METASTATES IN THE HOPFIELD MODEL IN THE REPLICA SYMMETRIC REGIME[#]

Anton Bovier [1]

Weierstraß–Institut
für Angewandte Analysis und Stochastik
Mohrenstrasse 39, D-10117 Berlin, Germany

Véronique Gayrard[2]

Centre de Physique Théorique - CNRS
Luminy, Case 907
F-13288 Marseille Cedex 9, France

Abstract: We study the finite dimensional marginals of the Gibbs measure in the Hopfield model at low temperature when the number of patterns, M, is proportional to the volume with a sufficiently small proportionality constant $\alpha > 0$. It is shown that even when a single pattern is selected (by a magnetic field or by conditioning), the marginals do not converge almost surely, but only in law. The corresponding limiting law is constructed explicitly. We fit our result in the recently proposed language of "metastates" which we discuss in some length. As a byproduct, in a certain regime of the parameters α and β (the inverse temperature), we also give a simple proof of Talagrand's [T1] recent result that the replica symmetric solution found by Amit, Gutfreund, and Sompolinsky [AGS] can be rigorously justified.

Keywords: Hopfield model, neural networks, metastates, replica symmetry, Brascamp-Lieb inequalities

AMS Subject Classification: 82B44, 60K35, 82C32

Work partially supported by the Commission of the European Communities under contract CHRX-CT93-0411

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

2 e-mail: gayrard@cpt.univ-mrs.fr

1. Introduction

Strongly disordered systems such as spin glasses represent some of the most interesting and most difficult problems of statistical mechanics. Amongst the most remarkable achievements of theoretical physics in this field is the exact solution of some models of mean field type via the replica trick and Parisi's replica symmetry breaking scheme (For an exposition see [MPV]; the application to the Hopfield model [Ho] was carried out in [AGS]). The replica trick is a formal tool that allows to eliminate the difficulty of studying disordered systems by integrating out the randomness at the expense of having to perform an analytic continuation of some function computable only on the positive integers to the value zero[1]. Mathematically, this procedure is highly mysterious and has so far resisted all attempts to be put on a solid basis. On the other hand, its apparent success is a clear sign that something ought to be understood better in this method. An apparently less mysterious approach that yields the same answer is the cavity method [MPV]. However, here too, the derivation of the solutions involves a large number of intricate and unproven assumptions that seem hard or impossible to justify in general.

However, there has been some distinct progress in understanding the approach of the cavity method at least in simple cases where no breaking of the replica symmetry occurs. The first attempts in this direction were made by Pastur and Shcherbina [PS] in the Sherrington-Kirkpatrick model and Pastur, Shcherbina and Tirozzi [PST] in the Hopfield model. Their results were conditional: They assert to show that the replica symmetric solution, holds under certain unverified assumption, namely the vanishing of the so-called Edwards-Anderson parameter. A breakthrough was achieved in a recent paper by Talagrand [T1] where he proved the validity of the replica symmetric solution in an explicit domain of the model parameters in the Hopfield model. His approach is purely by induction over the volume (i.e. the cavity method) and uses only some a priori estimates on the support properties of the distribution of the so-called overlap parameters as first proven in [BGP1,BGP2] and in sharper form in [BG1].

Let us recall the definition of the Hopfield model and some basic notations. Let $\mathcal{S}_N \equiv \{-1,1\}^N$ denote the set of functions $\sigma : \{1,\dots,N\} \to \{-1,1\}$, and set $\mathcal{S} \equiv \{-1,1\}^{\mathbb{N}}$. We call σ a spin configuration and denote by σ_i the value of σ at i. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an abstract probability space and let ξ_i^μ, $i,\mu \in \mathbb{N}$, denote a family of independent identically distributed random variables on this space. For the purposes of this paper we will assume that $\mathbb{P}[\xi_i^\mu = \pm 1] = \frac{1}{2}$. We will write $\xi^\mu[\omega]$ for the N-dimensional random vector whose i-th component is given by $\xi_i^\mu[\omega]$ and call such

