A NEW REM CONJECTURE

G. BEN AROUS, VÉRONIQUE GAYRARD, AND A. KUPTSOV

Abstract. We introduce here a new universality conjecture for levels of random Hamiltonians, in the same spirit as the local REM conjecture made by S. Mertens and H. Bauke. We establish our conjecture for a wide class of Gaussian and non-Gaussian Hamiltonians, which include the p-spin models, the Sherrington-Kirkpatrick model and the number partitioning problem. We prove that our universality result is optimal for the last two models by showing when this universality breaks down.

1. Introduction

S. Mertens and H. Bauke recently observed ([Mer00], [BM04], see also [BFM04]) that the statistics of energy levels for very general random Hamiltonians are Poissonian, when observed micro-canonically, i.e. in a small window in the bulk. They are universal and identical to those of the simplest spin-glass model, the Random Energy Model or REM, hence the name of this (numerical) observation: the REM conjecture or more precisely the *local REM conjecture*.

This local REM conjecture was made for a wide class of random Hamiltonians of statistical mechanics of disordered systems, mainly spin-glasses (mean field or not), and for various combinatorial optimization problems (like number partitioning). Recently, two groups of mathematicians have established this conjecture in different contexts : C. Borgs, J. Chayes, C.Nair, S.Mertens, B.Pittel for the number partitioning question (see [BCP01], [BCMN05a],[BCMN05b]), and A. Bovier, I. Kurkova for general spin-glass Hamiltonians (see [BK06a], [BK06b]).

We introduce here a new kind of universality for the energy levels of disordered systems. We believe that one should find universal statistics for the energy levels of a wide class of random Hamiltonians if one re-samples the energy levels, i.e. draws a random subset of these energies. Put otherwise, our conjecture is thus that the level statistics should also be universal, i.e Poissonian, when observed on a large random subset of the configuration space rather than in a micro-canonical window. We establish this new universality result (which could be called *the re-sampling REM universality* or *the REM universality by dilution*) for general (mean-field) spin-glass models, including the case of number partitioning and for large but sparse enough subsets. This approach has the following interesting property: the range of energies involved is not reduced to a small window as in the local REM conjecture. Thus we can study the extreme value distribution on the random subset, by normalizing the energies properly. Doing so we establish that the Gibbs measure restricted to a sparse enough random subset of configuration space has a universal distribution which is thus the same as for the REM, i.e. a Poisson-Dirichlet measure.

Date: May 11, 2007.
2000 *Mathematics Subject Classification.* 82B44, 60F99.

To be more specific, we specialize our setting to the case of random Hamiltonians $(H_N(\sigma))_{\sigma\in S_N}$ defined on the hypercube $S_N = \{-1,1\}^N$. We want to consider a sparse random subset of the hypercube, say X, and the restriction of the function H_N to X. We introduce the random point process

$$\mathcal{P}_N = \sum_{\sigma\in X} \delta_{H'_N(\sigma)}, \tag{1.1}$$

with a normalization:

$$H'_N(\sigma) = \frac{H_N(\sigma) - a_N}{b_N} \tag{1.2}$$

to be chosen. Our conjecture specializes to the following: the asymptotic behavior of the random point process $\mathcal{P}_N$ is universal, for a large class of Hamiltonians H_N, for appropriate sparse random subsets X, and appropriate normalization.

We will only study here the simplest possible random subset, i.e. a site percolation cluster $X = \{\sigma \in S_N : X_\sigma = 1\}$, where the random variables $(X_\sigma)_{\sigma\in S_N}$ are i.i.d. and Bernoulli:

$$\mathrm{P}(X_\sigma = 1) = 1 - \mathrm{P}(X_\sigma = 0) =: p_N = \frac{2^M}{2^N}. \tag{1.3}$$

Thus the mean size of X is 2^M and we will always assume that X is not too small, e.g. that $\log N = o(2^M)$. We will sometimes call X a *random cloud*.

In order to understand what the universal behavior should be, let us examine the trivial case where the $H_N(\sigma)$ are i.i.d centered standard Gaussian random variables, i.e. the case of the Random Energy Model. Then, standard extreme value theory proves that if

$$a_N = \sqrt{2M\log 2 + 2\log b_N - \log 2} \text{ and } b_N = \sqrt{\frac{1}{M}}, \tag{1.4}$$

then $\mathcal{P}_N$ converges to a Poisson point process with intensity measure

$$\mu(dt) = \frac{1}{\sqrt{\pi}} e^{-t\sqrt{2\log 2}} dt. \tag{1.5}$$

We will now fix the normalization needed in (1.2) by choosing a_N and b_N as in (1.4). The basic mechanism of the REM universality we propose is that the influence of correlations between the random variables $H'_N(\sigma)$ should be negligible when the two-point correlation (the covariance) is a decreasing function of the Hamming distance

$$d_H(\sigma,\sigma') = \#\{i \le N : \sigma_i \ne \sigma'_i\} \tag{1.6}$$

and when the random cloud is sparse enough. We first establish this universality conjecture for a large class of Gaussian Hamiltonians. This class contains the Sherrington-Kirkpatrick (SK) model as well as the more general p-spin models. It also contains the Gaussian version of the number partitioning problem.

Consider a Gaussian Hamiltonian H_N on the hypercube $S_N = \{-1,1\}^N$ such that the random variables $(H_N(\sigma))_{\sigma\in S_N}$ are centered and whose covariance is a smooth decreasing function of the Hamming distance or, equivalently a smooth increasing function ν of the overlap $R(\sigma,\sigma') = \frac{1}{N}\sum_{i=1}^N \sigma_i\sigma'_i$:

$$\mathrm{cov}(H_N(\sigma), H_N(\sigma')) = \nu(R(\sigma,\sigma')) = \nu\Big(1 - \frac{2d_H(\sigma,\sigma')}{N}\Big). \tag{1.7}$$

We will always assume that $\nu(0) = 0$ and that $\nu(1) = 1$. The first assumption is crucial, since it means that the correlation of the Hamiltonian vanishes for pairs of points on the hypercube which are at a typical distance. The second assumption simply normalizes the variance of $H_N(\sigma)$ to 1.

This type of covariance structure can easily be realized for ν real analytic, of the form:

$$\nu(r) = \sum_{p\geq 1} a_p^2 r^p. \tag{1.8}$$

Indeed such a covariance structure can be realized by taking mixtures of p-spin models. Let $H_{N,p}$ be the Hamiltonian of the p-spin model given by

$$H_{N,p}(\sigma) = \frac{1}{N^{(p-1)/2}} \sum_{1\leq i_1,i_2,\ldots,i_p\leq N} g_{i_1,i_2,\ldots,i_p}\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_p}, \tag{1.9}$$

where random variables $(g_{i_1,i_2,\ldots,i_p})_{1\leq i_1,i_2,\ldots,i_p\leq N}$ are independent standard Gaussians defined on a common probability space $(\Omega^g, \mathcal{F}^g, \mathbb{P})$. Then

$$H_N(\sigma) = \frac{1}{\sqrt{N}} \sum_{p\geq 1} a_p H_{N,p}(\sigma) \tag{1.10}$$

has the covariance structure given in (1.7)-(1.8) Let us recall that the case where $\nu(r) = r$ is the Gaussian version of the number partitioning problem ([BCP01]), the case where $\nu(r) = r^2$ is the SK model, and more generally, $\nu(r) = r^p$ defines the pure p-spin model.
Let us normalize H_N as above (see (1.4)):

$$H'_N(\sigma) = \frac{H_N(\sigma) - a_N}{b_N}, \tag{1.11}$$

and consider the sequence of point processes

$$\mathcal{P}_N = \sum_{\sigma\in X} \delta_{H'_N(\sigma)}. \tag{1.12}$$

Theorem 1.1 (Universality in the Gaussian case)**.** *Assume that $M = o(\sqrt{N})$ if $\nu'(0) \neq 0$, and that $M = o(N)$ if $\nu'(0) = 0$. Then,* P*-almost surely, the distribution of the point process $\mathcal{P}_N$ converges weakly to the distribution of a Poisson point process $\mathcal{P}$ on $\mathbb{R}$ with intensity given by*

$$\mu(dt) = \frac{1}{\sqrt{\pi}} e^{-t\sqrt{2\log 2}} dt. \tag{1.13}$$

Remark 1.2. The condition $\log N = o(2^M)$ is needed in order to get P-almost sure results.

We extend this result, in Section 5, to a wide class of non-Gaussian Hamiltonians (introduced in [BCMN05a] for the case of number partitioning).
The theorem has the following immediate corollary. Let us fix the realization of the random cloud X. For configurations σ belonging to the cloud we consider the Gibbs' weights $G_{N,\beta}(\sigma)$ of the re-scaled Hamiltonian $H'_N(\sigma)$

$$G_{N,\beta}(\sigma) = \frac{e^{-\beta H'_N(\sigma)}}{\sum_{\varrho\in X} e^{-\beta H'_N(\varrho)}} = \frac{e^{-\beta\sqrt{M} H_N(\sigma)}}{\sum_{\varrho\in X} e^{-\beta\sqrt{M} H_N(\varrho)}}. \tag{1.14}$$

Reordering the Gibbs' weights $G_{N,\beta}(\sigma)$ of the configurations $\sigma \in X$ as a non-increasing sequence $(w_\alpha)_{\alpha \leq |X|}$ and defining $w_\alpha = 0$ for $\alpha > |X|$ we get a random element w of the space $\mathcal{S}$ of non increasing sequences of non negative real numbers with sum less than one.

Corollary 1.3 (Convergence to Poisson-Dirichlet)**.** *If $\beta > \sqrt{2\log 2}$ then* P*-almost surely under the assumptions of Theorem 1.1 the law of the sequence $w = (w_\alpha)_{\alpha \geq 1}$ converges to the Poisson-Dirichlet distribution with parameter $m = \frac{\sqrt{2\log 2}}{\beta}$ on $\mathcal{S}$.*

The fact that Theorem 1.1 implies Corollary 1.3 is well-known, see for instance [Tal03] (pp.13-19) for a good exposition.

It is then a natural question to know if our sparseness assumption is optimal. When the random cloud is denser can this universality survive? We show that our sparseness condition is indeed optimal for the number partitioning problem and for the SK model, and that the universality does break down.

Theorem 1.4. *[Breakdown of Universality for the number partitioning problem]*
(i) Let $\nu(r) = r$. Suppose that $\limsup \frac{M(N)}{\sqrt{N}} < \infty$. Then P*-almost surely, the distribution of the point process $\mathcal{P}_N$ converges to the distribution of a Poisson point process if and only if $M = o(\sqrt{N})$.*

[Breakdown of Universality for the Sherrington-Kirkpatrick model]
(ii) Let $\nu(r) = r^2$. Suppose that $\limsup \frac{M(N)}{N} < \frac{1}{8\log 2}$. Then P*-almost surely, the distribution of the point process $\mathcal{P}_N$ converges to the distribution of a Poisson point process if and only if $M = o(N)$.*

We prove this theorem in Section 4 by showing that the second factorial moment of the point process does not converge to the proper value. The case of pure p-spin models, with $p \geq 3$, or more generally the case where $\nu'(0) = \nu''(0) = 0$ differs strongly (see Theorem 4.7). The asymptotic behavior of the first three moments is compatible with a Poissonian convergence. Proving or disproving Poissonian convergence (or REM universality) in this case is still open, as it is for the local REM conjecture.

The paper is organized as follows. In Section 2 we establish important combinatorial estimates about maximal overlaps of ℓ-tuples of points on the random cloud. We give a particular care to the case of pairs ($\ell = 2$) and triples ($\ell = 3$) which are important for the breakdown of universality results. In Section 3 we establish the universality in the Gaussian case (Theorem 1.1). We then prove, in Section 4, the breakdown of universality given in Theorem 1.4. Finally we extend the former results to a wide non-Gaussian setting in Section 5.

2. Combinatorial estimates

In this section we fix an integer $\ell \geq 1$ and study the maximal overlap

$$R_{\max}(\sigma^1, \ldots, \sigma^\ell) = \max_{1 \leq i < j \leq \ell} |R(\sigma^i, \sigma^j)|. \tag{2.1}$$

For fixed $N \in \mathbb{N}$ and $R \in [0, 1)$ let us define the following subsets of S_N^ℓ

$$U_\ell(R) = \{(\sigma^1, \ldots, \sigma^\ell) : R_{\max}(\sigma^1, \ldots, \sigma^\ell) \leq R\} \tag{2.2}$$

and

$$V_\ell(R) = \{(\sigma^1, \ldots, \sigma^\ell) : R_{\max}(\sigma^1, \ldots, \sigma^\ell) = R\}. \tag{2.3}$$

More generally, let the sequence $R_N \in [0,1)$ be given (respectively, the corresponding sequence of Hamming distances $d_N = \frac{N}{2}(1-R_N)$) and introduce the sequence of sets $U_\ell(R_N)$ and $V_\ell(R_N)$ which we denote for simplicity of notation by $U_{N,\ell}$ and $V_{N,\ell}$ respectively.

For a set $Y \subset S_N^\ell$ we denote by Y^X its intersection with X^ℓ. In the following theorem we study the properties of the sets

$$U_{N,\ell}^X = \{(\sigma^1, \ldots, \sigma^\ell) \in X^\ell : R_{\max}(\sigma^1, \ldots, \sigma^\ell) \leq R_N\} \tag{2.4}$$

and

$$V_{N,\ell}^X = \{(\sigma^1, \ldots, \sigma^\ell) \in X^\ell : R_{\max}(\sigma^1, \ldots, \sigma^\ell) = R_N\}. \tag{2.5}$$

In order to state the main result of this section (Theorem 2.1) we define the function

$$\mathcal{J}(x) = \begin{cases} \frac{1-x}{2}\log(1-x) + \frac{1+x}{2}\log(1+x) & \text{if } x \in [-1,1], \\ +\infty & \text{otherwise.} \end{cases} \tag{2.6}$$

Theorem 2.1. *Let the sequence R_N be such that $NR_N^2 \to \infty$.*

(i) Then P*-almost surely*

$$|U_{N,\ell}^X| = \mathrm{E}|U_{N,\ell}^X|(1+o(1)). \tag{2.7}$$

(ii) If $R_N = o(1)$ and $M(N) \geq \log N$ then there exists $\alpha \in [0,1)$ and $C > 0$, depending only on ℓ, such that P*-almost surely*

$$|V_{N,\ell}^X| \leq C\, e^{\alpha N \mathcal{J}(R_N)} \mathrm{E}|V_{N,\ell}^X|. \tag{2.8}$$

Proof. The proof is based on standard inequalities for i.i.d. random variables that result from exponential Chebychev inequality. We formulate them without proof: let $(X_i)_{1\leq i\leq n}$ be i.i.d. Bernoulli rv's with $\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = p$ and let $Z = \sum_{i=1}^n X_i$. Then, for $t > 0$,

$$\mathbb{P}(Z - \mathbb{E}Z \geq t\mathbb{E}Z) \leq e^{-n\left(p(1+t)\log(1+t) + (1-p(1+t))\log\frac{1-p(1+t)}{1-p}\right)}, \tag{2.9}$$

$$\mathbb{P}(Z - \mathbb{E}Z \leq -t\mathbb{E}Z) \leq e^{-n\left(p(1-t)\log(1-t) + (1-p(1-t))\log\frac{1-p(1-t)}{1-p}\right)}. \tag{2.10}$$

If $p = p(n) \to 0$ and $t = t(n) \to 0$ as $n \to \infty$, the above inequalities imply that, for large enough n,

$$\mathbb{P}\big(|Z - \mathbb{E}Z| \geq t\mathbb{E}Z\big) \leq 2e^{-npt^2/4}, \tag{2.11}$$

whereas if $p(n) \to 0, t(n) \to \infty$, and $p(n)t(n) \to 0$ we get from (2.9) that for large enough n

$$\mathbb{P}(Z - \mathbb{E}Z \geq t\mathbb{E}Z) \leq e^{-\frac{1}{2}np(1+t)\log(1+t)}. \tag{2.12}$$

The proof of Theorem 2.1 relies on the following elementary lemma that again we state without proof.

Lemma 2.2. *(i) For any sequence $R_N \in [0,1)$*

$$|U_{N,\ell}| \geq 2^{N\ell}\left(1 - 2\binom{\ell}{2}e^{-\frac{1}{8}NR_N^2}\right). \tag{2.13}$$

(ii) Suppose R_N satisfies $NR_N^2 \to \infty$ and $R_N = o(1)$. Then for some $C > 0$ depending only on ℓ,

$$|V_{N,\ell}| = 2^{N\ell}\frac{C}{\sqrt{N}}e^{-N\mathcal{J}(R_N)}(1+o(1)). \tag{2.14}$$

As an elementary consequence of part (i) of Lemma 2.2 one can prove that:

Corollary 2.3. P-*a.s.* $\max_{\sigma,\sigma'\in X}|R(\sigma,\sigma')| \le \delta_N$, *where* $\delta_N \equiv 4\sqrt{\frac{M(N)\log 2}{N}+\frac{\log N}{N}}$.

The proof of part (i) of Theorem 2.1 then proceeds as follows. Let us first express the size of the random cloud $|X|$ as a sum of i.i.d. random variables

$$|X| = \sum_{\sigma\in S_N}\mathbf{1}_{X_\sigma=1}. \tag{2.15}$$

Using (2.11) and the assumption that $\log N = o(2^M)$, we see that P-almost surely $|X|$ is given by its expected value, i.e. $|X| = 2^M(1+o(1))$. Therefore $|X^\ell| = 2^{M\ell}(1+o(1))$.

