Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Much ado about Derrida's GREM

A. Bovier[1], I. Kurkova[2]

submitted: November 24, 2004

[1] Weierstraß-Institut
für Angewandte Analysis
und Stochastik
Mohrenstr. 39
10117 Berlin
Germany
and
Technische Universität Berlin
Institut für Mathematik
Str. des 17. Juni 136
10623 Berlin
Germany
E-Mail: bovier@wias-berlin.de

[2] Laboratoire de Probabilités
et Modèles Aléatoires
Université Paris 6
B.C. 188
4, place Jussieu
75252 Paris Cedex 5
France
E-Mail: kourkova@ccr.jussieu.fr

No. 981
Berlin 2004

Key words and phrases. Gaussian processes, spin-glasses, Generalised random energy model, Poisson point processes, branching processes, coalescence .

A.B. is partially supported by the DFG in the Dutch-German bilateral research group "Mathematical models from physics and biology". We acknowledge support from the European Science Foundation though the programme RDSES.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

ABSTRACT. We provide a detailed analysis of Derrida's Generalised Random Energy Model (GREM). In particular, we describe its limiting Gibbs measure in terms Ruelle's Poisson cascades. Next we introduce and analyse a more general class of Continuous Random Energy Models (CREMs) which differs from the well-known class of Sherrington-Kirkpatrick models only in the choice of distance on the space of spin configurations : the Hamming distance defines the later class while the ultrametric distance corresponds to the former one. We express explicitly the geometry of its limiting Gibbs measure in terms of genealogies of Neveu's Continuous State branching Process via an appropriate time change. We also identify the distances between replicas under the limiting CREM's Gibbs measure with those between integers of Bolthausen-Sznitman coalescent under the same time change.

1. INTRODUCTION

Through the remarkable progress achieved recently through the work of Guerra [G], Aizenman, Sims, and Starr [ASS], and Talagrand [T5, T6] (see also this volume) towards a rigorous justification of the Parisi solution in the Sherrington-Kirkpatrick models, we have now a clear understanding of how Parisi's replica symmetry breaking solution for the free energy emerges. With the exception of a regime in the p-spin SK models where one-step replica symmetry breaking occurs [T2, T4], however, these result only justify the formula for the free energy. The question of how the asymptotics of the Gibbs measure is described in general, and whether it conforms to the picture suggested by the replica theory remains open.

In this situation it may still be instructive to see how a picture like the one predicted by replica theory emerges in another class of spin glass models, the Generalised Random Energy models of B. Derrida and Gardner [D3, DG1, DG2]. This is reinforced by the fact that these structures play a crucial rôle in the Parisi solution. In this article we give a concise review of a detailed rigorous analysis of the asymptotics of the Gibbs measures in this class of models that we carried out recently [BK1, BK2, BK3].

The class of models we consider here can be described as follows. Consider the N dimensional hypercube $\Sigma_N = \{-1,1\}^N$ endowed with the (normalized) ultrametric distance

$$d_N(\sigma,\tau) = 1 - N^{-1}(\min\{i : \sigma_i \neq \tau_i\} - 1). \tag{1.1}$$

Define a centered normal Gaussian process X on Σ_N with covariance given by

$$\mathsf{E}\, X_\sigma X_\tau = A(1 - d_N(\sigma,\tau)) \tag{1.2}$$

for some non-decreasing right-continuous function $A : [0,1] \to [0,1]$.

The principal objects of interest are the Gibbs measures on Σ_N:

$$\mu_{\beta,N}(\sigma) \equiv \frac{e^{\beta\sqrt{N}X_\sigma}}{Z_{\beta,N}}, \qquad \sigma \in \Sigma_N, \tag{1.3}$$

where the partition function, $Z_{\beta,N}$, is

$$Z_{\beta,N} = \sum_{\sigma\in\Sigma_N} e^{\beta\sqrt{N}X_\sigma}. \tag{1.4}$$

This class of models differs from the Sherrington-Kirkpatrick (SK) models *only* in the choice of the distance (1.1). In fact, the SK models are defined in th same way, but instead of the ultrametric distance d_N one uses the Hamming distance,

$$d_N^H(\sigma,\tau) = N^{-1}\#\{i : \ \sigma_i \neq \tau_i\}.$$

Then the Hamiltonian of the class of SK models has a covariance structure $\mathsf{E}\, X_\sigma X_\tau = A(d_N^H(\sigma,\tau))$ with any function A such that the matrix of $A(d_N^H(\sigma,\tau))$ is positively defined.

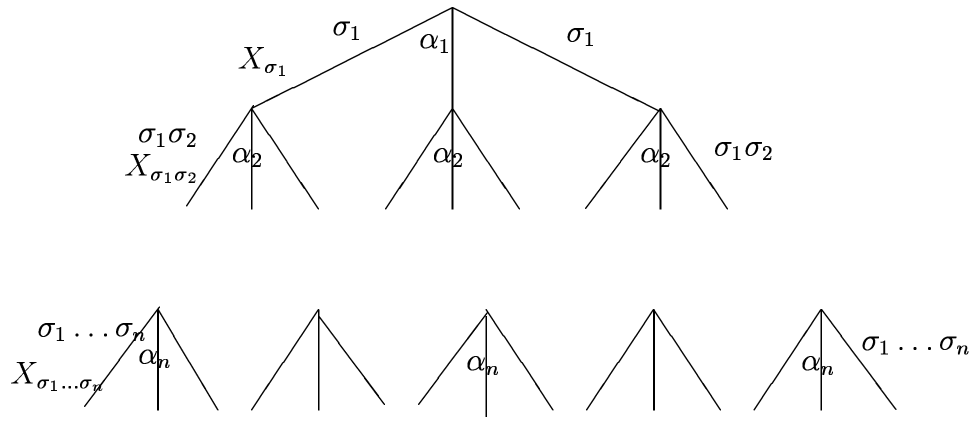

FIGURE 1. Structure of the Hamiltonian of the GREM.

Since $N^{-1}\sum_i \sigma_i\tau_i = 1 - 2d_N^H(\sigma,\tau)$, the choice of $A(x) = (1-2x)^2$ corresponds to the original SK model [SK].

1.1. **History of the models.** In 1980 B. Derrida proposed the simplest spin-glass model where the standard Gaussian random variables X_σ are independent [D1, D2]. It was called the *Random Energy Model* (REM). Note that this is a particular case of the model (1.2) with $A(x) = 1_{\{x=1\}}$.

B. Derrida also introduced later [D3] the *Generalized Random Energy Model* (GREM) in view of keeping dependence while simplifying it to a hierarchical structure in order to obtain a more tractable model. The Hamiltonian of the GREM can be constructed explicitly in terms of i.i.d. Gaussian random variables. Namely, chose the parameters $n \geq 1$ (number of hierarchies), $a_1, a_2, \dots, a_n \in [0,1]$ with $\sum_{i=1}^n a_i = 1$, and $\alpha_1, \alpha_2, \dots, \alpha_n \in [1,2]$ with $\prod_{i=1}^n \alpha_i = 2$. Let us represent the hypercube Σ_N as a product $\Sigma_N = \prod_{i=1}^n \Sigma_{N \ln \alpha_i/\ln 2}$ and write $\sigma = \sigma_1 \dots \sigma_n$ where $\sigma_i \in \Sigma_{N\ln\alpha_i/\ln 2}$. Let $X_{\sigma_1}, X_{\sigma_1\sigma_2}, \dots, X_{\sigma_1\cdots\sigma_n}$ be $\alpha_1^N + \alpha_1^N\alpha_2^N + \cdots + \alpha_1^N \cdots \alpha_n^N$ independent standard Gaussian random variables. Then the Hamiltonian of the GREM is given by:

$$X_\sigma = \sqrt{a_1}X_{\sigma_1} + \sqrt{a_2}X_{\sigma_1\sigma_2} + \cdots + \sqrt{a_n}X_{\sigma_1\cdots\sigma_n} \qquad \text{if } \sigma = \sigma_1 \dots \sigma_n. \tag{1.5}$$

To get some intuition in (1.5), one could imagine a tree illustrated on Figure 1 : α_1^N branches of the first level are indexed by σ_1. Each of these branches supports α_2^N branches of the second level indexed by $\sigma_1\sigma_2$: thus on the second level there are $(\alpha_1\alpha_2)^N$ branches etc. Each configuration $\sigma = \sigma_1 \dots \sigma_n$ is represented uniquely as a path on this tree going from the top to the bottom through the branches σ_1, $\sigma_1\sigma_2$, $\dots$, $\sigma_1 \dots \sigma_n$. If, moreover, we associate to each of branches $\sigma_1 \dots \sigma_k$ a random variable $X_{\sigma_1 \dots \sigma_k}$, then X_σ is the linear combination of these random variables taken along the path associated with σ and multiplied by coefficients $\sqrt{a_1}, \dots, \sqrt{a_n}$.

As can be verified by computing the covariance of X_σ, this model is a special case of the models (1.2), where $A(x)$ is a step function given as

$$A(x) = \sum_{i=0}^{k} a_k, \quad \text{for } x \in [\ln(\alpha_0 \cdots \alpha_k)/\ln 2, \ln(\alpha_0 \cdots \alpha_{k+1})/\ln 2), \tag{1.6}$$

$k = 0, 1, \dots, n$, where $a_0 = 0$, $\alpha_0 = 1$; see Figure 2. The GREM was analyzed by Derrida and Gardner [D3, DG1, DG2, DG3]. A rigorous computation of the free energy, $N^{-1}\ln\sum_\sigma e^{\beta\sqrt{N}X_\sigma}$, in full generality was later given [CCP]. Derrida and Gardner also

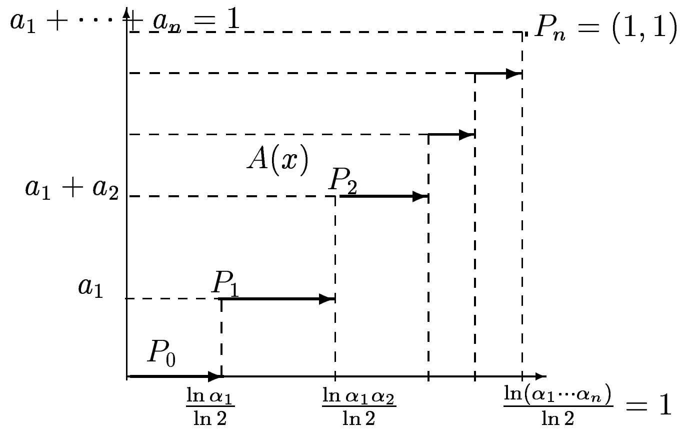

FIGURE 2. The function $A(x)$ of the GREM

considered limits of their results as the number of steps tended to infinity, and interpreted these results as corresponding to continuous functions A [DG1].

While there were very few further rigorous results on these models (but see [Ei, GMP]), Ruelle in a seminal paper of 1988 [Ru] introduced a new class of models based on *Poisson cascades* (to which we will henceforth refer to as "Ruelle's REM and GREM"), which he understood to be the appropriate models to describe the limiting Gibbs measures of Derrida's GREMs. Ruelle noted a number of remarkable features of these models, and in particular observed that it was possible to construct limits as the number of steps went to infinity in terms of projective limits.

Shortly after that, Neveu [Ne] observed a connection between Ruelle's models and *continuous state branching processes*. Unfortunately, this remark appeared only in a preliminary internal report that was never published. Later, Bolthausen and Sznitman in [BS] interpreted the results of the replica theory of spin glasses in terms of a coalescent process, now known as the *Bolthausen-Sznitman coalescent.* Following this paper, Bertoin and Le Gall [BLG] gave a precise and complete form of the relation between Neveu's continuous state branching processes, Ruelle's GREM, and the Bolthausen-Sznitman coalescent.

Around the time when these fascinating results appeared, we began to investigate more closely the link to the original spin glass models with Ruelle's models. In the REM, this connection was made in a paper with M. Löwe [BKL] (see also [Bo] and [BS2, T3, T4]). These results were extended to the GREMs in [BK1], using essentially elementary methods. We observed, however, that the use of the so-called Ghirlanda-Guerra identities [GG] allowed for a different approach that circumvents parts of these explicit computations (this fact was first observed in the REM by Talagrand [T3], who also exploited these identities heavily in his work on the p-spin SK models [T1, T2, T3, T4]). It allowed us in [BK2] to extend our convergence results to the general class of models defined above with general right-continuous non-decreasing functions $A(x) : [0,1] \to [0,1]$. We called this class *Continuous Random Energy Models* (CREM). Finally, combining the results of [BK2] and those of J. Bertoin and J.F. Le Gall [BLG], we concluded our investigation in [BK3] by linking our results to the continuous state branching process of Neveu. More precisely, we identified the geometry of the limiting Gibbs measure proven to exist in [BK2] explicitly in terms of the genealogy of Neveu's branching process, which were defined in [BLG]. The rôle played by these random genealogies in the Parisi solution can be most clearly seen in the paper by Aizenman, Sims, and Starr [ASS] (see also this volume). We hope that these examples help to explain to a mathematical audience what physicist describe when they talk about "continuous replica symmetry breaking".

1.2. **Geometry of Gibbs measures.** The central problem one is faced with when analyzing mean field spin glasses is to describe the geometric structure of a random probability measure (1.3) on a set Σ_N as $N \to \infty$. Two scenarios can be expected:

(i) at high temperatures (small β) this measure will be spread over over an exponentially large set of configurations that is distributed rather uniformly over the hypercube;
(ii) at low temperatures (large β) this measure will concentrate on a very small subset of configurations σ, with a rather complicated structure, corresponding to the largest values of X_σ, while the mass of the enormous amount of all other configurations σ will be negligible.