[1] As a matter of fact, such an analytic continuation is not performed. What is done is much more subtle: The function at integer values is represented as some integral suitable for evaluation by a saddle point method. Instead of doing this, apparently irrelevant critical points are selected judiciously and the ensuing wrong value of the function is then continued to the correct value at zero.

a vector a 'pattern'. On the other hand, we use the notation $\xi_i[\omega]$ for the M-dimensional vector with the same components. When we write $\xi[\omega]$ without indices, we frequently will consider it as an $M \times N$ matrix and we write $\xi^t[\omega]$ for the transpose of this matrix. Thus, $\xi^t[\omega]\xi[\omega]$ is the $M \times M$ matrix whose elements are $\sum_{i=1}^N \xi_i^\mu[\omega]\xi_i^\nu[\omega]$. With this in mind we will use throughout the paper a vector notation with $(\cdot,\cdot)$ standing for the scalar product in whatever space the argument may lie. E.g. the expression (y, ξ_i) stands for $\sum_{\mu=1}^M \xi_i^\mu y_\mu$, etc.

We define random maps $m_N^\mu[\omega] : \mathcal{S}_N \to [-1,1]$ through[2]

$$m_N^\mu[\omega](\sigma) \equiv \frac{1}{N}\sum_{i=1}^N \xi_i^\mu[\omega]\sigma_i \tag{1.1}$$

Naturally, these maps 'compare' the configuration σ globally to the random configuration $\xi^\mu[\omega]$. A Hamiltonian is now defined as the simplest negative function of these variables, namely

$$\begin{aligned} H_N[\omega](\sigma) &\equiv -\frac{N}{2}\sum_{\mu=1}^{M(N)} \left(m_N^\mu[\omega](\sigma)\right)^2 \\ &= -\frac{N}{2}\left\|m_N[\omega](\sigma)\right\|_2^2 \end{aligned} \tag{1.2}$$

where $M(N)$ is some, generally increasing, function that crucially influences the properties of the model. $\|\cdot\|_2$ denotes the ℓ_2-norm in $I\!R^M$, and the vector $m_N[\omega](\sigma)$ is always understood to be $M(N)$-dimensional.

Through this Hamiltonian we define in a natural way finite volume Gibbs measures on $\mathcal{S}_N$ via

$$\mu_{N,\beta}[\omega](\sigma) \equiv \frac{1}{Z_{N,\beta}[\omega]} e^{-\beta H_N[\omega](\sigma)} \tag{1.3}$$

and the induced distribution of the overlap parameters

$$\mathcal{Q}_{N,\beta}[\omega] \equiv \mu_{N,\beta}[\omega] \circ m_N[\omega]^{-1} \tag{1.4}$$

The normalizing factor $Z_{N,\beta}[\omega]$, given by

$$Z_{N,\beta}[\omega] \equiv 2^{-N}\sum_{\sigma\in\mathcal{S}_N} e^{-\beta H_N[\omega](\sigma)} \equiv I\!E_\sigma e^{-\beta H_N[\omega](\sigma)} \tag{1.5}$$

is called the partition function. We are interested in the large N behaviour of these measures. In our previous work we have been mostly concerned with the limiting induced measures. In this paper we return to the limiting behaviour of the Gibbs measures themselves, making use, however, of the information obtained on the asymptotic properties of the induced measures.

[2] We will make the dependence of random quantities on the random parameter ω explicit by an added $[\omega]$ whenever we want to stress it. Otherwise, we will frequently drop the reference to ω to simplify the notation.

We pursue two objectives. Firstly, we give an alternative proof of Talagrand's result (with possibly a slightly different range of parameters) that, although equally based on the cavity method, makes more extensive use of the properties of the overlap-distribution that were proven in [BG1]. This allows, in our opinion, some considerable simplifications. Secondly, we will elucidate some conceptual issues concerning the infinite volume Gibbs states in this model. Several delicacies in the question of convergence of finite volume Gibbs states (or local specifications) in highly disordered systems, and in particular spin glasses, were pointed out repeatedly by Newman and Stein over the last years [NS1,NS2]. But only during the last year did they propose the formalism of so-called "metastates" [NS3,NS4,N] that seems to provide the appropriate framework to discuss these issues. In particular, we will show that in the Hopfield model, this formalism seems unavoidable for spelling out convergence results.