Since $U^X_{N,\ell}\subset X^\ell$ and $\mathrm{E}|U^X_{N,\ell}| = p_N^\ell|U_{N,\ell}| = 2^{M\ell}(1+o(1))$, then proving part (i) of Theorem 2.1 is equivalent to proving that the set $U^X_{N,\ell}$ coincides, up to an error of magnitude $o(2^{M\ell})$, with the set X^ℓ. Let us rewrite $U^X_{N,\ell}$ as

$$U^X_{N,\ell} = \bigcap_{1\le j<j'\le\ell}(U^X_{N,\ell})_{jj'}, \tag{2.16}$$

where we defined

$$(U^X_N)_{jj'} = \Big\{(\sigma^1,\dots,\sigma^\ell)\in X^\ell:\ |R(\sigma^j,\sigma^{j'})|\le R_N\Big\}. \tag{2.17}$$

If we prove that every set $(U^X_{N,\ell})_{jj'}$ coincides, up to an error of order $o(2^{M\ell})$, with the set X^ℓ, then the representation (2.16) implies part (i) of the theorem. We therefore concentrate on proving that

$$|(U^X_{N,\ell})_{jj'}| = \mathrm{E}|(U_{N,\ell})^X_{jj'}|(1+o(1)),\quad \text{P-a.s.} \tag{2.18}$$

Without loss of generality we can consider the case of $j=1$ and $j'=2$. By definition of $(U^X_{N,\ell})_{jj'}$ we get

$$\begin{aligned}|(U^X_{N,\ell})_{12}| &= \Bigg(\sum_{\substack{\sigma^1,\sigma^2\in S_N:\\ |R(\sigma^1,\sigma^2)|\le R_N}}\mathbf{1}_{X_{\sigma^1}=1}\mathbf{1}_{X_{\sigma^2}=1}\Bigg)\Bigg(\sum_{\sigma\in S_N}\mathbf{1}_{X_\sigma=1}\Bigg)^{\ell-2}\\ &= \Bigg(\sum_{\sigma^1\in S_N}\mathbf{1}_{X_{\sigma^1}=1}\sum_{\sigma^2:|R(\sigma^1,\sigma^2)|\le R_N}\mathbf{1}_{X_{\sigma^2}=1}\Bigg)\Bigg(\sum_{\sigma\in S_N}\mathbf{1}_{X_\sigma=1}\Bigg)^{\ell-2}.\end{aligned} \tag{2.19}$$

As we already noted, the sum in the second factor of (2.19) concentrates on its expected value and is equal to $2^{M(\ell-2)}(1+o(1))$. Let us thus turn to the first factor in (2.19).

Introduce the set $(U^X_{N,\ell})_{\sigma^1} = \{\sigma^2\in X : |R(\sigma^1,\sigma^2)|\le R_N\}$. Then

$$|(U^X_{N,\ell})_{\sigma^1}| = \sum_{\sigma^2:|R(\sigma^1,\sigma^2)|\le R_N}\mathbf{1}_{X_{\sigma^2}=1} \tag{2.20}$$

The summands in this sum are i.i.d. and it follows from part (i) of Lemma 2.2 that their number is at least $2^N(1-2e^{-\frac{1}{8}NR_N^2}) = 2^N(1+o(1))$. Applying (2.11) together with the assumption that $\log N = o(2^M)$ we obtain from Borel-Cantelli Lemma that P-a.s., for any $\sigma^1 \in S_N$,

$$|(U_{N,\ell}^X)_{\sigma^1}| = (1+o(1))\mathrm{E}|(U_{N,\ell}^X)_{\sigma^1}|. \tag{2.21}$$

From (2.19), (2.20) and (2.21) we immediately conclude that

$$|(U_{N,\ell}^X)_{12}| = (1+o(1))\mathrm{E}|(U_{N,\ell}^X)_{12}|, \quad \text{P-a.s.} \tag{2.22}$$

This finishes the proof of part (i) of Theorem 2.1.

The proof of part (ii) is quite similar to the proof of part (i). By definition of $V_{N,\ell}^X$ we get

$$V_{N,\ell}^X \subseteq \bigcup_{1\le j<j'\le \ell} (V_{N,\ell}^X)_{jj'}, \tag{2.23}$$

where

$$(V_{N,\ell}^X)_{jj'} = \Big\{(\sigma^1,\dots,\sigma^\ell) \in V_{N,\ell}^X : \; |R(\sigma^j,\sigma^{j'})| = R_N\Big\}. \tag{2.24}$$

We claim that it suffices to prove that P-almost surely

$$|(V_{N,\ell}^X)_{jj'}| \le e^{\alpha N \mathcal{J}(R_N)}\mathrm{E}|(V_{N,\ell}^X)_{jj'}|. \tag{2.25}$$

Indeed, from (2.23) and from the above inequality we obtain that

$$|V_{N,\ell}^X| \le \sum_{1\le j<j'\le\ell} |(V_{N,\ell}^X)_{jj'}| \le \sum_{1\le j<j'\le\ell} e^{\alpha N\mathcal{J}(R_N)}\mathrm{E}|(V_{N,\ell}^X)_{jj'}|. \tag{2.26}$$

Using part (i) of Lemma 2.2 it is easy to establish that for all $1 \le j < j' \le \ell$

$$|V_{N,\ell}| = |(V_{N,\ell})_{jj'}|\big(1+o(1)\big), \tag{2.27}$$

and therefore

$$\mathrm{E}|V_{N,\ell}^X| = \mathrm{E}|(V_{N,\ell}^X)_{jj'}|(1+o(1)). \tag{2.28}$$

Since (2.26) and (2.28) imply the result we concentrate on the proof of (2.25).

Without loss of generality we can take $j=1$ and $j'=2$. Then, by definition of $(V_{N,\ell}^X)_{jj'}$, we get

$$\begin{aligned} |(V_{N,\ell}^X)_{12}| &= \Bigg(\sum_{\substack{\sigma^1,\sigma^2\in S_N:\\ |R(\sigma^1,\sigma^2)|=R_N}} \mathbf{1}_{X_{\sigma^1}=1}\mathbf{1}_{X_{\sigma^2}=1}\Bigg)\Bigg(\sum_{\sigma\in S_N}\mathbf{1}_{X_\sigma=1}\Bigg)^{\ell-2} \\ &= \Bigg(\sum_{\sigma^1\in S_N}\mathbf{1}_{X_{\sigma^1}=1}\sum_{\sigma^2:|R(\sigma^1,\sigma^2)|=R_N}\mathbf{1}_{X_{\sigma^2}=1}\Bigg)\Bigg(\sum_{\sigma\in S_N}\mathbf{1}_{X_\sigma=1}\Bigg)^{\ell-2}. \end{aligned} \tag{2.29}$$

As in the proof of part (i) we see that the second part of (2.29) concentrates on its expected value and equals to $2^{M(\ell-2)}(1+o(1))$. We are thus left to treat the first part. Introducing the set

$$(V_{N,\ell}^X)_{\sigma^1} = \{\sigma^2 \in X : |R(\sigma^1,\sigma^2)| = R_N\} \tag{2.30}$$

it is clear that

$$|(V_{N,\ell}^X)_{\sigma^1}| = \sum_{\sigma^2:|R(\sigma^1,\sigma^2)|=R_N}\mathbf{1}_{X_{\sigma^2}=1} \tag{2.31}$$

There are $2\binom{N}{d_N}$ i.i.d. terms in the above sum. Applying (2.12) with $t+1 = e^{\alpha N \mathcal{J}(R_N)}$ where $\alpha \in [0,1)$ will be chosen later, we obtain

$$\mathrm{P}(|(V_{N,\ell}^X)_{\sigma^1}| \geq e^{\alpha N \mathcal{J}(R_N)} \mathrm{E}|(V_{N,\ell}^X)_{\sigma^1}|) \leq e^{-\frac{1}{2}\mathrm{E}|(V_{N,\ell}^X)_{\sigma^1}|(1+t)\log(1+t)}. \tag{2.32}$$

From Stirling's formula we see that for some $C > 0$,

$$\mathrm{E}|(V_{N,\ell}^X)_{\sigma^1}| = 2p_N \binom{N}{d_N} = 2^M \frac{C}{\sqrt{N}} e^{-N\mathcal{J}(R_N)}(1+o(1)), \tag{2.33}$$

and thus the exponent in (2.32) is

$$\begin{aligned}\frac{1}{2}\mathrm{E}|(V_{N,\ell}^X)_{\sigma^1}|(1+t)\log(1+t) &= \frac{C}{2\sqrt{N}} 2^M e^{-(1-\alpha)N\mathcal{J}(R_N)} \alpha N \mathcal{J}(R_N)(1+o(1)) \\ &= \frac{\alpha C}{2} e^{M\log 2 - (1-\alpha)N\mathcal{J}(R_N) - \frac{1}{2}\log N} N\mathcal{J}(R_N)(1+o(1)). \end{aligned} \tag{2.34}$$

If for every $\sigma^1 \in S_N$ the set $(V_{N,\ell}^X)_{\sigma^1}$ is empty then there is nothing to prove. Otherwise, by Corollary 2.3 we obtain that P-almost surely $R_N < \delta_N$, and since $\mathcal{J}(x) = \frac{1}{2}x^2(1+O(x^2))$ near the origin we can choose $\alpha \in [0,1)$ in such a way that

$$M\log 2 - (1-\alpha)N\mathcal{J}(R_N) - \frac{1}{2}\log N > \gamma \log N \tag{2.35}$$

for some $\gamma > 0$. Hence $\sum_N e^{-\frac{1}{2}\mathrm{E}|(V_{N,\ell}^X)_{\sigma^1}|(1+t)\log(1+t)} < \infty$, and we obtain from (2.32) and Borel-Cantelli Lemma that P-almost surely

$$|(V_{N,\ell}^X)_{\sigma^1}| \leq e^{\alpha N\mathcal{J}(R_N)} \mathrm{E}|(V_{N,\ell}^X)_{\sigma^1}|. \tag{2.36}$$

It is easy to see that (2.36) and (2.29) imply (2.25). This concludes the proof of Theorem 2.1. □

In part *(ii)* of Theorem 2.1 we studied the properties of the sets $V_{N,\ell}^X \subset X^\ell$ for arbitrary $\ell \geq 2$. For $\ell = 2$ we can improve on Theorem 2.1.

Theorem 2.4. *Suppose that* $\limsup \frac{M}{N} < 1$.

(i) If for some $c_1 > \frac{1}{2}$

$$\mathcal{J}(R_N) \leq \frac{M\log 2}{N} - \frac{c_1 \log N}{N}, \tag{2.37}$$

then P*-almost surely*

$$|V_{N,2}^X| = (1+o(1))\mathrm{E}|V_{N,2}^X|. \tag{2.38}$$

(ii) If for positive constants c_1, c_2

$$\frac{M\log 2}{N} - \frac{c_1\log N}{N} \leq \mathcal{J}(R_N) \leq \frac{M\log 2}{N} + \frac{c_2 \log N}{N} \tag{2.39}$$

then there is a constant c *such that* P*-almost surely*

$$|V_{N,2}^X| \leq N^c\ \mathrm{E}|V_{N,2}^X|. \tag{2.40}$$

(iii) If for some $c_2 > \frac{3}{2}$

$$\mathcal{J}(R_N) > \frac{M\log 2}{N} + \frac{c_2 \log 2}{N}, \tag{2.41}$$

then the set $V_{N,2}^X$ *is* P*-almost surely empty.*

Proof. (i) By definition of $V_{N,2}^X$,

$$|V_{N,2}^X| = \sum_{\sigma^1\in S_N} \mathbf{1}_{X_{\sigma^1}=1} \sum_{\sigma^2\in S_N:|R(\sigma^1,\sigma^2)|=R_N} \mathbf{1}_{X_{\sigma^2}=1}. \tag{2.42}$$

Using (2.30) the inner sum is $|(V_{N,2}^X)_{\sigma^1}|$. Since it is a sum of i.i.d. random variables then for all $t = t(N) = o(1)$ we get from (2.11) that

$$\mathrm{P}\left(||(V_{N,2}^X)_{\sigma^1}| - \mathrm{E}|(V_{N,2}^X)_{\sigma^1}|| \geq t\,\mathrm{E}|(V_{N,2}^X)_{\sigma^1}|\right) \leq 2e^{-t^2\mathrm{E}|(V_{N,2}^X)_{\sigma^1}|/4}. \tag{2.43}$$

Using Stirling's approximation

$$\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n+\frac{1}{12n+1}} < n! < \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n+\frac{1}{12n}} \tag{2.44}$$

we obtain that

$$\mathrm{E}|(V_{N,2}^X)_{\sigma^1}| = 2p_N\binom{N}{d_N} = O\Big(\frac{e^{M\log 2 - N\mathcal{J}(R_N)}}{\sqrt{N(1-R_N^2)}}\Big). \tag{2.45}$$

Further, from (2.37) and Corollary 2.3 we obtain that $\mathrm{E}|(V_{N,2}^X)_{\sigma^1}| \geq CN^{c_1-1/2}$ for some positive constant $C > 0$. Choosing $t = N^{-\gamma}$ for some small enough $\gamma > 0$ in (2.43), we conclude from Borel-Cantelli Lemma that

$$|(V_{N,2}^X)_{\sigma^1}| = \mathrm{E}|(V_{N,2}^X)_{\sigma^1}|(1+o(1)), \quad \text{P-a.s.} \tag{2.46}$$

Using (2.42), (2.46), and the fact that P-almost surely there are $2^M(1+o(1))$ configurations in the random cloud X proves (i).

(ii) The proof is similar to the proof of part (i). In particular, from the representation (2.42) it is easy to see that it suffices to prove that

$$|(V_{N,2}^X)_{\sigma^1}| \leq N^c\ \mathrm{E}|(V_{N,2}^X)_{\sigma^1}|, \quad \text{P-a.s.} \tag{2.47}$$

Choosing $1 + t = N^c$ we get from (2.45) that for some positive constant C

$$\frac{1}{2}E|(V_{N,2}^X)_{\sigma^1}|(1+t)\log(1+t) \geq CN^{c-c_2-1/2}\log N. \tag{2.48}$$

Choosing c large enough and applying (2.12) together with Borel-Cantelli Lemma proves part (ii).

(iii) We again use the representation (2.42). Clearly it is enough to prove that for all $\sigma^1 \in X$ the set $(V_{N,2}^X)_{\sigma^1}$ is P-almost surely empty. By (2.45) $\mathrm{E}|(V_{N,2}^X)_{\sigma^1}| \leq CN^{-c_2-1/2}$ for some positive constant C. Thus, choosing $1 + t = \frac{1}{E|(V_{N,2}^X)_{\sigma^1}|} \to \infty$ we obtain from (2.12)

$$\mathrm{P}\left(|(V_{N,2}^X)_{\sigma^1}| \geq 1\right) = \mathrm{P}\left(|(V_{N,2}^X)_{\sigma^1}| \geq (1+t)\mathrm{E}|(V_{N,2}^X)_{\sigma^1}|\right) \leq (t+1)^{-1/2}. \tag{2.49}$$

By definition of t and by condition (2.41) we get

$$(t+1)^{-1/2} \leq C^{1/2}N^{-c_2/2-1/4}. \tag{2.50}$$

Applying Borel-Cantelli Lemma proves (iii). □

In order to estimate the third moment in Theorems 4.1, 4.2, and 4.7 we give a result similar to Theorem 2.4 but for $\ell = 3$, i.e. for a given sequence of vectors $(R_{12}^N, R_{23}^N, R_{31}^N)$ we estimate the cardinal of the set

$$W_{N,3}^X = \Big\{(\sigma^1,\sigma^2,\sigma^3) \in X^3 : R(\sigma^1,\sigma^2) = R_{12}^N, R(\sigma^2,\sigma^3) = R_{23}^N, R(\sigma^3,\sigma^1) = R_{31}^N\Big\}. \tag{2.51}$$

Below we omit the explicit dependence of the sequence $(R_{12}^N, R_{23}^N, R_{31}^N)$ on N and instead of R_{12}^N, R_{23}^N and R_{31}^N will write R_{12}, R_{23} and R_{31} respectively. In order to formulate the theorem we introduce the following function on $\mathbb{R}^3$:

$$\mathcal{J}^{(2)}(x,y,z) = \frac{1+x+y+z}{4}\log(1+x+y+z) + \frac{1+x-y-z}{4}\log(1+x-y-z) + \frac{1-x+y-z}{4}\log(1-x+y-z) + \frac{1-x-y+z}{4}\log(1-x-y+z) \tag{2.52}$$

if $|1+x| \geq |y+z|$ and $|1-x| \geq |y-z|$ and $\mathcal{J}^{(2)}(x,y,z) = +\infty$ otherwise.

Theorem 2.5. *Suppose* $\limsup \frac{M}{N} < 1$.

