_j})|^{n_{\delta_j}} dv_{t_j} \times \int_{|v_{t_1}|>\mu_1} |\hat{\rho}(v_{t_1})|^{n_{\delta_1}} dv_{t_1} \qquad (5.27) \\
&\leq (C_0)^{k-1} \int_{|v_{t_1}|>\mu_1} |\hat{\rho}(v_{t_1})|^{n_{\delta_1}} dv_{t_1} \leq C_0^k e^{-c_1(n_{\min}-n_0)},
\end{aligned}$$

where C_0 and c_1 are as in (5.1) and (5.2). Since the number of choices for $\boldsymbol{\delta}_{t_1}$ is bounded by 2^k, the bound (5.27) implies that the contribution of the region where at least one of the $|f_j|$'s is larger than μ_1 is bounded by $O(e^{-c_1 n_{\min}})$, with the constant implicit in the O-symbol depending on k.

Noting that $q_{n,\ell(j)} = O(\sqrt{n}\xi_n)$, we finally observe that in the region where $|f_j| \leq \mu_1$ for all j, we may expand $\operatorname{sinc}(q_{n,\ell(j)} f_j)$ as $1+O(\sqrt{n}\xi_n)$ and $\cos(2\pi f_j t_n \sqrt{n})$ as $\cos(2\pi f_j \alpha_n \sqrt{n}) + O(\sqrt{n}\xi_n)$. Rewriting the product $\prod_s \hat{\rho}(v_s)$ as

$$\prod_{s=1}^{n} \hat{\rho}(v_s) = \prod_{\boldsymbol{\delta}\in\{-1,+1\}^k} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}}, \qquad (5.28)$$

this gives (5.25). □

Next we rewrite the integral in (5.25) as the average

$$\begin{aligned}
&\iiint_{-\mu_1}^{\mu_1} \prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}} \prod_{j=1}^{k} \cos(2\pi f_j \alpha_n \sqrt{n}) df_j \\
&\qquad = 2^{-k} \sum_{\mathbf{x}\in\{-\frac{\alpha_n}{\sqrt{n}}, \frac{\alpha_n}{\sqrt{n}}\}^k} \iiint_{-\mu_1}^{\mu_1} e^{2\pi i n \mathbf{f}\cdot\mathbf{x}} \prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}} \prod_{j=1}^{k} df_j
\end{aligned} \qquad (5.29)$$

where the sum goes over all sequences $\mathbf{x} = (x_1, \ldots, x_k) \in \{-\frac{\alpha_n}{\sqrt{n}}, \frac{\alpha_n}{\sqrt{n}}\}^k$ and $\mathbf{f}\cdot\mathbf{x}$ stands again for the scalar product, $\mathbf{f}\cdot\mathbf{x} = \sum_j f_j x_j$. In order to analyze the integrals on the right hand side, we will again use the method of steepest decent. To this end, we first prove the following analog of Lemma 5.1. As before, μ_0 is a constant such that $\hat{\rho}(f)$ is analytic in the strip $|\Im\mathrm{m} f| < \mu_0$.

Lemma 5.4. *Given $\mu_1 > 0$ there are constants $c_1 > 0$ and $\mu_2 \in (0, \mu_0)$ such that the following bound holds whenever $n_{\min} \geq 2^{-(k+1)}n$ and $\eta_1, \ldots, \eta_k$ is sequence of real numbers with $\sum_j |\eta_j| \leq \mu_2$ and $\eta_j x_j \geq 0$ for all j:*

$$\begin{aligned}
&\iiint_{-\mu_1}^{\mu_1} \prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}} \prod_{j=1}^{k} e^{2\pi i n x_j f_j} df_j \\
&= \iiint_{-\mu_1}^{\mu_1} e^{2\pi n(i\mathbf{x}\cdot\mathbf{f} - \mathbf{x}\cdot\boldsymbol{\eta})} \prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta} + i\boldsymbol{\eta}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}} \prod_{j=1}^{k} df_j + O(e^{-\frac{1}{2}c_1 n_{\min}}).
\end{aligned} \qquad (5.30)$$

Proof. For $j = 1, \ldots, k$, let $\mathcal{C}_j$ be the path consisting of the three line segments which join the point $-\mu_1$ to the point $-\mu_1 + i\eta_j$, the point $-\mu_1 + i\eta_j$ to the point $\mu_1 + i\eta_j$, and the point $\mu_1 + i\eta_j$ to the point μ_1, respectively. For $j = 1, \ldots, k$, we replace, one by one, the integrals over the variables f_j by integrals over the paths $\mathcal{C}_j$ and then bound the contribution over the part coming from the two line segments joining $-\mu_1$ to $-\mu_1 + i\eta_j$ and $\mu_1 + i\eta_j$ to μ_1, respectively. In each of step, we then have to bound integrals of the form

$$\iiint \prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}} \prod_{j'=1}^{k} e^{2\pi i n x_{j'} f_{j'}}\, df_{j'} \tag{5.31}$$

where $f_1, \dots, f_{j-1}$ run over the line segments from $-\mu_1 + i\eta_j$ to $\mu_1 + i\eta_j$, f_j runs either over the the line segment from $-\mu_1$ to $-\mu_1 + i\eta_j$ or the line segment from $\mu_1 + i\eta_j$ to μ_1, and $f_{j+1}, \dots, f_k$ run over the interval $[-\mu_1, \mu_1]$. To bound these integrals, we note that in this domain of integration the real part of f_j has absolute value μ_1, implying in particular that

$$\mu_1 \le \sum_i |\Re\mathrm{e} f_i| \le k\mu_1. \tag{5.32}$$

As a consequence, $|\Re\mathrm{e}\mathbf{f}\cdot\boldsymbol{\delta}| \le k\mu_1$ for all $\boldsymbol{\delta}$, and further there exists at least one $\boldsymbol{\delta}$ for which $|\Re\mathrm{e}\mathbf{f}\cdot\boldsymbol{\delta}| \ge \mu_1$. (In fact it is easy to see that there are at least 2^{k-1} such $\boldsymbol{\delta}$'s.) On the other hand, the assumption $\sum_j |\eta_j| \le \mu_2$ implies that $|\Im\mathrm{m}\mathbf{f}\cdot\boldsymbol{\delta}| \le \mu_2$ for all $\boldsymbol{\delta}$.

Consider first a vector $\boldsymbol{\delta}$ for which $\mu_1 \le |\Re\mathrm{e}\mathbf{f}\cdot\boldsymbol{\delta}| \le k\mu_1$, and assume that $\mu_2 \le \mu_0$ is chosen in such a way that (5.2) holds. Under this assumption, we have $|\hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})| \le e^{-c_1}$, implying that the term $\hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}}$ contributes a factor that is at most $e^{-n_{\min} c_1}$.

To bound $|\hat{\rho}(z)|$ when we only know that $|\Re\mathrm{e} z| \le k\mu_1$, we use continuity in conjunction with the fact that $|\hat{\rho}(z)| \le 1$ for all real z. Decreasing μ_2, if necessary, we conclude that $|\hat{\rho}(z)| \le e^{c_1 2^{-k-2}}$ whenever $|\Re\mathrm{e} z| \le k\mu_1$ and $|\Im\mathrm{m} z| \le \mu_2$. As a consequence, the first product in (5.31) can be bounded by

$$\prod_{\boldsymbol{\delta}} \Big|\hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}}\Big| \le e^{-n_{\min} c_1} e^{c_1 2^{-k-2} n} \le e^{-\frac{1}{2} n_{\min} c_1}. \tag{5.33}$$

Since the second product is bounded by one as before, and the integral only contributes a constant, this proves the lemma. □

Next we have to determine the values of the shifts $\eta_1, \dots, \eta_k$. According to the method of steepest decent, we again choose a saddle point of the integrand. Since the integrand is now a function of k variables, this now gives a system of k equations for $\boldsymbol{\eta} = (\eta_1, \dots, \eta_k)$:

$$\sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} \delta_j F'(i\boldsymbol{\delta}\cdot\boldsymbol{\eta}) = 2\pi i x_j. \tag{5.34}$$

Assume for a moment that this equation has a unique solution $\boldsymbol{\eta} = \boldsymbol{\eta}(\mathbf{x})$. We then define a function

$$G_{n,k}(\mathbf{x}) = \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} F(i\boldsymbol{\delta}\cdot\boldsymbol{\eta}(\mathbf{x})) + 2\pi\boldsymbol{\eta}(\mathbf{x})\cdot\mathbf{x}. \tag{5.35}$$

Lemma 5.5. *Let $0 \le \alpha_n = o(\sqrt{n})$, let $q_{\max} = o(1)$, and let $\mathbf{x} \in \mathbb{R}^k$ be such that $|x_i| = \alpha_n/\sqrt{n}$. Then the equation* (5.34) *has a unique solution $\boldsymbol{\eta} = \boldsymbol{\eta}(\mathbf{x})$,*

$$\eta_i(\mathbf{x}) = \Big(\frac{1}{2\pi} \sum_{j=1}^{k} C^{-1}_{ij} x_j\Big)\Big(1 + O\big(\frac{\alpha_n^2}{n}\big)\Big) \tag{5.36}$$

and $\eta_j(\boldsymbol{x})x_j \geq 0$ *for all* $j \in \{1, \ldots, k\}$. *In addition, for sufficiently small* μ_1,

$$\begin{aligned}\iiint_{-\mu_1}^{\mu_1} & e^{2\pi n(i\mathbf{x}\cdot\mathbf{f}-\mathbf{x}\cdot\boldsymbol{\eta})} \prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta} + i\boldsymbol{\eta}\cdot\boldsymbol{\delta})^{n_\delta} \prod_{j=1}^{k} df_j \\ &= e^{-nG_{n,k}(\mathbf{x})} \left(\frac{1}{2\pi n}\right)^{k/2} \Big(1 + O(n^{-1/2}) + O(\alpha_n^2/n) + O(q_{\max})\Big).\end{aligned} \tag{5.37}$$

Proof. For $\mathbf{x} = 0$, the equation (5.34) is obviously solved by $\boldsymbol{\eta} = 0$. To obtain existence and uniqueness of a solution in the neighborhood of zero, we consider the derivative matrix of the function on the left hand side,

$$A_{ij}(\boldsymbol{\eta}) = i \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} \delta_i \delta_j F''(i\boldsymbol{\delta}\cdot\boldsymbol{\eta}). \tag{5.38}$$