Let us formulate our main result in a slightly preliminary form (precise formulations require some more discussion and notation and will be given in Section 5).

Denote by $m^*(\beta)$ the largest solution of the mean field equation $m = \tanh(\beta m)$ and by e^μ the μ-th unit vector of the canonical basis of $I\!R^M$. For all $(\mu, s) \in \{-1,1\}\times\{1,\dots,M\}$ let $B_\rho^{(\mu,s)} \subset I\!R^M$ denote the ball of radius ρ centered at sm^*e^μ. For any pair of indices (μ, s) and any $\rho > 0$ we define the conditional measures

$$\mu_{N,\beta,\rho}^{(\mu,s)}[\omega](\mathcal{A}) \equiv \mu_{N,\beta}[\omega](\mathcal{A} \mid B_\rho^{(\mu,s)}), \quad \mathcal{A} \in \mathcal{B}(\{-1,1\}^N) \tag{1.6}$$

The so called "replica symmetric equations"[3] of [AGS] is the following system of equations in three unknowns m_1, r, and q, given by

$$\begin{aligned} m_1 &= \int d\mathcal{N}(g) \tanh(\beta(m_1 + \sqrt{\alpha r} g)) \\ q &= \int d\mathcal{N}(g) \tanh^2(\beta(m_1 + \sqrt{\alpha r} g)) \\ r &= \frac{q}{(1-\beta+\beta q)^2} \end{aligned} \tag{1.7}$$

With this notation we can state

Theorem 1.1: *There exists a nonempty connected set of parameters β, α bounded by the curves $\alpha = 0$, $\alpha = c(m^*(\beta))^4$ and $\beta = c'\alpha$, such that if $\lim_{N\uparrow\infty} M(N)/N = \alpha$ the following holds: For any finite $I \subset I\!N$, and for any $s_I \subset \{-1,1\}^I$,*

$$\mu_{N,\beta,\rho}^{(\mu,s)}(\{\sigma_I = s_I\}) \to \prod_{i\in I} \frac{e^{\beta s_i [m_1 \xi_i^1 + g_i\sqrt{\alpha r}]}}{2\cos\left(\beta\left[m_1\xi_i^1 + g_i\sqrt{\alpha r}\right]\right)} \tag{1.8}$$

[3] We cite these equations, (3.3-5) in [AGS] only for the case $k = 1$, where k is the number of the so-called "condensed patterns". One could generalize our results presumably measures conditioned on balls around "mixed states", i.e. the metastable states with more than one "condensed pattern", but we have not worked out the details.

as $N\uparrow\infty$, where the g_i, $i\in I$ are independent gaussian random variables with mean zero and variance one that are independent of the random variables ξ_i^1, $i\in I$. The convergence is understood in law with respect to the distribution of the gaussian variables g_i.

This theorem should be juxtaposed to our second result:

Theorem 1.2: *On the same set of parameter as in Theorem 1.1, the following is true with probability one: For any finite $I\subset I\!N$ and for any $x\in I\!R^I$, there exist subsequences $N_k[\omega]\uparrow\infty$ such that for any $s_I\subset\{-1,1\}^I$, if $\alpha>0$,*

$$\lim_{k\uparrow\infty}\mu^{(\mu,s)}_{N_k[\omega],\beta,\rho}[\omega](\{\sigma_I=s_I\})=\prod_{i\in I}\frac{e^{s_ix_i}}{2\cosh(x_i)} \tag{1.9}$$

The above statements may look a little bit surprising and need clarification. This will be the main purpose of Section 2, where we give a rather detailed discussion of the problem of convergence and the notion of metastates with the particular issues in disordered mean field models in view. We will also propose yet a different notion of a state (let us call it "superstate"), that tries to capture the asymptotic volume dependence of Gibbs states in the form of a continuous time measure valued stochastic process. We also discuss the issue of the "boundary conditions" or rather "external fields", and the construction of conditional Gibbs measures in this context. This will hopefully prepare the ground for the understanding of our results in the Hopfield case.