(i) If for some $c_1 > \frac{1}{2}$ *the sequence* R_{12} *satisfies* (2.37), *and if for some* $c_1^{(2)} > 1$

$$\mathcal{J}^{(2)}(R_{12}, R_{23}, R_{31}) \leq \frac{M\log 2}{N} + \mathcal{J}(R_{12}) - \frac{c_1^{(2)}\log N}{N}, \tag{2.53}$$

then P*-almost surely*

$$|W_{N,3}^X| = (1+o(1))\mathrm{E}|W_{N,3}^X|. \tag{2.54}$$

(ii) If for positive constants $c_1^{(2)}, c_2^{(2)}$,

$$\frac{M\log 2}{N} + \mathcal{J}(R_{12}) - \frac{c_1^{(2)}\log N}{N} \leq \mathcal{J}^{(2)}(R_{12}, R_{23}, R_{31}) \leq \frac{M\log 2}{N} + \mathcal{J}(R_{12}) + \frac{c_2^{(2)}\log N}{N}, \tag{2.55}$$

then there is a constant c *such that* P*-almost surely*

$$|W_{N,3}^X| \leq N^c\mathrm{E}|W_{N,3}^X|. \tag{2.56}$$

(iii) If for some $c_2^{(2)} > \frac{3}{2}$

$$\mathcal{J}^{(2)}(R_{12}, R_{23}, R_{31}) > \frac{M\log 2}{N} + \mathcal{J}(R_{12}) + \frac{c_2^{(2)}\log 2}{N}, \tag{2.57}$$

then P*-almost surely the set* $W_{N,3}^X$ *is empty.*

Proof. ¿From Lemma 2.6 stated below it follows that, for arbitrary configurations $\sigma^1, \sigma^2, \sigma^3 \in S_N$, the function $\mathcal{J}^{(2)}(R_{12}, R_{23}, R_{31})$ is well defined. This lemma, whose proof we will omit, is a direct consequence of the fact that the Hamming distance on S_N satisfies triangle inequality.

Lemma 2.6. *For arbitrary configurations* σ^1, σ^2 *and* $\sigma^3 \in S_N$ *we have* $1 + R_{12} \geq |R_{23} + R_{31}|$ *and* $1 - R_{12} \geq |R_{23} - R_{31}|$.

The proof of the Theorem 2.5 is similar to that of Theorem 2.4. We begin by writing the size of the set $W_{N,3}^X$ as

$$|W_{N,3}^X| = \sum_{\sigma^1\in S_N} \mathbf{1}_{X_{\sigma^1}=1} \sum_{\substack{\sigma^2\in S_N:\\ R(\sigma^1,\sigma^2)=R_{12}}} \mathbf{1}_{X_{\sigma^2}=1} \sum_{\substack{\sigma^3\in S_N:\\ R(\sigma^2,\sigma^3)=R_{23}\\ R(\sigma^1,\sigma^3)=R_{31}}} \mathbf{1}_{X_{\sigma^3}=1} \tag{2.58}$$

Let us first estimate the number of terms in the last sum. This means that, given $\sigma^1, \sigma^2 \in S_N$ with overlap $R(\sigma^1, \sigma^2) = R_{12}$, we have to calculate the number of configurations σ^3 with $R(\sigma^2, \sigma^3) = R_{23}$ and $R(\sigma^3, \sigma^1) = R_{31}$. Without loss of generality we can assume that all the spins of σ^1 are equal to 1. Further, let $C(\sigma^1, \sigma^2, \sigma^3)$ be a $3 \times N$ matrix with rows $\sigma^1, \sigma^2, \sigma^3$. For a column vector $\boldsymbol{\delta} \in \{-1, 1\}^3$ we let $n_{\boldsymbol{\delta}}$ be the number of columns of the matrix C that are equal to $\boldsymbol{\delta}$, i.e.

$$n_{\boldsymbol{\delta}} = \left|\{j \leq N : (\sigma_j^1, \sigma_j^2, \sigma_j^3) = \boldsymbol{\delta}\}\right|. \tag{2.59}$$

Then the overlaps can be written in terms of $n_{\boldsymbol{\delta}}$, namely

$$\begin{cases} n_{(1,1,1)} + n_{(1,1,-1)} - n_{(1,-1,1)} - n_{(1,-1,-1)} = NR_{12}, \\ n_{(1,1,1)} - n_{(1,1,-1)} - n_{(1,-1,1)} + n_{(1,-1,-1)} = NR_{23}, \\ n_{(1,1,1)} - n_{(1,1,-1)} + n_{(1,-1,1)} - n_{(1,-1,-1)} = NR_{31}, \\ n_{(1,1,1)} + n_{(1,1,-1)} + n_{(1,-1,1)} + n_{(1,-1,-1)} = N. \end{cases} \tag{2.60}$$

Solving this system of linear equations we find

$$\begin{cases} n_{(1,1,1)} & = \frac{1}{4}N(1 + R_{12} + R_{23} + R_{31}), \\ n_{(1,1,-1)} & = \frac{1}{4}N(1 + R_{12} - R_{23} - R_{31}), \\ n_{(1,-1,1)} & = \frac{1}{4}N(1 - R_{12} - R_{23} + R_{31}), \\ n_{(1,-1,-1)} & = \frac{1}{4}N(1 - R_{12} + R_{23} - R_{31}). \end{cases} \tag{2.61}$$

We notice that specifying the configuration σ^3 is equivalent to specifying the numbers $n_{(1,1,1)}, n_{(1,1,-1)}, n_{(1,-1,1)}$, and $n_{1,-1,-1}$. Therefore the number of configurations $\sigma^3 \in S_N$ with overlaps $R(\sigma^2, \sigma^3) = R_{23}$, and $R(\sigma^3, \sigma^1) = R_{31}$ is

$$\begin{aligned} &\binom{n_{(1,1,1)} + n_{(1,1,-1)}}{n_{(1,1,1)}} \binom{n_{(1,-1,1)} + n_{(1,-1,-1)}}{n_{(1,-1,1)}} \\ &\quad = \frac{\big(n_{(1,1,1)} + n_{(1,1,-1)}\big)!}{n_{(1,1,1)}!\, n_{(1,1,-1)}!} \frac{\big(n_{(1,-1,1)} + n_{(1,-1,-1)}\big)!}{n_{(1,-1,1)}!\, n_{(1,-1,-1)}!}. \end{aligned} \tag{2.62}$$

Applying Stirling's approximation to (2.62) one obtains that the number of terms in the last summation in (2.58) is of order

$$\frac{2^N e^{N\mathcal{J}(R_{12}) - N\mathcal{J}^{(2)}(R_{12}, R_{23}, R_{31})} \sqrt{1 - R_{12}^2}}{N\sqrt{P(R_{12}, R_{23}, R_{31})}},$$

where $P(x, y, z) = (1 + x + y + z)(1 + x - y - z)(1 - x + y - z)(1 - x - y + z)$. The rest of the proof essentially is a rerun of the proof of Theorem 2.4. We skip the details. □

3. Proof of Theorem 1.1

As in [BK06a] and [BCMN05a],[BCMN05b], the proof of the Poisson convergence is based on the analysis of factorial moments of the point processes $\mathcal{P}_N$ defined in (1.12).

In general, let ξ_N be a sequence of point processes defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ and let ξ be a Poisson point process with intensity measure μ. Define the ℓ^{th} factorial moment $\mathbb{E}(Z)_\ell$ of the random variable Z to be $\mathbb{E}Z(Z-1)\dots(Z-\ell+1)$. The following is a classical lemma that is a direct consequence of Theorem 4.7 in [Kal83].

Lemma 3.1. *If for every $\ell \geq 1$ and every Borel set A*

$$\lim_{N\to\infty} \mathbb{E}_{\mathbb{Q}}(\xi_N(A))_\ell = (\mu(A))^\ell, \tag{3.1}$$

then the distribution of $(\xi_N)_{N\geq 1}$ converges weakly to the distribution of ξ.

Applying Lemma 3.1 to the sequence of point processes $\mathcal{P}_N = \sum_{\sigma \in X} \delta_{H'_N(\sigma)}$ the following result proves Theorem 1.1.

Theorem 3.2. *Under the assumptions of Theorem 1.1, for every $\ell \in \mathbb{N}$ and every bounded Borel set A*

$$\lim_{N\to\infty} \mathbb{E}(\mathcal{P}_N(A))_\ell = (\mu(A))^\ell, \quad \text{P-}a.s., \tag{3.2}$$

where μ is defined in (1.13).

Proof. We start with the computation of the first moment of $\mathcal{P}_N(A)$:

$$\mathbb{E}\,\mathcal{P}_N(A) = \sum_{\sigma \in X} \mathbb{P}(H'_N(\sigma) \in A). \tag{3.3}$$

As we saw in the proof of Theorem 2.1 the size of the random cloud $|X|$ is P-almost surely $2^M(1+o(1))$. Since $H'_N(\sigma), \sigma \in S_N$, are identically distributed normal random variables with mean $-a_N/b_N$ and variance $1/b_N$, the sum in (3.3) can be written as

$$2^M \frac{e^{-a_N^2/2}}{\sqrt{2\pi}} b_N \int_A e^{-x^2 b_N^2/2 - a_N b_N x} dx\,(1+o(1)), \quad P\text{-a.s.} \tag{3.4}$$

By the dominated convergence theorem and the definition of a_N it follows from (3.4) that the limit of the first moment is $\mu(A)$.

To calculate factorial moments of higher order we follow [BCMN05a] and rewrite the ℓ^{th} factorial moment of $\mathcal{P}_N(A)$ as

$$\mathbb{E}(\mathcal{P}_N(A))_\ell = \sum_{\sigma^1,\dots,\sigma^\ell} \mathbb{P}\big(H'_N(\sigma^1) \in A, \dots, H'_N(\sigma^\ell) \in A\big) \tag{3.5}$$

where the sum runs over all ordered sequences of distinct configurations $(\sigma^1, \dots, \sigma^\ell) \in X^\ell$. To analyze it we decompose the set S_N^ℓ into three non-intersecting subsets

$$S_N^\ell = U_\ell(R_N) \cup \big(U_\ell(\delta_N)\backslash U_\ell(R_N)\big) \cup (U_\ell(\delta_N))^c, \tag{3.6}$$

where U_ℓ is defined in (2.2), δ_N is defined in Corollary 2.3 and the sequence R_N is chosen such that:

$$R_N \to 0, M\nu(R_N) \to 0, NR_N^2 \to \infty. \tag{3.7}$$

This is possible since we have assumed that $M = o(\sqrt{N})$ if $\nu'(0) \neq 0$ and $M = o(N)$ if $\nu'(0) = 0$. We recall here that the function ν is defined in (1.7)-(1.8).

Having specified R_N, let us analyze the contribution to the sum (3.5) coming from the intersection of X^ℓ with the sets $U_\ell(R_N), U_\ell(\delta_N)\backslash U_\ell(R_N)$, and $(U_\ell(\delta_N))^c$. Firstly, by Corollary 2.3 the intersection of the set $(U_\ell(\delta_N))^c$ with X^ℓ is P-a.s. empty and therefore its contribution to the sum (3.5) is zero. Next, let us show that P-a.s.

$$\lim_{N\to\infty} \sum_{\sigma^1,\dots,\sigma^\ell} \mathbb{P}\left(H'_N(\sigma^1) \in A, \dots, H'_N(\sigma^\ell) \in A\right) = (\mu(A))^\ell, \tag{3.8}$$

where the sum is over all the sequences $(\sigma^1, \dots, \sigma^\ell) \in U_\ell(R_N) \cap X^\ell$.

For every $\ell \in \mathbb{N}$ let $B(\sigma^1, \dots, \sigma^\ell)$ denote the covariance matrix of random variables $H_N(\sigma^1), \dots, H_N(\sigma^\ell)$. By (1.7) its elements, b_{ij}, are given by

$$b_{ij} = \nu(R_{ij}) \tag{3.9}$$

where we wrote $R(\sigma^i,\sigma^j) = R_{ij}$. Since $R_N = o(1)$ the matrix $B(\sigma^1,\dots,\sigma^\ell)$ is non-degenerate. We therefore get for $(\sigma^1,\dots,\sigma^\ell) \in U_\ell(R_N)\cap X^\ell$ that

$$\mathbb{P}\big(H'_N(\sigma^1)\in A,\dots,H'_N(\sigma^\ell)\in A\big) = \frac{b_N^\ell}{(2\pi)^{\ell/2}\sqrt{\det B}} \times \int_A\cdots\int_A e^{-(\vec{x},B^{-1}\vec{x})b_N^2/2-a_Nb_N(\vec{x},B^{-1}\vec{1})-a_N^2(\vec{1},B^{-1}\vec{1})/2}d\vec{x}, \tag{3.10}$$

where $B = B(\sigma^1,\dots,\sigma^\ell)$, $\vec{x} = (x_1,\dots,x_\ell)^t$ and $\vec{1} = (1,\dots,1)^t$. From the definition of a_N we further get from (3.10) that

$$\mathbb{P}\big(H'_N(\sigma^1)\in A,\dots,H'_N(\sigma^\ell)\in A\big) = \frac{1}{2^{M\ell}}\frac{e^{(\ell-(\vec{1},B^{-1}\vec{1}))(M\log 2-1/2\log M)}}{\sqrt{\det B}} \times \frac{1}{(\sqrt{\pi})^\ell}\int_A\cdots\int_A e^{-(\vec{x},B^{-1}\vec{x})b_N^2/2-a_Nb_N(\vec{x},B^{-1}\vec{1})}d\vec{x}. \tag{3.11}$$

Since the matrix B^{-1} is positive definite and since $a_Nb_N \to \sqrt{2\log 2}$ we conclude from the dominated convergence theorem that for all bounded Borel sets A, uniformly in $(\sigma^1,\dots,\sigma^\ell)\in U_\ell(R_N)\cap X^\ell$,

$$\frac{1}{(\sqrt{\pi})^\ell}\int_A\cdots\int_A e^{-(\vec{x},B^{-1}\vec{x})b_N^2/2-a_Nb_N(\vec{x},B^{-1}\vec{1})}d\vec{x} \to \frac{1}{(\sqrt{\pi})^\ell}\int_A\cdots\int_A e^{-\sqrt{2\log 2}(\vec{x},\vec{1})}d\vec{x} = (\mu(A))^\ell. \tag{3.12}$$

To evaluate $e^{(\ell-(\vec{1},B^{-1}\vec{1}))(M\log 2-1/2\log M)}$ in (3.11) we look at $(\ell-(\vec{1},B^{-1}\vec{1}))\det B$ as a multivariate function of $(b_{ij})_{1\le i<j\le\ell}$. It is a polynomial of degree ℓ with coefficients depending only on ℓ and without constant term. It implies that $|\ell-(\vec{1},B^{-1}\vec{1})| = O(\nu(R_N))$ and therefore, by (3.7),

$$e^{(\ell-(\vec{1},B^{-1}\vec{1}))(M\log 2-1/2\log M)} = 1+o(1). \tag{3.13}$$

Combining (3.11), (3.12), and (3.13) we may rewrite the sum (3.8) as

$$\sum_{\sigma^1,\dots,\sigma^\ell}\frac{1}{2^{M\ell}}\left(\mu(A)\right)^\ell(1+o(1)) = \frac{|U_\ell(R_N)\cap X^\ell|}{2^{M\ell}}\left(\mu(A)\right)^\ell(1+o(1)). \tag{3.14}$$

Now it follows from Theorem 2.1 (i) and Lemma 2.2 (i) that $|U_\ell(R_N)\cap X^\ell|$ concentrates around its expected value, namely $2^{M\ell}(1+o(1))$, so that (3.14) implies (3.8).