These statements are easily proven in the REM, using the classical theory of extremes of i.i.d. random variables [LLR]. Let us briefly recall these results. To be able to embed all hypercubes Σ_N, $N \in \mathbb{N}$, in the same compact space, it is convenient to map them to the unit interval via the canonical maps $r_N : \Sigma_N \to [0,1]$:

$$r_N(\sigma) = 1 - \sum_{i=1}^{N} 2^{-i-1}(1+\sigma_i). \tag{1.7}$$

For finite N, the Gibbs measure is then mapped to a discrete measure on $[0,1]$ concentrated on 2^N points:

$$\widetilde{\mu}_{\beta,N} = \sum_{\sigma\in\Sigma_N} \mu_{\beta,N}(\sigma)\delta_{r_N(\sigma)} \tag{1.8}$$

with distribution function

$$\theta_{\beta,N}(x) = \int_0^x \mathrm{d}\,\widetilde{\mu}_{\beta,N}. \tag{1.9}$$

It was proved in [Bo] that

$$\theta_{\beta,N} \xrightarrow{\mathcal{D}} \begin{cases} y = x & \text{if } \beta \le \sqrt{2\ln 2} \\ \dfrac{S_{\beta/\sqrt{2\ln 2}}(x)}{S_{\beta/\sqrt{2\ln 2}}(1)} & \text{if } \beta > \sqrt{2\ln 2}. \end{cases} \tag{1.10}$$

This means that $\widetilde{\mu}_{\beta,N}$ converges to the Lebesgue measure on $[0,1]$ at high temperatures, confirming scenario (i). The random function $S_{\beta/\sqrt{2\ln 2}}(x)$ is a stable subordinator with the index $\beta/\sqrt{2\ln 2}$, i.e. a step function that jumps at random points, t_i, $i = 1,2,\ldots$, which are distributed uniformly on $[0,1]$. The values of jumps w_i are also random and can be expressed as

$$w_i = \frac{e^{(\beta/\sqrt{2\ln 2})x_i}}{\sum_j e^{(\beta/\sqrt{2\ln 2})x_j}} \tag{1.11}$$

where $x_1 > x_2 > \cdots$ are the atoms of the Poisson point process $\mathcal{P}$ on $\mathbb{R}$ with intensity measure $e^{-x}dx$. This confirms scenario (ii): at low temperatures the limiting Gibbs measure concentrates on a countable number of randomly chosen configurations corresponding to points $t_i \in [0,1]$.

This description of the limiting Gibbs measure does not give any information about its geometry. But to define the geometry of a measure on the infinite dimensional hypercube, it is necessary, first of all, to specify a topology. The conventional choice of the product topology is not suitable to capture the fact that these measure tend to concentrate on *individual random* configurations. To resolve this problem we introduce the following construction. Let

$$m_\sigma(t) = \mu_{\beta,N}(\tau : \ d_N(\sigma,\tau) \le 1-t) \tag{1.12}$$

be the picture of the landscape of the Gibbs measure taken from a given configuration σ. The function $1 - m_\sigma(t)$ is a random distribution function on $[0,1]$. In this way we get 2^N

different pictures of the landscape of the Gibbs measure taken from different configurations σ. It seems reasonable to subject the importance of each of these pictures to the Gibbs mass, $\mu_{\beta,N}(\sigma)$, of its starting point, σ. In this way we construct the random probability measure

$$\mathcal{K}_{\beta,N} \equiv \sum_{\sigma\in\Sigma_N} \mu_{\beta,N}(\sigma)\delta_{m_\sigma(\cdot)} \tag{1.13}$$

on these distribution functions $m_\sigma(t)$ that attributes to each function $m_\sigma(t)$ the weight $\mu_{\beta,N}(\sigma)$. We call $\mathcal{K}_{\beta,N}$ the *empirical distance distribution function.* It has a very appealing physical interpretation: it tells, for a fixed realization of the disorder, with which probability an observer, that is himself distributed with the Gibbs measure, will see a given distribution of mass around himself. Convergence results for the Gibbs measures will be formulated in term of convergence of the law, under the Gaussian process X_σ, of $\mathcal{K}_{\beta,N}$. A key object is the first moment of $\mathcal{K}_{\beta,N}$:

$$\int \mathcal{K}_{\beta,N}(dm)m(\cdot) = \mu_{\beta,N}^{\otimes 2}(\sigma,\tau : d_N(\sigma,\tau)\in\cdot), \tag{1.14}$$

which is the probability that two configurations, σ,τ, drawn independently from the Gibbs sample satisfy $d_N(\sigma,\sigma')\in\cdot$.

In the case of the REM, the limit of $\mathcal{K}_{\beta,N}$ is a rather simple object :

$$\mathcal{K}_{\beta,N} \stackrel{\mathcal{D}}{\to} \begin{cases} \delta_{\delta(0)} & \beta\le\sqrt{2\ln 2} \\ \sum_{w_i} w_i\delta_{w_i\delta(0)+(1-w_i)\delta(1)} & \beta>\sqrt{2\ln 2}. \end{cases} \tag{1.15}$$

It will manifest much more rich and interesting structure in the case of the GREM and CREM, as we will see.

1.3. **Point process of extremes.** To describe efficiently the behavior of the limiting Gibbs measure according to scenario (ii), it is necessary to know the maximal values of the Gaussian process X_σ. In the case of *independent* variables the corresponding result is well known. First of all $\max_{\sigma\in\Sigma_N} X_\sigma N^{-1/2}\to\sqrt{2\ln 2}$ a.s. from where for any $\varepsilon>0$

$$\mathsf{P}(\forall\sigma: \quad X_\sigma<\sqrt{N}(\sqrt{2\ln 2}+\varepsilon))\to 1, \quad \mathsf{P}(\forall\sigma: \quad X_\sigma<\sqrt{N}(\sqrt{2\ln 2}-\varepsilon))\to 0.$$

To get the limiting value here between 0 and 1, one should take $\varepsilon=\varepsilon_N\to 0$ as $N\to\infty$. It turns out that the right function depending on the parameter $x\in\mathbb{R}$ is

$$u_{\ln\alpha,N}(x)=\sqrt{2N\ln\alpha}+\frac{x}{\sqrt{2N\ln\alpha}}-\frac{\ln N+\ln\ln\alpha+\ln 4\pi}{2\sqrt{2N\ln\alpha}}, \tag{1.16}$$

with $\alpha=2$ as

$$\mathsf{P}(\forall\sigma: \quad X_\sigma<u_{\ln 2,N}(x))\to e^{-e^{-x}}, \quad N\to\infty.$$

Thus, we come to the classical result on the convergence of extreme value statistics in the case where X_σ are 2^N independent Gaussian random variables. It says that the point process

$$\sum_{\sigma\in\Sigma_N}\delta_{u^{-1}_{\ln 2,N}(X_\sigma)}\stackrel{\mathcal{D}}{\to}\mathcal{P} \tag{1.17}$$

converges weakly to the Poisson point process $\mathcal{P}$ on $\mathbb{R}$ with the intensity measure $e^{-x}dx$, see e.g. [LLR]. This result is the crucial ingredient in the proof of (1.10) and clarifies the meaning of (1.11).

To start the analysis of the GREM, we need an analogous result in the case of correlated Gaussian random variables. Results of this kind in the correlated case are much more scarce. Most of them establish conditions under which the same limiting point process arises as for the independent random variables. We will see that this is in general not the case for the random variables (1.5) correlated as in the Hamiltonian of the GREM.

1.4. **Organization of the paper.** The remainder of the paper is organized as follows. Section 2 is devoted to convergent point processes associated with the Hamiltonian of the GREM. Namely, we find the point process of extreme value statistics of its Hamiltonian. These results can be viewed as those on convergence of extreme value statistics for correlated Gaussian random variables independently of the context of spin glasses. In Section 3 we study the GREM (1.5) with finitely many hierarchies. In particular we identify the limit of $\mathcal{K}_{\beta,N}$ for this model with Ruelle's probability cascades. In Section 4 we analyze the general case of CREM's (1.1), (1.2) with a "continuum of hierarchies". We prove the existence of the limit of $\mathcal{K}_{\beta,N}$ by the so-called "Ghirlanda-Guerra" identities, i.e. identifying limits of all its moments. In Section 5 we describe explicitly the limit of $\mathcal{K}_{\beta,N}$ in terms of the genealogical structure of Neveu's continuous state branching process modulo an appropriate time change depending only on β and on the concave hull of A.

Notations. When A is a step-function as on Figure 2, we will denote by $\underline{A}(x)$ its linear interpolation. Its graph consists of the segments $[P_0, P_1]$, $[P_1, P_2], \ldots,$ $[P_{n-1}, P_n]$ where $P_k = (\sum_{i=0}^k a_k,\ \ln(\alpha_0 \cdots \alpha_k)/\ln 2)$ for $k = 0, \ldots, n$, with $a_0 = 0$, $\alpha_0 = 1$ so that $P_0 = (0,0)$ and $P_n = (1,1)$, see Figure 2.

We will denote by $\widehat{A}(x)$ the concave hull of the function $A(x)$ and by $\widehat{A}'(x)$ the right derivative of the concave hull of A, see Figure 5.

2. Convergent point processes associated to the GREM.

In Theorem 2.1 below we give a necessary and sufficient condition on the parameters a_i, α_i which assure that point process of extreme value statistics of GREM's Hamiltonian (1.5) is the same as in the case of independent random variables (1.17). This condition is the convexity of the linear interpolation $\underline{A}(x)$. In other words, the concave hull of $\underline{A}(x)$ should be the straight line $y = x$. It is illustrated on Figure 3(a). This condition is strictly weaker than the sufficient condition implied by Slepian's Lemma on the comparison of Gaussians. (Theorem 4.2.1 in [LLR]). We use the notation (1.16).

Theorem 2.1. [BK1] *Let $n \in \mathbb{N}$, $n \geq 1$, $0 < a_i < 1$ with $\sum_{i=1}^n a_i = 1$, $\alpha_i > 1$, $i = 1, 2, \ldots, n$. The point process*

$$\sum_{\sigma = \Sigma_N} \delta_{u_{\ln 2, N}^{-1}(\sqrt{a_1} X_{\sigma_1} + \sqrt{a_2} X_{\sigma_1 \sigma_2} + \cdots + \sqrt{a_n} X_{\sigma_1 \sigma_2 \ldots \sigma_n})}$$

converges weakly to the Poisson point process $\mathcal{P}$ on $\mathbb{R}$ with the intensity measure $Ke^{-x}dx$, $K \in \mathbb{R}$, iff the linear interpolation $\underline{A}(x)$ is convex, that is

$$a_i + a_{i+1} + \cdots + a_n \geq \ln(\alpha_i \alpha_{i+1} \cdots \alpha_n)/\ln \bar{\alpha} \quad \text{for all } i = 2, 3, \ldots, n, \tag{2.1}$$

see Figure 3(a). If all inequalities in (2.1) are strict, then $K = 1$. If some of the relations are equalities, the $0 < K < 1$[1].

The next lemma gives a sufficient condition for the convergence of the multidimensional point process to the point process of Poisson cascades defined by Ruelle in [Ru]. This is a generalization of Theorem 3 of [GMP]: we do not specify the law of the vectors $Y_{\sigma_1 \ldots \sigma_i}$, neither assume their independence.

Lemma 2.1. [BK1] *Let $\alpha_i \geq 1$, $i = 1, 2, \ldots, k$, $\bar{\alpha} \equiv \prod_{i=1}^k \alpha_i$. Let $Y_{\sigma_1}, Y_{\sigma_1 \sigma_2}, \ldots, Y_{\sigma_1 \ldots \sigma_k}$ be $\alpha_1^N + \cdots + (\alpha_1 \cdots \alpha_k)^N$ identically distributed random variables. Assume that $1 + \alpha_1^N + \cdots + (\alpha_1 \cdots \alpha_{k-1})^N$ vectors $(Y_{\sigma_1})_{\sigma_1 \in \{-1,1\}^{N \ln \alpha_1/\ln \bar{\alpha}}}$, $(Y_{\sigma_1 \sigma_2})_{\sigma_2 \in \{-1,1\}^{N \ln \alpha_2/\ln \bar{\alpha}}}$ $\forall \sigma_1 \in \{-1,1\}^{N \ln \alpha_1/\ln \bar{\alpha}}$,*
$\ldots$, $(Y_{\sigma_1 \sigma_2 \ldots \sigma_k})_{\sigma_k \in \{-1,1\}^{N \ln \alpha_k/\ln \bar{\alpha}}}$ $\forall\, \sigma_1 \ldots \sigma_{k-1} \in \{-1,1\}^{N \ln(\alpha_1 \cdots \alpha_{k-1})/\ln \bar{\alpha}}$ are independent.

[1] Explicit expressions for K are given in [BK1].

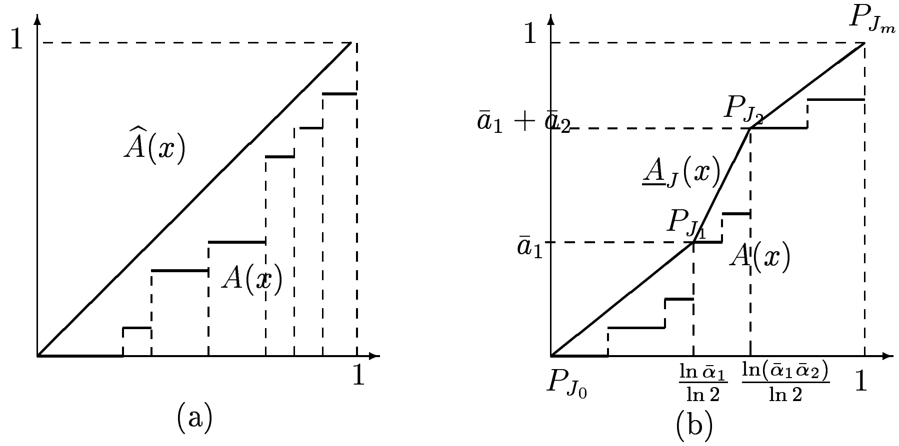

FIGURE 3. (a) Condition (2.1), (b) Condition (2.5)

Let $v_{N,1}(x),\dots,v_{N,k}(x)$ be functions on $\mathbb{R}$ such that the following point processes

$$\begin{aligned}\sum_{\sigma_1}\delta_{v_{N,1}(Y_{\sigma_1})} &\stackrel{\mathcal{D}}{\to} \mathcal{P}_1\\ \sum_{\sigma_2}\delta_{v_{N,2}(Y_{\sigma_1\sigma_2})} &\stackrel{\mathcal{D}}{\to} \mathcal{P}_2 \quad \forall\sigma_1\\ &\cdots\\ \sum_{\sigma_k}\delta_{v_{N,k}(Y_{\sigma_1\sigma_2\dots\sigma_k})} &\stackrel{\mathcal{D}}{\to} \mathcal{P}_k \quad \forall\sigma_1\dots\sigma_{k-1}\end{aligned} \tag{2.2}$$

converge weakly to the Poisson point processes $\mathcal{P}_1,\dots,\mathcal{P}_k$ on $\mathbb{R}$ with the intensity measures $K_1e^{-x}dx,\dots,K_ke^{-x}dx$ with some constants $K_1,\dots,K_k>0$ respectively. Then the following point process on $\mathbb{R}^k$

$$\mathcal{P}_N^{(k)} \equiv \sum_{\sigma_1}\delta_{v_{N,1}(Y_{\sigma_1})}\sum_{\sigma_2}\delta_{v_{N,2}(Y_{\sigma_1\sigma_2})}\cdots\sum_{\sigma_k}\delta_{v_{N,k}(Y_{\sigma_1\sigma_2\dots\sigma_k})}\stackrel{\mathcal{D}}{\to}\mathcal{P}^{(k)}$$

converges weakly to a point process, $\mathcal{P}^{(k)}$, called a k-level Poisson cascade, on $\mathbb{R}^k$.

