Derrida's Generalized Random Energy models 3:
Models with continuous hierarchies.

Anton Bovier[1][2]
Weierstraß–Institut
für Angewandte Analysis und Stochastik
Mohrenstrasse 39, D-10117 Berlin, Germany

Irina Kurkova[3]
Laboratoire de Probabilités et Modèles Aléatoires
Université Paris 6
4, place Jussieu, B.C. 188
75252 Paris, Cedex 5, France

Abstract: This is the third of a series of three papers in which we present a rigorous analysis of Derrida's Generalized Random Energy Models (GREM). Here we study the general case of models with a "continuum of hierarchies". We prove the convergence of the free energy and give explicite formulas for the free energy and the two-replica distribution function. Then we introduce the empirical distance distribution to describe effectively the Gibbs measures. We shhow that its limit is uniquely determined via the Ghirlanda-Guerra identities up to the mean of the replica distribution function. Finally, we show that suitable discretizations of the limiting random measure can be described by the same objects in suitably constructed GREM's.

Keywords: Gaussian processes,generalized random energy model, continuous hierarchies, spin glasses, Poisson cascades, probability cascades, Ghirlanda-Guerra identities.

AMS Subject Classification: 82B44, 60G70,60K35

---

[1] e-mail: bovier@wias-berlin.de

[2] Research supported in part by the DFG in the Concentration program "Interacting stochastic systems of high complexity".

[3] e-mail: kourkova@ccr.jussieu.fr

1. Introduction.

In two companion papers [BK1,BK2] we have given a rather complete analysis of thermodynamic limit of Derrida's Generalized Random Energy models [D,DG1,DG2,DG3]. We have seen that these are naturally considered as a special class of Gaussian models defined in terms of Gaussian processes X_σ on the hypercube $\mathcal{S}_N \equiv \{-1,1\}^N$ whose covariance depends only on the ultra-metric valuation $d_N(\sigma,\sigma')$ defined by

$$d_N(\sigma,\sigma') \equiv \frac{1}{N}\left(\min(i : \sigma_i \neq \sigma_i') - 1\right) \tag{1.1}$$

i.e.

$$\mathbb{E}X_\sigma X_{\sigma'} = A(d_N(\sigma,\sigma')) \tag{1.2}$$

where A is a probability distribution function on the interval $[0,1]$. In [BK1,BK2] this function was assumed to be a step function with finitely many steps. We refer to this situation as the standard GREM. In this paper we return to the general class of models where the covariance function $A(x)$ is no longer considered to be discrete. To avoid complications that are not of interest here, we will assume that A is a piecewise smooth function throughout this paper. For some of our results we will moreover have to assume that A is "non-critical", in the sense that it is equal to its convex hull *only* on the set of extremal points of the convex hull. We will of course draw advantage from the fact that any such function can be approximated arbitrarily well by discrete distributions, for which we have computed everything quite explicitly. The task is then to show that these approximants can be used to capture precisely the asymptotic properties of the processes of interest. Such an approach is of course classical in extreme value theory where it is based on comparison lemmata of Slepian.

We will see that this program can be carried out to a surprising extent and that it allows to compute essentially all the physical objects that are commonly studied in the theory of mean field spin glasses. Let us note that what we are doing here goes considerably beyond the earlier investigations by Ruelle [Ru] and Derrida and Gardner [D,DG1,DG2], who considered limits of GREMs when the number of hierarchies tends to infinity. Here, in contrast we start in finite volume with models with arbitrary covariances and *prove* that the thermodynamic limit can be described using certain sequences of GREMs. Moreover, the construction of Ruelle using inductive limits does not capture, in our view, the physics of these models in an appropriate way.

The class of models we introduce here is in principle as rich as the more commonly considered models where the covariance is a function of the overlap $R_N(\sigma,\sigma') \equiv N^{-1}\sum_{i=1}^N \sigma_i\sigma_j$.

We call this latter class the SK-models, since it comprises in particular the standard SK-model and the p-spin SK models (for recent progress on these models, see [T1,T2,GT,Gu]. Interestingly, we sill see that in our class of models, upon suitable choices of the covariance function A *any* structure that emerges as the solution of a mean field spin glass in Parisi's theory can be obtained. In particular, for some choice of A the ensuing limiting object will have the same properties as the standard SK model. However, in contrast to the standard class of spin glasses defined via the Hamming distance, where we have no clue on how Parisi's solution arises even on a very qualitative level (with the exception of a rare glimpse at the emergence of non-trivial structure in the p-spin models [T1,T2] or the Hopfield models (see e.g. [Bo])), in our models all this can be perfectly understood both on a mathematical and intuitive level. From that perspective, we feel that our results constitute a major contribution towards a partial de-mystification of the Parisi solution. Moreover, as had been pointed out already by Derrida and Gardner [DG1], one may argue that the particular choice of the distance concept used may not have dramatic consequences on the properties of the system, and that the real key parameter of the model might just be the entropy as a function of the covariance. With such a universality hypothesis, our results suggest that indeed the Parisi solution is very natural and that indeed the structures we exhibit here ought to be canonical for any (Gaussian) mean field spin glass. The natural relation between models of the SK class and the Derrida class would be that if a model in the SK class has a covariance function B, then the analogue model in the Derrida class must have covariance function $A(x) = B(I^{-1}(\ln 2x))$, where I is Cramér's entropy function. In particular, the classical p-spin SK models correspond to Derrida models with $A(x) = (I^{-1}(\ln 2x))^p$. Note for instance that this function is convex if $p = 2$, while for $p \geq 3$, it is concave in a neighborhood of the origin. At least on a qualitative level, the properties of these models are indeed those that are expected in the corresponding SK-models. However, as was noted by Derrida and Gardner [DG1,DG2], on closer inspection one finds numerical agreement with the replica solutions of the true SK models unsatisfactory, and in particular the temperature dependence of the free energy does not have the predicted qualitative behaviour. In the SK model, in particular, the replica theory appears to predict "massive" pure states at the edge of the overlap distribution, a fact that does not hold in the corresponding approximation. Thus, a fundamental difference between the two model classes remains and a profound understanding of the SK models is still amiss.

Acknowledgements: We are indebted to Erwin Bolthausen for helpful discussions. A.B. thanks Francis Comets and the Université Paris 7 for kind hospitality during the period when

this work was completed.

2. Limiting formula and heuristic considerations.

As had already been noted by Derrida and Gardner [DG1,DG2], thermodynamic quantities in the GREMs possess limits as the number of levels tends to infinity. In fact, the formulae for the free energy and the overlap distribution can be written in a closed form in terms of the function A that makes this fact manifest.

Let us denote the convex hull of the function $A(x)$ by $\bar{A}(x)$. We will also need the left-derivative of this function, $\bar{a}(x) \equiv \lim_{\epsilon\downarrow 0} \epsilon^{-1}(\bar{A}(x) - \bar{A}(x-\epsilon))$ which exists for all values of $x \in (0,1]$.