Using the fact that $F''(f) = (2\pi)^2 + O(f^2)$ we may expand $A_{ij}(\boldsymbol{\eta})$ as

$$A_{ij}(\boldsymbol{\eta}) = A_{ij}(0) + O(\|\boldsymbol{\eta}\|_2^2), \tag{5.39}$$

and for $\boldsymbol{\eta} = 0$ we have

$$A_{ij}(0) = (2\pi)^2 i \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} \delta_i \delta_j = (2\pi)^2 i C_{ij}, \tag{5.40}$$

where C_{ij} is the overlap matrix $C_{ij} = q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)})$. If the maximal off-diagonal overlap $q_{\max}$ is $o(1)$, the matrix C is invertible, implying in particular that $A_{ij}(\boldsymbol{\eta})$ is non-singular in a neighborhood of zero. By the implicit function theorem, we conclude that for $\mathbf{x}$ sufficiently small, the equation (5.34) had a unique solution $\boldsymbol{\eta}(\mathbf{x})$, and by (5.39), we may expand $\boldsymbol{\eta}(\mathbf{x})$ as

$$\begin{aligned}\eta_i(\mathbf{x}) &= 2\pi i \sum_{j=1}^{k} A_{ij}^{-1}(0) x_j + O(\|\mathbf{x}\|_2^3) \\ &= \frac{1}{2\pi} \sum_{j=1}^{k} C_{ij}^{-1} x_j + O(\|\mathbf{x}\|_2^3).\end{aligned} \tag{5.41}$$

For $q_{\max} = o(1)$, the matrix C^{-1} can be approximated as $C_{ij}^{-1} = \delta_{ij} + o(q_{\max})$, implying in particular that for $|x_j| = \alpha_n/\sqrt{n}$, the leading term on the right hand side is of order $\alpha_n/\sqrt{n}$. As a consequence, we can convert the additive error into a multiplicative error $1 + O(\alpha_n^2/n)$, giving the desired bound (5.36). Using once more that $C_{ij}^{-1} = \delta_{ij} + o(q_{\max})$, the fact that $x_j\eta_j(\boldsymbol{x}) \geq 0$ is an immediate corollary of (5.36).

We are left with the proof of (5.37). To this end, we rewrite the left hand side as

$$\begin{aligned}\iiint_{-\mu_1}^{\mu_1} & e^{2\pi n(i\mathbf{x}\cdot\mathbf{f}-\mathbf{x}\cdot\boldsymbol{\eta})} e^{-\sum_{\boldsymbol{\delta}} n_{\boldsymbol{\delta}} F(\mathbf{f}\cdot\boldsymbol{\delta}+i\boldsymbol{\eta}\cdot\boldsymbol{\delta})} \prod_{j=1}^{k} df_j \\ &= e^{-nG_k(\mathbf{x})} \iiint_{-\mu_1}^{\mu_1} e^{-n\tilde{G}_{n,k}(\mathbf{f})} \prod_{j=1}^{k} df_j\end{aligned} \tag{5.42}$$

with

$$\tilde{G}_{n,k}(\mathbf{f}) = \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} \Big(F(\mathbf{f}\cdot\boldsymbol{\delta} + i\boldsymbol{\eta}\cdot\boldsymbol{\delta}) - F(i\boldsymbol{\eta}\cdot\boldsymbol{\delta})\Big) - 2\pi i\mathbf{x}\cdot\mathbf{f}. \tag{5.43}$$

Observing that the derivatives of $\tilde{G}_{n,k}(\mathbf{f})$ at $\mathbf{f}=0$ vanish by the definition of $\boldsymbol{\eta}(\mathbf{x})$, we now expand $\tilde{G}_{n,k}(\mathbf{f})$ as

$$\tilde{G}_{n,k}(\mathbf{f}) = \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n}\Big(\frac{1}{2}(\mathbf{f}\cdot\boldsymbol{\delta})^2 F''(i\boldsymbol{\eta}\cdot\boldsymbol{\delta}) + \frac{1}{3!}(\mathbf{f}\cdot\boldsymbol{\delta})^3 F^{(3)}(i\boldsymbol{\eta}\cdot\boldsymbol{\delta}) + O((\mathbf{f}\cdot\boldsymbol{\delta})^4)\Big). \tag{5.44}$$

Arguing as in the proof of Proposition 5.2 we now have that

$$\Re\mathrm{e}\tilde{G}_{n,k}(\mathbf{f}) \ge \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n}(\mathbf{f}\cdot\boldsymbol{\delta})^2 = \sum_{i,j} f_i C_{ij} f_j \tag{5.45}$$

provided μ_1 is sufficiently small and n is sufficiently large. Since the maximal off-diagonal overlap $q_{\max}$ is of order $o(1)$, we conclude that $\Re\mathrm{e}\tilde{G}_{n,k}(\mathbf{f}) \ge \frac{1}{2}\|\mathbf{f}\|_2^2$ provided μ_1 is sufficiently small and n is sufficiently large. As a consequence, the integral in (5.42) is dominated by configurations for which $\|\mathbf{f}\|_2$ is smaller than, say, $\log n/\sqrt{n}$. More quantitatively, we have

$$\begin{aligned}
&\iiint_{-\mu_1}^{\mu_1} e^{-n\tilde{G}_{n,k}(\mathbf{f})} \prod_{j=1}^k df_j \\
&\quad = \iiint_{\|\mathbf{f}\|_2 \le \log n/\sqrt{n}} e^{-n\tilde{G}_{n,k}(\mathbf{f})} \prod_{j=1}^k df_j + O\Big(\iiint_{\|\mathbf{f}\|_2 \ge \log n/\sqrt{n}} e^{-n\|\mathbf{f}\|^2/2} \prod_{j=1}^k df_j\Big) \\
&\quad = \iiint_{\|\mathbf{f}\|_2 \le \log n/\sqrt{n}} e^{-n\tilde{G}_{n,k}(\mathbf{f})} \prod_{j=1}^k df_j + O\Big(e^{-\frac{1}{2}\log^2 n}\Big).
\end{aligned} \tag{5.46}$$

For $\|\mathbf{f}\|_2 \le \log n/\sqrt{n}$ we expand $e^{-n\tilde{G}_{n,k}(\mathbf{x})}$ as

$$\begin{aligned}
\tilde{G}_{n,k}(\mathbf{f}) &= \exp\Big(-\frac{1}{2}\sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n}(\mathbf{f}\cdot\boldsymbol{\delta})^2 F''(i\boldsymbol{\eta}\cdot\boldsymbol{\delta}) + O(n\|\mathbf{f}\|^3)\Big) \\
&= e^{-\frac{1}{2}\sum_{ij} f_i M_{ij} f_j}\Big(1 + O(n\|\mathbf{f}\|^3)\Big)
\end{aligned} \tag{5.47}$$

where M is the matrix with matrix elements

$$\begin{aligned}
M_{ij} &= \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n}\delta_i\delta_j F''(i\boldsymbol{\eta}\cdot\boldsymbol{\delta}) = (2\pi)^2 C_{ij} + O(\|\boldsymbol{\eta}\|^2) \\
&= (2\pi)^2\delta_{ij} + O(\alpha_n^2/n) + O(q_{\max}).
\end{aligned} \tag{5.48}$$

As a consequence,

$$\iiint_{\|\mathbf{f}\|_2 \le \log n/\sqrt{n}} e^{-n\tilde{G}_{n,k}(\mathbf{f})} \prod_{j=1}^k df_j = \Big(\frac{1}{2\pi n}\Big)^{\frac{k}{2}}\Big(1+O(n^{-1/2})+O(\alpha_n^2/n)+O(q_{\max})\Big). \tag{5.49}$$

Combined with (5.42) and (5.46), this implies the desired bound (5.37). □

Lemma 5.3, Lemma 5.4, the relation (5.29), Lemma 5.4 and Lemma 5.5 we now easily prove the following proposition.

Proposition 5.6. *Let $0 \le \alpha_n = o(\sqrt{n})$ and let $0 \le \lambda_n = o(\sqrt{n})$. If n is sufficiently large and $\boldsymbol{\sigma}^{(1)} \ldots, \boldsymbol{\sigma}^{(k)}$ obey the condition (3.13), then*

$$\begin{aligned} \mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for } j=1,\ldots,k\Big) &= \Big(\prod_{\ell=1}^m \gamma_\ell^{k_\ell}\Big) \\ \times \left(\frac{\xi_n}{\sqrt{2\pi}}\right)^k \sum_{\mathbf{x}\in\{-\frac{\alpha_n}{\sqrt{n}}, \frac{\alpha_n}{\sqrt{n}}\}^k} & e^{-nG_{n,k}(\mathbf{x})}\Big(1+o(1)\Big). \end{aligned} \tag{5.50}$$

If we strengthen the condition $\alpha_n = o(\sqrt{n})$ to $\alpha_n = o(n^{1/4})$, we have

$$\begin{aligned} &\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j=1,\ldots,k\Big) \\ &\qquad = \prod_{i=1}^k \mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}]\Big) e^{O(\alpha_n^2 q_{max})+o(1)}. \end{aligned} \tag{5.51}$$

Proof. Observe that under the condition (3.13), we have that $n_{\min} = n2^{-k} + o(n)$ and $q_{\max} = o(1)$. We may therefore use Lemma 5.3, Lemma 5.4, the relation (5.29), Lemma 5.4 and Lemma 5.5 to conclude that under the conditions of the proposition, there is a constant $c > 0$ such that

$$\begin{aligned} &\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for } j=1,\ldots,k\Big) = \prod_{\ell=1}^m (2q_{n,\ell})^{k_\ell} \\ &\times 2^{-k} \sum_{\mathbf{x}\in\{-\frac{\alpha_n}{\sqrt{n}}, \frac{\alpha_n}{\sqrt{n}}\}^k} \left(e^{-nG_{n,k}(\mathbf{x})} \left(\frac{1}{2\pi n}\right)^{k/2} \Big(1+o(1)\Big) + O(e^{-cn}) + O(\sqrt{n}\xi_n) \right). \end{aligned} \tag{5.52}$$

Next we expand $G_{n,k}(\mathbf{x})$ with the help of the bounds (5.14) and (5.36), yielding the approximation

$$\begin{aligned} G_{n,k}(\mathbf{x}) &= -\frac{1}{2}\sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n}(2\pi\boldsymbol{\delta}\cdot\boldsymbol{\eta})^2 + 2\pi\boldsymbol{\eta}\cdot\mathbf{x} + O(\|\mathbf{x}\|_2^4) \\ &= -\frac{(2\pi)^2}{2}(\boldsymbol{\eta}, C\boldsymbol{\eta}) + 2\pi\boldsymbol{\eta}\cdot\mathbf{x} + O(\|\mathbf{x}\|_2^4) \\ &= \frac{1}{2}(\mathbf{x}, C^{-1}\mathbf{x}) + O\Big(\frac{\alpha_n^4}{n^2}\Big). \end{aligned} \tag{5.53}$$

Note that this implies in particular that $nG_{n,k}(\mathbf{x}) = O(\alpha_n^2) + O(\alpha_n^4/n) = o(n)$. The leading term on the right hand side of (5.52) is therefore much larger than the additive error terms, which both decay exponentially in n. These additive error terms can therefore be converted into a multiplicative error term $(1 + o(1))$. Inserting the value of $q_{n,\ell}(= \gamma_\ell \xi_n \sqrt{n})$ and using the fact that $\sum_\ell k_\ell = k$, this yields the approximation (5.50).