Structure of $\mathcal{P}^{(k)}$. The Poisson cascades $\mathcal{P}^{(k)}$ can be characterized in terms of their Laplace transfroms, see [BK1]. Informally, they are best described as follows [Ru]: If $k=1$, it is a ordinary Poisson point process on $\mathbb{R}$ with intensity measure $K_1e^{-x}\mathrm{d}x$. To construct $\mathcal{P}^2$ on $\mathbb{R}^2$, we place the process $\mathcal{P}^1$ for $k=1$ on the axis of the first coordinate and through each of its points draw a straight line parallel to the axis of the second coordinate. Then we put on each of these lines independently a Poisson point process with intensity $K_2e^{-x}dx$. These points on $\mathbb{R}^2$ form the process $\mathcal{P}^2$. This procedure is now simply iterated k times.

Theorem 2.1 and Lemma 2.1 combined give a first important result, that establishes which convergent point processes may be constructed in the GREM: one can group together the hierarchies between the levels $J_0,J_1,\dots,J_m$, if condition (2.5) is verified. This condition is illustrated in Figure 3(b): it means the convexity of the function $\underline{A}(x)$ between the levels $J_0,J_1,\dots,J_m$.

Theorem 2.2. [BK1] *Let $\alpha_i\geq 1$, $0<a_i<1$, $i=1,2,\dots,n$, $\prod_{i=1}^n\alpha_i=2$, $\sum_{i=1}^n a_i=1$. Let $J_1,J_2,\dots,J_m\in\mathbb{N}$ be the indices $0=J_0<J_1<J_2<\dots<J_m=n$. We denote by $\bar{a}_l\equiv\sum_{i=J_{l-1}+1}^{J_l}a_i$, $\bar{\alpha}_l\equiv\prod_{i=J_{l-1}+1}^{J_l}\alpha_i$, $l=1,2,\dots,m$, and introduce the standard*

Gaussian random variables

$$\bar{X}^{\sigma_1\ldots\sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1}\sigma_{J_{l-1}+2}\cdots\sigma_{J_l}} \equiv \Big(\sqrt{a_{J_{l-1}+1}}X_{\sigma_1\ldots\sigma_{J_{l-1}}\sigma_{J_{l-1}+1}} + \sqrt{a_{J_{l-1}+2}}X_{\sigma_1\ldots\sigma_{J_{l-1}}\sigma_{J_{l-1}+1}\sigma_{J_{l-1}+2}} + \cdots + \sqrt{a_{J_l}}X_{\sigma_1\ldots\sigma_{J_{l-1}}\sigma_{J_{l-1}+1}\cdots\sigma_{J_l}}\Big)/\sqrt{\bar{a}_l}. \tag{2.3}$$

Assume that a partition $J_1, J_2, \ldots, J_m$ satisfies the following condition : for all $l = 1, 2, \ldots, m$ and all k such that $J_{l-1} + 2 \le k \le J_l$

$$(a_k + a_{k+1} \cdots + a_{J_l-1} + a_{J_l})/\bar{a}_l \ge \ln(\alpha_k\alpha_{k+1}\cdots\alpha_{J_l-1}\alpha_{J_l})/\ln(\bar{\alpha}_l). \tag{2.4}$$

If $\underline{A^J}(x)$ is the linear interpolation of the points $(0,0), P_{J_1}, P_{J_2}, \ldots, P_{J_m} = (1,1)$, condition (2.4) is equivalent to

$$\underline{A}(x) \le \underline{A^J}(x) \quad \forall x \in [0,1], \tag{2.5}$$

(see Figure 3(b)), then the point process

$$\mathcal{P}_N^{(m)} \equiv \sum_{\sigma_1\ldots\sigma_{J_1}} \delta_{u^{-1}_{\ln\bar{\alpha}_1,N}(\bar{X}_{\sigma_1\ldots\sigma_{J_1}})} \sum_{\sigma_{J_1+1}\ldots\sigma_{J_2}} \delta_{u^{-1}_{\ln\bar{\alpha}_2,N}(\bar{X}^{\sigma_1\ldots\sigma_{J_1}}_{\sigma_{J_1+1}\ldots\sigma_{J_2}})} \cdots \sum_{\sigma_{J_{m-1}+1}\ldots\sigma_{J_m}} \delta_{u^{-1}_{\ln\bar{\alpha}_m,N}(\bar{X}^{\sigma_1\ldots\sigma_{J_{m-1}}}_{\sigma_{J_{m-1}+1}\ldots\sigma_{J_m}})} \tag{2.6}$$

converges weakly in distribution to the point process $\mathcal{P}^{(m)}$ on $\mathbb{R}^m$, defined in Lemma 2.1, with constants $K_1, \ldots, K_m$. Moreover, $K_l = 1$, if all $J_l - J_{l-1} - 1$ inequalities in (2.4) for $k = J_{l-1} + 2, \ldots, J_l$ are strict. Otherwise, $0 < K_l < 1$[2].

It is clear that the point process of extreme values of the Hamiltonian can be constructed from one of the partitions of Theorem 2.2. This is the one that allows to group together the maximal number of hierarchies: among all series of indices $J_1, \ldots, J_m$ satisfying (2.4) one should choose the one with the largest differences $J_1 - J_0, \ldots, J_m - J_{m-1}$. To define it, we set $J_0 \equiv 0$ and

$$J_l \equiv \min\{J > J_{l-1} : A_{J_{l-1}+1,J} > A_{J+1,k}\ \forall k \ge J+1\} \quad \text{where } A_{j,k} \equiv \frac{\sum_{i=j}^k a_i}{2\ln(\prod_{i=j}^k \alpha_i)}. \tag{2.7}$$

The sequence $J_1, \ldots, J_m$, defined by (2.7), verifies (2.4), for all k, such that $J_{l-1}+2 \le k \le J_l$ and all $l = 1, 2, \ldots, m$. This choice of the partition $J_1, J_2, \ldots, J_m$ (2.7) has a beautiful geometric interpretation: the linear interpolation $\underline{A}^J(x)$ of $(0,0), P_{J_1}, \ldots, P_{J_m} = (1,1)$ is *the concave hull* of the function $\underline{A}(x)$, see Figure 4.

We set

$$\bar{a}_l \equiv \sum_{i=J_{l-1}+1}^{J_l} a_i, \quad \bar{\alpha}_l \equiv \prod_{i=J_{l-1}+1}^{J_l} \alpha_i, \quad \gamma_l \equiv \sqrt{\frac{\bar{a}_l}{2\ln\bar{\alpha}_l}} = \sqrt{\frac{(\widehat{A})'(P_{J_{l-1}})}{2\ln 2}}, \quad l = 1, 2, \ldots, m, \tag{2.8}$$

see Figure 4. Next, let us define the function $U_{J,N}$ as

$$U_{J,N}(x) \equiv \sum_{l=1}^m \Big(\sqrt{2N\bar{a}_l\ln\bar{\alpha}_l} - N^{-1/2}\gamma_l(\ln(N(\ln\bar{\alpha}_l)) + \ln 4\pi)/2\Big) + N^{-1/2}x \tag{2.9}$$

and the point process

$$\mathcal{E}_N \equiv \sum_{\sigma\in\{-1,1\}^N} \delta_{U^{-1}_{J,N}(\sqrt{a_1}X_{\sigma_1}+\cdots+\sqrt{a_n}X_{\sigma_1\cdots\sigma_n})}. \tag{2.10}$$

Theorem 2.3. [BK1] *(i) The point process $\mathcal{E}_N$ converges weakly, as $N \uparrow \infty$, to the point process on $\mathbb{R}$*

$$\mathcal{E} \equiv \int_{\mathbb{R}^m} \mathcal{P}^{(m)}(dx_1, \ldots, dx_m)\delta_{\sum_{l=1}^m \gamma_l x_l} \tag{2.11}$$

[2] Explicit expressions for K are given in [BK1].

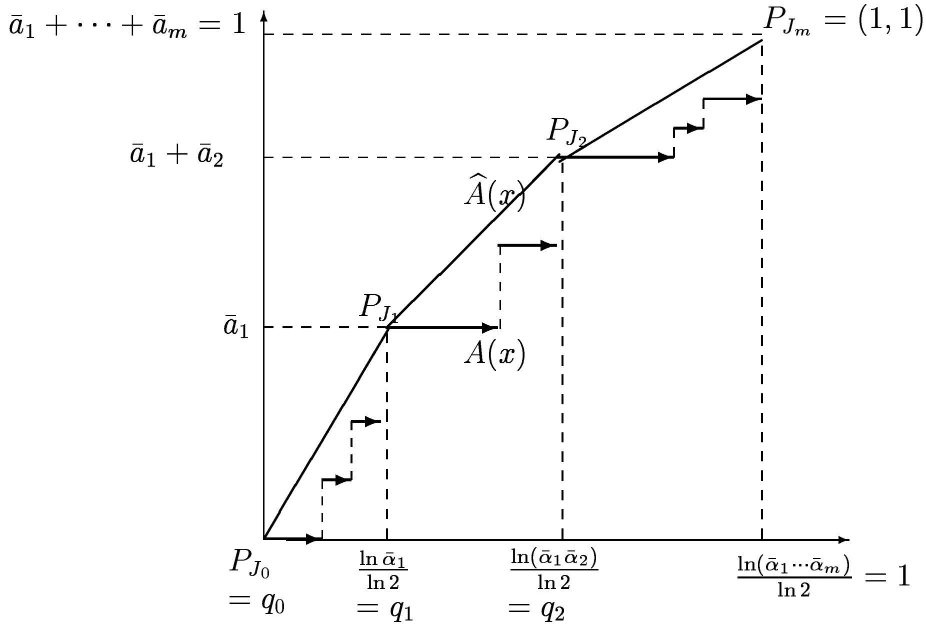

FIGURE 4. The function $\widehat{A}(x)$, with parameters (2.8), (3.6).

where $\mathcal{P}^{(m)}$ is a Poisson cascade (introduced in Lemma 2.1) with constants $K_1,\ldots,K_m$, as defined in Theorem 2.2 according to the partition $J_1,\ldots,J_m$ of (2.7) and the parameters $\gamma_1,\ldots,\gamma_m$ defined by (2.8).

(ii) The inequalities $\gamma_1>\cdots>\gamma_m$ imply the existence of $\mathcal{E}$.

(iii) We have $\max_\sigma(X_\sigma/\sqrt{N})\to\sqrt{\bar a_1 2\ln\bar\alpha_1}+\cdots+\sqrt{\bar a_m 2\ln\bar\alpha_m}$ *a.s. and also* $\mathsf{E}(\max_\sigma X_\sigma/\sqrt{N})\to\sqrt{2\bar a_1\ln\bar\alpha_1}+\cdots+\sqrt{2\bar a_m\ln\bar\alpha_m}$.

3. GREM: DETAILED ANALYSIS

3.1. **Fluctuations of the partition function.** For *any* sequence of indices $0<J_1<\cdots<J_m=n$, the partition function (1.4) of the GREM can be written as:

$$Z_{\beta,N} = e^{\sum_{j=1}^m\left(\beta N\sqrt{2\bar a_j\ln\bar\alpha_j}-\beta\gamma_j[\ln(N\ln\bar\alpha_j)+\ln 4\pi]/2\right)} \times\sum_{\sigma_1\ldots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\ldots\sigma_{J_1}})}\cdots\sum_{\sigma_{J_{m-1}+1}\ldots\sigma_{J_m}} e^{\beta\gamma_m u^{-1}_{\ln\bar\alpha_m,N}(\bar X^{\sigma_1\ldots\sigma_{J_{m-1}}}_{\sigma_{J_{m-1}+1}\ldots\sigma_{J_m}})} \tag{3.1}$$

where $\bar a_l\equiv\sum_{i=J_{l-1}+1}^{J_l}a_i$, $\bar\alpha_l\equiv\prod_{i=J_{l-1}+1}^{J_l}\alpha_i$, $\gamma_l\equiv\sqrt{\bar a_l}/\sqrt{2\ln\bar\alpha_l}$, $l=1,2,\ldots,m$, and the random variables $\bar X^{\sigma_1\ldots\sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1}\ldots\sigma_{J_l}}$ are defined in (2.3). For any sequence $J_1,\ldots,J_m$, satisfying conditions (2.4), the point process (2.6) in the exponent of (3.1) converges to the corresponding Poisson cascade by Theorem 2.2. The sequence constructed according to (2.7) gives the correct scale of fluctuations of $Z_{\beta,N}$ via (3.1). Nevertheless it should be cut at a certain level $J_{l(\beta)}$ that depends on the temperature: using the sequence $\gamma_1>\gamma_2>\cdots>\gamma_m$ defined in (2.8), we set

$$l(\beta)\equiv\max\{l\geq 1:\beta\gamma_l>1\} \tag{3.2}$$

and $l(\beta)\equiv 0$ if $\beta\gamma_1\leq 1$. This definition (3.2) has a simple geometric interpretation:

$$l(\beta)\equiv\max\left\{l\geq 1:\beta\sqrt{\frac{(\widehat{A})'(P_{J_{l-1}})}{2\ln 2}}>1\right\}.$$

In [CCP], the limit of the free energy has been computed in terms of (2.8) and (3.2):

$$\lim_{N\to\infty} N^{-1}\ln Z_{N,\beta} = \beta\big(\sqrt{2\bar a_1 \ln\bar\alpha_1}+\cdots+\sqrt{2\bar a_{l(\beta)}\ln\bar\alpha_{l(\beta)}}\big)\big)+\sum_{i=J_{l(\beta)}+1}^{n}(\beta^2 a_i/2+\ln\alpha_j),\ \text{a.s.} \tag{3.3}$$

We see that the domain $\{\beta : l(\beta)=0\}=\{\beta:\beta\le 1/\gamma_1\}$ is the high temperature region, where $\lim_{N\to\infty}=\frac{1}{N}\mathsf{E}\ln Z_{\beta,N}=\lim_{N\to\infty}\frac{1}{N}\ln\mathsf{E}\,Z_{\beta,N}$. The next theorem gives the fluctuations of the partition function.

Theorem 3.1. [BK1] *Let $\alpha_i\ge 1$, $0<a_i<1$, $i=1,2,\dots,n$, $\prod_{i=1}^n\alpha_i=2$, $\sum_{i=1}^n a_i=1$. Let $J_1,J_2,\dots,J_m\in\mathbb{N}$ be the sequence of indices defined by (2.7), the parameters $\bar a_i$, $\bar\alpha_i$, γ_i be defined by (2.8) and $l(\beta)$ be defined by (3.2).*

If $l(\beta)=0$, then $\frac{Z_{\beta,N}}{2^N e^{\beta^2N/2}}\to C(\beta)$.