It is a very simple matter to realize that the formula of Capocaccia et al. [CaCaPi] (see Theorem 1.2 in [BK2]) can be written as

$$F_\beta = \sqrt{2\ln 2}\beta \int_0^{x(\beta)} \sqrt{\bar{a}(x)}dx + \frac{\beta^2}{2}(1 - \bar{A}(x(\beta))) \tag{2.1}$$

where

$$x(\beta) \equiv \sup\left(x|\bar{a}(x) > \frac{2\ln 2}{\beta^2}\right) \tag{2.2}$$

The mean distance distribution function

$$f_\beta(x) \equiv \lim_{N\uparrow\infty} \mu_{\beta,N}^{\otimes 2}\left(d_N(\sigma,\sigma') < x\right) \tag{2.3}$$

as given in Derrida and Gardner [DG2] or Proposition 1.7 of [BK2] can be re-written as

$$\mathbb{E}f_\beta(x) = \begin{cases} \beta^{-1}\sqrt{2\ln 2}/\sqrt{\bar{a}(x)}, & \text{if}\, x \le x_\beta \\ 1, & \text{if}\, x > x_\beta \end{cases} \tag{2.4}$$

These appealing formulas of course suggest that they might be true also when A is not a step function with finitely many steps. We will show that this is indeed the case.

Note that one cannot write such closed expressions for the sub-leading *corrections* to the free energy. In fact, it is to be expected that these will in general not only depend on $\bar{A}$ (see e.g. the analysis of the maximum of branching Brownian motion by Bramson [BR,BDG]). In particular, we cannot construct immediately sensible limits of the Poisson cascades $\mathcal{P}^{(k)}$ (see [BK1]). This might discourage us from inquiring into the finer properties of the Gibbs measures that were analysed in [BK2]. However, the situation is actually better than might be feared.

Let us look at the interpretation of the processes $\mathcal{W}_\beta^{(m)}$ that we considered in [BK2]. We have thought of them as point processes on $(0,1]^m$. However, the vectors $(\mu_{\beta,N}(B_1(\sigma)),\dots,\mu_{\beta,N}(B_m(\sigma)))$ are in fact more reasonably thought of as the values of the mass distribution

$$m_\sigma(x) \equiv \mu_{\beta,N}\left(d_N(\sigma,\sigma') < x\right) \tag{2.5}$$

at the values q_i for which the limiting distribution will have jumps. Seen in this light, it would be most appropriate to introduce the following objects, that we may call *empirical distance distribution function*

$$\mathcal{K}_{\beta,N} \equiv \sum_{\sigma\in\mathcal{S}_N} \mu_{\beta,N}(\sigma)\delta_{m_\sigma(\cdot)} \tag{2.6}$$

as a random measure on the set of probability measures on $[0,1]$. This object is an element of the *compact* space $\mathcal{M}_1(\mathcal{M}_1(\mathcal{M}_1([0,1])))$ where all sets of measures are endowed with the topology of weak convergence. Note that the first moment of $\mathcal{K}_{\beta,N}$ is the distance distribution function $f_{\beta,N}$.

In the case of the GREM with finitely many levels, the results in [BK2] imply readily the convergence of $\mathcal{K}_{\beta,N}$ to a measure that is concentrated on atomic distributions with jumps at the values q_i those values are controlled by the point process $\mathcal{W}_\beta^{(m)}$, and whence given finally in terms of the processes $\mathcal{P}^{(m)}$. This approach clearly fails in the continuous case. However, we have already seen in the standard GREM, that the *Ghirlanda-Guerra* identities [GG,AC] provide an alternative approach to the construction of the infinite volume limit. Here we will exploit this fully. Let us briefly outline the strategy that we will follow in the remainder of the paper.

(i) First we will prove that the free energy and the mean distance distribution function will converge to the limits given by (2.1) and (2.2). This will rely on the computations in the model with finitely many hierarchies, and standard comparison theorems that allow to control convergence of the free energy of a general model through those of standard GREM's with finitely many hierarchies.

(ii) Next we observe that the random measures $\mathcal{K}_{\beta,N}$ are completely determined by the so called 'multi-overlap distribution functions'. We will then show that the latter satisfy the Ghirlanda-Guerra identities, and that these allow to characterize the possible limit points up to function $\mathbb{E}f_\beta$. By compactness, this implies convergence of $\mathcal{K}_{\beta,,N}$ to a $\mathcal{K}_\beta$ that is completely determined by $\mathbb{E}f_\beta$.

(iii) Finally, we will introduce certain marginals of the measures $\mathcal{K}_\beta$ and show that, again due to the Ghirlanda-Guerra identities, these are identical to the corresponding objects in certain GREM's with finitely many hierarchies, and are thus explicitly constructable in terms of Poisson cascades.

As a result, we obtain a complete and quite explicit description of the asymptotics of the Gibbs measures in this class of models.

3. Ground state energy and free energy.

The basis of all our results is control of the convergence of the free energy. As a warm-up let us first consider the ground state energy density. Recall that the leading term in the function $U_{J,N}(x)$ takes the form

$$\sqrt{2N\ln 2}\int_0^1 \sqrt{\bar a(x)}dx \tag{3.1}$$

which provides a closed form for the ground state energy for our models.

In fact, this formula holds in the general case.

Theorem 3.1: *Let X_σ be a centered Gaussian process on $\mathcal{S}_N$ with covariance given by (1.2). Then*

$$\lim_{N\uparrow\infty} N^{-1/2}\mathbb{E}\max_{\sigma\in\mathcal{S}_N} X_\sigma = \sqrt{2\ln 2}\int_0^1 \sqrt{\bar a(x)}dx \tag{3.2}$$

Proof: Let A_n denote a sequence of piecewise constant distributions functions such that $A_n(x)\leq A(x)$ for all $x\in[0,1]$ that converge to A in the sup-norm. Denote the corresponding processes by $X_\sigma^{(n)}$. Then, by Slepian's lemma [LLR],

$$\mathbb{E}\max_{\sigma\in\mathcal{S}_N} X_\sigma^{(n)} \geq \mathbb{E}\max_{\sigma\in\mathcal{S}_N} X_\sigma \tag{3.3}$$

On the other hand, for all n,

$$\lim_{N\uparrow\infty} N^{-1/2}\mathbb{E}\max_{\sigma\in\mathcal{S}_N} X_\sigma = \sqrt{2\ln 2}\int_0^1 \sqrt{\bar a(x)}dx \tag{3.4}$$

Moreover, if A_n converges to A, then $\bar a_x(x)$ converges to $\bar a(x)$. Thus

$$\limsup_{N\uparrow\infty} N^{-1/2}\mathbb{E}\max_{\sigma\in\mathcal{S}_N} X_\sigma \leq \sqrt{2\ln 2}\int_0^1 \sqrt{\bar a(x)}dx \tag{3.5}$$

It remains to prove a corresponding lower bound. To do so, we need to construct a sequence of piecewise constant distribution functions B_n such that $B_n(x) \geq A(x)$ for all $x \in [0,1]$. This introduces only one small problem, namely that in general such an upper approximation will have to have a positive atom at zero, and the last "atom", b_n, may have to be zero. The latter fact has no consequence, while the former requires us to slightly modify the representation formula Eq. (1.4) in [BK1] into

$$\widetilde{X}_\sigma = \sqrt{b_0}X_0 + \sqrt{b_1}X_{\sigma_1} + \cdots + \sqrt{b_{n-1}}X_{\sigma_1\dots\sigma_n} \tag{3.6}$$

where now $\sum_{k=1}^{n-1} b_k = 1 - b_0$, and X_0 is a new standard Gaussian, independent of all others. Thus the mean of the maximum of $\widetilde{X}_\sigma$ is equal to the mean of the maximum of $\widetilde{X}_\sigma - \sqrt{b_0}X_0$. This variable has the same form as before, except that the total variance is diminished by b_0. Then the formula for its mean takes the same form as before. Thus

$$\lim_{N\uparrow\infty} N^{-1/2}\mathbb{E}\max_{\sigma\in\mathcal{S}_N} \widetilde{X}_\sigma = \sqrt{2\ln 2}\int_0^1 \sqrt{\bar{b}_n(x)}dx \tag{3.7}$$

and Slepian's lemma now shows that

$$\liminf_{N\uparrow\infty} N^{-1/2}\mathbb{E}\max_{\sigma\in\mathcal{S}_N} X_\sigma \geq \sqrt{2\ln 2}\int_0^1 \sqrt{\bar{a}(x)}dx \tag{3.8}$$