Next, let us establish that the contribution from the second set in (3.6) is negligible, i.e. let us prove that

$$\lim_{N\to\infty}\sum_{\sigma^1,\dots,\sigma^\ell}\mathbb{P}\left(H'_N(\sigma^1)\in A,\dots,H'_N(\sigma^\ell)\in A\right) = 0, \tag{3.15}$$

where the sum runs over all the sequences $(\sigma^1,\dots,\sigma^\ell)\in\big(U_\ell(\delta_N)\backslash U_\ell(R_N)\big)\cap X^\ell$. To do this we first bound the righthand side of (3.11). By definition (2.1) and Corollary 2.3, $R_{\max}(\sigma^1,\dots,\sigma^\ell)\le\delta_N = o(1)$. Therefore following the same reasoning as above we obtain from (3.11) that for some constant $C>0$ and for all

$(\sigma^1,\dots,\sigma^\ell)\in\big(U_\ell(\delta_N)\backslash U_\ell(R_N)\big)\cap X^\ell$

$$\mathbb{P}\big(H'_N(\sigma^1)\in A,\dots,H'_N(\sigma^\ell)\in A\big)\le \frac{e^{CM\nu(R_{\max})}}{2^{M\ell}}(\mu(A))^\ell(1+o(1)). \tag{3.16}$$

For fixed N the overlap takes only a discrete set of values

$$K_N=\Big\{1-\frac{2k}{N}:k=0,1,\dots,N\Big\}. \tag{3.17}$$

We represent the set $U_\ell(\delta_N)\backslash U_\ell(R_N)$ as a union of sets $V_\ell(R_{N,k})$, where we denoted $R_{N,k}=1-\frac{2k}{N}\in K_N\cap(R_N,\delta_N)$. Let us fix k and bound the contribution from the set $V_\ell(R_{N,k})\cap X^\ell$, i.e.

$$\sum_{\substack{(\sigma^1,\dots,\sigma^\ell)\in\\ V_\ell(R_{N,k})\cap X^\ell}}\mathbb{P}\left(H'_N(\sigma^1)\in A,\dots,H'_N(\sigma^\ell)\in A\right). \tag{3.18}$$

We obtain from Lemma 2.2 (ii) that

$$|V_\ell(R_{N,k})|=2^{N\ell}\frac{C}{\sqrt{N}}e^{-N\mathcal{J}(R_{N,k})}(1+o(1)). \tag{3.19}$$

In the case $M(N)\le\log N$ we can choose the sequence R_N in such a way that the set $U_\ell(\delta_N)\backslash U_\ell(R_N)$ is empty. Therefore we can assume without loss of generality that $M(N)\ge\log N$. Applying part (ii) of Theorem 2.1 we further conclude that

$$|V_\ell(R_{N,k})\cap X^\ell|\le 2^{M\ell}\frac{C}{\sqrt{N}}e^{-(1-\alpha)N\mathcal{J}(R_{N,k})},\quad \text{P-a.s.} \tag{3.20}$$

Using (3.16) and the last inequality we bound the sum (3.18) by

$$\frac{C}{\sqrt{N}}e^{-(1-\alpha)N\mathcal{J}(R_{N,k})}e^{CM\nu(R_{N,k})}. \tag{3.21}$$

One can easily check that $M\nu(R_{N,k})=o(NR^2_{N,k})$ for $R_{N,k}\in(R_N,\delta_N)$. Together with $\mathcal{J}(x)\ge x^2/2$ it implies that for some positive constants C_1 and C_2 we can further bound (3.21) by

$$\frac{C_1}{\sqrt{N}}e^{-C_2NR^2_{N,k}}. \tag{3.22}$$

As a consequence, we obtain an almost sure bound

$$\begin{aligned}\sum_{\substack{R_{N,k}\in\\ K_N\cap(R_N,\delta_N)}}\ \sum_{\substack{(\sigma^1,..,\sigma^\ell)\in\\ V_\ell(R_{N,k})\cap X^\ell}}&\mathbb{P}\left(H'_N(\sigma^1)\in A,\dots,H'_N(\sigma^\ell)\in A\right)\\ &\le\frac{C_1}{\sqrt{N}}\sum_{\substack{R_{N,k}\in\\ K_N\cap(R_N,\delta_N)}}e^{-C_2NR^2_{N,k}}.\end{aligned} \tag{3.23}$$

Introducing new variables $y_{N,k}=\sqrt{N}R_{N,k}$ we rewrite the above sum as

$$\frac{C_1}{2}\sum_{y_{N,k}}e^{-C_2y^2_{N,k}}\frac{2}{\sqrt{N}}, \tag{3.24}$$

where the summation is over the discrete set $\sqrt{N}K_N \cap (\sqrt{N}R_N, \sqrt{N}\delta_N)$. Since $NR_N^2 \to \infty$, then for arbitrary $C > 0$ and for large N we can further bound this sum by

$$\frac{C_1}{2} \sum_{y_{N,k} \geq C} e^{-C_2 y_{N,k}^2} \frac{2}{\sqrt{N}}. \tag{3.25}$$

Interpreting the last sum as a sum of areas of nonintersecting rectangles with one side equal $e^{-C_2 y_{N,k}}$ and the other $\frac{2}{\sqrt{N}}$ we bound it with the integral

$$\int_{C-\frac{2}{\sqrt{N}}}^{\infty} e^{-C_2 y^2} dy \tag{3.26}$$

Since the constant C is arbitrary we get that (3.15) is $o(1)$. This finishes the proof of Theorem 3.2 and therefore of Theorem 1.1. □

4. Proof of Theorem 1.4.

In order to prove the breakdown of universality in Theorem 1.4 we use a strategy similar to that used in [BCMN05b] to disprove the local REM conjecture for the number partitioning problem and for the Sherrington-Kirkpatrick model when the energy scales are too large. We prove that P-a.s. for every bounded Borel set A

(1) the limit of the first factorial moment exists and equals $\mu(A)$;
(2) the second factorial moment $\mathbb{E}(\mathcal{P}_N(A))_2$ does not converge to $(\mu(A))^2$;
(3) the third moment is bounded.

These three facts immediately imply that the sequence of random variables $\mathcal{P}_N(A)$ does not converge weakly to a Poisson random variable and so the sequence of point processes $\mathcal{P}_N$ does not converge weakly to a Poisson point process. Part (i) of Theorem 1.4 is thus obviously implied by the following

Theorem 4.1 (Breakdown of Universality for the number partitioning problem)**.** *Let $\nu(r) = r$. For every bounded Borel set A*

$$\lim \mathbb{E}(\mathcal{P}_N(A))_1 = \mu(A), \quad \text{P-}a.s. \tag{4.1}$$

Moreover, if $\limsup \frac{M(N)}{\sqrt{N}} = \varepsilon < \infty$ *then* P-*a.s.*

(i) $\limsup \mathbb{E}(\mathcal{P}_N(A))_2 = e^{2\varepsilon^2 \log^2 2}(\mu(A))^2$,
(ii) $\limsup \mathbb{E}(\mathcal{P}_N(A))_3 < \infty$.

Similarly, part (ii) of Theorem 1.4 is implied by the following

Theorem 4.2 (Breakdown of Universality for the Sherrington-Kirkpatrick model)**.** *Let $\nu(r) = r^2$. For every bounded Borel set A*

$$\lim \mathbb{E}(\mathcal{P}_N(A))_1 = \mu(A), \quad \text{P-}a.s. \tag{4.2}$$

Moreover, if $\limsup \frac{M(N)}{N} = \varepsilon < \frac{1}{8\log 2}$ *then* P-*a.s.*

(i) $\limsup \mathbb{E}(\mathcal{P}_N(A))_2 = \frac{(\mu(A))^2}{\sqrt{1-4\varepsilon \log 2}}$;
(ii) $\limsup \mathbb{E}(\mathcal{P}_N(A))_3 < \infty$.

Remark 4.3. Condition $\varepsilon < \frac{1}{8\log 2}$ in not optimal and could be improved. The reason for such a choice is that for $\varepsilon < \frac{1}{8\log 2}$ the third moment estimate is quite simple.

We will prove in detail Theorem 4.2 but omit the proof of Theorem 4.1 as it is very similar and much simpler.

Proof of Theorem 4.2. We successively prove the statement on the first, second, and third moment.

1. **First moment estimate.**
Since all the random variables $H'_N(\sigma), \sigma \in S_N$, are identically distributed then

$$\mathbb{E}(\mathcal{P}_N(A))_1 = |X|\mathbb{P}(H'_N(\sigma) \in A). \tag{4.3}$$

We saw in the proof of Theorem 2.1 that for $M(N)$ satisfying $\log N = o(2^M)$, $|X| = 2^M(1+o(1))$ P-a.s. Combined with (3.11), the definition of a_N, and the dominated convergence theorem this fact implies that P-a.s.

$$\mathbb{E}(\mathcal{P}_N(A))_1 = 2^M(1+o(1))\frac{1}{2^M}\frac{1}{\sqrt{\pi}}\int\limits_A e^{-x^2b_N^2/2-a_Nb_Nx}dx = \mu(A)(1+o(1)). \tag{4.4}$$

Hence (4.2) is proven.

2. **Second moment estimate.** Next assume that $\limsup \frac{M(N)}{N} = \varepsilon$. We now want to calculate $\limsup \mathbb{E}(\mathcal{P}_N(A))_2$. For this we rewrite the second factorial moment as

$$\mathbb{E}(\mathcal{P}_N(A))_2 = \sum_{\sigma^1,\sigma^2} \mathbb{P}\left(H'_N(\sigma^1) \in A, H'_N(\sigma^2) \in A\right), \tag{4.5}$$

where the summation is over all pairs of distinct configurations $(\sigma^1,\sigma^2) \in X^2$. We then split the set S_N^2 into four non-intersecting subsets and calculate the contributions from these subsets separately (these calculations are similar to those of Theorem 3.2).

(1) We begin by calculating the contribution from the set $\mathcal{S}_1^X$, where

$$\mathcal{S}_1 = \left\{(\sigma^1,\sigma^2) \in S_N^2 :\ |R(\sigma^1,\sigma^2)| \le \tau_N\right\} \tag{4.6}$$

and where the sequence τ_N is chosen in such a way that $N\tau_N^4 \to 0$ and $e^{-N\tau_N^2}$ decays faster than any polynomial: this can be achieved by choosing e.g. $\tau_N = \frac{1}{N^{1/4}\log N}$. First, we get from (3.11) for $\ell = 2$

$$\begin{aligned}\mathbb{P}\big(H'_N(\sigma^1) \in A, H'_N(\sigma^2) \in A\big) &= \frac{1}{2^{2M}}\frac{e^{\frac{2b_{12}}{1+b_{12}}(M\log 2 - 1/2\log M)}}{\sqrt{1-b_{12}^2}} \\ &\times \frac{1}{(\sqrt{\pi})^2}\int\limits_A\int\limits_A e^{-(\vec{x},B^{-1}\vec{x})b_N^2/2-a_Nb_N(\vec{x},B^{-1}\vec{1})}dx_1dx_2,\end{aligned} \tag{4.7}$$

where $b_{12} = \mathrm{cov}(H'_N(\sigma^1), H'_N(\sigma^2)) = R_{12}^2$. If $(\sigma^1,\sigma^2) \in \mathcal{S}_1$ then $b_{12} = o(1)$ and we get from (3.12) that uniformly in $(\sigma^1,\sigma^2) \in \mathcal{S}_1$ the second line in (4.7) is just $(\mu(A))^2(1+o(1))$.

Next, we represent the set $\mathcal{S}_1$ as a union of sets $V_2(R_{N,k})$ with $R_{N,k} = 1 - \frac{2k}{N} \in K_N \cap [0, \tau_N]$. Applying Theorem 2.4 (i) we get that

$$|V_2(R_{N,k}) \cap X^2| = 2\sqrt{\frac{2}{\pi N}} 2^{2M} e^{-N\mathcal{J}(R_{N,k})}(1 + o(1)). \tag{4.8}$$

Therefore, up to a multiplicative term of the form $1 + o(1)$, the contribution from the set $\mathcal{S}_1^X$ to the sum (4.5) is

$$\sum_{R_{N,k}} \frac{1}{2^{2M}} e^{\frac{2b_{12}}{1+b_{12}}(M\log 2 - 1/2\log M)} (\mu(A))^2 2\sqrt{\frac{2}{\pi N}} 2^{2M} e^{-N\mathcal{J}(R_{N,k})}$$

$$= 2\sqrt{\frac{2}{\pi N}} (\mu(A))^2 \sum_{R_{N,k}} e^{-N\mathcal{J}(R_{N,k})} e^{\frac{2b_{12}}{1+b_{12}}(M\log 2 - 1/2\log M)}. \tag{4.9}$$

Since $b_{12}(R_{N,k}) = R_{N,k}^2 > 0$ the last sum is monotone in M. As we will see below due to this fact it is sufficient to calculate the upper limit of $\mathbb{E}(\mathcal{P}(A))_2$ for sequences of the form $M(N) = \varepsilon N$ with $\varepsilon \in (0, \frac{1}{8\log 2})$. Thus let $M = \varepsilon N, \varepsilon \in (0, \frac{1}{8\log 2})$. We obtain

$$\frac{2b_{12}}{1 + b_{12}}(M\log 2 - 1/2\log M) = 2\varepsilon N R_{N,k}^2 \log 2 + O(NR_{N,k}^4). \tag{4.10}$$

The choice of τ_N guarantees that for $R_{N,k} \le \tau_N$ the last term in the rhs of (4.10) is of order $o(1)$. Moreover, for $x < 1$ we observe that $\mathcal{J}(x) = \frac{1}{2}x^2 + O(x^4)$. Thus

$$N\mathcal{J}(R_{N,k}) = \frac{NR_{N,k}^2}{2} + o(1). \tag{4.11}$$

Using (4.10) and (4.11) the sum (4.9) becomes

$$2\sqrt{\frac{2}{\pi N}} (\mu(A))^2 \sum_{R_{N,k}} e^{-\frac{1}{2}NR_{N,k}^2(1-4\varepsilon\log 2)}(1 + o(1)), \tag{4.12}$$

where the summation is over $R_{N,k} \in K_N \cap [0, \tau_N]$. Introducing new variables $y_{N,k} = \sqrt{N} R_{N,k}$ we further rewrite (4.12) as

$$\sqrt{\frac{2}{\pi}} \sum_{y_{N,k}} \frac{2}{\sqrt{N}} e^{-\frac{y_N^2}{2}(1-4\varepsilon\log 2)}(1 + o(1)). \tag{4.13}$$

It is not difficult to see that for $\varepsilon < \frac{1}{4\log 2}$ the sum in (4.13) converges to the integral

$$\int_0^\infty e^{-\frac{y^2}{2}(1-4\varepsilon\log 2)} dy = \frac{1}{\sqrt{1 - 4\varepsilon\log 2}} \sqrt{\frac{\pi}{2}}. \tag{4.14}$$

Therefore the contribution from the set S_1^X is $\frac{(\mu(A))^2}{\sqrt{1-4\varepsilon\log 2}}(1 + o(1))$.

We can extend this result to the case when $M(N) = \varepsilon N + o(N)$ with $\varepsilon < \frac{1}{8\log 2}$. Indeed, assume that $\varepsilon_n \uparrow \varepsilon$ as $n \to \infty$. Then using the above calculation for $M = \varepsilon_n N$ together with the monotonicity argument we have that for all $n \ge 1$

$$\mathbb{E}(\mathcal{P}_N(A))_2 \ge \frac{(\mu(A))^2}{\sqrt{1 - 4\varepsilon_n \log 2}}. \tag{4.15}$$

Taking the limit in n we obtain a lower bound. With exactly the same argument we prove the corresponding upper bound.

(2) Next, we estimate the contribution from the set $\mathcal{S}_2^X$, where

$$\mathcal{S}_2 = \big\{(\sigma^1,\sigma^2) \in S_N^2 :\ |R(\sigma^1,\sigma^2)| > \tau_N \text{ and } R(\sigma^1,\sigma^2) \text{ satisfies (2.37)}\big\}. \quad (4.16)$$

Since the set A is bounded an elementary computation yields that for a constant $C = C(A)$, uniformly in $(\sigma^1,\sigma^2) \in \mathcal{S}_2$,

$$\int_A \int_A e^{-(\vec{x},B^{-1}\vec{x})\ b_N^2/2 - a_N b_N (\vec{x},B^{-1}\vec{1})} dx_1 dx_2 \leq C. \quad (4.17)$$

For fixed N we let R_N^1 to be the largest value of the overlap satisfying condition (2.37). Then representing $\mathcal{S}_2$ as a union of sets $V_2(R_{N,k})$, where $R_{N,k} = 1 - \frac{2k}{N} \in K_N \cap [\tau_N, R_N^1]$, and using Theorem 2.4 (i), we conclude that the contribution from the set $\mathcal{S}_2^X$ is, up to a multiplicative term of the form $1+o(1)$, bounded by

$$\sum_{R_{N,k}} 2\sqrt{\frac{2}{\pi N(1-R_{N,k}^2)}}\ \frac{C}{\sqrt{1-b_{12}^2}}\ e^{-N\mathcal{J}(R_{N,k}) + \frac{2b_{12}}{1+b_{12}} M \log 2}. \quad (4.18)$$

This quantity is monotone in M. As a consequence, to show that it is negligible in the limit $N \to \infty$ it suffices to show this fact under the assumption $M(N) = \varepsilon N$. Thus, letting $M = \varepsilon N$ and using that $\mathcal{J}(x) \geq \frac{1}{2}x^2$, we can bound the exponent in (4.18) by

$$-N\left(\mathcal{J}(R_{N,k}) - \frac{2b_{12}}{1+b_{12}}\varepsilon \log 2\right) \leq -\frac{1}{2}Nx^2(1-4\varepsilon\log 2). \quad (4.19)$$

Since $1 - b_{12}^2 \geq 1/N$ and since the number of terms in (4.18) is at most N, we can further bound it, for some positive constant $C > 0$, by

$$CN\ 2\sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}N\tau_N^2(1-4\varepsilon\log 2)}. \quad (4.20)$$

Since $\varepsilon < \frac{1}{8\log 2}$ the contribution from the set $\mathcal{S}_2^X$ is negligible by definition of τ_N.

(3) We now analyze the contribution from the set $\mathcal{S}_3^X$, where

$$\mathcal{S}_3 = \big\{(\sigma^1,\sigma^2) \in S_N^2 : R(\sigma^1,\sigma^2) \text{ satisfies (2.39)}\big\}. \quad (4.21)$$

Let R_N^2 be the largest overlap value satisfying condition (2.39). To bound the contribution from the set $\mathcal{S}_3$ we represent it as a union of sets $V_2(R_{N,k})$, where $R_{N,k} = 1 - \frac{2k}{N}$ runs over the set $K_N \cap [R_N^1, R_N^2]$. Proceeding as in (2) and using Theorem 2.4 (ii), we bound it by

$$\sum_{R_{N,k}} 2\sqrt{\frac{2}{\pi N}} N^c \frac{C}{\sqrt{1-b_{12}^2}} e^{\frac{2b_{12}}{1+b_{12}} M\log 2 - N\mathcal{J}(R_{N,k})}. \quad (4.22)$$

Again, the sum is monotone in M and thus it is enough to bound it for $M = \varepsilon N$. For $R_{N,k}$ satisfying condition (2.39) we can bound the exponent in (4.22) as follows:

$$\begin{aligned} \frac{2b_{12}}{1+b_{12}}\varepsilon N\log 2 - N\mathcal{J}(R_{N,k}) &\leq \frac{2b_{12}}{1+b_{12}}\varepsilon N \log 2 - N\varepsilon\log 2 - c_1\log N \\ &= -\frac{1-b_{12}}{1+b_{12}}\varepsilon N\log 2 - c_1 \log N \leq -CN\varepsilon\log 2 - c_1\log N, \end{aligned} \quad (4.23)$$

where C is some positive constant.