To infer the bound (5.51), we use the assumption $\alpha_n = o(n^{1/4})$ to expand the factor $(\xi_n/\sqrt{2\pi})^k$ as $2^{-nk}e^{k\alpha_n^2/2}(1 + o(1))$, and the factor $e^{-nG_{n,k}(\mathbf{x})}$ as

$$\begin{aligned} e^{-\frac{n}{2}(\mathbf{x},C^{-1}\mathbf{x})+O(\alpha_n^4/n)} &= e^{-\frac{n}{2}\|\mathbf{x}\|_2^2+O(n\|x\|_2^2 q_{\max})}\Big(1+O(\alpha_n^4/n)\Big) \\ &= e^{-\frac{k}{2}\alpha_n^2+O(\alpha_n^2 q_{\max})}\Big(1+o(1)\Big). \end{aligned} \tag{5.54}$$

Inserting these estimates into (5.50) and taking into account the bound (5.12), this gives (5.51). □

5.4. **Failure of Poisson convergence for faster growing $\boldsymbol{\alpha}_{\mathbf{n}}$.** In this section, we prove the last statement of Theorem 2.1 for general distributions. Throughout this section, we will assume that $\alpha_n = O(n^{1/4})$ and $\lambda_n = o(\sqrt{n})$. Near the end of the section, we will specialize to $\alpha_n = \Theta(n^{1/4})$, for which we will prove absence of Poisson convergence.

We start from (5.50), but instead of (5.53) we use a more accurate approximation for $G_{n,k}(\mathbf{x})$. To this end, we use the fact that $\hat{\rho}(f) = \hat{\rho}(-f)$ is analytic in a neighborhood of zero to infer the existence of a constant c_4 such that

$$F(z) = \frac{(2\pi)^2}{2} z^2 + c_4 (2\pi)^4 z^4 + O(|z|^6)$$

and

$$F'(z) = (2\pi)^2 z + 4c_4 (2\pi)^4 z^3 + O(|z|^5).$$

Remark 5.7. Using the fact that $\mathbb{E}(X^4) \geq \mathbb{E}(X^2)^2 = 1$ and expanding the $\log \hat{\rho(f)}$ one can see that $c_4 \leq \frac{1}{12}$, with equality holding if and only if $X^2 = 1$ with probability one. Thus for all random variables whose density satisfies assumption (2.1) we have $c_4 < 1/12$.

Using these expressions and the fact that $C^{-1}_{ij} = \delta_{ij} + O(q_{\max}) = \delta_{ij} + O(\lambda_n n^{-1/2})$ when $\boldsymbol{\sigma}^{(1)} \dots, \boldsymbol{\sigma}^{(k)}$ obey the condition (3.13), we then expand the solution of (5.34) as

$$\begin{aligned} \eta_i &= \frac{1}{2\pi}(C^{-1}\mathbf{x})_i + \frac{4c_4}{2\pi} \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} (C^{-1}\boldsymbol{\delta})_i (\boldsymbol{\delta}, C^{-1}\mathbf{x})^3 + O(\|\mathbf{x}\|^5) \\ &= \frac{1}{2\pi}(C^{-1}\mathbf{x})_i + \frac{4c_4}{2\pi} \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} \delta_i (\boldsymbol{\delta} \cdot \mathbf{x})^3 + O(\|\mathbf{x}\|^5) + O(\|\mathbf{x}\|^3 \lambda_n n^{-1/2}). \end{aligned} \tag{5.55}$$

In order to analyze the sum over $\boldsymbol{\delta}$, we use the condition (3.13) to estimate

$$\sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} \delta_i \delta_j \delta_k \delta_l = 2^{-k} \sum_{\boldsymbol{\delta}} \delta_i \delta_j \delta_k \delta_l + O(\lambda_n n^{-1/2}).$$

The first term is zero unless i, j, k, l are such that either all of them are equal or are pairwise equal for some pairing of i, j, k, l. Observing that there are 3 possible ways to pair four numbers into two pairs of two, this leads to the estimate

$$\begin{aligned} \eta_i &= \frac{1}{2\pi}(C^{-1}\mathbf{x})_i + \frac{4c_4}{2\pi} x_i \big(x_i^2 + 3 \sum_{k \neq i} x_k^2 \big)\Big) + O(\|\mathbf{x}\|^3 \lambda_n n^{-1/2}) + O(\|\mathbf{x}\|^5) \\ &= \frac{1}{2\pi}(C^{-1}\mathbf{x})_i + (1 + 3(k-1)) \frac{4c_4 \alpha_n^2}{2\pi n} x_i + O(\|\mathbf{x}\|^3 \lambda_n n^{-1/2}) + O(\|\mathbf{x}\|^5) \end{aligned} \tag{5.56}$$

where we used the fact that $|x_i| = |x_j| = \alpha_n/\sqrt{n}$ in the last step. Inserting this expression into the definition (5.35) and expanding the result in a similar way as we expanded η above, this leads to the approximation

$$
\begin{aligned}
G_{n,k}(\mathbf{x}) &= -\frac{(2\pi)^2}{2}\sum_{\boldsymbol{\delta}}\frac{n_{\boldsymbol{\delta}}}{n}(\boldsymbol{\delta}\cdot\boldsymbol{\eta})^2 + c_4 2^{-k}\sum_{\boldsymbol{\delta}}(\boldsymbol{\delta}\cdot\mathbf{x})^4 + 2\pi\boldsymbol{\eta}\cdot\mathbf{x} \\
&\quad + O(\|x\|^6) + O(\|x\|^4\lambda_n n^{-1/2}) \\
&= -\frac{(2\pi)^2}{2}(\boldsymbol{\eta}, C\boldsymbol{\eta}) + c_4 k(1+3(k-1))\frac{\alpha_n^4}{n^2} + 2\pi\boldsymbol{\eta}\cdot\mathbf{x} \\
&\quad + O(\|x\|^6) + O(\|x\|^4\lambda_n n^{-1/2}) \\
&= -\frac{1}{2}(\mathbf{x}, C^{-1}\mathbf{x}) - k(1+3(k-1))\frac{4c_4\alpha_n^4}{n^2} + c_4 k(1+3(k-1))\frac{\alpha_n^4}{n^2} \\
&\quad + (\mathbf{x}, C^{-1}\mathbf{x}) + k(1+3(k-1))\frac{4c_4\alpha_n^4}{n^2} + O(\|x\|^6) \\
&\quad + O(\|x\|^4\lambda_n n^{-1/2}) \\
&= \frac{1}{2}(\mathbf{x}, C^{-1}\mathbf{x}) + c_4 k(1+3(k-1))\frac{\alpha_n^4}{n^2} + O\Big(\frac{\alpha_n^6}{n^3}\Big) + O(\frac{\alpha_n^4}{n^2}\lambda_n n^{-1/2}).
\end{aligned}
\tag{5.57}
$$

Using that $\alpha_n = O(n^{1/4})$ and $\lambda_n = o(\sqrt{n})$, this gives

$$
nG_{n,k}(\mathbf{x}) = \frac{n}{2}(\mathbf{x}, C^{-1}\mathbf{x}) + c_4 k(1+3(k-1))\frac{\alpha_n^4}{n} + o(1).
$$

As a consequence, we have that

$$
\begin{aligned}
&\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for } j=1,\dots,k\Big) = \Big(\prod_{\ell=1}^m \gamma_\ell^{k_\ell}\Big) \\
&\times\left(\frac{\xi_n}{\sqrt{2\pi}}\right)^k \sum_{\mathbf{x}\in\{-\frac{\alpha_n}{\sqrt{n}},\frac{\alpha_n}{\sqrt{n}}\}^k} e^{-\frac{n}{2}(\mathbf{x},C^{-1}\mathbf{x})} e^{-c_4 k(1+3(k-1))\frac{\alpha_n^4}{n}}\Big(1+o(1)\Big)
\end{aligned}
\tag{5.58}
$$

whenever $\boldsymbol{\sigma}^{(1)}\dots,\boldsymbol{\sigma}^{(k)}$ obey the condition (3.13).

Next we expand ξ_n, using that $\xi_n = (2^{(n-1)}g_n(\alpha_n))^{-1}$ with $g_n(\alpha_n)$ given by (5.10). To this end, we approximate $\eta(x)$ as $\eta(x) = \frac{x}{2\pi}(1+4c_4x^2) + O(x^5)$ and $G(x)$ as

$$
G(x) = -\frac{1}{2}(2\pi\eta(x))^2 + c_4(2\pi\eta(x))^4 + 2\pi x\eta(x) + O(x^6) = \frac{x^2}{2} + c_4x^4 + O(x^6),
$$

giving the expansion

$$
\xi_n = \sqrt{\frac{\pi}{2}}2^{-(n-1)}\exp\Big(\frac{1}{2}\alpha_n^2 + c_4\frac{\alpha_n^4}{n}\Big)\big(1+o(1)\big). \tag{5.59}
$$

Inserting this expansion into (5.58) we then continue as in the proof of Theorem 3.3 to get that

$$
\mathbb{E}[\prod_{\ell=1}^m (Z_n(a_n^\ell, b_n^\ell))_{k_\ell}] = \Big(\prod_{\ell=1}^m \gamma_\ell^{k_\ell}\Big)\exp\Big(\frac{\alpha_n^4}{n}\Big(\frac{k(k-1)}{4} - 3c_4k(k-1)\Big)\Big)\Big(1+o(1)\Big). \tag{5.60}
$$

Specializing now to $\alpha_n = \kappa n^{1/4}$, this implies in particular that all moments of $Z_n(t)$ are bounded, and

$$
\lim_{n\to\infty}\mathbb{E}[(Z_n(t))_k] = t^k\exp\Big(\kappa^4(k-1)k\Big(1/4 - 3c_4\Big)\Big). \tag{5.61}
$$

In view of Lemma 3.2 and the fact that $c_4 < 1/12$ by Remark 5.7, this is incompatible with weak convergence to a Poisson random variable. This establishes the failure of the modified REM conjecture for $\lim_{n\to\infty} \alpha_n^{-1/4} = \kappa > 0$.