This proves the theorem. $\diamondsuit$

In the same way, we get a corresponding result for the free energy in the general case. Essentially all we need is to replace Slepian's lemma by the more general result given as Theorem 3.11 in [LT], Chapter 3, page 74

Lemma 3.2: *Let X and Y be n-dimensional Gaussian vectors. Let D_1 and D_2 be subsets of $\{1,\dots,n\}\times\{1,\dots,n\}$. Assume that*

$$\begin{aligned} \mathbb{E}X_iX_j &\leq \mathbb{E}Y_iY_j, \quad \text{if} \quad (i,j)\in D_1 \\ \mathbb{E}X_iX_j &\geq \mathbb{E}Y_iY_j, \quad \text{if} \quad (i,j)\in D_2 \\ \mathbb{E}X_iX_j &= \mathbb{E}Y_iY_j, \quad \text{if} \quad (i,j)\notin D_1\cup D_2 \end{aligned} \tag{3.9}$$

Let f be a function on $\mathbb{R}^n$ such that its second derivatives satisfy

$$\begin{aligned} \frac{\partial^2}{\partial x_i\partial x_j}f(x) &\geq 0 \quad \text{if} \quad (i,j)\in D_1 \\ \frac{\partial^2}{\partial x_i\partial x_j}f(x) &\leq 0 \quad \text{if} \quad (i,j)\in D_2 \end{aligned} \tag{3.10}$$

Then

$$\mathbb{E}F(X) \leq \mathbb{E}F(Y) \tag{3.11}$$

Theorem 3.3: *Under the assumptions of Theorem 3.1, let*

$$x(\beta) \equiv \sup \left(x | \bar{a}(x) > \frac{2\ln 2}{\beta^2} \right)$$

Then

$$\lim_{N\uparrow\infty} N^{-1}\mathbb{E}\ln Z_{\beta,N} = \sqrt{2\ln 2}\beta \int_0^{x(\beta)} \sqrt{\bar{a}(x)}dx + \frac{\beta^2}{2}(1 - \bar{A}(x(\beta))) \tag{3.12}$$

Note that this formula implies that the critical temperature β_c *is given by*

$$\beta_c = \sqrt{\frac{2\ln 2}{\lim_{x\downarrow 0} \bar{a}(x)}}$$

Proof: We may easily check that the function $\ln \sum_\sigma e^{\beta\sqrt{N}X_\sigma}$ satisfies the hypothesis of Lemma 3.2 with $D_1 = \{\sigma, \sigma' \in \mathcal{S}_N^2 : \sigma \neq \sigma'\}$ and $D_2 = \emptyset$. Thus taking $A_n^{\pm}(1) = A(1)$ and $A_n^{\pm}(x)$ smaller respectively larger than A elsewhere, we can indeed construct upper and lower bounds for F_β that converge to the same limit when $n \uparrow \infty$. $\diamondsuit$

As we have seen in the discrete case, a crucial observation is that the distance distribution function can be expressed as a derivative of the free energy as a function of the covariance. In the continuous case, this will be a little more subtle and will require to introduce some more structure.

In the discrete case we needed to be able to differentiate with respect to the value of one atom, and then to use an integration by parts formula. Both is not immediately obvious in the general case. In fact, to do so it will be necessary to add some temporal structure to our Gaussian process X_σ and to define the Gaussian process $X_\sigma(t)$, $\sigma \in \mathcal{S}_N$, $t \in [0,1]$ with covariance

$$\operatorname{cov}\left(X_\sigma(t), X_{\sigma'}(s)\right) = A(t \wedge s \wedge [d_N(\sigma, \sigma')]) \tag{3.13}$$

Note that $X_\sigma(t)$ is a martingale in the variable t and that its increments with respect to t are independent. It is also useful to realize that we may represent $X_\sigma(t)$ as

$$X_\sigma(t) = Y_\sigma(A(t)) \tag{3.14}$$

where $Y_\sigma(t)$ is a continuous Gaussian martingale with covariance

$$\operatorname{cov}\left(Y_\sigma(t), Y_{\sigma'}(s)\right) = t \wedge s \wedge A([d_N(\sigma,\sigma')]) \tag{3.15}$$

Observe that there is the following integration by parts formula:

Lemma 3.4: *For any $t \in (0,1]$ and $\epsilon > 0$,*

$$\mathbb{E}\mathbb{E}_\sigma \frac{(X_\sigma(t) - X_\sigma(t-\epsilon))e^{\beta\sqrt{N}X_\sigma}}{\mathbb{E}_{\sigma'} e^{\beta\sqrt{N}X_{\sigma'}}} = \beta\sqrt{N}\int_{t-\epsilon}^{t} dA(s)\mathbb{E}\mu_{\beta,N}^{\otimes 2}[d_N(\sigma,\sigma') < s] \tag{3.16}$$

Proof: Let us introduce the infinitesimal increments of the process $Y_\sigma(t)$. Clearly

$$\mathbb{E}dY_\sigma(t)dY_{\sigma'}(s) = dtds\delta(s-t)\mathbb{1}_{A(d_N(\sigma,\sigma'))\geq t} \tag{3.17}$$

The proof makes use of the Gaussian integration by parts formula

$$\begin{aligned}\mathbb{E}dY_\sigma(t) f\left(\int dY_{\sigma'}(s)\right) &= \mathbb{E}f'\left(\int dY_{\sigma'}(s)\right)\int \mathbb{E}dY_\sigma(t)dY_{\sigma'}(s)\\ &= \mathbb{E}f'\left(Y_{\sigma'}\right)\mathbb{1}_{A(d_N(\sigma,\sigma'))\geq t}dt\end{aligned} \tag{3.18}$$

where f is any differentiable function. Using (3.18) in (3.17), we get

$$\mathbb{E}\mathbb{E}_\sigma \frac{dY_\sigma(t)e^{\beta\sqrt{N}Y_\sigma}}{\mathbb{E}_{\sigma'} e^{\beta\sqrt{N}Y_{\sigma'}}} = dt\mathbb{E}\mathbb{E}_\sigma\mathbb{E}_{\sigma'}\frac{\left(1-\mathbb{1}_{A(d_N(\sigma,\sigma'))\geq t}\right)e^{\beta\sqrt{N}(X_\sigma + X_{\sigma'})}}{\left(\mathbb{E}_{\sigma'} e^{\beta\sqrt{N}X_{\sigma'}}\right)^2}\beta\sqrt{N} \tag{3.19}$$