Since $1 - b_{12}^2 \geq 1/N$ and since there are at most N terms in the sum (4.22), we can bound the latter by $N^{c+1-c_1} \exp\big(-CN\varepsilon \log 2\big)$. Therefore, P-almost surely the contribution from the set $\mathcal{S}_3^X$ is negligible as $N \to \infty$.

(4) To finish the second moment estimate it remains to treat the set

$$\mathcal{S}_4 = \big\{(\sigma^1, \sigma^2) \in S_N^2 : R(\sigma^1, \sigma^2) \text{ satisfies (2.41)}\big\}. \tag{4.24}$$

But by Theorem 2.4 (iii), the set $\mathcal{S}_4^X$ is P-almost surely empty. This finishes the proof of assertion (i) of Theorem 4.2.

3. **Third moment estimate.** To analyze the third factorial moment we use that, by formula (3.5) and definition (2.51), it can be written as

$$\mathbb{E}\big(\mathcal{P}_N(A)\big)_3 = \sum_{R_{12}, R_{23}, R_{13}} \frac{|W_{N,3}^X|}{2^{3M}\sqrt{\det B}} e^{(3-(\vec{1}, B^{-1}\vec{1}))(M \log 2 - 1/2 \log M)}$$
$$\times \int\limits_A \int\limits_A \int\limits_A e^{-(\vec{x}, B^{-1}\vec{x}) b_N^2/2 - a_N b_N (\vec{x}, B^{-1}\vec{1})} d\vec{x}, \tag{4.25}$$

where $B = B(\sigma^1, \sigma^2, \sigma^3)$, the covariance matrix of the vector $(H_N(\sigma^1), H_N(\sigma^2), H_N(\sigma^3))$, is the matrix with elements $b_{ij} = R_{ij}^2, 1 \leq i,j \leq 3$, and where the summation runs over all triplets of overlaps $(R_{12}, R_{23}, R_{31}) \in K_N^3$. To estimate this sum we rely on three auxiliary lemmas whose proofs we skip since they are simple.

Lemma 4.4. *If* $\varepsilon < \frac{1}{8 \log 2}$ *then*

$$\limsup_{N\to\infty} \max_{\sigma^1, \sigma^2, \sigma^3 \in X} \int\limits_A \int\limits_A \int\limits_A e^{-(\vec{x}, B^{-1}\vec{x}) b_N^2/2 - a_N b_N (\vec{x}, B^{-1}\vec{1})} d\vec{x} < \infty, \quad \text{P-}a.s. \tag{4.26}$$

Lemma 4.5. P*-almost surely for all configurations* $\sigma^1, \sigma^2, \sigma^3 \in X$

$$3 - (\vec{1}, B^{-1}\vec{1}) \leq 2(R_{12}^2 + R_{23}^2 + R_{31}^2). \tag{4.27}$$

Lemma 4.6. *For all* $\sigma^1, \sigma^2, \sigma^3 \in S_N$

$$\mathcal{J}^{(2)}(R_{12}, R_{23}, R_{31}) \geq \frac{1}{4}\Big(R_{12}^2 + R_{23}^2 + R_{31}^2\Big). \tag{4.28}$$

We are now ready to estimate sum (4.25). In the same spirit as for the second moment calculation, we will split (4.25) into four parts and show that every each of them is bounded. Moreover, by monotonicity argument similar to that used in calculation of the second moment we can restrict our attention to the case when $M = \varepsilon N$.

(1) We first calculate the contribution to (4.25) coming from the set

$$\mathcal{S}_1 = \big\{(\sigma^1, \sigma^2, \sigma^3) \in S_N^3 : \; \max\big\{|R_{12}|, |R_{23}|, |R_{31}|\big\} \leq \tau_N\big\}, \tag{4.29}$$

where $\tau_N = \frac{1}{N^{1/4} \log N}$. Using Theorem 2.5 (i) and Lemma 4.5, we obtain that the contribution from $\mathcal{S}_1$ is at most of order

$$\frac{1}{N^{3/2}} \sum_{R_{12}, R_{23}, R_{31}} e^{-N\mathcal{J}^{(2)}(R_{12}, R_{23}, R_{31}) + 2(R_{12}^2 + R_{23}^2 + R_{31}^2)M \log 2}, \tag{4.30}$$

where the summation is over $R_{12}, R_{23}, R_{31} \in K_N \cap [-\tau_N, \tau_N]$. Expanding in Taylor series we obtain that for $|R_{12}|, |R_{23}|, |R_{31}| \le \tau_N$

$$\mathcal{J}^{(2)}(R_{12}, R_{23}, R_{31}) = \frac{1}{2}\big(R_{12}^2 + R_{23}^2 + R_{31}^2\big) + O(\tau_N^4), \tag{4.31}$$

and thus (4.30) is bounded by

$$\frac{1}{N^{3/2}} \sum_{R_{12},R_{23},R_{31}} e^{-\frac{1}{2}N(R_{12}^2+R_{23}^2+R_{31}^2)+2\varepsilon N(R_{12}^2+R_{23}^2+R_{31}^2)\log 2}$$
$$= \left(\frac{1}{\sqrt{N}} \sum_{R_N \in K_N \cap [-\tau_N, \tau_N]} e^{-\frac{1}{2}NR^2(1-4\varepsilon\log 2)}\right)^3 < \infty. \tag{4.32}$$

(2) We next calculate the contribution from the set

$$\mathcal{S}_2 = \big\{(\sigma^1, \sigma^2, \sigma^3) \in S_N^3 : R_{12}, R_{23}, R_{21} \text{ satisfy (2.53)} \\ \text{and } \max\big\{|R_{12}|, |R_{23}|, |R_{31}|\big\} > \tau_N\big\}. \tag{4.33}$$

Without loss of generality we can assume that $|R_{12}| > \tau_N$. Then, using Theorem 2.5 (i) and Lemma 4.6, the contribution from this set is at most of order

$$\sum_{R_{12},R_{23},R_{31}} \frac{e^{-N\mathcal{J}^{(2)}(R_{12},R_{23},R_{31})}}{N^{3/2}P(R_{12}, R_{23}, R_{31})} e^{2(R_{12}^2+R_{23}^2+R_{31}^2)M\log 2}, \tag{4.34}$$

where $P(x,y,z) = (1+x+y+z)(1+x-y-z)(1-x+y-z)(1-x-y+z)$ and where the sum is over the triplets $(R_{12}, R_{23}, R_{31}) \in K_N^3$ satisfying (2.53) and $|R_{12}| > \tau_N$. Then, from Lemma 4.5, we further get that the sum (4.34) is bounded by

$$\sum_{|R_{12}|>\tau_N} \frac{e^{-\frac{1}{4}N(R_{12}^2+R_{23}^2+R_{31}^2)}}{N^{3/2}P(R_{12}, R_{23}, R_{31})} e^{2(R_{12}^2+R_{23}^2+R_{31}^2)M\log 2}$$
$$= \sum_{|R_{12}|>\tau_N} \frac{e^{-\frac{1}{4}N(1-8\varepsilon\log 2)(R_{12}^2+R_{23}^2+R_{31}^2)}}{N^{3/2}P(R_{12}, R_{23}, R_{31})} \tag{4.35}$$

Since the number of terms in the sum is at most $(N+1)^3$, and since $e^{-\frac{1}{4}(1-8\varepsilon\log 2)N\tau_N^2}$ decreases faster than any polynomial it follows that the sum is of order $o(1)$.

(3) We now turn to the contribution from the set

$$\mathcal{S}_3 = \big\{(\sigma^1, \sigma^2, \sigma^3) \in S_N^3 :\ R_{12}, R_{23}, R_{31} \text{ satisfy (2.55)}\big\}. \tag{4.36}$$

By Lemma 2.5 (ii) the contribution from this set is at most of order

$$\sum_{R_{12},R_{23},R_{31}} \frac{N^c\,\mathrm{E}|W_{N,3}^X|}{2^{3M}\sqrt{\det B}} e^{(3-(\vec{1},B^{-1}\vec{1}))(M\log 2-1/2\log M)}, \tag{4.37}$$

where the summation is over the triplets $(R_{12}, R_{23}, R_{31}) \in K_N$ satisfying (2.55). Since $\mathrm{E}|W_{N,3}^X|$ is of order

$$\frac{2^{3M}e^{-N\mathcal{J}^{(2)}(R_{12},R_{23},R_{31})}}{N^{3/2}P(R_{12}, R_{23}, R_{31})} \tag{4.38}$$

we obtain, using Lemmas 4.5 and 4.6, that the sum (4.37) is bounded by

$$\sum_{R_{12},R_{23},R_{31}} \frac{N^c e^{-\frac{1}{4}N(R_{12}^2+R_{23}^2+R_{31}^2)}}{N^{3/2}P(R_{12},R_{23},R_{31})} e^{2\varepsilon N \log 2(R_{12}^2+R_{23}^2+R_{31}^2)}. \tag{4.39}$$

It is easy to show that the triplet (R_{12}, R_{23}, R_{31}) that satisfy (2.55) must satisfy either $|R_{23}| > \tau_N$ or $|R_{31}| > \tau_N$. Therefore we can further bound the contribution from the set $\mathcal{S}_3$ by the sum

$$\sum_{|R_{23}|>\tau_N \text{ or } |R_{31}|>\tau_N} \frac{N^c e^{-\frac{1}{4}N(R_{12}^2+R_{23}^2+R_{31}^2)}}{N^{3/2}P(R_{12},R_{23},R_{31})} e^{2\varepsilon N \log 2(R_{12}^2+R_{23}^2+R_{31}^2)}, \tag{4.40}$$

which is $o(1)$ by the same argument as in part (2).

(4) To finish the estimate of the third factorial moment we have to estimate the contribution to (4.25) coming from the set

$$\mathcal{S}_4 = \big\{(\sigma^1,\sigma^2,\sigma^3) \in S_N^3 : R_{12}, R_{23}, R_{21} \text{ satisfy (2.57)}\big\}. \tag{4.41}$$

By Theorem 2.5 (iii) the set $\mathcal{S}_4^X$ is P-a.s. empty and therefore its contribution is P-a.s. zero. This finishes the proof of assertion (ii) of Theorem 4.2. The proof of Theorem 4.2 is now complete. □

For comparison with the cases $\nu(r) = r^p$ for $p = 1$ and $p = 2$, we give here the asymptotic behavior of the first three factorial moments for the case $\nu(0) = \nu'(0) = 0$, i.e for instance for the case $\nu(r) = r^p$ of pure p-spins when $p \geq 3$. This behavior is compatible with a Poisson convergence theorem.

Theorem 4.7. *Assume that $\nu(0) = \nu'(0) = 0$. For every bounded Borel set A*

$$\lim \mathbb{E}(\mathcal{P}_N(A))_1 = \mu(A), \quad \text{P-}a.s. \tag{4.42}$$

Moreover, if $\limsup \frac{M(N)}{N} < \frac{1}{8\log 2}$ *then* P-*a.s.*

(i) $\lim \mathbb{E}(\mathcal{P}_N(A))_2 = (\mu(A))^2$;

(ii) $\lim \mathbb{E}(\mathcal{P}_N(A))_3 = (\mu(A))^3 < \infty$.

We do not include a proof of this last statement, which again follows the same strategy as the proof of Theorem 4.2.

5. Universality for Non-Gaussian Hamiltonians

In this section we extend the results of the previous sections to the case of non-Gaussian Hamiltonians. We are able to make this extension only for the pure p-spin models, i.e. $\nu(r) = r^p$. In this case we recall that the Hamiltonian is defined as

$$H_N(\sigma) = \frac{1}{\sqrt{N}} H_{N,p} = \frac{1}{\sqrt{N^p}} \sum_{1\leq i_1,\ldots,i_p\leq N} g_{i_1,\ldots,i_p}\sigma_{i_1}\ldots\sigma_{i_p}. \tag{5.1}$$

Our assumptions on the random variables $(g_{i_1,\ldots,i_p})_{1\leq i_1,\ldots,i_p\leq N}$ in (5.1) are the same as were made in [BCMN05a] and [BCMN05b] for the number partitioning problem. That is, we assume that their distribution function admits a density $\rho(x)$ that satisfies the following conditions:

(1) $\rho(x)$ is even;
(2) $\int x^2\rho(x)dx = 1$;

(3) for some $\epsilon > 0$

$$\int_{-\infty}^{\infty} \rho(x)^{1+\epsilon} dx < \infty; \tag{5.2}$$

(4) $\rho(x)$ has a Fourier transform that is analytic in some neighborhood of zero. We write

$$-\log \hat{\rho}(z) = \frac{1}{2}(2\pi)^2 z^2 + c_4 (2\pi)^4 z^4 + O(|z|^6). \tag{5.3}$$

Note that the inequality $\mathbb{E}(X^4) \geq \mathbb{E}(X^2)^2$ implies that necessarily $c_4 < \frac{1}{12}$.

Under these assumptions we will show, using the method introduced by C. Borgs, J. Chayes, S. Mertens and C. Nair in [BCMN05b], that Theorems 1.1 and 1.4 still hold.

5.1. **Proof of Universality.** In this subsection we fix $p \geq 1$ and prove the analog of Theorem 1.1 in the non-Gaussian case assuming that the Hamiltonian is given by (5.1), and that the random variables $(g_{i_1,\dots,i_p})_{1 \leq i_1,\dots,i_p \leq N}$ satisfy conditions (1)–(4) above.

Theorem 5.1 (Universality in the Non-Gaussian case)**.** *Assume $M(N) = o(\sqrt{N})$ for $p = 1$ and $M = o(N)$ for $p \geq 2$. Then* P*-almost surely the sequence of point processes $\mathcal{P}_N$ converges weakly to a Poisson point process $\mathcal{P}$ on $\mathbb{R}$ with intensity given by*

$$\mu(dt) = \frac{1}{\sqrt{\pi}} e^{-t\sqrt{2\log 2}} dt. \tag{5.4}$$

To prove Theorem 5.1 we essentially prove a local limit theorem. More precisely, for any fixed N let us introduce the Gaussian process Z_N on S_N that has the same mean and covariance matrix as the process $H'_N(\sigma)$ defined in (1.2). We will prove in Theorem 5.2 that P-a.s., for all sequences $(\sigma^1, \dots, \sigma^\ell) \in X^\ell$, the joint density of the random variables $H'_N(\sigma^1), \dots, H'_N(\sigma^\ell)$ is well approximated by the joint density of $Z_N(\sigma^1), \dots, Z_N(\sigma^\ell)$.

Theorem 5.2. *Assume $M(N) = o(\sqrt{N})$ for $p = 1$ and $M = o(N)$ for $p \geq 2$. Then* P*-almost surely for every $\ell \geq 1$ and every bounded Borel set A there exists $c > 0$ such that uniformly in $(\sigma^1, \dots, \sigma^\ell) \in X^\ell$*

$$\begin{aligned} \mathbb{P}(H'_N(\sigma^j) \in A, j = 1, \dots, \ell) &= \mathbb{P}(Z_N(\sigma^j) \in A, j = 1, \dots, \ell) \\ &\times \Big(1 + O(R_{\max}(\sigma^1, \dots, \sigma^\ell)) + O\Big(\frac{M^2}{N^p}\Big)\Big) + O(e^{-cN^p}). \end{aligned} \tag{5.5}$$

Applying Theorem 5.2 together with Theorem 1.1 we get from formula (3.5) that

$$\mathbb{E}(\mathcal{P}_N(A))_\ell = (\mu(A))^\ell (1 + o(1)) + O(2^{M\ell} e^{-cN^p}) \to (\mu(A))^\ell, \tag{5.6}$$

which, by Lemma 3.1, implies weak convergence of the sequence of point processes $\mathcal{P}_N$ to a Poisson point process with intensity measure μ, thus implying Theorem 5.1. We therefore focus on the proof of Theorem 5.2.

Proof. First, we obtain from the definition of $H'_N(\sigma)$ that

$$\begin{aligned} \Big\{H'_N(\sigma) \in (x, x + \triangle x)\Big\} = &\Big\{ \sum_{1 \leq i_1,\dots,i_p \leq N} g_{i_1,\dots,i_p} \sigma_{i_1} \dots \sigma_{i_p} \in \\ &(a_N + x b_N, a_N + (x + \triangle x) b_N) \sqrt{N^p} \Big\}. \end{aligned} \tag{5.7}$$

Following [BCMN05b] we get an integral representation of the indicator function

$$\mathbf{1}_{H'_N(\sigma)\in(x,x+\triangle x)} = \triangle x\, b_N\sqrt{n}\int_{-\infty}^{\infty}\operatorname{sinc}\big(f\triangle x b_N\sqrt{n}\big)\,e^{2\pi i f\sum g_{i_1,\dots,i_p}\sigma_{i_1}\dots\sigma_{i_p}-2\pi i f\alpha_N\sqrt{n}}df, \tag{5.8}$$

where, for brevity, we wrote $n = N^p, \alpha_N = a_N+b_N(x+\triangle x/2), \operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$, and where the sum in the exponent runs over all possible sequences $1 \le i_1, \dots, i_p \le N$.