If $l(\beta)>1$, then

$$e^{\sum_{j=1}^{l(\beta)}\left(-\beta N\sqrt{2\bar a_j\ln\bar\alpha_j}+\beta\gamma_j[\ln(N\ln\bar\alpha_j)+\ln 4\pi]/2\right)-N\sum_{i=J_{l(\beta)}+1}^{n}(\beta^2a_i/2+\ln\alpha_j)}Z_{\beta,N}$$
$$\stackrel{\mathcal{D}}{\to}\ C(\beta)\int_{\mathbb{R}^{l(\beta)}} e^{\beta\gamma_1x_1+\beta\gamma_2x_2+\cdots+\beta\gamma_{l(\beta)}x_{l(\beta)}}\mathcal{P}^{(l(\beta))}(dx_1\dots dx_{l(\beta)}). \tag{3.4}$$

This integral is computed over the Poisson cascades $\mathcal{P}^{(l(\beta))}$ on $\mathbb{R}^{l(\beta)}$, defined in Lemma 2.1, with the constants K_j of Theorem 2.2. The constant

$$C(\beta)=1,\quad \textit{if } \beta\gamma_{l(\beta)+1}<1, \tag{3.5}$$

and $0<C(\beta)<1$, if if $\beta\gamma_{l(\beta)+1}=1$[3].

3.2. **Gibbs measure: approach via Ruelle's probability cascades.** We consider everywhere below $\bar a_i$, $\bar\alpha_i$, γ_i defined by (2.8) according to (2.7) and $l(\beta)$ defined by (3.2). Let us denote the jump points of the derivative of the concave hull $\widehat{A}'(x)$ by

$$q_l\equiv\sum_{n=1}^{l}\frac{\ln\bar\alpha_n}{\ln 2},\quad l=1,2,\dots,m, \tag{3.6}$$

with the convention $q_0=0$. They are illustrated on Figure 4. Let $B_l(\sigma)$ be the ball in Σ_N with center σ and radius $1-q$:

$$B_l(\sigma)\equiv\{\sigma'\in\Sigma_N: d_N(\sigma,\sigma')\le 1-q_l\}=\{\sigma':\sigma'_1\dots\sigma'_{J_l}=\sigma_1\dots\sigma_{J_l}\},\quad l=1,2,\dots,l(\beta). \tag{3.7}$$

Let us define the point process $\mathcal{W}^{(m)}_{\beta,N}$ on $(0,1]^m$ as

$$\mathcal{W}^{(m)}_{\beta,N}\equiv\sum_\sigma\delta_{(\mu_{\beta,N}(B_1(\sigma)),\dots,\mu_{\beta,N}(B_m(\sigma)))}\frac{\mu_{\beta,N}(\sigma)}{\mu_{\beta,N}(B_m(\sigma))} \tag{3.8}$$

and its projection on the last coordinate

$$\mathcal{R}^{(m)}_{\beta,N}\equiv\sum_\sigma\delta_{\mu_{\beta,N}(B_m(\sigma))}\frac{\mu_{\beta,N}(\sigma)}{\mu_{\beta,N}(B_m(\sigma))}. \tag{3.9}$$

It is easy to see that $\mathcal{W}^{(m)}_{\beta,N}$ satisfy the following relation:

$$\mathcal{W}^{(m)}_{\beta,N}(dw_1,\dots,dw_m)=\int_0^1\mathcal{W}^{(m+1)}_{\beta,N}(dw_1,\dots,dw_m,dw_{m+1})\frac{w_{m+1}}{w_m}$$

where the integral is taken over the last coordinate w_{m+1}. The next theorem gives the limits of these point processes for all $m\le l(\beta)$.

[3]Explicit formulae for $C(\beta)$ are given in [BK1].

Theorem 3.2. [BK1] *Let $l(\beta) \geq 1$ i.e. $\beta > \sqrt{2\ln 2/(\widehat{A})'(0)}$. If $m \leq l(\beta)$, then the point process $\mathcal{W}_{\beta,N}^{(m)}$ on $(0,1]^m$ converges weakly, as $N \to \infty$, to the point process $\mathcal{W}_{\beta\vec{\gamma}}^{(m)}$, whose atoms $w(i)$ are expressed through the points $(x_1(i), \ldots, x_m(i))$ of the Poisson cascade $\mathcal{P}^{(m)}$ of Lemma 2.1, with constants K_j of Theorem 2.2, as follows:*

$$(w_1(i), \ldots, w_m(i)) \tag{3.10}$$
$$= \Big(\frac{\int \mathcal{P}^{(m)}(dy)\delta(y_1 - x_1(i))e^{\beta(\gamma,y)}}{\int \mathcal{P}^{(m)}(dy)e^{\beta(\gamma,y)}}, \ldots, \frac{\int \mathcal{P}^{(m)}(dy)\delta(y_1 - x_1(i))\ldots\delta(y_m - x_m(i))e^{\beta(\gamma,y)}}{\int \mathcal{P}^{(m)}(dy)e^{\beta(\gamma,y)}}\Big).$$

The vector $\vec{\gamma} = (\gamma_1, \ldots, \gamma_m)$ is defined by (2.8) according to (2.7). The process $\mathcal{R}_{\beta,N}^{(m)}$ converges to the process $\mathcal{R}_{\beta}^{(m)}$, where the atoms are the last components of the atoms of (3.10).

The balls $B_{l(\beta)}(\sigma)$ are the smallest ones that have positive mass, $\mu_{\beta,N}$, as $N \to \infty$: For $m > l(\beta)$, $\mu_{\beta,N}(B_m(\sigma)) \to 0$ for any $\sigma \in \Sigma_N$. Iff $J_{l(\beta)} = n$, i.e. $\beta > 1/\gamma_m = \sqrt{2\ln 2/\lim_{x\to 1}(\widehat{A})'(x)}$, these balls consist of a single configuration, σ. In this case the mass of the Gibbs measure is concentrated on certain randomly chosen individual configurations. Otherwise, these balls consist of all configurations having the same spins as σ starting from the first sit up to the $J_{l(\beta)}$th site.

Definition 3.1. [Ru] *The process $\mathcal{W}_{\beta\vec{\gamma}}^{(m)}$ defined in Theorem 3.2 is called the process of probability cascades on $[0,1]^m$ with m levels and parameters $\beta\gamma_1 > \cdots > \beta\gamma_m > 1$. I*

The most complete object of Theorem 3.2 is of course the process $\mathcal{W}_{\beta,N}^{(l(\beta))}$. Thus, Theorem 3.2 asserts the convergence of the point process $\mathcal{W}_{\beta,N}^{(l(\beta))}$ of Derrida's model with parameters $n \geq 1$, a_i, α_i to the point process of probability cascades of Ruelle's model with parameters $l(\beta)$ and $\beta\gamma_1, \ldots, \beta\gamma_{l(\beta)}$ defined by (2.8) and (3.2). Let us also emphasize the fact that the parameters of the limiting process of probability cascades depend only on the concave hull $\widehat{A}(x)$ and on β.

3.3. **Distribution of the overlaps.** One of the most important physical objects is the distribution of the overlap

$$\frac{\sigma \cdot \sigma'}{N} = \frac{\sum_{i=1}^N \sigma_i \sigma_i'}{N} \tag{3.11}$$

of two spin configurations under the Gibbs measure:

$$\widetilde{f}_{\beta,N}(q) \equiv \mu_{\beta,N}^{\otimes 2}\Big(\frac{(\sigma \cdot \sigma')}{N} \leq q\Big). \tag{3.12}$$

In the context of the GREM it appears more natural to consider the ultrametric distance

$$f_{\beta,N}(q) \equiv \mu_{\beta,N}^{\otimes 2}\big(d_N(\sigma, \sigma') \geq 1 - q\big). \tag{3.13}$$

The next theorem asserts the remarkable fact that the laws of these two objects coincides in the thermodynamic limit.

Theorem 3.3. [BK1] *The distribution functions $f_{\beta,N}$ et $\widetilde{f}_{\beta,N}$ converge in law to the same distribution function f_β as $N \to \infty$. Moreover $\mathsf{E}\, f_{\beta,N} \to \mathsf{E}\, f_\beta$ and $\mathsf{E}\, \widetilde{f}_{\beta,N} \to \mathsf{E}\, f_\beta$ where*

$$\mathsf{E}\, f_\beta(q) = \min\Big\{\beta^{-1}\sqrt{\frac{2\ln 2}{(\widehat{A})'(q)}},\ 1\Big\} = \begin{cases} \beta^{-1}\sqrt{\frac{2\ln \bar{\alpha}_j}{\bar{a}_j}} & \text{if } q \in [q_{j-1}, q_j), j \leq l(\beta) \\ 1 & \text{if } q \geq q_{l(\beta)}. \end{cases} \tag{3.14}$$

The function f_β is a step function that jumps at points $\{0, q_1, \ldots, q_{l(\beta)}\}$. For any $q \in [q_{i-1}, q_i)$

$$f_\beta(q) = \int \mathcal{W}^{(l(\beta))}_{\beta\vec{\gamma}}(dw_1, \ldots, dw_{l(\beta)}) w_{l(\beta)}(1 - w_i), \quad i = 1, \ldots, l(\beta); \tag{3.15}$$

$f_\beta(q) = 1$ for $q \geq q_{l(\beta)}$.

Rather then just considering the distribution of the total overlap, we can give a more precise description of the Gibbs measure by considering the vector of overlaps within each hierarchy. Let

$$\begin{aligned} \Delta_l &= [-\ln \bar{\alpha}_l / \ln 2, \ln \bar{\alpha}_l / \ln 2], \quad \text{for } l = 1, 2, \ldots, l(\beta) \\ \Delta_{l(\beta)+1} &= [-\ln(\alpha_{J_{l(\beta)}+1} \cdots \alpha_n) / \ln 2, \ln(\alpha_{J_{l(\beta)}+1} \cdots \alpha_n) / \ln 2]. \end{aligned} \tag{3.16}$$

It is clear that $(\bar{\sigma}_l \cdot \bar{\sigma}'_l)/N \in \Delta_l$, $l = 1, \ldots, l(\beta) + 1$. We introduce the measure $f^{\otimes l(\beta)+1}_{\beta,N}$ on $\Delta_1 \times \Delta_2 \times \cdots \times \Delta_{l(\beta)+1}$ induced by $\bar{\sigma}_l \cdot \bar{\sigma}'_l$ on all levels of limiting probability cascades: for any $I_l \in \Delta_l$, $l = 1, \ldots, l(\beta) + 1$, we put

$$\begin{aligned} f^{\otimes l(\beta)+1}_{\beta,N}(I_1 \times \cdots \times I_{l(\beta)+1}) &\equiv \frac{\mathsf{E}_{\sigma,\sigma'} \prod_{l=1}^{l(\beta)+1} \mathbb{1}_{(\bar{\sigma}_l \cdot \bar{\sigma}'_l)/N \in I_l} e^{\beta\sqrt{N}(X_\sigma + X_{\sigma'})}}{Z^2_{\beta,N}} \\ &= \mu^{\otimes 2}_{\beta,N}\Big(\prod_{l=1}^{l(\beta)+1} \mathbb{1}_{(\bar{\sigma}_l \cdot \bar{\sigma}'_l)/N \in I_l} \Big). \end{aligned} \tag{3.17}$$

Theorem 3.4. [BK1] *The measure $f^{\otimes l(\beta)+1}_{\beta,N}$ converges in law to the following measure on $\Delta_1 \times \Delta_2 \times \cdots \times \Delta_{l(\beta)+1}$:*

$$f^{\otimes l(\beta)+1}_{\beta,N} \to Q_0 \delta_{(0,0,\ldots,0)} + \sum_{j=1}^{l(\beta)} Q_j \delta_{(\ln \bar{\alpha}_1 / \ln 2, \ldots, \ln \bar{\alpha}_j / \ln 2, 0, \ldots, 0)} \qquad N \to \infty.$$

The random variables $Q_1, \ldots, Q_{l(\beta)}$ are defined as

$$Q_j(\beta) \equiv \int \mathcal{W}^{(l(\beta))}_{\beta\vec{\gamma}}(dw_1, \ldots, dw_{l(\beta)}) w_{l(\beta)}(w_j - w_{j+1} \mathbb{1}_{\{j \leq l(\beta)-1\}}), \quad j = 1, \ldots, l(\beta).$$

3.4. **Ghirlanda-Guerra identities.** The process $\mathcal{W}^{(l(\beta))}_{\beta\vec{\gamma}}$ has been constructed explicitly in Theorem 3.2 in terms of Ruelle's probability cascades. This allows to compute all its characteristics. Now we present a different approach that determines $\mathcal{W}^{(l(\beta))}_{\beta\vec{\gamma}}$ completely, without the use of Ruelle's probability cascades. This amounts to the computation of all moments of $\mathcal{W}^{(l(\beta))}_{\beta\vec{\gamma}}$ by recursion, starting from the second one. This approach will bear its full fruits in the analysis of the CREM.