Note that the original REM conjecture also fails if $\lim_{n\to\infty} \alpha_n n^{-1/4} = \kappa > 0$. Indeed, in this range, the original scaling (1.3) and the modified scaling differ by the asymptotically constant factor $e^{-c_4\kappa^4}(1+o(1))$. This makes things worse, since now even the first moment does not converge to the desired value unless $c_4 = 0$, in which case the original and the modified REM conjecture remain equivalent for $\alpha_n = O(n^{1/4})$. This completes the proof of the last statement of Theorem 2.1.

Remark 5.8. It is not hard to see that for $\alpha_n = \kappa n^{1/4}$ the distribution of the overlap $Q_{n,t}$ converges again to a superposition of two shifted Gaussians, see Remark 4.2, where this was discussed for the case where $X_1, \dots, X_n$ where standard normals. Indeed, the only difference with the situation discussed in that remark is the extra factor of $e^{-c_4(k-1)(1+3k)\kappa^4}$. But this factor does not depend on the overlap, and hence does not influence the overlap distribution.

6. Analysis of the SK model

For $\alpha_n = O(n^\eta)$ and $\eta < 1$, the local REM conjecture for the SK model has been proved in [BK05a]. To extend the proof up to the threshold $\alpha_n = o(1)$, a little bit more care is needed, but given the analysis of the Npp with Gaussian noise from the Section 4, this is still relatively straightforward, even though the details vary at several places. But for α_n of order n we have to be quite careful, since now several error terms which went to zero before are not vanishing anymore. It turns out, however, that for $\alpha_n/n \le \epsilon_0$ and ϵ_0 sufficiently small, we can at least control the first three moments, which is enough to prove absence of Poisson convergence if $\limsup \alpha_n/n > 0$.

We start with the analysis of the first moment.

6.1. **First moment.** For the SK model, the energy $E(\boldsymbol{\sigma})$ is a Gaussian random variable with density given by (1.5). For $\alpha_n = O(n)$ and $\tilde{\xi}_n = o(1)$, the first moment can therefore be approximated as

$$\begin{aligned}\mathbb{E}[Z_n(a_n^\ell, b_n^\ell)] &= \frac{2^{n-1}}{\sqrt{2\pi n}} \int_{a_n^\ell}^{b_n^\ell} e^{-\frac{1}{2n}x^2} dx \\ &= \frac{\gamma_\ell \tilde{\xi}_n 2^{n-1}}{\sqrt{2\pi n}} e^{-\frac{1}{2n}\alpha_n^2} \Big(1 + O\Big(\frac{\alpha_n \tilde{\xi}_n}{n}\Big) + O\Big(\frac{\tilde{\xi}_n^2}{n}\Big)\Big) \\ &= \gamma_\ell (1 + o(1)).\end{aligned} \tag{6.1}$$

This proves the convergence of the first moment for $\alpha_n \le cn$ and $c < \sqrt{2\log 2}$.

6.2. **Families of linearly independent configurations.** When analyzing the factorization properties of the joint distribution of $E(\boldsymbol{\sigma}^{(1)}), \dots, E(\boldsymbol{\sigma}^{(k)})$, we will want to use the representation (3.32), which at a minimum, requires that the covariance matrix C defined in (3.33) is invertible. The following Lemma 6.1 shows that a family of linearly independent configurations leads to an invertible covariance matrix, and gives a bound on the contribution of the families which are not linearly independent. It serves the same purpose as Lemma 3.4 (2) served for the

NPP. Note that statements similar to those in Lemma 6.1 were proved in [BK05a] under the more restrictive condition that $\alpha_n = O(n^{-\eta})$ with $\eta < 1$.

To state the lemma, we define $\tilde{R}_{n,k}$ as

$$\tilde{R}_{n,k} = \frac{1}{2^k} \sum_{\boldsymbol{\sigma}^{(1)},\cdots,\boldsymbol{\sigma}^{(k)}}{}' \mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j = 1, \ldots, k\Big), \tag{6.2}$$

where the sum runs over pairwise distinct configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$ which are linearly dependent.

Lemma 6.1. *Let $k < \infty$. Then there exists a constant $\epsilon_k > 0$ such that the following statements are true.*

(1) If $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$ are linearly independent, the matrix C defined in (3.33) *is invertible, and*

$$\tilde{g}^{(k)}(\mathbf{x}) \le \Big(\frac{n}{2\pi}\Big)^{k/2} e^{-\frac{1}{2}(\mathbf{x}, C^{-1}\mathbf{x})}. \tag{6.3}$$

(2) If $\alpha_n \le n\epsilon_k$, then

$$\tilde{R}_{n,k} \to 0 \qquad \text{as} \qquad n \to \infty. \tag{6.4}$$

Proof. (1) This statement is quite easy and must be known to most experts in the field. First, we rewrite the matrix elements of C as $C_{ij} = nq(\boldsymbol{\eta}^{(i)}, \boldsymbol{\eta}^{(j)})$, where $\boldsymbol{\eta}^{(1)}, \ldots, \boldsymbol{\eta}^{(k)}$ are vectors in $\{-1,+1\}^{n^2}$. Indeed, setting $\eta^{(i)}_{r,s} = \sigma^{(i)}_s \sigma^{(i)}_r$, where $r, s = 1, \ldots n$ and $i = 1, \ldots k$, we see that $q(\boldsymbol{\eta}^{(i)}, \boldsymbol{\eta}^{(j)}) = \big(q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)})\big)^2$, implying the above representation for C. Next we observe that that the linear independence of $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$ implies linear independence of $\boldsymbol{\eta}^{(1)}, \ldots, \boldsymbol{\eta}^{(k)}$, which in turn can easily be seen to imply linear independence of the row vectors of the matrix with matrix elements $q(\boldsymbol{\eta}^{(i)}, \boldsymbol{\eta}^{(j)})$. This gives that C is non-degenerate. But C is a $k \times k$ matrix with entries which are multiples of n^{-1}, so if $\det C \ne 0$, we must have $|\det C| \ge n^{-k}$. This implies the bound (6.3).

(2) To prove (2), we decompose the sum in (6.2) according to the rank of the matrix M formed by the vectors $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$. Assume that the rank of M is equal to $u < k$. Reordering the vectors $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$, if necessary, let us further assume that $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(u)}$ are linearly independent. With the help of (1), we then bound the probability on the right hand side of (6.2) by

$$\begin{aligned} &\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j = 1, \ldots, k\Big) \\ &\quad \le \mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j = 1, \ldots, u\Big) \\ &\quad \le \Big(\frac{n}{2\pi}\Big)^{u/2} \prod_{j=1}^{u} (\tilde{\xi}_n \gamma_{\ell(j)}) = O((\sqrt{n}\tilde{\xi}_n)^u) \end{aligned} \tag{6.5}$$

To continue, we use the following two facts, proven, e.g., in [BCP01] (see also Lemma 3.9 in [BCMN05]:

(1) If $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$ are pairwise distinct, the rank of M is $u < k$ and $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(u)}$ are linearly independent, then $n_{\boldsymbol{\delta}}(\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(u)}) = 0$ for at least one $\boldsymbol{\delta} \in \{-1,+1\}^u$.
(2) Given $u < k$ linearly independent vectors $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(u)}$, there are at most $2^{u(k-u)}$ ways to choose $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$ in such a way that the rank of M is u.

As a consequence, of (1), we have

$$|\max_{\boldsymbol{\delta}} |n_{\boldsymbol{\delta}}(\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(u)}) - 2^{-u}n| \geq 2^{-u}n.$$

Combined with Lemma 3.4 (1) and the property (2) above, we conclude that there are at most $O(2^{nu}e^{-2^{-(2u+1)}n}) = O(2^{nu}e^{-2^{-(2k-1)}n})$ ways to choose k pairwise distinct configurations $\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}$ such that the rank of M is u. Using the fact that $(\sqrt{n}\tilde{\xi}_n)^u = O(n^u e^{\frac{u}{2n}\alpha_n^2} 2^{-nu})$, we immediately see that for $\alpha_n \leq \epsilon_k n$ and ϵ_k sufficiently small, the contribution of all configurations $\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}$ such that the rank of M is smaller than k decays exponentially in n, so in particular it is $o(1)$. □

Consider now a family of linearly independent configurations $\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}$. We claim that for such a family and $\alpha_n \leq cn$ with $c < \sqrt{2\log 2}$, we have

$$\begin{aligned}\mathbb{P}\Big(E(\boldsymbol{\sigma}^{(j)}) \in [a_n^{\ell(j)}, b_n^{\ell(j)}] \text{ for all } j=1,\dots,k\Big) \\ = \prod_{\ell=1}^{m} (\tilde{\xi}_n\gamma_\ell)^{k_\ell} \frac{1}{(2\pi)^{k/2}} \frac{1}{(\det C)^{1/2}} e^{-\frac{1}{2}(\boldsymbol{\alpha}, C^{-1}\boldsymbol{\alpha})}(1+o(1)),\end{aligned} \tag{6.6}$$

where $\boldsymbol{\alpha}$ is the vector $(\alpha_n,\dots,\alpha_n) \in \mathbb{R}^k$ and C is the covariance matrix defined in (3.33). To prove this approximation, we have to show that $(\mathbf{x}, C^{-1}\mathbf{x}) = (\boldsymbol{\alpha}, C^{-1}\boldsymbol{\alpha}) = o(1)$ whenever $\mathbf{x}$ is a vector with $x_i = \alpha_n + O(\tilde{\xi}_n)$. This in turn requires an upper bound on the inverse of C. To prove such a bound we use that the matrix elements of C are bounded by n, while $\det C$ is bounded from below by n^{-k}. Using Cramer's rule, we conclude that the norm of C^{-1} is $O(n^{2k-1})$, which in turn implies that $(\mathbf{x}, C^{-1}\mathbf{x}) = (\boldsymbol{\alpha}, C^{-1}\boldsymbol{\alpha}) + O(n^{2k-1}\alpha_n\tilde{\xi}_n) + O(n^{2k-1}\tilde{\xi}_n^2)$. For $\alpha_n \leq cn$ with $c < \sqrt{2\log 2}$, the error term is $o(1)$, as desired.