Changing the integration variable in (5.8) from f to $-f$ and applying the resulting formula to the product of indicator functions we arrive at the following representation

$$\prod_{j=1}^{\ell}\mathbf{1}_{H'_N(\sigma^j)\in(x_j,x_j+\triangle x_j)} = \prod_{j=1}^{\ell}\triangle x_j\, b_N\sqrt{n} \times \iiint_{-\infty}^{\infty}\prod_{j=1}^{\ell}\operatorname{sinc}\big(f_j\triangle x_j b_N\sqrt{n}\big)\,e^{-2\pi i f_j\sum g_{i_1\dots i_p}\sigma^j_{i_1}\dots\sigma^j_{i_p}+2\pi i f_j\alpha_N^{(j)}\sqrt{n}}df_j, \tag{5.9}$$

where $\alpha_N^{(j)} = a_N + b_N(x_j + \triangle x_j/2)$. Introducing the variables

$$v_{i_1,\dots,i_p} = \sum_{j=1}^{\ell} f_j\sigma^j_{i_1}\dots\sigma^j_{i_p}, \tag{5.10}$$

we rewrite the integral in the above formula as

$$\iiint_{-\infty}^{\infty}\prod_{1\le i_1,\dots,i_p\le N} e^{-2\pi i g_{i_1\dots i_p}v_{i_1,\dots,i_p}}\prod_{j=1}^{\ell}\operatorname{sinc}\big(f_j\triangle x_j\sqrt{n}\big)e^{2\pi i f_j\alpha_N^{(j)}\sqrt{n}}df_j. \tag{5.11}$$

To get an integral representation of the joint density

$$\mathbb{P}\big(H'_N(\sigma^1)\in(x_1,x_1+dx_1),\dots,H'_N(\sigma^\ell)\in(x_\ell,x_\ell+dx_\ell)\big) \tag{5.12}$$

we have to take the expectation of (5.9) and then let $\triangle x_j \to 0$ for all $j = 1, 2, \dots \ell$. As was proved in [BCMN05a] (see Lemma 3.4), the exchange of expectation and integration for $p = 1$ is justified when the rank of the matrix formed by the row vectors $\sigma^1, \dots, \sigma^\ell$, is ℓ. To justify the exchange in our case we introduce an ℓ by N^p matrix, $C_p(\sigma^1, \dots, \sigma^\ell)$, defined as follows: for any given set of configurations $\sigma^1, \dots, \sigma^\ell$, the j-th row is composed of all N^p products, $\sigma^j_{i_1}\sigma^j_{i_2}\dots\sigma^j_{i_p}$ over all subsets $1 \le i_1, \dots, i_p \le N$. By generalizing the arguments from [BCMN05a] the exchange can then be justified provided that the rank of the matrix $C_p(\sigma^1, \dots, \sigma^\ell)$ is ℓ. As we will see in Lemma 5.3 below this holds true P-almost surely when $M = o(N)$.

Given a vector $\boldsymbol{\delta} \in \{-1, 1\}^\ell$ let $n_{\boldsymbol{\delta}}$ be the number of times the column vector $\boldsymbol{\delta}$ appears in the matrix C_p :

$$n_{\boldsymbol{\delta}} = n_{\boldsymbol{\delta}}(\sigma^1, \dots, \sigma^\ell) = \big|\big\{j \le N^p : (\sigma^1_j, \dots, \sigma^\ell_j) = \boldsymbol{\delta}\big\}\big|\,. \tag{5.13}$$

With this notation we have:

Lemma 5.3. *Suppose $M = o(N)$. Then there exists a sequence $\lambda_N = o(1)$ such that* P*-almost surely for all collections $(\sigma^1, \dots, \sigma^\ell) \in X^\ell$*

$$\max_{\boldsymbol{\delta}\in\{-1,1\}^\ell}\Big|n_{\boldsymbol{\delta}} - \frac{n}{2^\ell}\Big| \le n\lambda_N. \tag{5.14}$$

Proof. We first prove by induction that the following simple fact holds true: if for a given sequence of configurations $(\sigma^1,\dots,\sigma^\ell)\in S_N^\ell$ the matrix $C_1(\sigma^1,\dots,\sigma^\ell)$ satisfies condition (5.14) then necessarily the matrix $C_p(\sigma^1,\dots,\sigma^\ell)$ satisfies (5.14) for all $p\ge 1$.

For $p=1$ there is nothing to prove. We now assume that the statement is true for the matrix $C_{p-1}(\sigma^1,\dots,\sigma^\ell)$ and prove it for $C_p(\sigma^1,\dots,\sigma^\ell)$.

Let $\sigma_\mu, 1\le\mu\le N$ denote the columns of the matrix $C_1(\sigma^1,\dots,\sigma^\ell)$. For every column vector σ_μ let us construct a matrix $C^\mu_{p-1}=C^\mu_{p-1}(\sigma^1,\dots,\sigma^\ell)$ with entries

$$(C^\mu_{p-1})_{ij}=(\sigma_\mu)_j\ (C_{p-1})_{ij}. \tag{5.15}$$

For future convenience let $n^\mu_{\boldsymbol\delta}$ denote the variable $n_{\boldsymbol\delta}$ for the matrix C^μ_{p-1}. ¿From the inductive assumption it follows that for all $1\le\mu\le N$

$$\max_{\boldsymbol\delta}\Big|n^\mu_{\boldsymbol\delta}-\frac{N^{p-1}}{2^\ell}\Big|\le N^{p-1}\lambda_N. \tag{5.16}$$

Now note that the $\ell\times N^p$ matrix $C_p(\sigma^1,\dots,\sigma^\ell)$ can be obtained by concatenating N matrices $C^\mu_{p-1}(\sigma^1,\dots,\sigma^\ell)$ each of size $\ell\times N^{p-1}$. Therefore for any sequence of configurations $(\sigma^1,\dots,\sigma^\ell)$ with matrix $C_1(\sigma^1,\dots,\sigma^\ell)$ satisfying (5.14) we have

$$\begin{aligned}\Big|n_{\boldsymbol\delta}-\frac{N^p}{2^\ell}\Big| &\le \Big|n^{\mu_1}_{\boldsymbol\delta}-\frac{N^{p-1}}{2^\ell}\Big|+\dots+\Big|n^{\mu_N}_{\boldsymbol\delta}-\frac{N^{p-1}}{2^\ell}\Big|\le\\ &\le N^{p-1}\lambda_N+\dots+N^{p-1}\lambda_N=N^p\lambda_N\end{aligned} \tag{5.17}$$

and the induction is complete.

To prove Lemma 5.3 it is thus enough to demonstrate that P-almost surely there are no sequences $(\sigma^1,\dots,\sigma^\ell)\in X^\ell$ such that $C_1(\sigma^1,\dots,\sigma^\ell)$ violates condition (5.14). Let us prove this by induction in ℓ.

For $\ell=1$ let us introduce the sets

$$\mathcal{T}_N=\Big\{\sigma\in S:\max_{\boldsymbol\delta\in\{-1,1\}}\Big|n_{\boldsymbol\delta}-\frac{N}{2}\Big|\le N\lambda_N\Big\}. \tag{5.18}$$

Then by Chernoff bound

$$|\mathcal{T}_N^c|=2\sum_{i\ge N\lambda_N}\binom{N}{\frac{N}{2}+i}\le 2^{N+1}e^{-\frac12 N((1+\lambda_N)\log(1+\lambda_N)-\lambda_N)}. \tag{5.19}$$

Let us choose $\lambda_N=o(1)$ in such a way that $M=o(N\lambda_N^2)$ and $\log N=o(N\lambda_N^2)$. Then using (2.12) we obtain that

$$\sum_{\sigma\in\mathcal{T}_N^c}\mathbf{1}_\sigma=0,\quad \text{P-a.s.} \tag{5.20}$$

which proves the statement for $\ell=1$. Now assume that P-a.s. for all sequences $(\sigma^1,\dots,\sigma^{\ell-1})\in X^{(\ell-1)}$ the matrix $C_1(\sigma^1,\dots,\sigma^{\ell-1})$ satisfies condition (5.14). Since there is only a countable number of sequences $(\sigma^1,\dots,\sigma^{\ell-1})$ we fix $(\sigma^1,\dots,\sigma^{\ell-1})\in X^{(\ell-1)}$ and prove that P-almost surely there are no configurations σ^ℓ such that $C_1(\sigma^1,\dots,\sigma^\ell)$ violates (5.14).

Let $\boldsymbol\delta\in\{-1,1\}^\ell$ be given. Define $\boldsymbol\delta_1(\boldsymbol\delta)\in\{-1,1\}^{\ell-1}$ as $(\boldsymbol\delta_1)_i=(\boldsymbol\delta)_i$ for $1\le i\le \ell-1$ and also define $\boldsymbol\delta_2(\boldsymbol\delta)\in\{-1,1\}$ as $\boldsymbol\delta_2=(\boldsymbol\delta)_\ell$. Let us also introduce, for given $\boldsymbol\delta_1\in\{-1,1\}^{\ell-1}$, the set

$$N_{\boldsymbol\delta_1}=\{j\le N:(\sigma^1_j,\dots,\sigma^{\ell-1}_j)=\boldsymbol\delta_1\}. \tag{5.21}$$

By the inductive assumption we conclude that for all $\boldsymbol{\delta}_1 \in \{-1,1\}^{\ell-1}$

$$\Big| |N_{\boldsymbol{\delta}_1}| - \frac{N}{2^{\ell-1}} \Big| \le N\lambda_N. \tag{5.22}$$

From the definition of $N_{\boldsymbol{\delta}}$ it is not hard to see that for all $\boldsymbol{\delta} \in \{-1,1\}^{\ell}$

$$n_{\boldsymbol{\delta}} = \big| \{ j \in N_{\boldsymbol{\delta}_1(\boldsymbol{\delta})} : \sigma_j^{\ell} = \boldsymbol{\delta}_2(\boldsymbol{\delta}) \} \big|. \tag{5.23}$$

Using the above relation together with (2.11) and the assumptions on λ_N we get that P-almost surely

$$\Big| n_{\boldsymbol{\delta}} - \frac{|N_{\boldsymbol{\delta}_1(\boldsymbol{\delta})}|}{2} \Big| \le \frac{N\lambda_N}{2}. \tag{5.24}$$

Therefore from (5.22) and (5.24)

$$\Big| n_{\boldsymbol{\delta}} - \frac{N}{2^{\ell}} \Big| \le \Big| n_{\boldsymbol{\delta}} - \frac{|N_{\boldsymbol{\delta}_1(\boldsymbol{\delta})}|}{2} \Big| + \Big| \frac{|N_{\boldsymbol{\delta}_1(\boldsymbol{\delta})}|}{2} - \frac{N}{2^{\ell}} \Big| \le \frac{N\lambda_N}{2} + \frac{N\lambda_N}{2} = N\lambda_N. \tag{5.25}$$

The induction is now complete and the lemma is proved. □

Lemma 5.3 implies that P-almost surely, for all $(\sigma^1, \dots, \sigma^{\ell}) \in X^{\ell}$,

$$n_{\min}(\sigma^1, \dots, \sigma^{\ell}) = \min_{\boldsymbol{\delta} \in \{-1,1\}^{\ell}} n_{\boldsymbol{\delta}} = \frac{n}{2^{\ell}}(1 + O(\lambda_N)), \tag{5.26}$$

and hence, for sufficiently large N, the rank of the matrix $C_p(\sigma^1, \dots, \sigma^{\ell})$ is ℓ. The exchange of integration and expectation is thus justified.

Using once again Lemma 5.3, condition (5.2), and the dominated convergence theorem we obtain that the joint density is

$$\mathbb{P}\big(H_N'(\sigma^j) \in (x_j, x_j + dx_j) \text{ for } j = 1, \dots, \ell\big) = \prod_{j=1}^{\ell} b_N \sqrt{n}\, dx_j$$
$$\times \iiint_{-\infty}^{\infty} \prod_{1 \le i_1, \dots, i_p \le N} \hat{\rho}(v_{i_1,\dots,i_p}) \prod_{j=1}^{\ell} e^{2\pi i f_j \alpha_N^{(j)} \sqrt{n}} df_j, \tag{5.27}$$

where we redefined $\alpha_N^{(j)} = a_N + b_N x_j$. We remark for future use that $\alpha_N^{(j)} = O(a_N)$ for all $1 \le j \le \ell$.

It is straightforward at this point to adapt the saddle point analysis used in [BCMN05b] to calculate the integrals of such type. The only difference is that instead of the matrix $C_1(\sigma^1, \dots, \sigma^{\ell})$ with rows formed by row vectors $\sigma^1, \dots, \sigma^{\ell}$, we use the matrix $C_p(\sigma^1, \dots, \sigma^{\ell})$. By analogy with Lemma 5.3 from [BCMN05b] we first approximate the integral in (5.27) by an integral over a bounded domain, i.e. for some $c_1 > 0$ depending on $\mu_1 > 0$

$$\iiint_{-\infty}^{\infty} \prod_{i_1,\dots,i_p} \hat{\rho}(v_{i_1,\dots,i_p}) \prod_{j=1}^{\ell} e^{2\pi i f_j \alpha_N^{(j)} \sqrt{n}} df_j =$$
$$\iiint_{-\mu_1}^{\mu_1} \prod_{i_1,\dots,i_p} \hat{\rho}(v_{i_1,\dots,i_p}) \prod_{j=1}^{\ell} e^{2\pi i f_j \alpha_N^{(j)} \sqrt{n}} df_j + O(e^{-c_1 n_{\min}}). \tag{5.28}$$

We next rewrite the integral in the r.h.s. of (5.28) as

$$\iiint_{-\mu_1}^{\mu_1} e^{2\pi i n \mathbf{f}\cdot\boldsymbol{\alpha}} \prod_{\boldsymbol{\delta}\in\{-1,1\}^\ell} \hat\rho(\mathbf{f}\cdot\boldsymbol{\delta})^{n_\delta} \prod_{j=1}^{\ell} df_j, \tag{5.29}$$

where $\boldsymbol{\alpha} = \big(\frac{\alpha_N^{(1)}}{\sqrt{n}}, \dots, \frac{\alpha_N^{(\ell)}}{\sqrt{n}}\big)$, $\boldsymbol{f} = (f_1, \dots, f_\ell)$, and where $\boldsymbol{\alpha}\cdot\boldsymbol{f} = \alpha_1 f_1 + \dots + \alpha_\ell f_\ell$ is the standard scalar product.

Using Lemma 5.3 again we can apply Lemma 5.4 from [BCMN05b] to conclude that given μ_1 there are constants $c_1(\mu_1) > 0$ and $\mu_2 > 0$ such that the following equality holds whenever $\eta_1, \dots, \eta_\ell$ is a sequence of real numbers with $\sum_j |\eta_j| \le \mu_2$ and $\eta_j \alpha_N^{(j)} \ge 0$ for all $j = 1, \dots, \ell$

$$\iiint_{-\mu_1}^{\mu_1} \prod_{\boldsymbol{\delta}} \hat\rho(\mathbf{f}\cdot\boldsymbol{\delta})^{n_\delta} \prod_{j=1}^{\ell} e^{2\pi i n f_j \alpha_N^{(j)}} df_j \tag{5.30}$$

$$= \iiint_{-\mu_1}^{\mu_1} e^{2\pi n (i\mathbf{f}\cdot\boldsymbol{\alpha} - \boldsymbol{\eta}\cdot\boldsymbol{\alpha})} \prod_{\boldsymbol{\delta}} \hat\rho(\mathbf{f}\cdot\boldsymbol{\delta} + i\boldsymbol{\eta}\cdot\boldsymbol{\delta})^{n_\delta} \prod_{j=1}^{\ell} df_j + O(e^{-\frac{1}{2}c_1 n_{\min}}).$$

The values of the shifts $\eta_1, \dots, \eta_\ell$ are determined by the following system:

$$\sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} \delta_j F'(i\boldsymbol{\delta}\cdot\boldsymbol{\eta}) = 2\pi i \frac{\alpha_N^{(j)}}{\sqrt{n}}, \quad j = 1, \dots, \ell, \tag{5.31}$$

where we wrote $F = -\log\hat\rho$.