Thus

$$\mathbb{E}\mathbb{E}_\sigma \frac{dX_\sigma(t)e^{\beta\sqrt{N}X_\sigma}}{\mathbb{E}_{\sigma'} e^{\beta\sqrt{N}X_{\sigma'}}} = dA(t)\mathbb{E}\mu_{\beta,N}^{\otimes 2}\left(\mathbb{1}_{A(d_N(\sigma,\sigma'))<A(t)}\right)\beta\sqrt{N} \tag{3.20}$$

which yields the result upon integration and realizing that $A(d_N(\sigma,\sigma')) < A(t)$ is equivalent to $d_N(\sigma,\sigma') < t$ whenever $A(t)$ is not constant. ♢

Next we want to express the right hand side of (3.16) as a derivative of the free energy. To that end consider for $t \in [0,1]$ and $\epsilon > 0$ fixed the random process

$$X_\sigma^u \equiv X_\sigma + u[X_\sigma(t) - X_\sigma(t-\epsilon)] \tag{3.21}$$

Clearly

$$\operatorname{cov}\left(X_\sigma^u, X_{\sigma'}^u\right) = A^u(d_N(\sigma,\sigma')) \tag{3.22}$$

where

$$A^u(x)=\begin{cases} A(x), & \text{if } x<t-\epsilon \\ A(x)+(2u+u^2)(A(x)-A(t-\epsilon)), & \text{if } t-\epsilon\leq x<t \\ A(x)+(2u+u^2)(A(t)-A(t-\epsilon)), & \text{if } x\geq t \end{cases} \tag{3.23}$$

Let us denote the partition function corresponding to the process with covariance A^u by $Z^u_{\beta,N}$, etc. Clearly we have that

$$\beta\sqrt{N}\mathbb{E}\mathbb{E}_\sigma\frac{(X_\sigma(t)-X_\sigma(t-\epsilon))e^{\beta\sqrt{N}X_\sigma}}{\mathbb{E}_{\sigma'}e^{\beta\sqrt{N}X_{\sigma'}}}=\frac{d}{du}\left(\mathbb{E}\ln Z^u_{\beta,N}\right)_{u=0} \tag{3.24}$$

This yields the generalization of Proposition 4.1 of [BK2] in the discrete case:

Lemma 3.5: *With the notation introduced above we have for any $t\in(0,1]$ and any $\epsilon>0$ that*

$$\beta^{-2}N^{-1}\frac{d}{du}\left(\mathbb{E}\ln Z^u_{\beta,N}\right)_{u=0}=\int_{t-\epsilon}^{t}dA(s)\mathbb{E}\mu^{\otimes 2}_{\beta,N}[d_N(\sigma,\sigma')<s] \tag{3.25}$$

This allows us to obtain an explicit formula for the distance distribution function.

Theorem 3.6: *Under the assumptions of Theorem 3.1,*

$$\lim_{N\uparrow\infty}\mathbb{E}\mu^{\otimes 2}_{\beta,N}(d_N(\sigma,\sigma')<x)=\begin{cases}\beta^{-1}\sqrt{2\ln 2}/\sqrt{\bar{a}(x)}, & \text{if} x\leq x_\beta \\ 1, & \textit{if} x>x_\beta\end{cases} \tag{3.26}$$

Proof: Observe that $F^u_{\beta,N}\equiv N^{-1}\mathbb{E}\ln Z^u_{\beta,N}$ is a convex function of u. A trivial extension of Theorem 3.1 shows that $F^u_{\beta,N}$ converges to the function F^u_β given by the expression (3.12) when A is replaced by A^u. By convexity, this implies that $\lim_{N\uparrow\infty}\frac{d}{du}F_{\beta,N}=\frac{d}{du}F_\beta$ at all points u where F^u_β is differentiable. Thus we only have to compute this derivative. We can write

$$\begin{aligned}F^u_\beta-F_\beta&=\sqrt{2\ln 2}\beta\int_0^{x_\beta}dx\left(\sqrt{\bar{a}^u(x)}-\sqrt{\bar{a}(x)}\right)\\&+\sqrt{2\ln 2}\beta\int_{x_\beta}^{x^u_\beta}dx\sqrt{\bar{a}^u(x)}+\frac{\beta^2}{2}\left(\bar{A}^u(x_\beta)-\bar{A}^u(x^u_\beta)\right)\\&+\frac{\beta^2}{2}\left((\bar{A}^u(1)-\bar{A}^u(x_\beta))-(\bar{A}(1)-\bar{A}(x_\beta))\right).\end{aligned} \tag{3.27}$$

If $x_\beta<t$, then $x_\beta<t-\epsilon$ for $\epsilon>0$ small enough and (3.27) simplifies to

$$\begin{aligned}F^u_\beta-F_\beta&=\frac{\beta^2}{2}\left((\bar{A}^u(1)-\bar{A}^u(x_\beta))-(\bar{A}(1)-\bar{A}(x_\beta))\right)\\&=\frac{\beta^2}{2}\left(\bar{A}^u(1)-\bar{A}(1)\right)=(2u+u^2)\frac{\beta^2}{2}[A(t)-A(t-\epsilon)]\end{aligned} \tag{3.28}$$

Hence in this case

$$\frac{d}{du}\left(F^u_\beta\right)_{u=0}=\beta^2[A(t)-A(t-\epsilon)] \tag{3.29}$$

Inserting this into (3.25) and letting ϵ tend to zero, we obtain that for $t>x_\beta$, $\lim_{N\uparrow\infty}\mathbb{E}\mu^{\otimes 2}_{N,\beta}[d_N(\sigma,\sigma')<t]=1$.

If $x_\beta>t$, we must distinguish two cases: (a) A is strictly convex in a left neighborhood of t, then $\bar{A}(x)=A(x)$ for $x\in[t-\epsilon,t]$; (b) A is linear or concave in a left neighborhood of t, then $\bar{A}(x)$ is linear for $x\in[t-\epsilon,t]$.

(a) In this case the function $A^u(x)$ is not convex in a neighborhood of $t-\epsilon$, since for $x<t-\epsilon$ its derivative $a^u(x)=a(x)$ and for $x>t-\epsilon$ $a^u(x)=a(x)(1+2u+u^2)$. To construct its convex hull $\bar{A}^u(x)$, one should find the points $z_1(u)$, $z_2(u)$, $z_1(u)<t-\epsilon<z_2(u)\le t$ such that the straight line passing through $A(z_1(u))$ and $A^u(z_2(u))$ is tangent to $A(x)$ at $x=z_1(u)$ and to $A^u(x)$ at $x=z_2(u)$. In other words $a(z_1(u))=a(z_2(u))(1+2u+u^2)$ and $A(z_2(u))+(2u+u^2)(A(z_2(u))-A(t-\epsilon))=A(z_1(u))+(z_2(u)-z_1(u))a(z_1(u))$. Then $\bar{A}^u(x)$ coincides with this straight line for $x\in[z_1(u),z_2(u)]$, $\bar{A}^u(x)=\bar{A}(x)$ for $x\in[0,z_1(u))$, $\bar{A}^u(x)=A^u(x)$ for $x\in(z_2(u),t]$, $\bar{A}^u(x)=\bar{A}(x)+(2u+u^2)(A(t)-A(t-\epsilon))$ for $x\in(t,1]$. Then the last terms in (3.27) are zero and

$$\begin{aligned}
F^u_\beta-F_\beta&=\sqrt{2\ln 2}\beta\int_0^{x_\beta}dx\left(\sqrt{\bar{a}^u(x)}-\sqrt{\bar{a}(x)}\right)\\
&=\sqrt{2\ln 2}\beta\int_{z_1(u)}^{t-\epsilon}dx\left(\sqrt{a(z_1(u))}-\sqrt{a(x)}\right)\\
&+\sqrt{2\ln 2}\beta\int_{t-\epsilon}^{z_2(u)}dx\left(\sqrt{a(z_2(u))(1+2u+u^2)}-\sqrt{a(x)(1+2u+u^2)}\right)\\
&+\sqrt{2\ln 2}\beta\int_{t-\epsilon}^{t}dx\left(\sqrt{a(x)(1+2u+u^2)}-\sqrt{a(x)}\right)
\end{aligned} \tag{3.30}$$