Assume that $\alpha_n \leq cn$ with $c < \min\{\sqrt{2\log 2}, \epsilon_k\}$. Using first the representation (3.11), then Lemma 6.1 and the bound (6.6), and finally the explicit formula (1.6) for $\tilde{\xi}_n$, we now approximate the k^{th} factorial moment as

$$\begin{aligned}\mathbb{E}\Big[\prod_{\ell=1}^{m}(Z_n(a_n^\ell, b_n^\ell))_{k_\ell}\Big] = \Big(\prod_{\ell=1}^{m}\gamma_\ell^{k_\ell}\Big) \\ \times 2^{-nk} \sum_{\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}}'' \frac{n^{k/2}}{(\det C)^{1/2}} e^{\frac{1}{2n}\|\boldsymbol{\alpha}\|_2^2} e^{-\frac{1}{2}(\boldsymbol{\alpha}, C^{-1}\boldsymbol{\alpha})}\Big(1+o(1)\Big) + o(1)\end{aligned} \tag{6.7}$$

where the sum goes over families of linearly independent configurations $\boldsymbol{\sigma}^{(1)}$, …, $\boldsymbol{\sigma}^{(k)}$.

6.3. **Poisson convergence for $\boldsymbol{\alpha_n = o(n)}$.** In this section, we prove Theorem 3.6. To this end, we again choose λ_n in such a way that $\alpha_n = o(\lambda_n\sqrt{n})$, $\lambda_n = o(\sqrt{n})$ and $e^{-\lambda_n^2/2}$ decays faster than any power of n. Recalling Lemmas 3.4 (1) and (3), we may then restrict the sum in (6.7) to a sum over configurations with $q_{\max} = O(\lambda_n/\sqrt{n})$. Expanding the inverse of C as $C_{ij}^{-1} = \frac{1}{n}(\delta_{ij} + O(q_{\max}^2))$, we then approximate $\det C$ as $\det C = n^k(1+o(1))$, and $(\boldsymbol{\alpha}, C^{-1}\boldsymbol{\alpha})$ as $(\boldsymbol{\alpha}, C^{-1}\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_2^2 + O(n^{-1}\alpha_n^2 q_{\max}^2)$. Using Lemma 3.4 and Lemma 6.1 a second time to extend the sum over families of configurations back to a sum over all families of configurations in $\{-1,+1\}^n$, the proof of Theorem 3.6 is therefore reduced to the proof of the bound

$$2^{-nk} \sum_{\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}} e^{O(\frac{1}{n}\alpha_n^2 q_{\max}^2)} = 1 + o(1). \tag{6.8}$$

Since typical configurations lead to a maximal off-diagonal overlap $q_{\max}$ of order $O(n^{-1/2})$, we expect that such a bound holds as long as $\alpha_n = o(n)$. Lemma 7.1 in Section 7.2 implies that this is indeed the case.

6.4. **Absence of Poisson convergence for faster growing $\boldsymbol{\alpha_n}$.** In this section, we prove Theorem 3.7. We start by proving the following lemma.

Lemma 6.2. *Given a positive integer k, there are constants $\tilde\epsilon_k > 0$ and $C_k < \infty$ such that*

$$\mathbb{E}\Big[\prod_{\ell=1}^{m}(Z_n(a_n^\ell, b_n^\ell))_{k_\ell}\Big] \le C_k \prod_{\ell=1}^{m} \gamma_\ell^{k_\ell} + o(1) \tag{6.9}$$

whenever $\alpha_n \le \tilde\epsilon_k n$, $k = \sum_\ell k_\ell$ and n is large enough.

Proof. Starting from the approximation (6.7), we will again restrict ourselves to configurations obeying the condition (3.13), but this time we will choose λ_n of the form $\lambda_n = \sqrt{2kn\tilde\epsilon_k}$ where $\tilde\epsilon_k > 0$ will be chosen in such a way that $\tilde\epsilon_k < \sqrt{2\log 2}$ and $\tilde\epsilon_k \le \epsilon_k$. Note that our choice of λ_n guarantees that for $\alpha_n \le n\tilde\epsilon_n$, the sum over families of configurations violating the condition (3.13) decays exponentially with n. Note further that the condition (3.13) with $\lambda_n = \sqrt{2kn\tilde\epsilon_k}$ guarantees that $q_{\max}^2 \le 2^{2k+1}k\tilde\epsilon_k^2$. Expanding $\det C$ as $\det C = n^k(1 + O(q_{\max}^2))$, we therefore have that $\det C \ge \frac{1}{2}n^k$ provided $\tilde\epsilon_k$ is chosen small enough.

Let us finally write the matrix C as $C = \frac{1}{n}(\mathbb{I}_k + A)$, where $\mathbb{I}_k$ is the identity matrix of size k, and A is the matrix with zero diagonal and off-diagonal entries $q^2(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)})$. Expanding C^{-1} as $C^{-1} = \frac{1}{n}(\mathbb{I}_k - A + \frac{A^2}{1+A})$, we see that

$$(\boldsymbol{\alpha}, C^{-1}\boldsymbol{\alpha}) \ge \frac{1}{n}\Big(\|\boldsymbol{\alpha}\|_2^2 - (\boldsymbol{\alpha}, A\boldsymbol{\alpha})\Big) \ge \frac{1}{n}\Big(\|\boldsymbol{\alpha}\|_2^2 - k^2 q_{\max}^2 \alpha_n^2\Big). \tag{6.10}$$

Together with our lower bound on $\det C$, we see that the proof of the lemma now reduces to the bound

$$2^{-nk} \sum_{\boldsymbol{\sigma}^{(1)},\ldots,\boldsymbol{\sigma}^{(k)}} \sqrt{2} e^{\frac{1}{2n}(k\alpha_n q_{\max})^2} \le C_k. \tag{6.11}$$

For $\tilde\epsilon_k$ small enough, this bound again follows from Lemma 7.1 in Section 7.2. □

Note that Lemma 6.2 implies statement (ii) of Theorem 3.7, while the bound (6.1) implies statement (i). We are therefore left with the proof of (iii), i.e., the statement that $\limsup_{n\to\infty} \mathbb{E}[(Z_n(a,b))_2] > \gamma^2$ when $\limsup \alpha_n/n > 0$. Recalling the representation, this in turn requires us to prove that

$$\limsup_{n\to\infty} 2^{-2n} \sum_{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}}{}'' \frac{n}{(\det C)^{1/2}} e^{\frac{1}{2n}\|\boldsymbol{\alpha}\|_2^2} e^{-\frac{1}{2}(\boldsymbol{\alpha}, C^{-1}\boldsymbol{\alpha})} > 1. \tag{6.12}$$

Let $q = q(\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)})$ be the off-diagonal overlap of the two configurations $\boldsymbol{\sigma}^{(1)}$ and $\boldsymbol{\sigma}^{(2)}$. Since C is now just a 2×2 matrix, both its inverse and its determinant can be easily calculated, giving $\det C = n^2(1-q^4)$, and, using that $\|\boldsymbol{\alpha}\|_2^2 = k\alpha_n^2 = 2\alpha_n^2$,

$$(\boldsymbol{\alpha}, C^{-1}\boldsymbol{\alpha}) = \frac{1}{n(1-q^4)}(2\alpha_n^2 - 2q^2\alpha_n^2) = \frac{2\alpha_n^2}{n(1+q^2)} = \frac{2\alpha_n^2}{n}(1 - \frac{q^2}{1+q^2}). \tag{6.13}$$

As a consequence, we will have to prove a lower bound on the expression

$$2^{-2n} \sum''_{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}} \frac{1}{\sqrt{1-q^4}} e^{\frac{\alpha_n^2}{n}\frac{q^2}{1+q^2}}. \tag{6.14}$$

Taking subsequence, if necessary, let us assume that α_n/n converges to some κ, with $0 < \kappa < \min\{\epsilon_2, \sqrt{2\log 2}\}$. We then bound the sum $\sum''$ from below by restricting it to all configurations for which $|q| \leq 2\kappa$. Under this restriction, the summand is bounded from below by $\frac{1}{\sqrt{1-16\kappa^4}} e^{\beta_n q^2}$, with $\beta_n = \alpha_n^2/(n(1+4\kappa^2))$. Observing that $q^2 \leq 1$ we may then use Lemma 3.4 (1) and (3) to extend the sum back to a sum which runs over all configurations $\boldsymbol{\sigma}^{(1)}$ and $\boldsymbol{\sigma}^{(2)}$, leading to the lower bound

$$2^{-2n} \sum''_{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}} \frac{1}{\sqrt{1-q^4}} e^{\frac{\alpha_n^2}{n}\frac{q^2}{1+q^2}} \geq \frac{2^{-2n}}{\sqrt{1-16\kappa^4}} \sum_{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}} e^{\beta_n q^2} + o(1). \tag{6.15}$$

Let $\mathbb{E}_2$ denote expectations with respect to the uniform measure over families of configurations $\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)} \in \{-1,+1\}^n$. Using Hölder and the fact that $\mathbb{E}_2[q^2] = \frac{1}{n}$, we then lower bound the right hand side by

$$\begin{aligned} &\frac{2^{-2n}}{\sqrt{1-16\kappa^4}} \sum_{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}} e^{\beta_n q^2} + o(1) \\ &\qquad \geq \frac{e^{\beta_n/n}}{\sqrt{1-16\kappa^4}} + o(1) = \frac{1}{\sqrt{1-16\kappa^4}} \exp\Big(\frac{\kappa^2}{1+4\kappa^2}\Big) + o(1). \end{aligned} \tag{6.16}$$

For κ sufficiently small, the right hand side is asymptotically larger than 1, proving the desired lower bound (6.12), which completes the proof of Theorem 3.7.

7. Auxiliary results

7.1. **Proof of Lemma 3.5.**

Proof of Lemma 3.5. We will have to prove that

$$2^{-nk} \sum_{\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}} e^{f(\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)})} = 1 + o(1). \tag{7.1}$$

Recalling the assumption $|f(\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)})| \leq c\alpha_n^2 q_{\max}$, let $\theta_n = o(n)$ be a sequence of positive integers such that $\theta_n \to \infty$ as $n \to \infty$ and $\alpha_n^2\theta_n = o(\sqrt{n})$.