Lemma 3.1. [BK1] *Assume that the parameters α_i and a_i are such that the inequalities (2.4) are strict. Then for any bounded function $h : \Sigma^n_N \to \mathbb{R}$ and for any $i = 1, \ldots, n$*

$$\begin{aligned} \lim_{N \uparrow \infty} \Big| \mathsf{E}\, \mu^{\otimes n+1}_{\beta,N}\big(h(\sigma^1, \ldots, \sigma^n) \mathbb{1}_{\sigma^k_1 \ldots \sigma^k_i = \sigma^{n+1}_1 \ldots \sigma^{n+1}_i}\big) \\ - \frac{1}{n} \mathsf{E}\, \mu^{\otimes n}_{\beta,N}\Big(h(\sigma^1, \ldots, \sigma^n)\Big(\sum_{l \neq k}^{n} \mathbb{1}_{\sigma^l_1 \ldots \sigma^l_i = \sigma^k_1 \ldots \sigma^k_i} + \mathsf{E}\, \mu^{\otimes 2}_{\beta,N}(\mathbb{1}_{\sigma^1_1 \ldots \sigma^1_i = \sigma^2_1 \ldots \sigma^2_i})\Big)\Big) \Big| = 0 \end{aligned} \tag{3.18}$$

The proof of this lemma is based on the integration by parts of Gaussian random variables coupled with a concentration of measure argument. This lemma determines the so-called *Ghirlanda-Guerra identities* for the GREM: it allows to compute the expected distance distribution function between n replicas under the Gibbs measure by the recurrence procedure (3.18) for $n = 3, 4 \ldots$ subsequently starting from $n = 2$. To see this, it suffices to put the function h equal to the indicator function of distances between $n+1$ replicas and to note that by (3.18) the term with $n+1$ replicas is completely determined by the terms with n replicas and by the one with two replicas,

$$\lim_{N\uparrow\infty} \mathsf{E}\,\mu^{\otimes 2}_{\beta,N}(\mathbb{1}_{\sigma^1_1\dots\sigma^1_i=\sigma^2_1\dots\sigma^2_i}) = 1 - \mathsf{E}\,f_\beta\Big(\sum_{j=1}^{i} \ln\alpha_j/\ln 2\Big),$$

that has been already computed in (3.14). In fact, let $\underline{\underline{J}} \equiv (\underline{J}_0,\dots,\underline{J}_N)$ be a set of subsets of $1,\dots,n+1$ that determines the distances between $n+1$ replicas: each element $\underline{J}_r = (J_{r,1},\dots,J_{r,j_r})$ is a collection of subsets of $1,\dots,n+1$ that reassembles the numbers of configurations for which the first r coordinated of the spin variables are equal. Then, for any $J_{r,i}$, there exists $J_{r-1,k}$, such that $J_{r,i}\subset J_{r-1,k}$. Assume that $J_{r,i}$ is the set of numbers $\{j^{r,i}_1,\dots,j^{r,i}_{|J_{r,i}|}\}$. We can then define the function:

$$\mathcal{A}_{\underline{\underline{J}}} \equiv \prod_{r=1}^{N}\prod_{i=1}^{j_r} \mathbb{1}_{\{\sigma_1^{j_1^{r,i}}\dots\sigma_r^{j_1^{r,i}}=\dots=\sigma_1^{j^{r,i}_{|J_{r,i}|}}\dots\sigma_r^{j^{r,i}_{|J_{r,i}|}}\}}. \tag{3.19}$$

The length of $\underline{\underline{J}}$ is $\|\underline{\underline{J}}\| = n+1$. Let us construct a set $\underline{\underline{J}}'$ of length n by erasing everywhere in $\underline{\underline{J}}$ the integer $n+1$. Indeed, there exists $r\in\{1,\dots,N\}$, such that there exists $l\in\{1,\dots,n\}$, such that $n+1$ and l belong to the same subset, $J_{r,i}$, of $\underline{\underline{J}}$, i.e. their first r coordinates coincide. If we choose the maximal r with this property, this determines uniquely the participation of $n+1$ everywhere in $\underline{\underline{J}}$: for any $p=1,2,\dots,r-1$ it belongs to the same subset $J_{p,i}$ as l. In other words, once the ultrametric distances between n replicas are fixed, it suffices to specify the distance of the $(n+1)$th replica to the closest to it, in order to determine completely its distance to all other replicas. This implies

$$\mathcal{A}_{\underline{\underline{J}}} = \mathcal{A}_{\underline{\underline{J}}'}\,\mathbb{1}_{\sigma^l_1\dots\sigma^l_r=\sigma^{n+1}_1\dots\sigma^{n+1}_r}. \tag{3.20}$$

Hence, substituting $h = A_{\underline{\underline{J}}'}$ in Lemma 3.1, we can compute $\lim_{N\to\infty}\mathsf{E}\,\mu^{\otimes n}_{\beta,N}(A_{\underline{\underline{J}}})$ subsequently for $n=3,4,\dots$, starting from $n=2$, given by (3.14).

From the other hand, in [BK1], we expressed all moments of $\mathcal{W}^{(m)}_\beta$ in terms of $\mathcal{A}_{\underline{\underline{J}}}$:

$$\begin{aligned}&\int \mathcal{W}^{(m)}_{\beta,N}(dw)w_1^{i_1}\dots w_l^{i_l}\dots w_m^{i_m}\\ &= \mu^{\otimes(i_1+\dots+i_m)}_{\beta,N}\Big(\mathbb{1}_{\{\sigma^1_1=\dots=\sigma_1^{i_1+\dots+i_m}\}}\dots\mathbb{1}_{\{\sigma_1^{i_1+\dots+i_{l-1}+1}\dots\sigma_l^{i_1+\dots+i_{l-1}+1}=\dots=\sigma_1^{i_1+\dots+i_m}\dots\sigma_l^{i_1+\dots+i_m}\}}\\ &\quad\dots\mathbb{1}_{\{\sigma_1^{i_1+\dots+i_{m-1}+1}\dots\sigma_m^{i_1+\dots+i_{m-1}+1}=\dots=\sigma_1^{i_1+\dots+i_m}\dots\sigma_m^{i_1+\dots+i_m}\}}\Big)\end{aligned} \tag{3.21}$$

where $i_m\geq 1$, otherwise this expression is infinite. This implies the following theorem.

Theorem 3.5. [BK1] *The process $\mathcal{W}^{(m)}_{\beta\vec\gamma}$ is completely determined by the relations (3.18) up to the mean value of the two-replica distance distribution function given by (3.14).*

Theorem 3.5 in the case of the REM has been first proven by M. Talagrand. Lemma 3.1 implies also the following result, that has been remarked by Ruelle in [Ru].

Corollary 3.1. *The l-th marginal of Ruelle's process of probability cascades $\mathcal{W}^{(m)}_{\beta\vec\gamma}$ with m levels and parameters $\beta\gamma_1>\dots>\beta\gamma_m>1$ has the same distribution as Ruelle's process of one level with parameter $\beta\gamma_l$, $l=1,\dots,m$*

To see this, we need to control all moments of this marginal that can be expressed via the quantities $\mu^{\otimes r}_{\beta,N}(\mathbb{1}_{\{\sigma^1_1\dots\sigma^1_l=\dots=\sigma^r_1\dots\sigma^r_l\}})$, which in turn satisfy the identities (3.18) for $r=3,\dots$, while for $r=2$ they are defined by $f_\beta(q_l) = (\beta\gamma_l)^{-1}$. But these identities are the same for the GREM with one hierarchy (i.e. the REM), with the same two replica

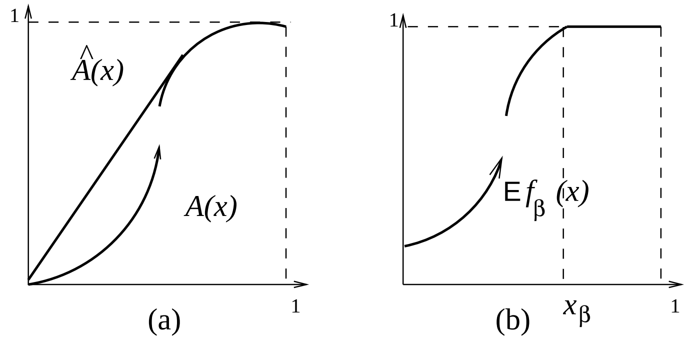

FIGURE 5. (a) Concave hull of $A(x)$, (b) The function (4.5).

distance distribution. Consequently, the l-th marginal of $\mathcal{W}_{\beta,N}^{(m)}$, in the limit $N \to \infty$ behaves as $\sum_\sigma \delta_{\mu_{\beta,N}(\sigma)}$ of the REM at temperature $\widetilde{\beta} = \beta\gamma_l\sqrt{2\ln 2}$.

3.5. **Empirical distance distribution function $\mathcal{K}_\beta$.** The process $\mathcal{W}_{\beta,N}^{(l(\beta))}$ is a point process on $[0,1]^m$. Its points $(\mu_{\beta,N}(B_1(\sigma)), \ldots, \mu_{\beta,N}(B_{l(\beta)}(\sigma)))$ can be considered as values of the ultrametric distance distribution function around σ

$$m_\sigma(x) = \mu_{\beta,N}(d_N(\sigma,\sigma') \leq 1-x) \tag{3.22}$$

at points $x = q_0, \ldots, q_{l(\beta)}$. The limit of this distribution function is a step-function that jumps precisely at these points. We could then consider $\mathcal{W}_{\beta,N}^{(l(\beta))}$ as a point process of these distribution functions : $\mathcal{W}_{\beta,N}^{(l(\beta))} = \sum_\sigma \delta_{m_\sigma(\cdot)}$.

This object is, however, not properly adapted to the CREM. In the analysis of the CREM it is essentially imperative to replace it by the probability measure on these distribution functions:

$$\mathcal{K}_{\beta,N} = \sum_\sigma \mu_{\beta,N}(\sigma)\delta_{m_\sigma(\cdot)}, \tag{3.23}$$

that we have introduced and discussed in the introduction. To conclude the analysis of the GREM, we give its asymptotic behavior in the following theorem.

Theorem 3.6. [BK4] *The process $\mathcal{K}_{\beta,N}$ converges weakly to the point process $\mathcal{K}_\beta$*

$$\mathcal{K}_\beta = \int_{\mathbb{R}^{l(\beta)}} \mathcal{W}_{\beta\vec{\gamma}}^{(l(\beta))}(dw)w(l(\beta))\delta_{m(w)} \tag{3.24}$$

where the measures $m(w)$ are defined by the formulas :

$$m(w) = (1-w(1))\delta_1 + (w(1)-w(2))\delta_{1-\ln\alpha_1/\ln 2} + \cdots + w(l(\beta))\delta_{1-\ln(\alpha_1\cdots\alpha_{l(\beta)})/\ln 2}. \tag{3.25}$$

4. CREM: IMPLICIT APPROACH

We start now the analysis of the CREM with covariances (1.2) where $A(x) : [0,1] \to [0,1]$ is a right-continuous distribution function with the concave hull $\widehat{A}(x)$ whose right right derivative we denote by $(\widehat{A})'(x)$; see Figure 5(a). We assume that A is non-critical in the sense that it is equal to its concave hull $\widehat{A}$ only on the set of extremal points of the convex hull.

4.1. **Maximum of the Hamiltonian. Limit of the free energy.**

Theorem 4.1. [BK2] *Let $\{X_\sigma\}$ be a family of 2^N standard Gaussian random variables with covariances (1.2). Then*

$$\lim_{N\to\infty} \mathsf{E} \max_{\sigma} \frac{X_\sigma}{\sqrt{N}} = \sqrt{2\ln 2} \int_0^1 \sqrt{(\widehat{A})'(x)}dx. \tag{4.1}$$

Theorem 4.2. [BK2] *Let $\{X_\sigma\}$ be a family of 2^N Gaussian random variables with covariances (1.2). Let*

$$x_\beta = \sup\Big\{x \;\Big|\; (\widehat{A})'(x) > \frac{2\ln 2}{\beta^2}\Big\}. \tag{4.2}$$

Then

$$\lim_{N\to\infty} N^{-1} \mathsf{E} \ln Z_{\beta,N} = \sqrt{2\ln 2}\beta \int_0^{x_\beta} \sqrt{(\widehat{A})'(x)}dx + \frac{\beta^2}{2}(1-\widehat{A}(x_\beta)) + \ln 2(1-x_\beta). \tag{4.3}$$

Consequently the critical temperature of the CREM defined as

$$\beta_0 = \sup\{\beta : \lim_{N\to\infty} N^{-1} E\ln Z_{\beta,N} = \lim_{N\to\infty} N^{-1} \ln \mathsf{E}\, Z_{\beta,N}\}$$

equals :

$$\beta_0 = \sqrt{\frac{2\ln 2}{\lim_{x\downarrow 0}(\widehat{A})'(x)}}. \tag{4.4}$$

The proofs of these theorems rely heavily on results already obtained for the GREM. Namely, we approximate $A(x)$ from above and below by step-functions for which corresponding results have been already established in Theorem 2.3 (ii) and (3.3) in the study of the GREM. Then the results announced in Theorems 4.1 and 4.2 follow from theorems about the comparison of the mean values of convex or concave functions of Gaussian processes implied by the comparison of the covariances of these processes (see Theorem 3.1 in [LT](Kahane's Theorem)).

We are not able to evaluate the fluctuations of the partition function of the CREM. We anticipate that they depend not only on $\widehat{A}(x)$ in view of the analysis of the maximum of branching Brownian motion by Bramson [Br]. But (4.3) suffices to deduce the following very important result which is in the basis of the description of the CREM's Gibbs measure.

4.2. **Two-replicas ultrametric distance distribution function.**

Theorem 4.3. [BK2] *Let $\{X_\sigma\}$ be a family of 2^N Gaussian random variables with covariances (1.2). Let x_β be defined by (4.2). Then*

$$\lim_{N\to\infty} \mathsf{E}\, \mu^{\otimes 2}_{\beta,N}(d_N(\sigma,\sigma') \geq 1-x) = \mathsf{E}\, f_\beta(x) = \begin{cases} \beta^{-1}\sqrt{\dfrac{2\ln 2}{(\widehat{A})'(x)}} & \text{if } x < x_\beta \\ 1 & \text{if } x \geq x_\beta \end{cases} \tag{4.5}$$

The function $\mathsf{E}\, f_\beta(x)$ is illustrated on Figure 5(b). Let us sketch the main points of the proof. The result of Theorem 4.2 allows to compute the limit of the free energy

$$\lim_{N\to\infty} N^{-1} \mathsf{E} \ln Z^u_{\beta,N} = F^u_\beta$$

for the CREM where the function $A(x)$ is slightly perturbed by a small parameter $u > 0$ in a neighborhood of the point x. Next, using the integration by parts of Gaussian random variables, we show that the desired quantity $\lim_{N\to\infty} \mathsf{E}\, \mu^{\otimes 2}_{\beta,N}(d_N(\sigma,\sigma') \geq 1-x)$ is equal to

$\lim_{N\to\infty}\frac{d}{du}N^{-1}\ln Z^u_{\beta,N}\Big|_{u=0}$ that can be computed as $\frac{d}{du}F^u_\beta\Big|_{u=0}$ by convexity and leads to (4.5).

4.3. **Ghirlanda-Guerra identities.** The next lemma is a generalization of Lemma 3.1: it proves Ghirlanda-Guerra identities in the case of the CREM.