The following two section collect technical preliminaries. Section 3 recalls some results on the overlap distribution from [BG1-3] that will be crucially needed later. Section 4 states and proves a version of the Brascamp-Lieb inequalities [BL] that is suitable for our situation.

Section 5 contains our central results. Here we construct explicitly the finite dimensional marginals of the Gibbs measures in finite volume and study their behaviour in the infinite volume limit. The results will be stated in the language of metastates. In this section we assume the convergence of certain thermodynamic functions which will be proven in Section 6. Modulo this, this section contains the precise statements and proofs of Theorems 1.1 and 1.2.

In Section 6 we give a proof of the convergence of these quantities and we relate them to the replica symmetric solution. This sections is largely based on the ideas of [PST] and [T1] and is mainly added for the convenience of the reader.

Acknowledgements: We gratefully acknowledge helpful discussions on metastates with Ch. Newman and Ch. Külske.

2. Notions of convergence of random Gibbs measures.

In this section we make some remarks on the appropriate picture for the study of limiting Gibbs measures for disordered systems, with particular regard to the situation in mean-field like systems. Although some of the observations we will make here arose naturally from the properties we discovered in the Hopfield model, our understanding has been greatly enhanced by the recent work of Newman and Stein [NS3,NS4,N] and their introduction of the concept of "metastates". We refer the reader to their papers for more detail and further applications. Some nice examples can also be found in [K,BGK]. Otherwise, we keep this section self-contained and geared for the situation we will describe in the Hopfield model, although part of the discussion is very general and not restricted to mean field situations. For this reason we talk about finite volume measures indexed by finite sets Λ rather then by the integer N.

Metastates. The basic objects of study are *finite volume Gibbs measures*, $\mu_{\Lambda,\beta}$ (which for convenience we will always consider as measures on the infinite product space $\mathcal{S}_\infty$). We denote by $(\mathcal{M}_1(\mathcal{S}_\infty),\mathcal{G})$ the measurable space of probability measures on $\mathcal{S}_\infty$ equipped with the sigma-algebra $\mathcal{G}$ generated by the open sets with respect to the weak topology on $\mathcal{M}_1(\mathcal{S}_\infty)$[4]. We will always regard Gibbs measures as random variables on the underlying probability space $(\Omega,\mathcal{F},I\!\!P)$ with values in the space $\mathcal{M}_1(\mathcal{S}_\infty)$, i.e. as measurable maps $\Omega\to\mathcal{M}_1(\mathcal{S}_\infty)$.

We are in principle interested in considering weak limits of these measures as $\Lambda\uparrow\infty$. There are essentially three things that may happen:

(1) Almost sure convergence: For $I\!\!P$-almost all ω,

$$\mu_\Lambda[\omega]\to\mu_\infty[\omega] \tag{2.1}$$

where $\mu_\infty[\omega]$ may or may not depend on ω (in general it will).

(2) Convergence in law:

$$\mu_\Lambda \stackrel{\mathcal{D}}{\to} \mu_\infty \tag{2.2}$$

(3) Almost sure convergence along random subsequences: There exist (at least for almost all ω) subsequences $\Lambda_i[\omega]\uparrow\infty$ such that

$$\mu_{\Lambda_i[\omega]}[\omega]\to\mu_{\infty,\{\Lambda_i[\omega]\}}[\omega] \tag{2.3}$$

In systems with compact single site state space, (3) holds always, and there are models with non-compact state space where it holds with the "almost sure" provision (see e.g. [BK]). However,

[4] Note that a basis of open sets is given by sets of the forms $\mathcal{N}_{f_1,\dots,f_k,\epsilon}(\mu)\equiv\{\mu'|\forall_{1\leq i\leq k}|\mu(f_i)-\mu'(f_i)|<\epsilon\}$, where f_i are continuous functions on $\mathcal{S}^\infty$; indeed, it is enough to consider cylinder functions.