Note that the straight line tangent to $A(x)$ at the point $x=z(u)<t-\epsilon$ such that $a(z(u))=a(t-\epsilon)(1+2u+u^2)$ does not cross $A^u(x)$ for $x>z(u)$. Then $z(u)<z_1(u)<t-\epsilon$. Since $z(u)=t-\epsilon+O(u)$, then $z_1(u)=t-\epsilon+O(u)$. It follows that $a(z_1(u))=a(t-\epsilon)+O(u)$, then $a(z_2(u))=a(t-\epsilon)+O(u)$ and finally $z_2(u)=t-\epsilon+O(u)$. Then the integrand in the first term satisfies

$$0\le\sqrt{a(z_1(u))}-\sqrt{a(x)}\le\sqrt{a(z_1(u))}-\sqrt{a(t-\epsilon)}=O(u) \tag{3.31}$$

and in the second one

$$\begin{aligned}
0\le&\sqrt{a(z_2(u))(1+2u+u^2)}-\sqrt{a(x)(1+2u+u^2)}\\
\le&\sqrt{a(z_2(u))(1+2u+u^2)}-\sqrt{a(t-\epsilon)}=O(u)
\end{aligned} \tag{3.32}$$

Since $z_2(u)-z_1(u) = O(u)$, both of these terms are of the order $O(u^2)$ and does not contribute to the derivative of F_β^u. The integral of the first term can be written as

$$(\sqrt{1+2u+u^2}-1)\int_{t-\epsilon}^t \sqrt{a(x)}dx = u(1+o(1))\int_{t-\epsilon}^t \frac{1}{\sqrt{\bar{a}(x)}}dA(x). \tag{3.33}$$

Therefore, by (3.25), we get that

$$\sqrt{2\ln 2}\beta^{-1}\int_{t-\epsilon}^t dA(x)\frac{1}{\sqrt{\bar{a}(x)}} = \int_{t-\epsilon}^t dA(x) \lim_{N\uparrow\infty} \mathbb{E}\mu_{N,\beta}^{\otimes 2}[d_N(\sigma,\sigma') < x]. \tag{3.34}$$

Since this is true for any $\epsilon > 0$, (3.26) follows.

(b) We now consider the case when $\bar{A}$ is linear to the left of t. The deformation of $\bar{A}$ now extends further down to the beginning of the linear piece of $\bar{A}$. Assume that $\bar{A}$ is linear on the interval $(y,z] \supset (t-\epsilon,t]$. Then, for small enough u, the slope of $\bar{A}^u$ will differ from that of $\bar{A}$ only in $(y,z]$. Moreover, $\bar{A}^u(y) = \bar{A}(y)$ and $\bar{A}^u(z) = \bar{A}(z) + (2u+u^2)(A(t)-A(t-\epsilon))$. Let

$$z^* \equiv \sup_{x\in(y,z]} \{\bar{A}^u(x) - \bar{A}(x) < (2u+u^2)(A(t)-A(t-\epsilon))\} \tag{3.35}$$

Obviously, $x_\beta \notin (y,z)$. If $t \le x_\beta$, we get that

$$\begin{aligned} F_\beta^u - F_\beta &= \sqrt{2\ln 2}\beta \int_y^{z^*} dx \left(\sqrt{\bar{a}(z) + (2u+u^2)(A(t)-A(t-\epsilon))/(z^*-y)} - \sqrt{\bar{a}(z)}\right) \\ &= \sqrt{2\ln 2}\beta\sqrt{\bar{a}(z)} \int_y^{z^*} dx \left(\sqrt{1+(2u+u^2)(A(t)-A(t-\epsilon))/(\bar{a}(z)(z^*-y))} - 1\right) \\ &= \sqrt{2\ln 2}\beta(\bar{a}(z))^{-1/2} u(A(t)-A(t-\epsilon)) + O(u^2) \end{aligned} \tag{3.36}$$

Again (3.26) follows now from (3.25). ◊

Remark: It is clear from the above consideration that we may repeat the same computation with $u < 0$ to compute the left-derivative of F_β^u at zero. The results coincide, except when $t = x_\beta$. Similarly, one shows that the second derivative of F_β^u is finite in a neighborhood of zero whenever $t \neq x_\beta$.

4. Ruelle's processes and Ghirlanda-Guerra identities.

We need to begin with some more notation. Let us denote by $\Sigma_\sigma(t)$ the sets

$$\Sigma_\sigma(t) \equiv \{\sigma' \in \mathcal{S}_N | d_N(\sigma,\sigma') > t\} \tag{4.1}$$

Note that by ultrametricity, we have that for all $\sigma' \in \Sigma_\sigma(t)$, $\Sigma_{\sigma'}(t) = \Sigma_\sigma(t)$. The set of all distinct sets $\Sigma_\sigma(t)$, $\sigma \in \mathcal{S}_N$ is isomorphic to $\{-1,1\}^{[tN]} \equiv \mathcal{S}_N^t$.

We will now reconsider the point process $\widehat{\mathcal{W}}_{\beta,N}$ introduced in [BK2] in a somewhat different way. Note that the atoms of the process $\widehat{\mathcal{W}}_{\beta,N}$, $\left(\mu_{\beta,N}(\Sigma_\sigma(q_1)),\dots,\mu_{\beta,N}(\Sigma_\sigma(q_{l(\beta)}))\right)$ can be seen as the increments of the mass distribution around σ measured in the hierarchical distance, i.e. if we define the distribution functions

$$m_\sigma(x) \equiv \mu_{\beta,N}(\sigma : d_N(\sigma',\sigma) > x) \tag{4.2}$$

we can think of the process as $\widehat{\mathcal{W}}_{\beta,N}$ as a point process on the space of probability distributions on $[0,1]$,

$$\widehat{\mathcal{W}}_{\beta,N} = \sum_{\sigma\in\mathcal{S}_N} \delta_{m_\sigma(\cdot)} \tag{4.3}$$

Of course, in the case of finitely many hierarchies, all measures are supported on the set of atomic distributions with atoms at fixed points. As a matter of fact, this definition turns out not to be too convenient. Already in the case of the discrete GREM, we have followed Ruelle's prescription only by convention. In the continuous case, it will become essentially imperative to replace $\widehat{\mathcal{W}}_{\beta,N}$ by the process

$$\mathcal{K}_{\beta,N} \equiv \sum_{\sigma\in\mathcal{S}_N} \mu_{\beta,N}(\sigma)\delta_{m_\sigma(\cdot)} \tag{4.4}$$

The advantage is that this object is a random *probability* measure on the space of mass distributions. Let us mention that in spirit this measure is rather close to the "metastates" introduced by Aizenman and Wehr [AW] and promoted in the context of spin glasses by Newman and Stein [N,NS1,NS2]. It has a very appealing physical interpretation: it tells, for a fixed realization of the disorder, with which probability an observer that is itself distributed with the Gibbs distribution, will see a given distribution of mass around himself.