Lemma 4.1. [BK2] *For any $n\in\mathbb{N}$, any bounded function $h(x)$ and $x\in[0,1]\setminus x_\beta$*

$$\lim_{N\uparrow\infty}\Big|\,\mathsf{E}\,\mu^{\otimes n+1}_{\beta,N}\left(h(\sigma^1,\ldots,\sigma^n)1\!\!1_{d_N(\sigma^k,\sigma^{n+1})>x}\right) - \frac{1}{n}\,\mathsf{E}\,\mu^{\otimes n+1}_{\beta,N}\Big(h(\sigma^1,\ldots,\sigma^n)\Big(\sum_{l\neq k}^{n}1\!\!1_{d_N(\sigma^k,\sigma^l)>x}+\mathsf{E}\,\mu^{\otimes 2}_{\beta,N}(1\!\!1_{d_N(\sigma^1,\sigma^2)>x})\Big)\Big)\Big| = 0. \quad (4.6)$$

One of the pillars of the proof of this lemma is the representation $X_\sigma = X_\sigma(1)$ where $X_\sigma(t)$ is the family of standard Gaussian processes on $[0,1]$ with covariances: cov $(X_\sigma(t),X_{\sigma'}(s)) = A(t\wedge s\wedge d_N(\sigma,\sigma'))$. Two other pillars are the same as in the case of the GREM : the integration by parts of Gaussians and a concentration of measure argument.

This lemma implies the following important Theorem 4.4, that determines implicitly the empirical distance distribution function $\mathcal{K}_{\beta,N}$. Let us define a family of measures $\mathcal{Q}^n_N$ on $[0,1]^{n(n-1)/2}$

$$\mathcal{Q}^{(n)}_{\beta,N}(\bar{d}_N\in\mathcal{C})\equiv\mathsf{E}\,\mu^{\otimes n}_{N,\beta}(\bar{d}_N\in\mathcal{C}) \quad (4.7)$$

where $\bar{d}_N$ is the vector of distances between n replicas with components $d^{k,l}_N = d_N(\sigma^l,\sigma^k)$, $1\le l<k\le n$, and $\mathcal{C}$ is a Borel subset of $[0,1]^{n(n-1)/2}$. We denote by $\mathcal{B}_k$ the sigma-field generated by the first $k(k-1)/2$ coordinates.

Theorem 4.4. [BK2] *For any $n\in\mathbb{N}$, the family of measures $\mathcal{Q}^{(n)}_{\beta,N}$ converges, as $N\uparrow\infty$, to the limiting measure $\mathcal{Q}^{(n)}_\beta$. All these measures are uniquely determined by (4.5). They satisfy the identities:*

$$\mathcal{Q}^{(n+1)}_\beta(d^{k,n+1}\in C|\mathcal{B}_n) = \frac{1}{n}\mathcal{Q}^{(2)}_\beta(C)+\frac{1}{n}\sum_{l=1,l\neq k}^{n}\mathcal{Q}^{(n)}_\beta(d^{k,l}\in C|\mathcal{B}_n) \quad (4.8)$$

for any Borel subset $C\subset[0,1]$. Consequently $\mathcal{K}_{\beta,N}$ defined by (3.23) and (3.22) converges in law to the limit $\mathcal{K}_\beta$ with generalized moments determined by $\mathcal{Q}^{(n)}_\beta$.

The recurrent formulas (4.8) come from (4.6) if we put h equal to the indicator function of any desired event of $\mathcal{B}_n$. Let us remark also that, due to the ultrametric structure, once the distances between n replicas are prescribed, it suffices to fix the distance from the $(n+1)$-th replica up to the closest to it among the n replicas $\{1,2,\ldots,n\}$, in order to determine its distance up to all other $n-1$ replicas. This fact is already formally explained in (3.19). Then the formulas (4.8) determine completely the measures $\lim_{N\to\infty}\mathcal{Q}^{(n)}_{\beta,N} = \mathcal{Q}^{(n)}_\beta$ up to the measure $Q^{(2)}_\beta$ already computed in (4.5). The moments $\mathcal{K}_{\beta,N}$ can be expressed in terms of the measures $\mathcal{Q}^{(n)}_{\beta,N}$. This implies the convergence of $\mathcal{K}_{\beta,N}$ to a limiting object $\mathcal{K}_\beta$ with moments expressed in terms of $\mathcal{Q}^{(n)}_\beta$.

4.4. **Marginals of $\mathcal{K}_\beta$ in terms of Ruelle's probability cascades.** In this subsection we give an explicit form of all marginals of $\mathcal{K}_{\beta,N}$. Let $0<t_1<t_2<\cdots<t_m<1$ be points of increase of the function (4.5), $t_0=0$. We can define then the marginal process:

$$\mathcal{K}_{\beta,N}(t_0,t_1,\ldots,t_m) = \sum_\sigma \mu_{\beta,N}(\sigma)\delta_{m_\sigma(t_0),m_\sigma(t_1),m_\sigma(t_2),\ldots,m_\sigma(t_m)}. \quad (4.9)$$

Theorem 4.5. [BK2] *Let $t_0 = 0 < t_1 < t_2 < \ldots < t_m \le 1 = t_{m+1}$ be points of increase of the function (4.5). Consider the GREM of $m+1$ hierarchies, with parameters α_i such that $\ln \alpha_i / \ln 2 = t_i - t_{i-1}$, $i = 1, \ldots, m+1$, a_i with $\sum_{i=1}^{m+1} a_i = 1$ at temperature $\widetilde{\beta}$ such that $\widetilde{\beta}^{-1}\sqrt{2 \ln \alpha_i / a_i} = \beta^{-1}\sqrt{2\ln 2/(\widehat{A})'(t_{i-1})}$, $i = 1, \ldots, m+1$. Then*

$$\lim_{N\to\infty} \mathcal{K}_{\beta,N}(t_0, t_1, \ldots, t_m) = \mathcal{K}_{\widetilde{\beta}}^{(m+1)}, \tag{4.10}$$

where $\mathcal{K}_{\widetilde{\beta}}^{(m+1)}$ is the empirical distance distribution function of the GREM computed in Theorem 3.6 in terms of Ruelle's probability cascades.

The second moments of $\lim_{N\to\infty} \mathcal{K}_{\beta,N}(t_0, t_1, \ldots, t_m)$ and of $\mathcal{K}_{\widetilde{\beta}}^{(m)}$ for the GREM in question are the same due to the choice of the parameters of the GREM. Then all their moments coincide by the Ghirlanda-Guerra identities. The parameters a_i and $\widetilde{\beta}$ explicitly are equal to :

$$a_i = \kappa \frac{\widehat{A}'(t_{i-1}) \ln \alpha_i}{\ln 2}, \quad \widetilde{\beta} = \kappa^{-1/2}\beta, \quad \kappa = \Big(\sum_{i=1}^{m+1} \frac{\widehat{A}'(t_{i-1}) \ln \alpha_i}{\ln 2} \Big)^{-1}, \quad i = 1, \ldots, m+1.$$

5. Genealogies and Neveu's branching process.

5.1. **Problems with the explicit description of limiting Gibbs measures.** We obtained an implicit description of the limiting Gibbs measure of the CREM via recursive computation of all moments of $\mathcal{K}_\beta$. Nevertheless, we would like to identify explicitly a limiting measure to which our Gibbs measures converge and that encodes the full geometric information contained in $\mathcal{K}_\beta$. This is not immediately possible for the following reason. In [Bo] one of us proposed to describe the infinite volume limit of the Gibbs measure for the REM by considering the image of the hypercube Σ_N on $[0,1]$ through the map $r_N : \Sigma_N \to (0,1]$ (1.7). However, the definition of $\mathcal{K}_{\beta,N}$ involves masses of sets $\{\sigma' : d_N(\sigma, \sigma') < 1-t\}$. If we map such sets on the unit interval via r_N, we obtain intervals $(r_{[Nt]} - 2^{-[tN]}, r_{[Nt]}]$ of length $2^{-[tN]}$. So, when $N = \infty$, these sets map to intervals of length $2^{-t\infty}$. We can not analyze the structure of the measure by looking at intervals of the size $2^{-t\infty}$.

What will however be possible, is the following. We will introduce the notion of a flow of compatible probability measures on $[0,1]$ indexed by pairs of parameters $s \le t \in I$ and with distribution functions satisfying the compatibility assumption (5.1). Next, we will associate to each of such flows a certain genealogical structure on $[0,1]$ described by a genealogical map, $K_T \in M_1(M_1([0,1]))$, which is an empirical distribution of family sizes of all individuals as functions of degree of relatedness. Then we will provide a flow of compatible probability measures for each finite N with the genealogy describing efficiently the geometry of the Gibbs measure of the CREM: its genealogical map, $K_T^{\beta,N}$, will equal the empirical distance distribution function $\mathcal{K}_{\beta,N}$. Finally, we will show that this flow of probability measures converges as $N \to \infty$ to the flow of compatible random probability measures with distribution functions that are normalized stable subordinators associated to Neveu's continuous state branching process via an appropriate deterministic time change. This convergence of flows is understood in the sense that their genealogical maps, $\mathcal{K}_{\beta,N} = K_T^{\beta,N}$, converge. Thus, the limiting geometry of the Gibbs measure of the CREM will be expressed in terms of the genealogy of Neveu's continuous state branching process modulo a time change determined only by $\mathsf{E}\, f_\beta(x)$ of (4.5).

5.2. **Genealogical map of a flow of probability measures.**

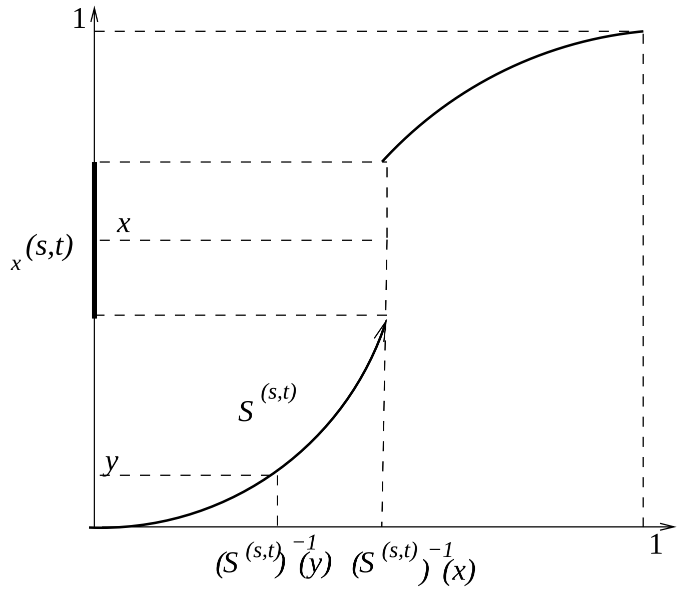

FIGURE 6. Genealogical structure induced by a flow $S^{(s,t)}$.

Definition 5.1. *A two-parameter family of measures with probability distribution functions $S^{(s,t)}$ on $[0,1]$, $s \le t$, $s,t \in I \subset \mathbb{R}$, is called a flow of compatible probability measures on I, if and only if for any collection $t_1 \le t_2 \le \cdots \le t_n \subset I$*

$$S^{(t_1,t_n)} = S^{(t_{n-1},t_n)} \circ S^{(t_{n-2},t_{n-1})} \circ \ldots S^{(t_2,t_3)} \circ S^{(t_1,t_2)} \tag{5.1}$$

holds.

Let us admit the following terminology. We say that each point $a \in [0,1]$ is an individual in generation s and its image $S^{(s,t)}(a) \in [0,1]$ is its offspring in generation t. Let us define for any distribution function $\Theta(x)$ its inverse function

$$\Theta^{-1}(x) = \inf\{a \mid \Theta(a) \ge x\}. \tag{5.2}$$

Then each individual $x \in [0,1]$ in generation t has an ancestor a in generation s which is $a = (S^{(s,t)})^{-1}(x)$. Given an individual $x \in [0,1]$ in generation t, let us look for individuals x' having the same ancestor as x in generation s:

$$m_x(s,t) \equiv \{x' : (S^{(s,t)})^{-1}(x') = (S^{(s,t)})^{-1}(x)\}. \tag{5.3}$$

If $S^{(s,t)}$ is continuous at $a = (S^{(s,t)})^{-1}(x)$, then any individual $x' \neq x$ has a different ancestor from the one of x. If $S^{(s,t)}$ makes a jump at $a = (S^{(s,t)})^{-1}(x)$, then the family (5.3) of the individual x having the same ancestor as x in generation s is the following interval :

$$m_x(s,t) = \lim_{\eta \downarrow 0} \Big(S^{(s,t)}\big((S^{(s,t)})^{-1}(x) - \varepsilon\big) \ , \ S^{(s,t)} \circ (S^{(s,t)})^{-1}(x) \Big] .$$

In Figure 5.2 the individual x in generation t has a family of "cousins" $m_x(s,t)$ having the same "grand-father" in generation s, while the individual y is the unique "grand-child" of his ancestor in generation s. We are mainly interested in a nontrivial case when functions $S^{(s,t)}$ make jumps. The next lemma justifies this terminology. It says that any individual having an ancestor in common with x in generation s has necessarily an ancestor in common with x in any generation $s' < s$. In other words, if we partition the interval $[0,1]$ into families $m_x(s',t)$ having the same ancestor in generation s', then the partition into families $m_x(s,t)$ having the same ancestor in generation $s > s'$ is a refinement of the previous one.

Lemma 5.1. *Let $S^{(s,t)}$ be distribution functions of a flow of measures according to Definition 5.1. Then for all $x \in [0,1]$*

$$m_x(s,t) \subset m_x(s',t) \quad \forall s' < s \le t \in I. \tag{5.4}$$

Whenever $t = T$ is fixed, the function $|m_x(\cdot,T)|$ is the family size of the individual x in generation T as a function of the degree of relatedness. By Lemma 5.1, it is a decreasing function on I. Finally, we define the associated empirical distribution of the functions $|m_x(\cdot,T)|$

$$K_T = \int_0^1 dx \delta_{|m_x(\cdot,T)|}. \tag{5.5}$$

This construction allows to associate to any flow of probability measures, in the sense of Definition 5.1, an empirical distribution K_T. If we assume, in addition, that $[0,T] \subset I$ and $|m_x(\cdot,T)|$ are right-continuous, then $1 - |m_x(\cdot,T)|$ are probability distribution functions. Then we will think of K_T as a map from flows of probability measures into $M_1(M_1([0,1]))$ which we call the genealogical map.

5.3. **Coalescent associated with a flow of probability measures.** Now, let us define the exact degree of relatedness between two individuals $x, y \in [0,1]$ with respect to a flow of measures (5.1) as

$$\gamma_T(x,y) \equiv \sup\left(s \in I : y \in m_x(s,T)\right). \tag{5.6}$$

Lemma 5.2. *$T - \gamma_T$ defines an ultrametric distance on the unit interval.*

We will be interested in cases where the flow $S^{(s,t)}$ of Definition 5.1 is *random*. We will now define the coalescent process on integers that completely characterizes a random genealogical map K_T in this case.