We now split the sum on the left hand side of (7.1) into two parts: the sum over all families of configurations $\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}$ that satisfy (3.13) with θ_n taking the place of λ_n and the sum over the configurations that violate the bound (3.13), again with θ_n replacing λ_n.

Consider the first sum. By Lemma 3.4, the number of terms in this sum is bounded between $2^{nk}(1-2^{k+1}e^{-\frac{\theta_n^2}{2}})$ and 2^{nk}. For all these configurations we know from Lemma 3.4 that $q_{max} \leq 2^k \frac{\theta_n}{\sqrt{n}}$ and hence $|f(\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)})| \leq c2^k\alpha_n^2\frac{\theta_n}{\sqrt{n}} = o(1)$ as $n \to \infty$. Using the fact that $\theta_n \to \infty$ as $n \to \infty$ we obtain that the contribution from the first summation is $1 + o(1)$ as $n \to \infty$.

Thus, to establish Lemma 3.5 it suffices to show that the contribution from the second sum is $o(1)$ as $n \to \infty$. To this end we partition the interval $[\theta_n, n]$ into subintervals $[\kappa_i, \kappa_{i+1}]$ of length one. Now consider the families of configurations $\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}$ that satisfy (3.13) when λ_n is replaced by κ_{i+1} but violate it if λ_n

is replaced by κ_i. It is easy to see that for n large enough, the contribution from these terms is bounded by

$$2^{k+1}e^{-\kappa_i^2/2}e^{c\alpha_n^2 2^k \frac{\kappa_{i+1}}{\sqrt{n}}} = 2^{k+1}e^{-\kappa_i^2/2}e^{o(\kappa_i)} \le e^{-\kappa_i^2/4}.$$

Adding up these error terms, we obtain that the second sum is of order $O(e^{-\theta_n^2/4}) = o(1)$. This completes the proof of Lemma 3.5. □

7.2. **Proof of Lemma 4.1.** The proof of Lemma 4.1 will be based on the following lemma.

Lemma 7.1. *Let k be a positive integer, and let $\beta < \frac{n}{k(k-1)}$ be a non-negative real. If $|g(\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)})| \le \beta q_{\max}^2$ for all families of configurations $\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}$, then*

$$\Big(1 - \frac{k(k-1)}{2}\frac{\beta}{n}\Big) \le 2^{-nk} \sum_{\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}} e^{g(\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)})} \le \Big(1 - k(k-1)\frac{\beta}{n}\Big)^{-1/2} \tag{7.2}$$

Proof. Let $\mathbb{E}_k[\cdot]$ denote expectations with respect to the uniform measure over all families of configurations $\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)} \in \{-1,+1\}^n$. With the help of Jensen's inequality, we then immediately obtain the lower bound

$$\begin{aligned}\mathbb{E}_k[e^g] &\ge \exp(-\beta\mathbb{E}_k[q_{\max}^2]) \ge \exp\Big(-\beta\sum_{i<j}\mathbb{E}_k[q_{ij}^2]\Big)\\ &= \exp\Big(-\beta\frac{k(k-1)}{2}\frac{1}{n}\Big) \ge 1-\beta\frac{k(k-1)}{2}\frac{1}{n}.\end{aligned} \tag{7.3}$$

To obtain an upper bound, we first use Hölder's inequality to obtain the estimate

$$\mathbb{E}_k[e^g] \le \mathbb{E}_k[e^{\sum_{i<j} q_{ij}^2}] \le \prod_{i<j}\mathbb{E}_k[e^{K\beta q_{ij}^2}]^{1/K} \le \max_{i<j}\mathbb{E}_k[e^{K\beta q_{ij}^2}] \tag{7.4}$$

with $K = k(k-1)/2$. Next we rewrite the expectation on the right hand side as

$$\mathbb{E}_k[e^{K\beta q_{ij}^2}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\mathbb{E}_k[e^{\sqrt{2K\beta}q_{ij}x}]dx. \tag{7.5}$$

But now we can calculate the expectation exactly. Together with the inequality $\cosh(y) \le \exp(y^2/2)$, this leads to the estimate

$$\mathbb{E}_k[e^{\sqrt{2K\beta}q_{ij}x}] = \Big(\cosh(\sqrt{2K\beta}xn^{-1})\Big)^n \le \exp\Big(K\beta x^2 n^{-1}\Big).$$

Inserting this into (7.5) and (7.4), we have

$$\mathbb{E}_k[e^g] \le \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(1-2K\beta n^{-1})\frac{x^2}{2}}dx = \frac{1}{\sqrt{1-2K\beta n^{-1}}} = \frac{1}{\sqrt{1-k(k-1)\beta n^{-1}}},$$

the desired upper bound. □

Having established the above lemma, we are now ready to prove Lemma 4.1.

Proof of Lemma 4.1. Let $f_1(\boldsymbol{\sigma}^{(1)},\dots,\boldsymbol{\sigma}^{(k)}) = \sum_{i\neq j}\tilde{q}_{ij}$ and $f = f_1 + R$, and define

$$F(\beta) = \mathbb{E}_k[e^{\beta f}], \quad F_1(\beta) = \mathbb{E}_k[e^{\beta f_1}], \text{ and } \quad F_2(\beta) = \mathbb{E}_k[e^{\beta R}].$$

Since (by (4.10))

$$F_1(\beta_n) = \exp\Big(\frac{k(k-1)}{n}\beta_n^2 + O(\beta_n^3 n^{-2})\Big), \tag{7.6}$$

we only have to prove that $F(\beta_n) = F_1(\beta_n)e^{O(\beta_n^3 n^{-2})+o(1)}$ whenever $\beta_n = o(n)$.

Let $\epsilon_n = 2k(k-1)C\frac{\beta_n}{n}$ where C is the constant from Lemma 4.1. Applying Hölder's inequality to $\mathbb{E}_k[e^f]$ we obtain

$$F(\beta_n) \leq F_1(\beta_n(1+\epsilon_n))^{\frac{1}{1+\epsilon_n}} F_2(\beta_n \frac{1+\epsilon_n}{\epsilon_n})^{\frac{\epsilon_n}{1+\epsilon_n}}$$

and hence

$$\frac{F(\beta_n)}{F_1(\beta_n)} \leq \frac{F_1(\beta_n(1+\epsilon_n))^{\frac{1}{1+\epsilon_n}}}{F_1(\beta_n)} F_2(\beta_n \frac{1+\epsilon}{\epsilon})^{\frac{\epsilon_n}{1+\epsilon_n}}.$$

Using equation (7.6) one obtains

$$\frac{F_1(\beta_n(1+\epsilon_n))^{\frac{1}{1+\epsilon_n}}}{F_1(\beta_n)} = \exp(\beta_n^2 \epsilon_n n^{-1} + O(\beta_n^3/n^2)) = \exp(O(\beta_n^3/n^2)).$$

Recalling that $|R| \leq Cq_{\max}^2$, we would like to use Lemma 7.1 to bound the second factor. To this end, we need to guarantee that $k(k-1)\frac{1+\epsilon_n}{\epsilon_n}C\frac{\beta_n}{n} < 1$. By our choice of ϵ_n this will be the case as soon as n is large enough to ensure that $\epsilon_n < 1$. For n large enough, we therefore have

$$F_2(\beta_n \frac{1+\epsilon}{\epsilon})^{\frac{\epsilon}{1+\epsilon}} \leq \Big(1 - k(k-1)\frac{1+\epsilon_n}{\epsilon_n}C\frac{\beta_n}{n}\Big)^{-\frac{\epsilon_n}{2(1+\epsilon_n)}} = \exp(O(\beta_n/n)).$$

Putting everything together, we get

$$\frac{F(\beta_n)}{F_1(\beta_n)} \leq \exp\Big(O(\beta_n^3/n^2) + O(\beta_n n^{-1})\Big) = \exp\Big(O(\beta_n^3/n^2) + o(1)\Big).$$

Applying Hölder's inequality to $\mathbb{E}_k[e^{(1+\epsilon_n)^{-1}\beta_n(f-R)}]$ we obtain that

$$F_1\Big(\frac{\beta_n}{1+\epsilon_n}\Big)^{1+\epsilon_n} \leq F(\beta_n)F_2\Big(-\frac{\beta_n}{\epsilon_n}\Big)^{\epsilon_n}.$$

This implies that

$$\frac{F(\beta_n)}{F_1(\beta_n)} \geq \frac{F_1((1+\epsilon_n)^{-1}\beta_n)^{1+\epsilon_n}}{F_1(\beta_n)} \frac{1}{F_2(-\beta_n\frac{1}{\epsilon_n})^{\epsilon_n}}.$$

Proceeding along similar lines as for the upper bound, we obtain that

$$\frac{F(\beta_n)}{F_1(\beta_n)} \geq \exp\Big(O(\beta_n^3/n^2) + o(1)\Big).$$

Combining the two applications of Hölder's inequality we finally obtain that

$$\frac{F(\beta_n)}{F_1(\beta_n)} = \exp\Big(O(\beta_n^3/n^2) + o(1)\Big).$$

This completes the proof of the lemma. □

8. Summary and Outlook

8.1. **Summary of Results.** In this paper, we considered the local REM conjecture of Bauke, Franz and Mertens, for both the Npp and the SK spin glass model. For the Npp we showed that the local REM conjecture holds for energy scales α_n of order $o(n^{1/4})$, and fails if α_n grows like $\kappa n^{1/4}$ with $\kappa > 0$. For the SK model we established a similar threshold, showing that the local REM conjecture holds for energies of order $o(n)$, and fails if the energies grow like κn with $\kappa > 0$ sufficiently small.

Although we believe that the local REM conjecture also fails for still faster growing energy scales, our analysis did not allow us to make this rigorous since we could not exclude that the moments of the energy spectrum diverge, while the spectrum itself undergoes a re-entrance transition and converges again to Poisson for faster growing energy scales.

8.2. **Proof Strategy.** Before discussing possible extensions, let us recall our proof strategy. Both the proof of Poisson convergence, and the proof that Poisson convergence fails after a certain point, relied on a precise asymptotic analysis of the factorial moments.