Let us say a word more on the interpretation of these processes. Note that the distance distribution function

$$\mu_{\beta,N}^{\otimes 2}(d_N(\sigma,\sigma') > q) = \int \mathcal{K}_{\beta,N}(dm)m(q) \tag{4.5}$$

is the first moment of $\mathcal{K}_{\beta,N}$. Thus $\mathcal{K}_{\beta,N}$ will be asymptotically concentrated on distributions for which $m(q) = 0$ if $q > x(\beta)$. In other words, the smallest blocks Σ_σ around any point σ that have positive mass are of size $2^{(1-x(\beta))N}$. Since the mass distribution around any point within such a "massive" block is identical, such a block contributes with mass $\mu_{\beta,N}(\sigma)$ with a Dirac measure on the empirical mass distribution around itself.

While in the discrete case, the convergence of these processes could be proven directly, it is a priori not clear how this could be achieved in the general case. But recall that in the discrete case, instead of constructing the limiting processes directly, the Ghirlanda-Guerra identities provide an alternative way. We will see that this path is still open in the general case.

Point processes on the space of probability distributions.

We will think of the quantities $\mathcal{K}_{\beta,N}$ as probability distributions on the space of probability distributions on $[0,1]$, that is to say as elements of the space $\mathcal{M}_1(\mathcal{M}_1(\mathcal{M}_1([0,1])))$. Equipping the spaces $\mathcal{M}_1([0,1])$ and $\mathcal{M}_1(\mathcal{M}_1([0,1]))$ with the topologies of weak convergence, there is no obstacle to define weak convergence of our objects in much the same way as in the discrete case (when $\mathcal{M}_1([0,1])$ was replaced by a finite dimensional simplex in $\mathbb{R}^n$). Just note that the continuous functions of a measure $m \in \mathcal{M}_1([0,1])$ can be approximated arbitrarily well by monomials in finite collections of integrals with respect to m of indicator functions of (disjoint) intervals $\Delta_1, \dots, \Delta_l \subset [0,1]$, and that in turn continuous functions of a measure $\mathcal{W} \in \mathcal{M}_1(\mathcal{M}_1([0,1]))$ can be approximated by polynomials in a collection of integrals of such functions. Thus, if we show that for any collection $\Delta_{ij} \subset [0,1]$ and integers q_i, r_{ij}, $i = 1, \dots, l$, $j = 1, \dots, k_i$,

$$\begin{aligned}\mathbb{E}\Bigg(&\left(\int \mathcal{K}_{\beta,N}(dm)m(\Delta_{11})^{r_{11}}\dots m(\Delta_{1j_1})^{r_{1j_1}}\right)^{q_1}\dots\\ &\dots\left(\int \mathcal{K}_{\beta,N}(dm)m(\Delta_{l1})^{r_{l1}}\dots m(\Delta_{lj_l})^{r_{lj_l}}\right)^{q_l}\Bigg)\end{aligned} \tag{4.6}$$

converges, then the point process $\mathcal{K}_{N,\beta}$ converges weakly to a limit $\mathcal{K}_\beta$.

A somewhat simpler way of looking at this process is to simply consider

$$K_{\beta,N}(t) \equiv \sum_\sigma \mu_{\beta,N}(\sigma)\delta_{m_\sigma(t)} \tag{4.7}$$

and to think of it as a measure valued process on $[0,1]$. This process is usually called the coalescent and has been studied in the spin glass context by Bolthausen and Sznitman [BS].

An alternative construction based on subordinators was later given by Bertoin and Le Gall [BL]. It clearly contains somewhat less information than $\mathcal{K}_\beta$ itself.

Multi-overlap distributions.

Our task is now to prove the Ghirlanda-Guerra identities in the general case.

Theorem 4.1: *For any $n \in \mathbb{N}$ and any $x \in [0,1]\backslash x_\beta$ for which the derivative of $A(x)$ does not vanish*

$$\begin{aligned}\lim_{N\uparrow\infty}\Bigg|&\mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\dots,\sigma^n)\mathbb{1}_{d_N(\sigma^k,\sigma^{n+1})\geq x}\right)\\ &-\frac{1}{n}\mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\dots,\sigma^n)\left(\sum_{l\neq k}^{n}\mathbb{1}_{d_N(\sigma^k,\sigma^l)\geq x}+\mathbb{E}\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{d_N(\sigma^1,\sigma^2)\geq x})\right)\right)\Bigg| = 0\end{aligned} \tag{4.8}$$

Proof: One of the pillars of the Ghirlanda-Guerra identities is concentration of measure for the free energy, that holds also here:

Lemma 4.2: *Assume that A is non-critical, i.e. whenever x is not in the set of extremal points of $\bar{A}$, $A(x) < \bar{A}(x)$.*

For any β and for any $\epsilon \geq 0$

$$\mathbb{P}\left[|\ln Z_{\beta,N} - \mathbb{E}\ln Z_{\beta,N}| > \epsilon\right] \leq 2\exp\left(-\frac{\epsilon^2}{2\beta^2 N}\right) \tag{4.9}$$

Proof: This follows, e.g., from the standard Gaussian concentration of measure theorem (see [LT], page 21) and the representation of $\ln Z_{\beta,N}$ as a Lipshitz function of $2^{N+1}-2$ independent standard Gaussian random variables with Lipshitz constant $\beta\sqrt{N}$. ◇

As a first step we need the generalization of Lemma 3.4 as in Lemma 4.2 of [BK2].

Lemma 4.3: *For any $t \in (0,1]$ and $\epsilon > 0$, and let $h : \mathcal{S}_N^n \to \mathbb{R}$ be any bounded function of n spin-configurations:*

$$\begin{aligned}&\frac{1}{\sqrt{N}}\mathbb{E}\mu_{\beta,N}^{\otimes n}\left((X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))h(\sigma^1,\dots,\sigma^n)\right)\\ &=\beta\int_{t-\epsilon}^{t} dA(s)\mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\dots,\sigma^n)\left(\sum_{l=1}^{n}\mathbb{1}_{d_N(\sigma^k,\sigma^l)\geq s}-n\mathbb{1}_{d_N(\sigma^k,\sigma^{n+1})\geq s}\right)\right)\end{aligned} \tag{4.10}$$

Proof: Exactly analogous to the proof of Lemma 3.4. ◊

The more important step is the proof of the analogue of Lemma 4.3 of [BK2].