Having defined a distance $T - \gamma_T$ on $[0,1]$, we can define in a very natural way the analogous distance on the integers. To do this, consider a family of i.i.d. random variables, $\{U_i\}_{i \in \mathbb{N}}$, distributed according to the uniform law on $[0,1]$. Given such a family, we set

$$\rho_T(i,j) = \rho_T(U_i, U_j). \tag{5.7}$$

Due to the ultrametric property of the ρ_T and the independence of the U_i, for fixed T, the sets $B_i(s) \equiv \{j : \rho_T(i,j) \le T - s\}$ form an exchangeable random partition of the integers. Moreover, the family of these partitions as a function of $T - s$ is a stochastic process on the space of integer partitions with the property that for any $s > s'$, the partition $B_i(s')$ is a coarsening of the partition $B_i(s)$. Such a process is called a *coalescent process.*

The key observation is the following lemma.

Lemma 5.3. *The genealogical map K_T of a flow $S^{(s,t)}$ is completely determined by its moments; they can be expressed through the probabilities*

$$\mathsf{P}(\rho_T(i,j) \le T - t_{m(i,j)}, \quad \forall i,j \in \{1,\dots,l\}) \tag{5.8}$$

of the corresponding coalescent, where $m(i,j) \in \{1,\dots,p\}$, $0 < t_1 < \dots < t_p \le T$, $l \ge 2$.

To illustrate this lemma, let us note that

$$\mathsf{E}\int m(t) K_T(dm) = \mathsf{E}\int_0^1 m_x(T,t)dx = \mathsf{P}(\rho_T(1,2) \le T - t). \tag{5.9}$$

5.4. **Finite N setting for the CREM..** We will now show that for finite N we can use the general construction from Subsections 5.2, 5.3 to relate the geometric description of the Gibbs measure on Σ_N to the genealogical description of a family of embedded measures on $[0,1]$.

Recall that we have already introduced the image measure $\widetilde{\mu}_{\beta,N}$ (1.8) of the Gibbs measure on the unit interval via the map r_N (1.7). Let $\theta_{\beta,N}$ be the probability distribution function of $\widetilde{\mu}_{\beta,N}$:

$$\theta_{\beta,N}(x) = \widetilde{\mu}_{\beta,N}(\sigma : r_{[N]}(\sigma) \leq x). \tag{5.10}$$

Let us take a parameter $s \in [0,1]$ and consider the map $r_{[sN]} : \Sigma_N \to [0,1]$. Clearly, its image consists of $2^{[sN]}$ points, and for any σ, σ' with $d_N(\sigma,\sigma') > s$ we have $r_{[sN]}(\sigma) = r_{[sN]}(\sigma')$. Now we define a family of compatible distribution functions in the sense of Definition 5.1:

$$S^{(s,t)}_{\beta,N}(a) = \sum_{\sigma} \mu_{\beta,N}(\sigma) \mathbb{1}_{\{\theta(r_{[sN]}(\sigma)) \leq a\}} \tag{5.11}$$

as states Lemma 5.4. To better understand the construction of (5.11), let us take configurations, $\sigma^1, \sigma^2, \ldots, \sigma^{2^{[sN]}}$, differing in the first $[sN]$ coordinates, i.e. with $d_N(\sigma^i,\sigma^j) \geq 1-s$, and arrange them in order such that $0 < r_{[sN]}(\sigma^1) < r_{[sN]}(\sigma^2) < \cdots < r_{[sN]}(\sigma^{2^{[sN]}}) = 1$. Let

$$x_i^s = \mu_{\beta,N}(\sigma' : d_N(\sigma',\sigma^i) < 1-s),\ i = 1,\ldots,2^{[sN]}, \quad x_0 = 0.$$

Define

$$y_i^s \equiv x_0^s + x_1^s + \cdots + x_i^s = \theta(r_{[sN]}(\sigma^i)), \quad i = 0,1,\ldots,2^{[sN]}.$$

Then we may write the representation

$$S^{(s,t)}_{\beta,N}(a) = \sum_{i=0}^{2^{[sN]}} y_i^s \mathbb{1}_{\{a \in [y_i^s, y_{i+1}^s)\}}. \tag{5.12}$$

Lemma 5.4. *The functions, $S^{(s,t)}_{\beta,N}$, defined in (5.11) satisfy the assumptions of Definition 5.1 with $I = [0,1]$.*

It follows from this observation that we are entitled to apply the construction of the previous section to $S^{(s,t)}_{\beta,N}$. Their genealogy is

$$m_x(s,t) = (y_{i-1}^s, y_i^s] \text{ with } |m_x(s,t)| = |x_i^s|, \text{ if } x \in (y_{i-1}^s, y_i^s],\ i = 1,\ldots,2^{[sN]}.$$

We may associate with this genealogy the genealogical map, K_T, and the coalescent process on the integers. The next lemma expresses the geometry of the Gibbs measure of the CREM contained in the empirical distance distribution function $\mathcal{K}_{\beta,N}$, defined in (1.13), in terms of the genealogy induced by the functions defined in (5.11).

Lemma 5.5. *We have*

$$\mathcal{K}_{\beta,N} = K_1^{\beta,N},$$

where the empirical distance distribution function $\mathcal{K}_{\beta,N}$ is defined in (1.13) and $K_1^{\beta,N}$ is the genealogical map defined in (5.5), with $T = 1$, of the flow of probability distribution functions (5.11).

5.5. **Genealogy of a continuous state branching process.** Another example of flows of probability measures satisfying Definition 5.1 arises in the context of continuous state branching process [BLG]. The basic object here is a continuous state branching process $X(t)$ on $\mathbb{R}^+$ characterized by its Laplace exponent $u_t(\lambda)$. The process started in $a \geq 0$ will be denoted by $X(\cdot,a)$. This can be extended to a genuine two parameter process $(X(t,a), t, a \geq 0)$ using the fundamental branching property that states that, if $X'(\cdot,b)$ and $X(\cdot,a)$ are independent copies, then $X(\cdot,a+b)$ has the same law as $X'(\cdot,b) + X(\cdot,a)$. The process $X(t,a)$ is characterized by the property that, for any $a,b \geq 0$, $X(\cdot,a+b)-X(\cdot,a)$ is independent of the processes $X(\cdot,c)$, for all $c \leq a$, and its law is the same as that of $X(\cdot,b)$. The right continuous version of $X(t,\cdot)$ is a subordinator. Bertoin and Le Gall [BLG] prove the following proposition, based on the Markov property of this process.

Proposition 5.1. *On some probability space there exists a process $(\widetilde{S}^{(s,t)}(a), 0 \leq s \leq t, a \geq 0)$, such that*

(i) For any $0 \leq s \leq t$, $\widetilde{S}^{(s,t)}$ is a subordinator with Laplace exponent $u_{t-s}(\lambda)$.

(ii) For any integer $p \geq 3$ and $0 \leq t_1 \leq t_2 \leq \cdots \leq t_p$, the subordinators $\widetilde{S}^{(t_1,t_2)}, \widetilde{S}^{(t_2,t_3)}, \ldots, \widetilde{S}^{(t_{p-1},t_p)}$ are independent, and

$$\widetilde{S}^{(t_1,t_p)}(a) = \widetilde{S}^{(t_{p-1},t_p)} \circ \widetilde{S}^{(t_{p-1},t_p)} \circ \cdots \circ \widetilde{S}^{(t_2,t_3)} \circ \widetilde{S}^{(t_1,t_2)}(a), \quad \forall a \geq 0, \text{ a.s.} \tag{5.13}$$

(iii) The processes $\widetilde{S}^{(0,t)}(a)$ and $X(t,a)$ have the same finite dimensional marginals.

The process $\widetilde{S}^{(s,t)}$ allows to construct a flow of probability distribution functions by setting

$$S^{(s,t)}(x) \equiv \frac{1}{X(t,1)} \widetilde{S}^{(s,t)}(X(s,1)x), \quad 0 \leq s \leq t \leq 1. \tag{5.14}$$

For I taken as any countable subset of $\mathbb{R}^+$, they satisfy the assumptions of Definition 5.1 a.s.

We are interested in a particular case of Neveu's continuous state branching process X_t with

$$E(e^{-\lambda X_t} \mid X_0 = a) = e^{-u_t(\lambda)a}, \quad u_t(\lambda) = \lambda^{e^{-t}}. \tag{5.15}$$

In this case $\widetilde{S}^{(s,t)}$ are *stable subordinators* with index e^{s-t}. Then the normalized stable subordinators $S^{(s,t)}$ of (5.14) is a family of random probability distribution functions satisfying Definition 5.1. Thus, the genealogical construction of Subsections 5.2, 5.3 applies to them.

Finally, note that if we take an increasing function $t(y) \geq 0$ for $y \in [0,1]$, then we may consider the time-changed flow $\bar{S}^{(y,z)} = S^{(t(y),t(z))}$, $0 \leq y \leq z$, satisfying again Definition 5.1 and therefore allowing the genealogical construction of Subsections 4.2, 4.3.

Bertoin and Le Gall [BLG] showed that the coalescent process on the integers induced by $S^{(s,t)}$ of (5.14) associated to Neveu's process (5.15) coincides with the coalescent process constructed by Bolthausen and Sznitman [BS]. They also proved the following remarkable result connecting the collection of subordinators to Ruelle's Generalized Random Energy Model: Take the parameters $0 < x_1 < \cdots < x_p < 1$ and $0 < t_1 < \cdots < t_p$ linked by the identities

$$t_k = \ln x_{k+1} - \ln x_1 \tag{5.16}$$

for $k = 0, \ldots, p-1$, and $t_p = -\ln x_1$. Then the law of the family of jumps of the normalized subordinators $S^{(t_k,t_p)}$, for $k = 0, \ldots, p-1$, is the same as the law of Ruelle's probability cascades $\mathcal{W}^{(p)}$ with parameters x_i, $i = 1, \ldots, p$, see Definition 3.1.

Now consider a GREM with finitely many hierarchies and parameters such that the points $y_0 = 0$ and $0 < y_1 < \ldots < y_p \leq 1$ are the extremal points of the concave hull of A. Recall that $\lim_{N\to\infty} \mathsf{E}\, f_{\beta,N}(y) = \mathsf{E}\, f_\beta(y)$ can be computed by (4.5) for any $y \in [0,1]$. Now set

$$\mathsf{E}\, f_\beta(y_{i-1}) = x_i, \quad i = 1, \ldots, p, \tag{5.17}$$

where all of the $x_i < 1$. In Theorem 3.2 we proved that the point process $\mathcal{W}_{N,\beta}^{(p)}$ in $[0,1]^p$ converge to Ruelle's probability cascades with parameters x_i, $i = 1, \ldots, p$. (The convergence of the marginals of the process $\mathcal{W}_{N,\beta}^{(p)}$ for the GREM under the assumption that for any given hierarchy $i = 1, \ldots, p$ and $N > 0$ the number of configurations $\{\sigma' :$

$d_N(\sigma,\sigma') < 1-y_i\}$ is the same for all $\sigma \in \Sigma_N$, has been also established in Proposition 9.6 of [BS2].) Combining these two results yields

Lemma 5.6. *Let $\mu_{\beta,N}$ be the Gibbs measure associated to a GREM with finitely many hierarchies satisfying (5.17) at the extremal points y_i, $i=1,\dots,p$ of the concave hull of the function A. Then the family of distribution functions $S_{\beta,N}^{(y_k,y_p)}$, $k=1,2,\dots,p$ defined according to (5.11) converges in law, and the limit has the same distribution as the family of normalized stable subordinators (5.14) $S^{(t_k,t_p)}$, $k=0,1,\dots,p-1$ in the sense that the joint distribution of their jumps has the same law, provided t_k are chosen according to (5.16), (5.17).*

5.6. **Main result.** From the preceding proposition we expect that Neveu's process will provide the universal limit for all of our CREMs. The dependence on the particular model (i.e. the function A) and on the temperature must come from a rescaling of time. Set

$$x(y) \equiv \mathsf{E}\, f_\beta(y) = \begin{cases} \dfrac{\sqrt{2\ln 2}}{\beta\sqrt{\widehat{A}'(y)}}, & \text{if } y < y_\beta \\ 1, & \text{if } y \geq y_\beta \end{cases} \tag{5.18}$$

where $y_\beta = \sup(y : \frac{\sqrt{2\ln 2}}{\beta\sqrt{\widehat{A}'(y)}} < 1)$ (here $\mathsf{E}\, f_\beta(y)$ is defined by the function A through (4.5)). Set also

$$T = -\ln x(0), \quad t(y) = T + \ln x(y). \tag{5.19}$$

Define the flow of probability distribution functions

$$\bar{S}^{(y,z)}(x) \equiv S^{(t(y),t(z))}(x) \tag{5.20}$$

where $S^{(s,t)}$ is the flow of functions (5.14) associated to Neveu's process (5.15). Let $\bar{K}_T^{t(y)}$ be the genealogical map (5.5) associated to this flow.

Theorem 5.1. *Consider Continuous Random Energy Model with general function A such that A does not touch its concave hull $\widehat{A}$ in the interior of any interval where $\widehat{A}$ is linear. Then*

$$\mathcal{K}_{\beta,N} = K_1^{\beta,N} \stackrel{\mathcal{D}}{\to} \bar{K}_1^{t(y)}. \tag{5.21}$$

Here $\mathcal{K}_{\beta,N}$ is the empirical distance distribution function (1.13), $K_1^{\beta,N}$ is the genealogical map (5.5) of the flow of probability distribution functions (5.11) and the equality $\mathcal{K}_{\beta,N} = K_1^{\beta,N}$ holds by Lemma 5.5. Theorem 5.1 is the main result of this paper. It expresses the geometry of the limiting Gibbs measure contained in $\mathcal{K}_{\beta,N}$ in terms of the genealogy of Neveu's branching process via the deterministic time change (5.19). We prove this theorem in the next subsection.

5.7. **Coalescence and Ghirlanda-Guerra identities.** As it was remarked in Subsection 5.3, K_T associated with a flow of measures is completely determined by its moments, and these can be expressed via genealogical distance distributions of the corresponding coalescent (5.8). So, we will prove that the moments of $\mathcal{K}_{\beta,N}$, which are the n-replica distance distributions in our spin glass model, converge to the genealogical distance distributions on the integers (5.8) constructed from the flow of compatible measures with distribution functions $\bar{S}^{(y,z)}$ (5.20). But the flow $\bar{S}^{(y,z)}$ is the time changed flow (5.14) of Neveu's branching process (5.15) that by [BLG] corresponds to the coalescent of Bolthausen-Sznitman. Therefore, its genealogical distance distributions on the integers are those of Bolthausen-Sznitman coalescent under this time change (5.19). Then the proof of Theorem 5.1 is reduced to the following Theorem 5.2 that gives in addition the connection between the n-replica distance distribution function of the CREM with the genealogical distance distribution function of the Bolthausen-Sznitman coalescent.