For the purposes of this discussion, let us restrict to the one-dimensional factorial moments $\mathbb{E}[(Z_n(a_n, b_n))_k]$ of the number of points in the energy spectrum between some a_n and b_n in the vicinity of α_n. We expressed these factorial moments in terms of the probability that the energies $E(\boldsymbol{\sigma}^{(1)}), \dots, E(\boldsymbol{\sigma}^{(k)})$ of k pairwise distinct configurations $\boldsymbol{\sigma}^{(1)}, \dots, \boldsymbol{\sigma}^{(k)}$ all lie in the interval $[a_n, b_n]$, see (3.11).

The technical meat of our proof then consisted of two steps: a bound on the sum over "atypical configurations" (see Lemmas 3.4 and 6.1), and a proof that for typical configurations, the probability that the energies of $\boldsymbol{\sigma}^{(1)}, \dots, \boldsymbol{\sigma}^{(k)}$ all lie in the interval $[a, b]$ is asymptotically equal to the product of the probabilities $\mathbb{P}(E(\boldsymbol{\sigma}^{(i)} \in [a_n, b_n]))$.

For both the NPP and the SK model, the value of the threshold can already be understood by considering the factorization properties of typical configurations, see (3.24). Taking, e.g., the case $k = 2$, our results say that for typical configurations in the NPP with Gaussian noise, we have factorization up to a factor

$$e^{\alpha_n^2 (q + O(q^2))}(1 + o(1)), \tag{8.1}$$

where q is the overlap between the two configurations. Since the overlap between typical configurations is of order $n^{-1/2}$, we obtain factorization if and only if $\alpha_n = o(n^{1/4})$. To obtain the same result for the NPP with general distribution was much more work, since it required establishing a large deviation density estimate for k *a priori* highly dependent variables, but the heuristic for the threshold of $n^{1/4}$ is still the same.

By contrast, the threshold for the SK model is easier to establish than that for the NPP. The main reason is the restriction to Gaussian noise, which made the proof of a large deviation density estimate unnecessary, in addition to simplifying the proof of the bound on atypical configurations. But on a heuristic level, there is not much of a difference: now we obtain factorization up to multiplicative factor of

$$e^{\frac{\alpha_n^2}{n} (q^2 + O(q^4))}(1 + o(1)), \tag{8.2}$$

giving the threshold of $\alpha_n = o(n)$ for Poisson convergence.

8.3. **p-Spin models.** In the physics literature, one often considers a generalization of the SK model which involves interactions between p different spins, instead of the two-body interaction of the standard SK model. For our purpose, these p-spin SK models are best defined as Gaussian fields indexed by the spin configurations $\boldsymbol{\sigma} \in \{-1, +1\}^n$. Recalling that a Gaussian field is uniquely defined by its mean and covariance matrix, we then define the p-spin SK-Hamiltonian $H^{(p)}(\boldsymbol{\sigma})$ as the Gaussian field with mean 0 and covariance matrix $\mathbb{E}[H^{(p)}(\boldsymbol{\sigma}) H^{(p)}(\tilde{\boldsymbol{\sigma}})] = nq(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})^p$ where $p = 1, 2, \dots,$ and $q(\cdot, \cdot)$ is the overlap defined in (1.2). Note that the energy of the NPP with Gaussian noise is nothing but the absolute value of the $p = 1$ SK Hamiltonian divided by $\sqrt{n}$. Up to a rescaling by $\sqrt{n}$, the energy spectrum of the

NPP with Gaussian noise is therefore identical to the positive energy spectrum of the $p = 1$ SK model.

It was shown in [BK05a] that the local REM conjecture holds for $p = 1$ if $\alpha_n = O(n^{3/4-\epsilon})$ for some $\epsilon > 0$, and for $p \geq 2$ if $\alpha_n = O(n^{1-\epsilon})$. For $p = 1, 2$ our results establish a little bit more, namely the validity of the REM conjecture for $\alpha_n = o(n^{3/4})$ if $p = 1$, and for $\alpha_n = o(n)$ if $p = 2$. More importantly, our results prove that these are actually the thresholds for the validity of the REM conjecture.

This raises the question about the true threshold for $p \geq 3$. Starting with the factorization properties for typical configurations, one easily sees that the joint density of $H(\boldsymbol{\sigma}^{(1)}), \ldots, H(\boldsymbol{\sigma}^{(k)})$ is again given by a formula of the form (3.32), where C is now the matrix with matrix element $C_{ij} = n(q(\boldsymbol{\sigma}^{(i)}, \boldsymbol{\sigma}^{(j)}))^p$. This then leads to factorization up to a multiplicative error term

$$e^{\frac{\alpha_n^2}{n} O(q_{\max}^p)}(1 + o(1)), \tag{8.3}$$

where $q_{\max}$ is the maximal off-diagonal overlap of $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(k)}$. If $\alpha_n = O(n)$, this gives factorization for typical configurations as long as $p > 2$. But unfortunately, our control over atypical configurations is not good enough to allow a proof of the REM conjecture for energies that grow that fast. Indeed, it is easy to see that Lemma 6.1 can be generalized to $p > 2$; but an application of the lemma to control the error for the k^{th} factorial moment requires the condition $\alpha_n \leq n\epsilon_k$ where ϵ_k is a positive constant that goes to zero as $k \to \infty$.

We therefore see that our methods can easily be used to prove the local REM conjecture for $\alpha_n = o(n^{3/4})$ and $p = 1$ as well as $\alpha_n = o(n)$ and all $p \geq 2$, but they are not strong enough to answer the question whether this is the actual threshold for $p \geq 3$, or whether the REM conjecture remains valid if $p \geq 3$ and α_n grows like $n\kappa$ for some small $\kappa > 0$.

An even more challenging question is the question of what happens after the local REM conjecture fails. This question seems quite hard, and so far it has only been answered for the hierarchical GREM-model. For this model it has been shown [BK05b, BK05c] that the suitably rescaled energy spectrum converges to a mixed Poisson process with density given in terms of a Poisson cascades on $\mathbb{R}^\ell$, where the dimension ℓ becomes larger and larger as κ passes through an infinite series of thresholds, the first threshold being the value where the local REM conjecture fails for the GREM.

8.4. **A Simple Heuristic.** Returning now to the NPP, we note that the rigorous moment analysis of the threshold is somewhat unintuitive and rather involved. Alternatively, let us present a simple heuristic to explain the threshold scale $n^{1/4}$. To this end it is useful to introduce the gauge invariant magnetization

$$M(\boldsymbol{\sigma}) = \frac{1}{n} \sum \sigma_i \text{sgn} X_i,$$

and consider the joint distribution of $\widetilde{M}(\boldsymbol{\sigma}) = \sqrt{n} M(\boldsymbol{\sigma})$ and the "signed" energy $H(\boldsymbol{\sigma})$ introduced in (3.7) with, as usual, $X_1, \ldots, X_n$ chosen i.i.d. with density ρ, and $\boldsymbol{\sigma}$ chosen uniformly at random.

The covariance matrix C of $\widetilde{M}(\boldsymbol{\sigma})$ and $H(\boldsymbol{\sigma})$ is easily calculated to be

$$C = \begin{bmatrix} 1 & \mu \\ \mu & 1 \end{bmatrix}$$

where $\mu = \mathbb{E}[|X_1|] < 1$, and its inverse C^{-1} is given by

$$C^{-1} = \frac{1}{1-\mu^2}\begin{bmatrix} 1 & -\mu \\ -\mu & 1 \end{bmatrix}.$$

Assuming, for the moment, that the joint density of $\widetilde{M}(\boldsymbol{\sigma})$ and $H(\boldsymbol{\sigma})$ obeys a local limit theorem, we now approximate this density by the Gaussian density

$$\begin{aligned} g(H,\widetilde{M}) &= \frac{1}{2\pi\sqrt{\det C}} e^{-\frac{1}{2}(H^2 C_{11}^{-1} + \widetilde{M}^2 C_{22}^{-1} + 2H\widetilde{M}C_{12})} \\ &= \frac{1}{2\pi\sqrt{\det C}} e^{-\frac{1}{2(1-\mu^2)}(H^2+\widetilde{M}^2-\mu H\widetilde{M})} \\ &= \frac{1}{2\pi\sqrt{\det C}} e^{-\frac{1}{2}H^2} e^{-\frac{1}{2(1-\mu^2)}(\widetilde{M}-\mu H)^2}. \end{aligned} \tag{8.4}$$

In this approximation, the distribution of $\widetilde{M}$ conditioned on $H = \alpha$ is therefore Gaussian with mean $\alpha\mu$ and covariance $1-\mu^2$, implying in particular that the expectation of M is equal to $\alpha\mu/\sqrt{n}$.

Consider now two configurations $\boldsymbol{\sigma}, \boldsymbol{\sigma'}$, both chosen uniformly at random among all configurations with magnetization M. Then the expected overlap between $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma'}$ is $M^2(1+o(1))$.

Finally, consider two configurations $\boldsymbol{\sigma}, \boldsymbol{\sigma'}$ both chosen uniformly at random among all configurations with signed energies in the range $H \in [\alpha, \alpha + d\alpha]$, for some small $d\alpha$. Since the energies of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma'}$ are correlated, it follows that their magnetizations are also correlated. Under the assumptions that (1) the joint distribution $g(H, \widetilde{M})$ obeys a local limit theorem, as in (8.4), and (2) the *only* correlation between these configurations is due to the correlation in their magnetizations, it would follow that the expected overlap $\mathbb{E}[q(\boldsymbol{\sigma}, \boldsymbol{\sigma'})]$ is given by $M^2(1+o(1)) = \alpha^2\mu^2/n(1+o(1))$.

Note that the above heuristic focuses on the "signed energy" $H(\boldsymbol{\sigma})$ rather than the true energy $E(\boldsymbol{\sigma}) = |H(\boldsymbol{\sigma})|$. Conditioning instead on $E(\boldsymbol{\sigma}), E(\boldsymbol{\sigma'}) \in [\alpha, \alpha+d\alpha]$, the above heuristic therefore suggests a bimodal distribution of $q(\boldsymbol{\sigma}, \boldsymbol{\sigma'})$ with peaks at $\pm\alpha^2\mu^2/n(1+o(1))$.

Now recall that the local REM conjecture says that the overlap, rescaled by $\sqrt{n}$, is asymptotically normal. However, our above heuristic says that $\sqrt{n}q(\boldsymbol{\sigma}, \boldsymbol{\sigma'})$ cannot be asymptotically normal for α growing like $n^{1/4}$ or faster, in agreement with our rigorous results. Thus this heuristic correctly predicts the scale at which the REM conjecture breaks down, and suggests that the breakdown could be due to correlations in the magnetization of configurations with similar energies of scale $n^{1/4}$ or greater.