Lemma 4.4: *Let h be as in the previous lemma. Except possibly when $t = x_\beta$,*

$$\begin{aligned}\lim_{N\uparrow\infty}\frac{1}{\sqrt{N}}\Big|&\mathbb{E}\mu_{\beta,N}^{\otimes n}\left((X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))h(\sigma^1,\dots,\sigma^n)\right)\\ &-\mathbb{E}\mu_{\beta,N}\left(X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon)\right)\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(h(\sigma^1,\dots,\sigma^n)\right)\Big|=0\end{aligned}\tag{4.11}$$

Proof: Let us write

$$\begin{aligned}&\Big(\mathbb{E}\mu_{\beta,N}^{\otimes n}\left((X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))\,h(\sigma^1,\dots,\sigma^n)\right)\\ &\qquad-\mathbb{E}\mu_{\beta,N}\left(X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon)\right)\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(h(\sigma^1,\dots,\sigma^n)\right)\Big)^2\\ &=\Big(\mathbb{E}\mu_{\beta,N}^{\otimes n}\Big(\big((X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))-\mathbb{E}\mu_{\beta,N}^{\otimes n}(X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))\big)h(\sigma^1,\dots,\sigma^n)\Big)\Big)^2\\ &\le\mathbb{E}\mu_{\beta,N}^{\otimes n}\Big((X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))-\mathbb{E}\mu_{\beta,N}^{\otimes n}(X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))\Big)^2\,\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(h(\sigma^1,\dots,\sigma^n)\right)^2\end{aligned}\tag{4.12}$$

where the last inequality is the Cauchy–Schwarz inequality applied to the joint expectation with respect to the Gibbs measure and the disorder. Obviously the first factor in the last line is equal to

$$\begin{aligned}&\mathbb{E}\mu_{\beta,N}\left((X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))-\mu_{\beta,N}(X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))\right)^2\\ &\quad+\mathbb{E}\left(\mu_{\beta,N}(X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))-\mathbb{E}\mu_{\beta,N}(X_{\sigma^k}(t)-X_{\sigma^k}(t-\epsilon))\right)^2\\ &=-\beta^{-2}\mathbb{E}\frac{d^2}{du^2}F_{\beta,N}^u+N\beta^{-2}\mathbb{E}\left(\frac{d}{du}F_{\beta,N}^u-\mathbb{E}\frac{d}{du}F_{\beta,N}^u\right)^2\end{aligned}\tag{4.13}$$

where we used the same notation as in the proof of Theorem 3.3. We know that $F_{\beta,N}^u$ converges as $N\uparrow\infty$ and that the limit is infinitely differentiable as a function of u, except possibly when $x_\beta = t$; moreover, $-F_{\beta,N}^u$ is convex in the variable u. Then a standard result of convex analysis (see [Ro], Theorem 25.7) imply that

$$\limsup_{N\uparrow\infty}(-\mathbb{E}\frac{d^2}{du^2}F_{\beta,N}^u)=-\frac{d^2}{du^2}\lim_{N\uparrow\infty}\mathbb{E}F_{\beta,N}^u\tag{4.14}$$

which is finite at zero except possibly if $x_\beta = t$. Thus, the first term in (4.13) will vanish when divided by N. To see that the coefficient of N of the second term gives a vanishing contribution, we use the general fact that if the variance of family of a convex (or concave) functions tends to zero, then the same is true for its derivative, provided the second derivative of the expectation is bounded (see e.g. Lemma 8.9 in [BG], or Proposition 4.3 in [T3]).

But by Lemma 4.2 the variance of $F_{\beta,N}$ tends to zero, and (4.14) implies that $\mathbb{E}\frac{d^2}{du^2}F^u_{\beta,N}$ is bounded for large enough N whenever $\frac{d^2}{du^2}\mathbb{E}F^u_\beta$ is finite. By the remark following the proof of Theorem 3.6 this is true except at the special value $t = x_\beta$. Hence the result of the Lemma is proven. $\diamondsuit$

The theorem now follows easily by using (4.11) on the left hand side of (4.10), and expressing the resulting term with the help of (3.16). Noting that the result holds for any $\epsilon > 0$ then yields (4.8). $\diamondsuit\diamondsuit$

Following [GG], we now define the family of measures $\mathbb{Q}^{(n)}_N$ on the space $[0,1]^{n(n-1)/2}$.

$$\mathbb{Q}^{(n)}_{\beta,N}(\underline{d_N} \in \mathcal{A}) \equiv \mathbb{E}\mu^{\otimes n}_{N,\beta}\left[\underline{d_N} \in \mathcal{A}\right] \tag{4.15}$$

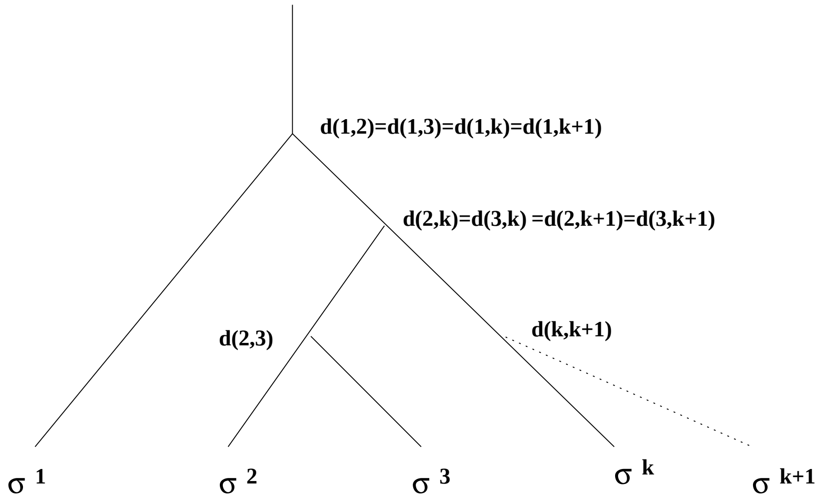
where $\underline{d_N}$ denotes the vector of replica distances whose components are $d_N(\sigma^l, \sigma^k)$, $1 \leq l < k \leq n$. Denote by $\mathcal{B}_k$ the sigma-algebra generated by the first $k(k-1)/2$ coordinates, and let A be a Borel set in $[0,1]$.

Theorem 4.5: *The family of measures $\mathbb{Q}^{(n)}_{\beta,N}$ converge to limiting measures $\mathbb{Q}^{(n)}_\beta$ for all finite n, as $N \uparrow \infty$. Moreover, these measures are uniquely determined by the distance distribution functions f_β. They satisfy the identities*

$$\mathbb{Q}^{(n+1)}_\beta\left(d_{k,n+1} \in A|\mathcal{B}_n\right) = \frac{1}{n}\mathbb{Q}^{(2)}_\beta(A) + \frac{1}{n}\sum_{l\neq k}^{n}\mathbb{Q}^{(n)}_\beta\left(d_{k,l} \in A|\mathcal{B}_n\right) \tag{4.16}$$

for any Borel set A. As a consequence, the random measure $\mathcal{K}_{\beta,N}$ converges in distribution to the random measure $\mathcal{K}_\beta$ whose generalized moments are given by $\mathbb{Q}_\beta$.

Proof: Choosing h as the indicator function of any desired event in $\mathcal{B}_k$, one sees that (4.8) implies (4.16). This actually implies that in the limit $N \uparrow \infty$, the family of measures $\mathbb{Q}^{(n)}_{\beta,N}$ is entirely determined by the two-replica distribution function. While this may not appear obvious, it follows when taking into account the ultrametric property of the function d_N. This is most easily seen by realising that the prescription of the mutual distances between k spin configurations amounts to prescribing a tree (start all k configurations at the origin and continue on top of each other as long as the coordinates coincide, then branch of). To determine the full tree of $k+1$ configurations, it is sufficient to know the overlap of configuration $\sigma^{(k+1)}$ with the configuration it has maximal overlap with, since then all overlaps with all other configurations are determined. But the corresponding probabilities can be computed recursively via (4.15).