Theorem 5.2. *Under the same assumptions as in Theorem 5.1, for any $n \in \mathbb{N}$,*

$$\lim_{N\uparrow\infty} \mathsf{E}\,\mu_{\beta,N}^{\otimes n}(d_N(\sigma^1,\sigma^2) \geq 1-y_1,,\ldots,d_N(\sigma^{n-1},\sigma^n) \geq 1-y_{n(n-1)/2}) \qquad (5.22)$$
$$= \mathsf{P}\left(\rho_T(1,2) \geq T - t(y_1),\ldots,\rho_T(n-1,n) \geq T - t(y_{n(n-1)/2})\right)$$

where $t(y)$ is defined in (5.19) via (5.18).

The distance ρ_T is the distance on integers for the Bolthausen-Sznitman coalescent, induced through (5.7) by the genealogical distance γ_T of the flow of measures $S^{(s,t)}$ (5.14) of Neveu's branching process (5.15). The fact that in Bolthausen-Sznitman coalescent $\mathsf{P}(\rho_T(1,2) \geq T-t) = e^{t-T}$ and the convergence (4.5) imply the statement of the theorem for $n=2$:

$$\mathsf{E}\,\mu_{\beta,N}^{\otimes 2}(d_N(\sigma,\sigma') \geq 1-y) \to x(y) = e^{t(y)-T} = \mathsf{P}(\rho_T(1,2) \geq T - t(y)).$$

The proof of the theorem for $n > 2$, and in fact the entire identification of the limiting processes with objects constructed from Neveu's branching process, relies on the Ghirlanda-Guerra identities [GG] that were derived in Theorem 4.4 for the left-hand side of (5.22). Thus we must show that the right-hand side of (5.22) satisfies the same identities, that is for $t < T$:

$$\mathsf{P}\left(\rho_T(1,n+1) \geq T-t \mid \mathcal{B}_n\right) = \frac{1}{n}e^{t-T} + \frac{1}{n}\sum_{k=2}^{n} \mathsf{P}\left(\rho_T(1,k) \geq T-t \mid \mathcal{B}_n\right) \qquad (5.23)$$

that can be equivalently written as

$$\mathsf{P}\left(\rho_T(k,n+1) < T-t \mid \mathcal{B}_n\right) = \frac{|l \in \{1,\ldots,n\} : \rho_T(k,l) < T-t| - e^{t-T}}{n} \qquad (5.24)$$

There are *two* ways to verify that (5.23) holds for the Bolthausen-Sznitman coalescent.

The first one is to observe that relation (5.23) involves only the marginals of the coalescent at a finite set of times. By Theorem 5 of Bertoin-Le Gall [BLG], these can be expressed in terms of Ruelle's probability cascades modulo the appropriate time change. Thus, by Theorem 3.2 these probabilities can be expressed as limits of a suitably constructed GREM (with finitely many hierarchies) for which the Ghirlanda-Guerra relations do hold by Lemma 3.1. Thus (5.23) is satisfied.

The second way is to verify directly that Ghirlanda-Guerra relations (5.24) hold for the Bolthausen-Sznitman coalescent.

This can be done by identifying its partitions with exchangeable random partitions called "Chinese restaurant process".

For that purpose, let us first give the following definition. Given the sequence of normalized jumps of the stable subordinator (Δ_i/T) with index x and given $U_1, U_2, \ldots$ independent uniform random variables on [0, 1], the partition of positive integers Π distributed as a partition of blocks of indices of U_i belonging to the same intervals $\Delta_i/T \in [0,1]$ is called $(x,0)$*-partition*, see [Pi].

Let us introduce an operation of coagulation on partitions, see [Pi1]: for a partition $\pi = (A_1, A_2, \ldots,)$ and $\Pi = (B_1, B_2, \ldots)$, the Π-coagulation of π consists of blocks of the form $\bigcup_{j \in B_i} A_j$.

By [BS] *the Markov kernels* $(e^{-t},0)$*-coagulation,* $t \geq 0$*, on partitions of* $\mathcal{N}$ *form a semi-group. The Markov process*

$$\mathsf{P}^{\pi}(\Pi(t+) \in \cdot) = (e^{t-T},0) - \textit{coagulation of } \pi \qquad (5.25)$$

is distributed as the Bolthausen-Sznitman coalescent. It starts from a partition of singletons at time T and finishes by a partition of one block $\mathcal{N}$ at time $-\infty$. (The semi-group property can be also seen from the fact that the limiting frequencies of $(e^{-t}, 0)$-partitions are distributed as normalized jumps of stable subordinators and from their matching condition (5.1).)

Next, consider exchangeable random partitions Π on $\mathbb{N}$, introduced by J. Pitman under the name of Chinese restaurant processes. For each parameter $0 < x < 1$ the partition called "Chinese restaurant process" can be constructed as follows. Let Π_n denote the restriction of Π to the first n positive integers. Then, conditionally given $\Pi_n = \{A_1, \dots, A_k\}$ for any particular partition of $\{1, 2, \dots, n\}$ into k subsets (tables) A_i of sizes n_i, $i = 1, \dots, k$, the partition Π_{n+1} is an extension of Π_n such that the number $n+1$ (new customer) is attached to the class (table) A_i with probability $(n_i - x)/n$, and forms a new class (sits at a new table) with probability kx/n. Let us denote by $p(n_1, \dots, n_k)$ the probability of partitions Π with Π_n a particular partition of k classes of sizes $n_1, \dots, n_k$ respectively. Then

$$p(n_1 + 1, n_2, \dots, n_k) = \frac{n_1 - x}{n} p(n_1, \dots, n_k) \tag{5.26}$$

The crucial fact is that *the partition* Π *of the Chinese restaurant process with parameter* x *is a* $(x, 0)-$ *partition.* This fact, noticed in [Pi], follows from the combination of the results of [Pi1] and [PPY] : On the one hand, in [Pi1] it is proven that the limiting relative frequencies, in order of appearance, P_i, in the Chinese restaurant process have the same distribution as the product $(1 - W_1)(1 - W_2) \cdots (1 - W_{i-1}) W_i$, with W_i independent beta random variables with parameters $(1 - x, ix)$. On the other hand, in [PPY] the following was proven: let $\Delta_{(i)}/T$ denote the reordering to the intervals Δ_i/T in order of appearance of the U_i, i.e. define $\Delta_{(i)}$ such that $U_1 \in \Delta_{(1)}/T$, $U_{\min\{j : U_j \notin \Delta_{(1)}/T\}} \in \Delta_{(2)}/T$ etc.. Then $|\Delta_{(i)}/T|$ has the same distribution as products, $(1 - W_1) \dots (1 - W_{i-1}) W_i$, where W_i are the independent beta random variables appearing above. Thus, the sequences $|\Delta_{(i)}/T|$ and P_i have the same distribution.

Therefore, by (5.25), the marginals of Bolthausen-Sznitman coalescent $\Pi(t)$ at times $0 = t_0 < t_1 < \cdots < t_{p-1} < t_p = T$ can be constructed as the following sequence of Chinese restaurant processes: let $x_i = e^{t_{i-1} - t_p}$, $0 < x_1 < x_2 < \cdots < x_p < 1$. Then $\Pi(t_{p-1}+)$ is distributed as a $(x_p, 0)-$ partition, i.e. as the Chinese restaurant process with parameter x_p. Next, we define the partition $\Pi(t_{p-2}+)$ as the Chinese restaurant process on the classes of partition $\Pi(t_{p-1}+)$ with parameter $x_{p-1}/x_p = e^{t_{p-2} - t_{p-1}}$; this means that, given the classes $A_1^{p-1}, \dots, A_k^{p-1}$ obtained from $A_1^p, \dots, A_l^p$, where A_i^{p-1} consists of l_i blocks of Π^p, $i = 1, \dots, k$, $l_1 + \cdots + l_k = l$, the block A_{l+1}^p joins A_i^{p-1} with probability $(l_i^{p-1} - x_{p-1}/x_p)/l$ and forms a new class with probability $kx_{p-1}/(x_p l)$. One iterates this procedure with parameters $x_{p-2}/x_{p-1}, \dots, x_1/x_2$ to construct the partitions $\Pi(t_{p-3}+), \dots, \Pi(t_0+)$. By the semi-group property of $(e^{-t}, 0)-$ coagulations, $\Pi(t_i+)$ is distributed as a Chinese restaurant process with parameter $x_{i+1} = e^{t_i - t_p}$ for all $i = 0, 1, \dots, p-1$, satisfying (5.26). Now (5.24) is immediate from the Chinese restaurant property (5.26).

References

[ASS] M. Aizenman, R. Sims, S.L. Starr. An extended variational principle for the SK spin-glass model. *Phys. Rev. B*, 6821: 4403 (2003).

[BLG] J. Bertoin, J-F. Le Gall. The Bolthausen-Sznitman coalescent and the genealogy of continuous state branching processes. *Probab. Theor. and Relat. Fields* **117**, 249–266 (2000).

[BS] E. Bolthausen, A-S. Sznitman. On Ruelle's probability cascades and an abstract cavity method. *Commun. Math. Physics.* **107**, 247–276 (1998).

[BS2] E. Bolthausen and A.-S. Sznitman, *Ten Lectures on Random Media* Birkhäuser Verlag, Basel-Boston-Berlin (2002).

[Bo] A. Bovier. Statistical mechanics of disordered systems, MaPhySto Lecture Notes 10, Aarhus, (2001).

[BKL] A. Bovier, I.A.Kurkova and M. Löwe Fluctuations of the free energy in the REM and p-spin SK models. *Annals Probab.* **30** (2), 605–651 (2002).

[BK1] A. Bovier, I. Kurkova. Derrida's Generalised Random Energy Models 1: Poisson cascades and Gibbs measures. *Annals de l'I.H.P.* **40** , 439–480, (2004).

[BK2] A. Bovier, I. Kurkova. Derrida's Generalised Random Energy Models 2: Models with continuous hierarchies. *Annals de l'I.H.P.* **40** , 481–495, (2004).

[BK3] A. Bovier, I. Kurkova. Gibbs measures of Derrida's Generalised Random Energy Models and the genealogy of Neveu's continuous state branching process. Preprint, University Paris 6 (2003).

[BK4] A. Bovier, I. Kurkova. Rigorous results on some simple spin glass models. *Markov Processes and Related Fields* **9** (2), 209–242 (2003).

[Br] M.D. Bramson. Maximal displacement of branching Brownian motion, *Commun. Pure and Appl. Math.* **31**, 531–581 (1978).

[CCP] D. Capocaccia, M. Cassandro and P. Picco. On the existence of thermodynamics for the Generalised Random Energy Models. *J. Stat. Phys.* **46**, 493–505 (1987).

[D1] B. Derrida. Random energy model: limit of a family of disordered models, *Phys. Rev. Letts.* **45**, 79-82 (1980).

[D2] B. Derrida. Random energy model: An exactly solvable model of disordered systems, *Phys. Rev. B* **24**, 2613-2626 (1981).

[D3] B. Derrida. A generalisation of the random energy model that includes correlations between the energies. *J. Physique. Lett.* **46**, 401–407 (1985).

[DG1] B. Derrida and E. Gardner. Solution of the Generalised Random Energy model. *J. Phys. C* **19**, 2253–2274, (1986).

[DG2] B. Derrida and E. Gardner. Magnetic properties and function $q(x)$ of the Generalised Random Energy Model. *J. Phys. C* **19**, 5783-5798 (1986).

[DG3] E. Gardner and B. Derrida. The probability distribution of the partition function of the random energy model. *J. Phys. A* **22**, 1975-1981 (1989).

[Ei] TH. Eisele. On a third order phase transition, *Commun. Math. Phys.* **90**, 125-159 (1983).

[GMP] A. Galvez, S. Martinez, and P. Picco. Fluctuations in Derrida's random energy and generalised random energy models. *J. Stat. Phys.* **54**, 515-529 (1989).

[GG] S. Ghirlanda, F. Guerra. General properties of the overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity, J. Phys. A **31**, 9144-9155 (1998).

[G] F. Guerra, Broken replica symmetry bounds in the mean field spin glass model, Comm. Math. Phys. **233**, 1-12 (2003).

[GT] F. Guerra and F. Toninelli, The thermodynamics limit in mean field spin glass models, Commun. Math. Phys. **230**, 71-79 (2002).

[LLR] M.R. Leadbetter, G. Lindgren, H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin-Heidelberg-New York (1983).

[Le] M. Ledoux, On the distribution of overlaps in the Sherrington-Kirkpatrick spin glass model, J. Statist. Phys. **100** (2000), 871-892.

[LT] M. Ledoux and M. Talagrand, Probability on Banach spaces, Ergebnisse der Mathematik und ihrer Grenzgebiete Vol. 3, Springer, Berlin, 1991.

[Ne] J. Neveu. A continuous state branching process in relation with the GREM model of spin-glass theory. Rapport interne No 267, Ecole Polytechnique.

[Pi] J. Pitman. Combinatorial stochastic processes. Lecture Notes of the 2002 Ecole de Probabilités de St. Flour, Springer Lecture Notes in Mathematics.

[Pi1] J. Pitman, Exchangeable and partially exchangeable random partitions, *Probab. Theor. Relat. Fields* **102**, 145–158 (1995).

[PPY] M. Perman, J. Pitman, M. Yor, Size-biased sampling of Poisson point processes and excursions, *Probab. Theor. Relat. Fields* **92**, 21–39 (1992).

[Ru] D. Ruelle. A mathematical reformulation of Derrida's REM and GREM. *Commun. Math. Phys.* **108**, 225-239 (1987).

[SK] D. Sherrington and S. Kirkpatrick. Solvable model of a spin glass. *Phys. Rev. Lett.* **35**, 1792–1796 (1972).

[T1] M. Talagrand, Rigorous low temperature results for mean field p-spin interaction models, Probab. Theor. Rel. Fields. **117**, 303-360 (2000).

[T2] M. Talagrand, Self organization in the low-temperature region of a spin glass model, Rev. Math. Phys. **15**, 1-78 (2003).

[T3] M. Talagrand, Mean field models for spin glasses: a first course, Lecture Notes of the 2000 École des Probabilités de St. Flour; to appear as Springer Lecture Notes in Mathematics.

[T4] M. Talagrand, Spin Glasses: a Challenge to Mathematicians, Ergebnisse der Mathematik und ihrer Grenzgebiete Vol. 46, Springer, Berlin-Heidelberg-New York, 2003.

[T5] M. Talagrand, The generalized Parisi formula. C. R. Math. Acad. Sci. Paris **337**, 111-114 (2003).

[T6] M. Talagrand, The Parisi formula. to appear in *Ann. Math.* (2005).