this contains little information, if the subsequences along which convergence holds are only known implicitly. In particular, it gives no information on how, for any given large Λ the measure μ_Λ "looks like approximately". In contrast, if (i) holds, we are in a very nice situation, as for any large enough Λ and for (almost) any realization of the disorder, the measure $\mu_\Lambda[\omega]$ is well approximated by $\mu_\infty[\omega]$. Thus, the situation would be essentially like in an ordered system (the "almost sure" excepted). It seems to us that the common feeling of most people working in the field of disordered systems was that this could be arranged by putting suitable boundary conditions or external fields, to "extract pure states". Newman and Stein [NS1] were, to our knowledge, the first to point to difficulties with this point of view. *In fact, there is no reason why we should ever be, or be able to put us, in a situation where (1) holds*, and this possibility should be considered as perfectly exceptional. With (3) uninteresting and (1) unlikely, we are left with (2). By compactness, (2) holds always at least for (non-random!) subsequences Λ_n, and even convergence without subsequences can be expected rather commonly. On the other hand, (2) gives us very reasonable information on our system, telling us what is the chance that our measure μ_Λ for large Λ will look like some measure μ_∞. This is much more than what (3) tells us, and baring the case where (1) holds, all we may reasonably expect to know.

We should thus investigate the case (2) more closely. As proposed actually first by Aizenman and Wehr [AW], it is most natural to consider an object K_Λ defined as a measure on the product space $\Omega \otimes \mathcal{M}_1(\mathcal{S}_\infty)$ (equipped with the product topology and the weak topology, respectively), such that its marginal distribution on Ω is $I\!P$ while the conditional measure, $\kappa_\Lambda(\cdot)[\omega]$, on $\mathcal{M}_1(\mathcal{S}_\infty)$ given $\mathcal{F}$[5]is the Dirac measure on $\mu_\Lambda[\omega]$; the marginal on $\mathcal{M}_1(\mathcal{S}_\infty)$ is then of course the law of μ_Λ. The advantage of this construction over simply regarding the law of μ_Λ lies in the fact that we can in this way extract more information by conditioning, as we shall explain. Note that by compactness K_Λ converges at least along (non-random!) subsequences, and we may assume that it actually converges to some measure K. Now the case (1) above corresponds to the situation where the conditional probability on $\mathcal{G}$ given $\mathcal{F}$ is degenerate, i.e.

$$\kappa(\cdot)[\omega] = \delta_{\mu_\infty[\omega]}(\cdot), \qquad \text{a.s.} \tag{2.4}$$

Thus we see that in general even the conditional distribution $\kappa(\cdot)[\omega]$ of K is a nontrivial measure on the space of infinite volume Gibbs measures, this latter object being called the (Aizenman-Wehr) metastate[6]. What happens is that the asymptotic properties of the Gibbs measures as the volume tends to infinity depend in a intrinsic way on the tail sigma field of the disorder variables, and even

[5] We write shorthand $\mathcal{F}$ for $\mathcal{M}_1(\mathcal{S}^\infty)\otimes\mathcal{F}$ whenever appropriate.

[6] It may be interesting to recall the reasons that led Aizenman and Wehr to this construction. In their analysis of the effect of quenched diorder on phase transition they required the existence of "translation-covariant" states. Such object could be constructed as weak limits of finite volume states with e.g. periodic or translation invariant

after all random variables are fixed, some "new" randomness appears that allows only probabilistic statements on the asymptotic Gibbs state.

A toy example: It may be useful to illustrate the passage from convergence in law to the Aizenman-Wehr metastate in a more familiar context, namely the ordinary central limit theorem. Let $(\Omega, \mathcal{F}, I\!P)$ be a probability space, and let $\{X_i\}_{i\in I\!N}$ be a family of i.i.d. centered random variables with variance one; let $\mathcal{F}_n$ be the sigma algebra generated by $X_1, \dots, X_n$ and let $\mathcal{F} \equiv \lim_{n\uparrow\infty} \mathcal{F}_n$. Define the real valued random variable $G_n \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$. We may define the joint law K_n of G_n and the X_i as a probability measure on $I\!R \otimes \Omega$. Clearly, this measure converges to some measure K whose marginal on $I\!R$ will be the standard normal distribution. However, we can say more, namely