The distance d(k,k+1) determines all other distances d(j,k+1)

Now we have already seen that $\mathbb{Q}^{(2)}_{\beta,N} = \mathbb{E}f_{\beta,N}$ converges. Therefore the relation (4.15) implies the convergence of all distributions $\mathbb{Q}^{(n)}_{\beta,N}$, and proves the relation (4.16) hold for the limiting measures.

Now it is clear that all expressions of the form (4.6) can be expressed in terms of the measures $\mathbb{Q}^{(k)}_{\beta,N}$ for k sufficiently large (we leave this as an exercise for the reader to write down). Thus all limit points of sequences of distributions of the measures $\mathcal{K}_{\beta,N}$ must coincide. By compactness of the space $\mathcal{M}_1(\mathcal{M}_1(\mathcal{M}_1([0,1])))$, this implies the convergence of the process $\mathcal{K}_{\beta,N}$ to a limit $\mathcal{K}_\beta$. $\diamondsuit\diamondsuit$

A remarkable feature takes place again if we are only interested in the marginal process $K_\beta(t)$ for fixed t. This process is a simple point process on $[0,1]$ and is fully determined in terms of the moments

$$\begin{aligned}&\mathbb{E}\left(\int K_{\beta,N}(t)(dx)x^{r_1}\cdots\int K_{\beta,N}(t)(dx)x^{r_j}\right)\\ &=\mathbb{E}\mu_{\beta,N}^{\otimes r_1+\dots+r_j+j}\Big(\mathbb{1}_{d_N(\sigma^1,\sigma^{j+1})>t}\cdots\mathbb{1}_{\dots,d_N(\sigma^1,\sigma^{j+r_1})>t}\cdots\\ &\qquad\cdots\mathbb{1}_{d_N(\sigma^j,\sigma^{j+r_1+\dots+r_{j-1}+1})>t}\cdots\mathbb{1}_{d_N(\sigma^j,\sigma^{j+r_1+\dots+r_j})>t}\Big)\end{aligned}\tag{4.17}$$

This restricted family of moments satisfies via the Ghirlanda-Guerra identities exactly the same recursion as in the case of the REM. This implies:

Theorem 4.6: *Assume that* t *is such that* $\mathbb{E}\mu_\beta^{\otimes 2}(d_N(\sigma,\sigma') < t) = 1/\alpha > 0$. *Then the random measure* $K_\beta(t)$ *is a Dirichlet-Poisson process (see e.g.* [Ru,T1]*) with parameter* α .

In fact much more is true. We can consider the processes on arbitrary finite dimensional marginals, i.e.

$$K_{\beta,N}(t_1,\dots,t_m) \equiv \sum_{\sigma\in\mathcal{S}_N} \mu_{\beta,N}(\sigma)\delta_{m_\sigma(t_1),\dots,m_\sigma(t_m)} \tag{4.18}$$

for $0 < t_1 < \cdots < t_m < 1$. The point is that this process is entirely determined by the expressions (4.6) with the Δ_{ij} all of the form $(t_i, 1]$ for t_i in the fixed set of values $t_1,\dots,t_m$. This in turn implies that the process is determined by the multi-replica distribution functions $\mathbb{Q}_{\beta,N}^{(n)}$ restricted to the discrete set of events $\{d_N(\sigma^i,\sigma^j) > t_k\}$. Since these numbers are totally determined through the Ghirlanda-Guerra identities, they are identically to those obtained in a GREM with m levels, i.e. a function A having steps at the values t_i, those two-replica distribution function takes the same values as that of the model with continuous A at the points t_i and is constant between those values. In fact

Theorem 4.7: *Let* $0 < t_1 < \cdots < t_k \leq q_{max}(\beta)$ *be points of increase of* $\mathbb{E}f_\beta$. *Consider a GREM with k levels and parameters* α_i, a_i *and temperature* $\tilde\beta$ *that satisfy* $\ln\alpha_i/\ln 2 = t_i - t_{i-1}$, $\tilde\beta^{-1}\sqrt{2\ln\alpha_i/a_i} = \mathbb{E}f_\beta(t_i)$. *Then*

$$\lim_{N\uparrow\infty} \mathcal{K}_{\beta,N}(t_1,\dots,t_k) = \mathcal{W}_{\tilde\beta}^{(k)} \tag{4.19}$$

Thus, if the t_i are chosen in such a way that for all of them $\mathbb{E}f_\beta(t_i) > 0$, then we can construct an explicit representation of the limiting marginal process $\mathcal{K}_\beta(t_1,\dots,t_m)$ in terms of a Poisson-cascade process via the corresponding formulae in the associated m-level GREM. In this sense we obtain an explicit description of the limiting mass distribution function $\mathcal{K}_\beta$.

References.

[AC] M. Aizenmann and P. Contucci, On the stability of the quenched state in mean field spin-glass models, J. Statist. Phys. **92**, 765-783 (1998).

[AW] M. Aizenman, J. Wehr, Rounding effects of quenched randomness on first-order phase transitions, Comm. Math. Phys. **130**, 489-528 (1990).

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

[BDG] E. Bolthausen, J.-D. Deuschel, and G. Giacomin, Entropic repulsion and the maximum of the two-dimensional harmonic crystal, Ann. Probab. **29**, 1670-1692 (2001).

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

[BK1] A. Bovier and I. Kurkova, Derrida's Generalized random Energy models 1. Poisson cascades and the extremal process, preprint (2002).

[BK2] A. Bovier and I. Kurkova, Derrida's Generalized Random Energy models 2. Gibbs measures and probability cascades, preprint (2002).

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

[BG] A. Bovier, V. Gayrard, The Hopfield model as a generalized random mean field model, in Mathematics of spin glasses and neural networks, A. Bovier, P. Picco. Eds., Progress in Probablity, Birkhäuser, Boston, (1997).

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

[CCR] D. Capocaccia, M. Cassandro, and P. Picco, On the existence of Thermodynamics for the Generalised Random Energy Model. J. Stat. Physics **46**, 493–505 (1987).

[D] B. Derrida, A generalization 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 generalized 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).

[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).

[Gu] F. Guerra, Broken replica symmetry bounds in the mean field spin glass model, preprint 2002.

[GT] F. Guerra and F. Toninelli, The thermodynamics limit in mean field spin glass models, to appear in Commun. Math. Phys. (2002).

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

[LT] M. Ledoux and L. Talagrand, Probability in Banach space. Spinger, Berlin-Heidelberg-New York, 1991.

[N] Ch.M. Newman, Topics in disordered systems, Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1997.

[NS1] Ch.M. Newman and D.L. Stein, Thermodynamic chaos and the structure of short range spin glasses, in "Mathematical aspects of spin glasses and neural networks", A. Bovier and P. Picco (Eds.), Progress in Probability, Birkhäuser, Boston (1997).

[NS2] C.M. Newman, D.L. Stein, Metastate approach to thermodynamic chaos. Phys. Rev. E **55**, 5194-5211 (1997).

[Ro] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, N.J. 1970.

[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", preprint 2002.

[T3] M. Talagrand, Rigorous results for the Hopfield model with many patterns, Probab. Theory Related Fields **110**, 177-276 (1998).