A detailed calculation, however, suggests that we exercise some caution in the application of this heuristic. Although the heuristic correctly predicts the scale of the threshold, and the double peak structure of the overlap at the threshold, it does not predict the correct position of the peaks when $\alpha = \kappa n^{1/4}$. Indeed, for this scaling, we rigorously showed that the rescaled overlap distribution converges to a convex combination of two Gaussians centered at $\pm\kappa^2$, in contrast to the heuristic prediction of $\pm\kappa^2\mu^2$. This, in turn, suggests that there are additional correlations besides those induced by the magnetization.

8.5. **Algorithmic Consequences.** Over twenty years ago, Karmarkar and Karp [KK82] gave a linear time algorithm for a suboptimal solution of the number partitioning problem. For i.i.d. weights with densities of the form studied in the current paper, the typical energy E_{KK} of the KK solution is of order $n^{-\theta(\log n)}$, [KK82, Yak96], while the minimal energy is known to be much smaller [KKLO86, Lue98, BCP01], namely of order $2^{-\theta(n)}$.

This raises the question of whether one can do better than the KK solution – a question which has received much study due to the numerous applications of the number partitioning problem. This work has led to many different heuristics, but to our knowledge no algorithm with *guaranteed* performance significantly better than KK has emerged. In the absence of a good global alternative to KK, one might try to base an improved solution of the problem on a local search algorithm starting from the KK solution. But such an approach is unlikely to produce better results, as the following argument shows.

Consider the random NPP as defined in this paper, and let $\boldsymbol{\sigma}$ be a partition with energy $E(\boldsymbol{\sigma})$ of the order of E_{KK}, i.e., $E(\boldsymbol{\sigma}) = n^{-\theta(\log n)}$. Let $\tilde{\boldsymbol{\sigma}}$ be a local perturbation of $\boldsymbol{\sigma}$, i.e., let $\tilde{\boldsymbol{\sigma}}$ be a configuration such that $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ differ on a small subset $K \subset \{1, \ldots, n\}$, with $k = |K|$ bounded uniformly in n. The signed energies $H(\boldsymbol{\sigma})$ and $H(\tilde{\boldsymbol{\sigma}})$ then differ by $n^{-1/2}\Delta_K(\boldsymbol{\sigma}_K)$, where $\Delta_K(\boldsymbol{\sigma}_K)$ is the random variable

$$\Delta_K(\boldsymbol{\sigma}_K) = 2\sum_{i\in K} \sigma_i X_i$$

and $\boldsymbol{\sigma}_K$ is the restriction of $\boldsymbol{\sigma}$ to K. Under mild assumptions on the probability density ρ of the weights $X_1, \ldots, X_n$, the density of $\Delta_K(\boldsymbol{\sigma}_K)$ is a continuous function near zero, and the probability that $|\Delta_K| \le \epsilon$ for some small ϵ is of order $\theta(\epsilon)$ with the constants implicit in the θ symbol depending only on k.

Obviously, any local improvement algorithm that changes exactly k bits will lead to a change in the unsigned energies that is bounded from below by

$$\delta_k^{(-)} = n^{-1/2} \min_{K:|K|=k} \min_{\boldsymbol{\sigma}_K} |\Delta_K(\boldsymbol{\sigma}_K)|.$$

Taking into account that there are only $\binom{n}{k} \le n^k$ possible choices for K, we see that the probability that $\delta_k^{(-)} \le \epsilon$ is bounded by $O(\epsilon n^{k+1/2})$. We conclude that with high probability, $\delta_k^{(-)}$ is at least $\theta(n^{-1/2-k})$, much larger than the energy $n^{-\theta(\log n)}$ of our starting configuration $\boldsymbol{\sigma}$. Thus any local improvement algorithm that changes k bits moves us with high probability far away from the starting configuration with energy of order $n^{-\theta(\log n)}$.

Note that this simple argument does not use very much; in particular, it is not related to REM conjecture, which suggests a much deeper reason for the apparent difficulty of the NPP. Indeed, applying our local REM Theorem to $\alpha = E_{KK}$, it says that in the vicinity of E_{KK}, the energy behaves like a random cost function of 2^{n-1} independent random variables. If the energy of the NPP were truly a random cost function of 2^{n-1} independent random variables, this would imply that there is no algorithm faster than exhaustive search, implying a running time exponential in n. But of course, we have not come near to proving anything as strong as this.

In fact, even on a non-rigorous level some caution is required when applying the above heuristic. Indeed, only n linear independent configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(n)}$ are needed to completely determine the random variables $X_1, \ldots, X_n$ from the energies $E(\boldsymbol{\sigma}^{(1)}), \ldots, E(\boldsymbol{\sigma}^{(n)})$, implying that the energy spectrum lies in a subspace of dimension n, not 2^n. Still, our REM Theorem proves that the energy spectrum behaves locally like that of a random cost function, suggesting several possible conjectures.

The first conjecture is best described in an oracle setting, where the oracle, O, keeps the n weights $X_1, \ldots, X_n$ secret from the algorithm A. The algorithm is given the KK-solution $\boldsymbol{\sigma}^{(KK)}$ and its energy E_{KK}, and successively asks the oracle for the energy of some configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(m)}$, where m is bounded. Given this information, the algorithm then calculates a new approximation $\tilde{\boldsymbol{\sigma}} \neq \pm\boldsymbol{\sigma}^{(KK)}$. Given our REM Theorem, we conjecture that with high probability (tending to one as $n \to \infty$), the energy of the new configuration $\tilde{\boldsymbol{\sigma}}$ is much larger than E_{KK}; in fact, we conjecture that with high probability $E(\tilde{\boldsymbol{\sigma}})/E_{KK} \to \infty$ as $n \to \infty$.

The second conjecture gives the algorithm much more input, namely the ℓ first energies above a threshold α and the configurations corresponding to these energies. The task of the algorithm is now to find a new partition $\tilde{\boldsymbol{\sigma}}$ whose energy is as near possible to α, which we assume to grow only slowly with n (say like $o(n^{-1/4})$, to stay in the realm of our Theorem 2.1). Again, the algorithm has no access to the original weights, but may ask the oracle O for the energies of m additional configurations $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(m)}$, adapting the next question to the answer of the preceding ones. For m and ℓ bounded uniformly in n, we then conjecture that with high probability the algorithm produces a configuration $\tilde{\boldsymbol{\sigma}}$ with $(E(\tilde{\boldsymbol{\sigma}}) - \alpha)\xi_n^{-1} \to \infty$, while the actual value of the next configuration above α is with high probability equal to $\alpha + O(\xi_n)$ by our Theorem 2.1.

One finally might want to consider the above conjectures in a setting where the oracle only reveals the relative order of the energies $E(\boldsymbol{\sigma}^{(1)}), \ldots, E(\boldsymbol{\sigma}^{(m)})$, but keeps the numerical values of $E(\boldsymbol{\sigma}^{(1)}), \ldots, E(\boldsymbol{\sigma}^{(m)})$ secret. In such a setting, there is no *a priori* reason to assume that $m < n$. Instead, it seems reasonable to conjecture inapproximability results for values of m and ℓ that are polynomial in n.

Acknowledgement: We thank Keith Ball, Amir Dembo, Laci Lovasz, Yuval Perez, David Wilson and Avi Widgerson for useful discussions. S.M. was supported in part by the German Science Council (grant ME2044/1-1).

References

[BCP01] C. Borgs, J.T Chayes, and B. Pittel, *Phase transition and finite-size scaling for the integer partitioning problem*, Rand. Struct. Alg. **19** (2001), no. 3–4, 247–288.

[BCMN05] C. Borgs, J.T. Chayes, S. Mertens and C. Nair, *Proof of the local REM conjecture for number partitioning I: constant energy scales*, arXiv.org/cond-mat/0501760.

[BFM04] H. Bauke, S. Franz, and S. Mertens, *Number partitioning as random energy model*, J. Stat. Mech.: Theor. Exp. (2004), P04003.

[BM04] H. Bauke and S. Mertens, *Universality in the level statistics of disordered systems*, Phys. Rev. E **70** (2004), 025102(R).

[Bi94] P. Billingsley, *Probability and Measure, 3^{rd} ed.*, Wiley series in Probability and Mathematical Statistics, 1994.

[BK05a] A. Bovier and I. Kurkova, *Local energy statistics in disorderd systems: a proof of the local REM conjecture*, Preprint, 2005.

[BK05b] A. Bovier and I. Kurkova, *A tomography of the GREM: beyond the REM conjecture*, Preprint, 2005.

[BK05c] A. Bovier and I. Kurkova, *Energy statistics in disordered systems: beyond the REM conjecture*, In: "Applications of random matrices in economics and other complex systems", Krakow, 2005, to appear.

[Der81] B. Derrida, *Random-energy model: An exactly solvable model of disordered systems*, Phys. Rev. B **24** (1981), no. 5, 2613–2626.

[Der85] B. Derrida, *A generalization of the random-energy model that includes correlations between energies*, J. Rev. Lett. **46** (1985), 401–407.

[Dan54] H.E. Daniels, *Saddlepoint approximations in statistics*, Ann. Math. Statist. **25** (1954) 631–650.

[Mer00] S. Mertens, *Random costs in combinatorial optimization*, Phys. Rev. Lett. **84** (2000), no. 7, 1347–1350.

[KK82] N. Karmarkar, R.M. Karp, *The differencing method of set partitioning* Technical Report UCB/CSD 82/113, Computer Science Division (EECS), University of California, Berkeley (1982)

[KKLO86] N. Karmarkar, R.M. Karp, G.S. Lueker and A.M. Odlyzko, *Probabilistic anaysis of optimum partitioning*, J. Appl. Prob. **23** (1986) 626–645.

[Lue98] G.S. Lueker, *Exponentially small bounds on the expected optimum of the partition and subset sum problem*, Rand. Struc. Alg. **12** (1998) 51–62.

[Yak96] B. Yakir, *The differencing algorithm LDM for partitioning: a proof of a conjecture of Karmakar and karp*, Math. of Operations Res., **21** (1996) 85–99

[1]Microsoft Research, One Microsoft Way, Redmond, WA 98052

[2]Inst. f. Theor. Physik, Otto-von-Guericke Universität, PF 4120, 39016 Magdeburg, Germany