Toy-Lemma 2.1 *In the example described above, the conditional measure $\kappa(\cdot)[\omega] \equiv K(\cdot|\mathcal{F})$ satisfies*

$$\kappa(\cdot)[\omega] = \mathcal{N}(0,1), \quad I\!P\text{-a.s.} \tag{2.5}$$

Proof: We need to understand what (2.5) means. Let f be a continuous function on $I\!R$. We claim that for almost all ω,

$$\int f(x)\kappa(dx)[\omega] = \int \frac{e^{-x^2/2}}{\sqrt{2\pi}} f(x)dx \tag{2.6}$$

Define the martingale $h_n \equiv \int f(x)K(dx, d\omega|\mathcal{F}_n)$. We may write

$$\begin{aligned} h_n &= \lim_{N\uparrow\infty} I\!E_{X_{n+1}} \dots I\!E_{X_N} f\left(\tfrac{1}{\sqrt{N}} \sum_{i=1}^N X_i\right) \\ &= \lim_{N\uparrow\infty} I\!E_{X_{n+1}} \dots I\!E_{X_N} f\left(\tfrac{1}{\sqrt{N-n}} \sum_{i=n+1}^N X_i\right), \quad \text{a.s.} \\ &= \int \frac{e^{-x^2/2}}{\sqrt{2\pi}} f(x)dx, \end{aligned} \tag{2.7}$$

where we used that for fixed N, $\frac{1}{\sqrt{N}}\sum_{i=1}^n X_i$ converges to zero as $N \uparrow \infty$ almost surely. Thus, for any continuous f, h_n is almost surely constant, while $\lim_{n\uparrow\infty} h_n = \int f(x)K(dx, d\omega|\mathcal{F})$, by the martingale convergence theorem. This proves the lemma. $\diamondsuit$

The CLT example may inspire the question whether one might not be able to retain more information on the convergence of the random Gibbs state than is kept in the Aizenman-Wehr metastate. The metastate tells us about the probability distribution of the limiting measure, but

---

boundary conditions, provided the corresponding sequences converge almost surely (and not via subsequences with possibly different limits). They noted that in a general disordered system this may not be true. The metastate provided a way out of this difficulty.

we have thrown out all information on how for a given ω, the finite volume measures behave as the volume increases.

Newman and Stein [NS3,NS4] have introduced a possibly more profound concept of the *empirical metastate* which captures more precisely the asymptotic volume dependence of the Gibbs states in the infinite volume limit. We will briefly discuss this object and elucidate its meaning in the above CLT context. Let Λ_n be an increasing and absorbing sequence of finite volumes. Define the random empirical measures $\kappa_N^{em}(\cdot)[\omega]$ on $(\mathcal{M}_1(\mathcal{S}^\infty))$ by

$$\kappa_N^{em}(\cdot)[\omega] \equiv \frac{1}{N}\sum_{n=1}^{N} \delta_{\mu_{\Lambda_n}[\omega]} \tag{2.8}$$

In [NS4] it was proven that for sufficiently sparse sequences Λ_n and subsequences N_i, it is true that almost surely

$$\lim_{i\uparrow\infty} \kappa_{N_i}^{em}(\cdot)[\omega] = \kappa(\cdot)[\omega] \tag{2.9}$$

Newman and Stein conjectured that in many situations, the use of sparse subsequences would not be necessary to achieve the above convergence. However, Külske [K] has exhibited some simple mean field examples where almost sure convergence only holds for very sparse (exponentially spaced) subsequences). He also showed that for more slowly growing sequences convergence in law can be proven in these cases.