Since $\max\limits_{\sigma,\sigma'\in X} |R(\sigma,\sigma')|$ is P-almost surely of order $o(1)$ when $M = o(N)$ we can apply Lemma 5.5 from [BCMN05b] and obtain that this system has a unique solution

$$\boldsymbol{\eta}(\boldsymbol{\alpha}) = \frac{1}{2\pi} B^{-1}\boldsymbol{\alpha}\left(1 + O\big(||\boldsymbol{\alpha}||_2^2\big)\right). \tag{5.32}$$

Moreover, for sufficiently small μ_1,

$$\iiint_{-\mu_1}^{\mu_1} e^{2\pi n (i\mathbf{f}\cdot\boldsymbol{\alpha} - \boldsymbol{\eta}\cdot\boldsymbol{\alpha})} \prod_{\boldsymbol{\delta}} \hat\rho(\mathbf{f}\cdot\boldsymbol{\delta} + i\boldsymbol{\eta}\cdot\boldsymbol{\delta})^{n_\delta} \prod_{j=1}^{\ell} df_j$$

$$= e^{-nG_{n,\ell}(\boldsymbol{\alpha})} \left(\frac{1}{2\pi n}\right)^{\ell/2} \big(1 + O(n^{-1/2}) + O(a_N^2/n) + O(R_{\max})\big), \tag{5.33}$$

where

$$G_{n,\ell}(\boldsymbol{\alpha}) = \sum_{\boldsymbol{\delta}} \frac{n_{\boldsymbol{\delta}}}{n} F(i\boldsymbol{\delta}\cdot\boldsymbol{\eta}(\boldsymbol{\alpha})) + 2\pi\boldsymbol{\eta}(\boldsymbol{\alpha})\cdot\boldsymbol{\alpha}. \tag{5.34}$$

Therefore

$$\mathbb{P}\big(H_N'(\sigma^j) \in (x_j, x_j + dx_j) \text{ for } j = 1, \dots, \ell\big) = \left(\frac{b_N}{\sqrt{2\pi}}\right)^{\ell} e^{-nG_{n,\ell}(\boldsymbol{\alpha})} \prod_{j=1}^{\ell} dx_j$$

$$\times \big(1 + O(n^{-1/2}) + O(a_N^2/n) + O(R_{\max})\big) + O(b_N^\ell n^{\ell/2} e^{-\frac{1}{2}c_1 n_{\min}}). \tag{5.35}$$

Expanding $G_{n,\ell}$ we get the approximation

$$nG_{n,\ell}(\boldsymbol{\alpha}) = \frac{n}{2}(\boldsymbol{\alpha}, B^{-1}\boldsymbol{\alpha}) + O\Big(\frac{a_N^4}{n}\Big). \tag{5.36}$$

By definition $a_N = O(\sqrt{M})$. Under the assumptions of Theorem 5.2 we get that $a_N = o(\sqrt[4]{N})$ for $p = 1$, that $a_N = o(\sqrt{N})$ for $p \geq 2$, and thus that $a_N^4 = o(n)$. It implies that asymptotically the joint density (5.12) is Gaussian. More precisely, it follows from the equations (5.35) and (5.36) that P-a.s., for all collections $(\sigma^1, \ldots, \sigma^\ell) \in X^\ell$,

$$\mathbb{P}\big(H_N'(\sigma^j) \in (x_j, x_j + dx_j) \text{ for } j = 1, \ldots, \ell\big) = \left(\frac{b_N}{\sqrt{2\pi}}\right)^\ell e^{-n(\boldsymbol{\alpha}, B^{-1}\boldsymbol{\alpha})/2}$$
$$\times \prod_{j=1}^{\ell} dx_j \left(1 + O(R_{\max}) + O\Big(\frac{a_N^4}{n}\Big)\right) + O(b_N^\ell n^{\ell/2} e^{-\frac{1}{2}c_1 n_{\min}}). \quad (5.37)$$

By Lemma 5.3 the term $O(b_N^\ell n^{\ell/2} e^{-\frac{1}{2}c_1 n_{\min}})$ is of order $o(e^{-cN^p})$ as $N \to \infty$. This finishes the proof of Theorem 5.2. □

5.2. **Breakdown of Universality.** In this subsection we follow the same strategy as we used in the proof of Theorem 1.4 – we fix a bounded set A and study the first three factorial moments of the random variable $\mathcal{P}_N(A)$. In the case of the number partitioning problem the following theorem implies that the Poisson convergence fails as soon as $\limsup M/\sqrt{N} > 0$.

Theorem 5.4 (Number partitioning problem)**.** *Fix* $p = 1$ *and let the Hamiltonian be given by* (5.1)*. For every bounded Borel set* A *we have*

$$\lim \mathbb{E}(\mathcal{P}_N(A))_1 = \mu(A) e^{-4c_4\varepsilon^2 \log^2 2}, \quad \text{P-}a.s. \quad (5.38)$$

Moreover, if $\limsup \frac{M(N)}{N} = \varepsilon < \infty$ *then* P-*a.s.*

(*i*) $\limsup \mathbb{E}(\mathcal{P}_N(A))_2 = e^{2\varepsilon^2 \log 2 - 32c_4\varepsilon^2 \log^2 2}(\mu(A))^2$;

(*ii*) $\limsup \mathbb{E}(\mathcal{P}_N(A))_3 < \infty$.

Therefore the limit of the ratio of the second factorial moment to the square of the first is

$$\frac{\mathbb{E}(\mathcal{P}_N(A))_2}{\mathbb{E}(\mathcal{P}_N(A))_1^2} \to e^{2\varepsilon^2 \log^2 2 - 24c_4\varepsilon^2 \log^2 2} = e^{2\varepsilon^2 \log^2 2(1 - 12c_4)}. \quad (5.39)$$

Taking into account that $c_4 < \frac{1}{12}$ we conclude that the ratio is strictly larger than one and thus there is no Poisson convergence for $\varepsilon > 0$.

And in the case of the Sherrington-Kirkpatrick model the failure of Poisson convergence follows from

Theorem 5.5 (Sherrington-Kirkpatrick model)**.** *Fix* $p = 2$ *and let the Hamiltonian be given by* (5.1)*. For every bounded Borel set* A

$$\lim \mathbb{E}(\mathcal{P}_N(A))_1 = \mu(A) e^{-4c_4\varepsilon^2 \log^2 2}, \quad \text{P-}a.s. \quad (5.40)$$

Moreover, if $\limsup \frac{M(N)}{N} = \varepsilon < \frac{1}{8 \log 2}$ *then* P-*a.s.*

(i) $\limsup \mathbb{E}(\mathcal{P}_N(A))_2 = \frac{e^{-32c_4\varepsilon^2 \log^2 2}}{\sqrt{1 - 4\varepsilon \log 2}}(\mu(A))^2$;

(ii) $\limsup \mathbb{E}(\mathcal{P}_N(A))_3 < \infty$.

The ratio of the second factorial moment to the square of the first moment is

$$\frac{\mathbb{E}(\mathcal{P}_N(A))_2}{\mathbb{E}(\mathcal{P}_N(A))_1^2} \to \frac{e^{-24c_4\varepsilon^2\log^2 2}}{\sqrt{1-4\varepsilon\log 2}}. \tag{5.41}$$

For $\varepsilon > 0$ the above ratio is strictly larger than one and thus convergence to a Poisson point process fails.

We will give the proof of Theorem 5.5 only since the case $p = 1$ is based on essentially the same computations.

Proof of Theorem 5.5. As in the proof of Theorem 4.2 we successively prove the statement on the first, second, and third moment. To simplify our computations we will assume that $M = \varepsilon N$ (Using the monotonicity argument, the case of general sequences $M(N)$ can be analyzed just as in Theorem 4.2 of Section 4.).

1. **First moment estimate.**

Following the same steps as in Subsection 5.1 we approximate the density of $H'_N(\sigma)$ by

$$e^{-nG_{n,1}(\alpha_N)}\frac{1}{\sqrt{2\pi}}\Big(1+O\Big(\frac{1}{\sqrt{N}}\Big)+O\Big(\frac{a_N^2}{n}\Big)\Big)+O(e^{-c_1 n}), \tag{5.42}$$

where according to the notation introduced above $n = N^2$.

In the case $M = \varepsilon N$ we need a more precise approximation of the function $G_{n,\ell}$ than given by formula (5.36). Expanding the solution of the system (5.31) as

$$\boldsymbol{\eta}(\boldsymbol{\alpha}) = \frac{1}{2\pi}B^{-1}\boldsymbol{\alpha} + \frac{4c_4}{2\pi}\sum_{\boldsymbol{\delta}}\frac{n_{\boldsymbol{\delta}}}{n}(\boldsymbol{\delta}, B^{-1}\boldsymbol{\alpha})^3 B^{-1}\boldsymbol{\delta} + O(||\alpha||^5) \tag{5.43}$$

and applying (5.43) we obtain from (5.34) and (5.3) that

$$\begin{aligned} G_{n,\ell}(\boldsymbol{\alpha}) &= -\frac{(2\pi)^2}{2}\sum_{\boldsymbol{\delta}}\frac{n_{\boldsymbol{\delta}}}{n}(\boldsymbol{\delta}\cdot\boldsymbol{\eta})^2 \\ &\quad + c_4(2\pi)^4\sum_{\boldsymbol{\delta}}\frac{n_{\boldsymbol{\delta}}}{n}(\boldsymbol{\delta}\cdot\boldsymbol{\eta})^4 + 2\pi(\boldsymbol{\eta}\cdot\boldsymbol{\alpha}) + O(||\boldsymbol{\alpha}||^6) \\ &= \frac{1}{2}(\boldsymbol{\alpha}, B^{-1}\boldsymbol{\alpha}) + c_4\sum_{\boldsymbol{\delta}}\frac{n_{\boldsymbol{\delta}}}{n}(\boldsymbol{\delta}, B^{-1}\boldsymbol{\alpha})^4 + O(||\boldsymbol{\alpha}||^6). \end{aligned} \tag{5.44}$$

Using the approximation for $G_{n,1}$ given by formula (5.44), we obtain

$$nG_{n,1} = \frac{\alpha_N^2}{2} + c_4\frac{\alpha_N^4}{n} + O\Big(\frac{a_N^6}{n^2}\Big), \tag{5.45}$$

where $\alpha_N = a_N + b_N\, x$. Since $\alpha_N^4/n = 4c_4\varepsilon^2\log^2 2(1+o(1))$, we see that up to an ε-dependent multiplier the density of $H'_N(\sigma)$ is given by the normal density, more precisely, the density of $H'_N(\sigma)$ is

$$\frac{1}{\sqrt{2\pi}}e^{-\alpha_N^2/2}\, e^{-4c_4\varepsilon^2\log^2 2}\Big(1+O\Big(\frac{1}{\sqrt{N}}\Big)+O\Big(\frac{\alpha_N^2}{n}\Big)\Big)+O(e^{-c_1 n}). \tag{5.46}$$

Therefore the first factorial moment of $\mathcal{P}_N(A)$ is

$$\mu(A)e^{-4c_4\varepsilon^2\log^2 2}(1+o(1)) + o(1) \tag{5.47}$$

and (5.40) is proven.

2. **Second moment estimate.** When analyzing the second moment

$$\mathbb{E}(\mathcal{P}_N(A))_2 = \sum_{(\sigma^1,\sigma^2)\in X^2} \mathbb{P}\left(H'_N(\sigma^1)\in A, H'_N(\sigma^2)\in A\right), \tag{5.48}$$

it is useful to distinguish between "typical" and "atypical" sets of configurations $(\sigma^1,\dots,\sigma^\ell)\in X^\ell$, a notion introduced in [BCMN05a].

Take some sequence $\theta_N\to 0$ such that $N\theta_N^2\to\infty$ and consider the $\ell\times n$ matrix $C_p(\sigma^1,\dots,\sigma^\ell)$, introduced in Subsection 5.1. Then all but a vanishing fraction of the configurations $(\sigma^1,\dots,\sigma^\ell)\in S_N^\ell$ obey the condition

$$\max_{\boldsymbol{\delta}\in\{-1,1\}^\ell}\left| n_{\boldsymbol{\delta}} - \frac{n}{2^\ell}\right| \le n\theta_N. \tag{5.49}$$

When $M=o(N)$, Lemma 5.3 guarantees that for a properly chosen sequences θ_N, P-almost surely, all the sampled sets $(\sigma^1,\dots,\sigma^\ell)\in X^\ell$ obey condition (5.49). It is no longer the case when $M=\varepsilon N$ and thus we have to consider the contribution from the sets violating (5.49).

Fix $p=2$ and $\ell=2$. Let a sequence $\theta_N\to 0$ be given, and define

$$I=\sum_{\sigma^1,\sigma^2}\mathbb{P}\big(H'_N(\sigma^1)\in A, H'_N(\sigma^2)\in A\big), \tag{5.50}$$

where the sum runs over all pairs of distinct configurations $(\sigma^1,\sigma^2)\in X^2$ satisfying condition (5.49) (the so called "typical" configurations). Also define

$$II=\sum_{\sigma^1,\sigma^2}\mathbb{P}\big(H'_N(\sigma^1)\in A, H'_N(\sigma^2)\in A\big), \tag{5.51}$$

where the sum is over all pairs of distinct configurations $(\sigma^1,\sigma^2)\in X^2$ violating (5.49) (the "atypical" configurations). For later use we introduce the quantities I_g and II_g – the analogs of the variables I and II in the case where the random variables $(g_{i_1,i_2})_{1\le i_1,i_2\le N}$ are i.i.d. standard normals.

Lemma 5.6. *Let* $\theta_N=\frac{1}{\sqrt{N}\log^2 N}$. *If* $\varepsilon\in(0,\frac{1}{4\log 2})$ *then* P*-almost surely* $II=o(1)$ *and*

$$I=\frac{e^{-32c_4\varepsilon^2\log^2 2}(\mu(A))^2}{\sqrt{1-4\varepsilon\log 2}}(1+o(1)). \tag{5.52}$$

Proof. To prove that II is $o(1)$ let us bound the quantity II by II_g and show that II_g is almost surely negligible for $\theta_N=\frac{1}{\sqrt{N}\log^2 N}$. We start with the proof of the second statement, for which we will need the following simple observation.

Consider a sequence $\lambda_N\to 0$ such that $N\lambda_N\to\infty$ and define the set of configurations $\sigma\in\{-1,1\}^N$ with almost equal number of spins equal to 1 and to -1 :

$$\mathcal{T}_N=\Big\{\sigma\in S_N : \big|\#\{\sigma_i=1\}-\#\{\sigma_i=-1\}\big|\le N\lambda_N\Big\}. \tag{5.53}$$

It is not hard to prove that configurations $\sigma^1,\sigma^2\in\mathcal{T}_N$ with overlap $|R(\sigma^1,\sigma^2)|\le\lambda_N$ must satisfy (5.49) with $\theta_N=\lambda_N^2$. Therefore the set of pairs $(\sigma^1,\sigma^2)\in X^2$ violating condition (5.49) with $\theta_N=\lambda_N^2$ is contained in the set

$$\Big\{(\sigma^1,\sigma^2)\in X^2 : |R(\sigma^1,\sigma^2)|>\lambda_N \text{ or } \sigma^1\in\mathcal{T}_N^c \text{ or } \sigma^2 >\in\mathcal{T}_N^c\Big\}. \tag{5.54}$$

Thus to prove that II_g is $o(1)$ for $\theta_N = \frac{1}{\sqrt{N}\log^2 N}$ it suffices to prove that

$$\sum_{\sigma^1,\sigma^2} \mathbb{P}(H'_N(\sigma^1) \in A, H'_N(\sigma^2) \in A) = o(1), \tag{5.55}$$

where the summation is over all pairs of distinct configurations contained in the set (5.54) with $\lambda_N = \sqrt{\theta_N} = \frac{1}{N^{1/4}\log N}$.

Let us prove (5.55). Since we already proved in Section 4 that the contribution from the set

$$\big\{(\sigma^1,\sigma^2) \in X^2 : |R(\sigma^1,\sigma^2)| > \lambda_N\big\} \tag{5.56}$$

to the sum (5.48) is negligible, it is enough to consider the sum (5.48) restricted to the set

$$\Big\{(\sigma^1,\sigma^2) \in X^2 : \sigma^1 \text{ or } \sigma^2 \in (\mathcal{T}_N^c)^X \text{ and } |R(\sigma^1,\sigma^2)| < \lambda_N\Big\}. \tag{5.57}$$

By Stirling's formula we obtain that $|\mathcal{T}_N^c| = \sqrt{\frac{2}{\pi N}} 2^N e^{-N\mathcal{J}(\lambda_N)}(1+O(\lambda_N^2))$ and using this fact one can prove, proceeding as in part (i) of Theorem 2.4, that

$$|(\mathcal{T}_N^c)^X| = \mathrm{E}|(\mathcal{T}_N^c)^X|(1+o(1)) = \sqrt{\frac{2}{\pi N}}\, 2^M e^{-N\mathcal{J}(\lambda_N)}(1+o(1)). \tag{5.58}$$

Thus for large enough N the contribution from the set (5.57) to the sum (5.48) is bounded by

$$2\sqrt{\frac{2}{\pi N}} 2^{2M} e^{-N\mathcal{J}(\lambda_N)} \frac{C}{2^{2M}} e^{\frac{2b12}{1+b_{12}}(M\log 2 - 1/2\log M)}, \tag{5.59}$$

where the constant C is from (4.17). Using that $N\mathcal{J}(\lambda_N) = \frac{1}{2}N\lambda_N^2 + O\Big(\frac{1}{\log^4 N}\Big)$ we can further bound (5.59) by

$$2C\sqrt{\frac{2}{\pi N}} e^{-\frac{1}{2}N\lambda_N^2(1-4\varepsilon\log 2)}, \tag{5.60}$$

which is $o(1)$ by the choice of λ_N and ε.

Our next step is to bound the sum II by II_g. For this purpose we need to give an estimate of the joint density of $H'_N(\sigma^1), H'_N(\sigma^2)$ that would be valid also for pairs (σ^1,σ^2) violating condition (5.49). As we already noted for such (σ^1,σ^2) the results of Subsection 5.1 cannot be applied directly since it is no longer true that $\max_{\sigma,\sigma'\in X} |R(\sigma,\sigma')|$ is $o(1)$. Fortunately, we have only 2×2 covariance matrix $B(\sigma^1,\sigma^2)$ and using this fact we can easily adapt the results of Subsection 5.1 to the case where $\max_{\sigma,\sigma'\in X} |R(\sigma,\sigma')|$ is not $o(1)$. We start with formula (5.27) which, in the case $\ell = 2$, can be rewritten as

$$\begin{aligned} &\mathbb{P}\big(H'_N(\sigma^j) \in (x_j, x_j+dx_j) \text{ for } j=1,2\big) \\ &\qquad = b_N^2 n\, dx_1 dx_2 \iint_{-\infty}^{\infty} \prod_{1\le i_1,i_2\le N} \hat\rho(v_{i_1,i_2}) e^{2\pi i\sqrt{n}\left(f_1\alpha_N^{(1)} + f_2\alpha_N^{(2)}\right)} df_1 df_2. \end{aligned} \tag{5.61}$$

We can rewrite the integral in the above expression as

$$\iint_{-\infty}^{\infty} \prod_{\boldsymbol{\delta}} \hat\rho(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}}\, e^{2\pi i n \mathbf{f}\cdot\boldsymbol{\alpha}} df_1 df_2, \tag{5.62}$$

where $\boldsymbol{\delta} \in \{-1,1\}^2$. Since the function $\hat{\rho}$ is even we obtain

$$\prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}} = \hat{\rho}(f_1+f_2)^{n_{(1,1)}+n_{(-1,-1)}}\hat{\rho}(f_1-f_2)^{n_{(1,-1)}+n_{(-1,1)}}. \tag{5.63}$$

One obvious relation between $n_{(1,1)}, n_{(-1,-1)}, n_{(1,-1)}$ and $n_{(-1,1)}$ is

$$n_{(1,1)} + n_{(-1,-1)} + n_{(1,-1)} + n_{(-1,1)} = n. \tag{5.64}$$

The other one we obtain by noting that

$$n_{(1,1)}+n_{(-1,-1)} - n_{(-1,1)} - n_{(1,-1)} = \sum_{i,j} \sigma_i^1\sigma_j^1\sigma_i^2\sigma_j^2 = nR_{12}^2. \tag{5.65}$$

Therefore

$$\begin{cases} n_{(1,1)} + n_{(-1,-1)} = \frac{1}{2}n(1+R_{12}^2), \\ n_{(1,-1)} + n_{(-1,1)} = \frac{1}{2}n(1-R_{12}^2). \end{cases} \tag{5.66}$$

By Theorem 2.4 we conclude that P-a.s. $\max_{\sigma,\sigma'\in X} |R(\sigma,\sigma')| < 1$ and therefore, for some positive constant c,

$$n_{(1,1)} + n_{(-1,-1)} \geq n_{(1,-1)} + n_{(-1,1)} \geq cn. \tag{5.67}$$

The above inequality allows us to approximate (5.62) by

$$\iint_{-\mu_1}^{\mu_1} e^{2\pi n(i\mathbf{f}\cdot\boldsymbol{\alpha}-\boldsymbol{\eta}\cdot\boldsymbol{\alpha})} \prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta}+i\boldsymbol{\eta}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}}\, df_1 df_2 + O\big(e^{-c_1 n}\big). \tag{5.68}$$

Adapting the proof of Lemma 5.5 from [BCMN05b] we get

$$\begin{aligned}\iint_{-\mu_1}^{\mu_1} & e^{2\pi n(i\mathbf{f}\cdot\boldsymbol{\alpha}-\boldsymbol{\eta}\cdot\boldsymbol{\alpha})} \prod_{\boldsymbol{\delta}} \hat{\rho}(\mathbf{f}\cdot\boldsymbol{\delta}+i\boldsymbol{\eta}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}}\, df_1 df_2 \\ &= e^{-nG_{n,2}(\boldsymbol{\alpha})}\frac{\sqrt{\det B(\sigma^1,\sigma^2)}}{2\pi n}\Big(1+O\Big(\frac{1}{\sqrt{n}}\Big)+O\Big(\frac{a_N^2}{n}\Big)\Big). \end{aligned} \tag{5.69}$$

Finally, we obtain that

$$\begin{aligned}\mathbb{P}\big(H_N'(\sigma^j) \in (x_j, x_j+dx_j) \text{ for } j=1,2\big) &= dx_1dx_2\, b_n e^{-nG_{n,2}(\boldsymbol{\alpha})} \\ \times \frac{\sqrt{\det B}}{2\pi}&\Big(1+O\Big(\frac{1}{\sqrt{n}}\Big)+O\Big(\frac{a_N^2}{n}\Big)\Big) + O\big(b_N^2 n e^{-c_1 n}\big). \end{aligned} \tag{5.70}$$

From (5.44) we see that for some constant C

$$nG_{n,2} = \frac{n}{2}(\boldsymbol{\alpha}, B^{-1}\boldsymbol{\alpha}) + \frac{16c_4}{(1+R_{12}^2)^3}\frac{\alpha_N^4}{n} + O\Big(\frac{a_N^6}{n^2}\Big) \geq \frac{n}{2}(\boldsymbol{\alpha}, B^{-1}\boldsymbol{\alpha}) + C, \tag{5.71}$$

and thus the joint density of $H_N'(\sigma^1), H_N'(\sigma^2)$ is bounded by

$$\frac{b_N^2\sqrt{\det B}}{2\pi}\, e^{-n(\boldsymbol{\alpha},B^{-1}\boldsymbol{\alpha})/2-C}\Big(1+O\Big(\frac{1}{\sqrt{n}}\Big)+O\Big(\frac{a_N^2}{n}\Big)\Big) + O\big(b_N^2 n e^{-c_1 n}\big). \tag{5.72}$$

This last bound for the joint density clearly implies that the sum II could be bounded by II_g plus an error resulting from the second term. But the cumulative error coming from the second term is of order $O\big(2^{2M} b_N^2 n e^{-c_1 n}\big)$ which is negligible even in the case $\limsup M/N > 0$.

To prove the second statement of the lemma we will approximate the sum I by I_g, which was already calculated in Section 4. We first notice that for sequences $(\sigma^1,\ldots,\sigma^\ell)$ satisfying condition (5.49) $R_{\max}(\sigma^1,\ldots,\sigma^\ell) = O(\theta_N)$. Furthermore, for configurations $(\sigma^1, \dots, \sigma^\ell)$ obeying condition (5.49) it is possible to derive from (5.44) that

$$G_{n,\ell}(\boldsymbol{\alpha}) = \frac{1}{2}(\boldsymbol{\alpha}, B^{-1}\boldsymbol{\alpha}) + c_4\ell(1+3(\ell-1))\frac{a_N^4}{n} + O\Big(\frac{a_N^6}{n^3}\Big) + O\Big(\frac{a_N^2}{n^2}\theta_N\Big). \quad (5.73)$$

For the details of the derivation we refer to Subsection 5.4 of [BCMN05b] and in particular to formula (5.57) in there. Using formula (5.73) with $\ell = 2$ and substituting it into (5.35) we obtain that

$$I = I_g\, e^{-32c_4\varepsilon^2\log^2 2}\Big(1 + O\Big(\frac{1}{\sqrt{n}}\Big) + O\Big(\frac{\alpha_N^2}{n}\Big)\Big). \quad (5.74)$$

This finishes the proof of Lemma 5.6. □

To conclude the calculation of the second moment we notice that summing I and II we get the second factorial moment

$$\mathbb{E}(\mathcal{P}_N(A))_2 = \frac{e^{-32c_4\varepsilon^2\log^2 2}(\mu(A))^2}{\sqrt{1-4\varepsilon\log 2}}(1+o(1)). \quad (5.75)$$

Assertion (i) of Theorem 5.5 is thus proven.

3. **Third moment estimate.** To deal with the third moment

$$\mathbb{E}(\mathcal{P}_N(A))_3 = \sum_{(\sigma^1,\sigma^2,\sigma^3)\in X^3} \mathbb{P}\left(H'_N(\sigma^1)\in A, H'_N(\sigma^2)\in A, H'_N(\sigma^3)\in A\right), \quad (5.76)$$

we use the same strategy as we used to calculate the second moment. In particular, we fix $\ell = 3$ and split the sum (5.76) in two parts:

$$I = \sum_{\sigma^1,\sigma^2,\sigma^3} \mathbb{P}\big(H'_N(\sigma^1)\in A, H'_N(\sigma^2)\in A, H'_N(\sigma^3)\in A\big), \quad (5.77)$$

where the sum runs over all sequences of distinct configurations $(\sigma^1,\sigma^2,\sigma^3)\in X^3$ satisfying condition (5.49), and

$$II = \sum_{\sigma^1,\sigma^2,\sigma^3} \mathbb{P}\big(H'_N(\sigma^1)\in A, H'_N(\sigma^2)\in A, H'_N(\sigma^3)\in A\big), \quad (5.78)$$

where the sum is over all sequences of distinct configurations $(\sigma^1,\sigma^2,\sigma^3)\in X^3$ violating (5.49).

By exactly the same argument as in the calculation of the second moment the contribution from the "typical" collections, I, is bounded. We therefore concentrate on the analysis of the contribution from the "atypical" collections, II. Since we are interested only in the estimate of the third moment from above it suffices to bound the joint density $\mathbb{P}\big(H'_N(\sigma^j)\in(x_j, x_j+dx_j) \text{ for } j=1,2,3\big)$. We start with formula (5.27) which, in the case $\ell = 3$, can be rewritten as

$$\mathbb{P}\big(H'_N(\sigma^j)\in(x_j,x_j+dx_j) \text{ for } j=1,2,3\big) = b_N^3 n\, dx_1 dx_2 dx_3 \\ \times \iiint_{-\infty}^{\infty} \prod_{i_1,i_2,i_3} \hat\rho(v_{i_1,i_2,i_3}) e^{2\pi i\sqrt{n}\left(f_1\alpha_N^{(1)}+f_2\alpha_N^{(2)}+f_3\alpha_N^{(3)}\right)} df_1 df_2 df_3. \quad (5.79)$$

We can rewrite the integral in the above expression as

$$\iiint_{-\infty}^{\infty} \prod_{\boldsymbol{\delta}} \hat\rho(\mathbf{f}\cdot\boldsymbol{\delta})^{n_\delta}\, e^{2\pi i n \mathbf{f}\cdot\boldsymbol{\alpha}} df_1 df_2 df_3, \quad (5.80)$$

where $\boldsymbol{\delta} \in \{-1,1\}^3$. Since the function $\hat\rho$ is even we obtain

$$\prod_{\boldsymbol{\delta}} \hat\rho(\mathbf{f}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}} = \hat\rho(f_1+f_2+f_3)^{n_1}\hat\rho(f_1+f_2-f_3)^{n_2} \times \hat\rho(f_1-f_2+f_3)^{n_3}\hat\rho(f_1-f_2-f_3)^{n_4}, \tag{5.81}$$

where

$$\begin{cases} n_1 = n_{(1,1,1)} + n_{(-1,-1,-1)}, \\ n_2 = n_{(1,1,-1)} + n_{(-1,-1,1)}, \\ n_3 = n_{(1,-1,1)} + n_{(-1,1,-1)}, \\ n_4 = n_{(-1,1,1)} + n_{(1,-1,-1)}. \end{cases} \tag{5.82}$$

By definition of the matrix $C_2(\sigma^1,\sigma^2,\sigma^3)$ we have

$$\begin{cases} n_1+n_2-n_3-n_4 = nR_{12}^2, \\ n_1-n_2-n_3+n_4 = nR_{23}^2, \\ n_1-n_2+n_3-n_4 = nR_{31}^2, \\ n_1+n_2+n_3+n_4 = n. \end{cases} \tag{5.83}$$

Solving the system

$$\begin{cases} n_1 = \frac14 n(1+R_{12}^2+R_{23}^2+R_{31}^2), \\ n_2 = \frac14 n(1+R_{12}^2-R_{23}^2-R_{31}^2), \\ n_3 = \frac14 n(1-R_{12}^2-R_{23}^2+R_{31}^2), \\ n_4 = \frac14 n(1-R_{12}^2+R_{23}^2-R_{31}^2). \end{cases} \tag{5.84}$$

From Theorem 2.4 we obtain that P-almost surely

$$\limsup_{N\to\infty} \max_{\sigma^1,\sigma^2\in X} \mathcal{J}\big(R(\sigma^1,\sigma^2)\big) \le \varepsilon \log 2. \tag{5.85}$$

Since the function $\mathcal{J}$ is monotone we obtain from (5.85) and from assumption $\varepsilon < \frac{1}{8\log 2}$ that P-a.s.

$$\limsup_{N\to\infty} \max_{\sigma^1,\sigma^2\in X} |R(\sigma^1,\sigma^2)| < \frac12. \tag{5.86}$$

It implies that $\min\{n_1,n_2,n_3,n_4\} \ge cn$ for some positive constant c. It allows us to approximate the integral in (5.79) by

$$\iiint_{-\mu_1}^{\mu_1} e^{2\pi n(i\mathbf{f}\cdot\boldsymbol{\alpha}-\boldsymbol{\eta}\cdot\boldsymbol{\alpha})} \prod_{\boldsymbol{\delta}} \hat\rho(\mathbf{f}\cdot\boldsymbol{\delta}+i\boldsymbol{\eta}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}}\, df_1 df_2 df_3 + O\big(e^{-cn}\big). \tag{5.87}$$

Adapting the proof of Lemma 5.5 from [BCMN05b] we get

$$\iiint_{-\mu_1}^{\mu_1} e^{2\pi n(i\mathbf{f}\cdot\boldsymbol{\alpha}-\boldsymbol{\eta}\cdot\boldsymbol{\alpha})} \prod_{\boldsymbol{\delta}} \hat\rho(\mathbf{f}\cdot\boldsymbol{\delta}+i\boldsymbol{\eta}\cdot\boldsymbol{\delta})^{n_{\boldsymbol{\delta}}}\, df_1 df_2 df_3 = e^{-nG_{n,3}(\boldsymbol{\alpha})} \frac{\sqrt{\det B(\sigma^1,\sigma^2)}}{(2\pi n)^{3/2}} \Big(1 + O\Big(\frac{1}{\sqrt n}\Big) + O\Big(\frac{a_N^2}{n}\Big)\Big). \tag{5.88}$$

Next, after a little algebra, one can derive from (5.44) that for some constant C

$$nG_{n,3} \ge \frac n2 (\boldsymbol{\alpha}, B^{-1}\boldsymbol{\alpha}) + C \tag{5.89}$$

and this estimate is enough to bound the joint density of $H_N'(\sigma^1), H_N'(\sigma^2), H_N'(\sigma^3)$ by the joint density of $Z_N(\sigma^1), Z_N(\sigma^2), Z_N(\sigma^3)$. Thus Theorem 5.5 is proved. □

Acknowledgements. We thank J.Cerny for a helpful reading of the first versions of the manuscript. V. Gayrard thanks the Chair of Stochastic Modeling of the École Polytechnique Fédérale of Lausanne for financial support.

References

[BFM04] H. Bauke, S. Franz, and S. Mertens, *Number partitioning random energy model.* Journal of Statistical Mechanics: Theory and Experiment, page P04003, 2004

[BM04] Heiko Bauke and Stephen Mertens, *Universality in the level statistics of disordered systems.* Physical Review E **70**, 025102(R), 2004

[BCMN05a] C. Borgs, J.T. Chayes, S. Mertens, and Ch. Nair, *Proof of the local REM conjecture for number partitioning I: Constant energy scales.* `http://arxiv.org/abs/cond-mat/0501760`, to appear in Random Structures & Algorithms.

[BCMN05b] C. Borgs, J.T. Chayes , S. Mertens, and Ch. Nair, *Proof of the local REM conjecture for number partitioning II: Growing energy scales.* `http://arxiv.org/abs/cond-mat/0508600`

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

[B06] A. Bovier, *Statistical mechanics of disordered systems.* Cambridge Series in Statistical and Probabilistic Mathematics 18, Cambridge University Press, 2006.

[BK06a] A. Bovier and I. Kurkova, *Local Energy Statistics in Disordered Systems: A Proof of the Local REM Conjecture.* Commun. Math. Phys. 263, 513–533 (2006).

[BK06b] A. Bovier and I. Kurkova, *Local energy statistics in spin glasses,* Journal of Statistical Physics, (Online).

[BK04] A. Bovier and I. Kurkova, *Poisson convergence in the restricted $k-$partitioning problem.* WIAS preprint 964, to appear in Random Structures & Algorithms, 2006.

[BKL02] A. Bovier, I. Kurkova, and M. Löwe, *Fluctuations of the free energy in the REM and the p-spin SK models.* Ann. Probab. 30, 605651 (2002)

[Kal83] O. Kallenberg, *Random measures.* 3rd, rev. and enl. ed. New York : Academic Press, 1983

[Mer00] Stephan Mertens, *Random costs in combinatorial optimization.* Phys. Rev. Lett. **84** (2000), no. 6, 1347-1350.

[Tal03] Michel Talagrand, *Spin glasses : a challenge for mathematicians : cavity and mean field models.* Berlin; New York : Springer, c2003

G. Ben Arous, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland and Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York NY 10012

E-mail address: `benarous@cims.nyu.edu`

Véronique Gayrard, Laboratoire d'Analyse, Topologie, Probabilités, CMI, 39 rue Joliot-Curie, 13453 Marseille Cedex

E-mail address: `gayrard@latp.univ-mrs.fr, veronique@gayrard.net`

A. Kuptsov, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland and Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York NY 10012

E-mail address: `kuptsov@cims.nyu.edu`