Toy example revisited: All this is easily understood in our example. We set $G_n \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$. Then the empirical metastate corresponds to

$$\kappa_N^{em}(\cdot)[\omega] \equiv \frac{1}{N}\sum_{n=1}^{N} \delta_{G_n[\omega]} \tag{2.10}$$

We will prove that the following Lemma holds:

Toy-Lemma 2.2 *Let G_n and $\kappa_N^{em}(\cdot)[\omega]$ be defined above. Let B_t, $t \in [0,1]$ denote a standard Brownian motion. Then*

(i) The random measures κ_N^{em} converge in law to the measure $\kappa^{em} = \int_0^1 dt \delta_{t^{-1/2}B_t}$

(ii)

$$I\!E\left[\kappa^{em}(\cdot)|\mathcal{F}\right] = \mathcal{N}(0,1) \tag{2.11}$$

Proof: Our main objective is to prove (i). We will see that quite clearly, this result relates to Lemma 2.1 as the CLT to the Invariance Principle, and indeed, its proof is essentially an immediate consequence of Donsker's Theorem. Donsker's theorem (see [HH] for a formulation in more generality than needed in this chapter) asserts the following: Let $\eta_n(t)$ denote the continuous function on $[0,1]$ that for $t = k/n$ is given by

$$\eta_n(k/n) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{k} X_i \tag{2.12}$$

and that interpolates linearly between these values for all other points t. Then, $\eta_n(t)$ converges in distribution to standard Brownian motion in the sense that for any continuous functional $F : C([0,1]) \to I\!R$ it is true that $F(\eta_n)$ converges in law to $F(B)$. From here the proof of (i) is obvious. We have to proof that for any bounded continuous function f,

$$\begin{aligned} &\frac{1}{N}\sum_{n=1}^{N} \delta_{G_n[\omega]}(f) \equiv \frac{1}{N}\sum_{n=1}^{N} f\left(\eta_n(n/N)/\sqrt{n/N}\right) \to \\ &\int_0^1 dt f(B_t/\sqrt{t}) \equiv \int_0^1 dt \delta_{B_t/\sqrt{t}}(f) \end{aligned} \tag{2.13}$$

To see this, simply define the continuous functionals F and F_N by

$$F(\eta) \equiv \int_0^1 dt f(\eta(t)/\sqrt{t}) \tag{2.14}$$

and

$$F_N(\eta) \equiv \frac{1}{N}\sum_{n=1}^{N} f(\eta(n/N)/\sqrt{n/N}) \tag{2.15}$$

We have to show that in distribution $F(B) - F_N(\eta_N)$ converges to zero. But

$$F(B) - F_N(\eta_N) = F(B) - F(\eta_N) + F(\eta_N) - F_N(\eta_N) \tag{2.16}$$

By the invariance principle, $F(B) - F(\eta_N)$ converges to zero in distribution while $F(\eta_N) - F_N(\eta_N)$ converges to zero since F_N is the Riemann sum approximation to F.

To see that (ii) holds, note first that as in the CLT, the brownian motion B_t is measurable with respect to the tail sigma-algebra of the X_i. Thus

$$I\!E\left[\kappa^{em}|\mathcal{F}\right] = \mathcal{N}(0,1) \tag{2.17}$$

$\diamondsuit$

Remark: It is easily seen that for sufficiently sparse subsequences n_i (e.g. $n_i = i!$),

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{G_{n_i}} \to \mathcal{N}(0,1), \quad \text{a.s} \tag{2.18}$$

but the weak convergence result contains in a way more information.

Superstates: In our example we have seen that the empirical metastate converges in distribution to the empirical measure of the stochastic process $B_t/\sqrt{t}$. It appears natural to think that the construction of the corresponding continuous time stochastic process itself is actually the right way to look at the problem also in the context of random Gibbs measures, and that the the empirical metastate could converge (in law) to the empirical measure of this process. To do this we propose the following, yet somewhat tentative construction.

We fix again a sequence of finite volumes Λ_n[7]. We define for $t \in [0,1]$

$$\mu^t_{\Lambda_n}[\omega] \equiv (t-[tn]/n)\mu_{\Lambda_{[tn]+1}}[\omega] + (1-t+[tn]/n)\mu_{\Lambda_{[tn]}}[\omega] \tag{2.19}$$

(where as usual $[x]$ denote the smallest integer less than or equal to x). Clearly this object is a continuous time stochastic process whose state space is $\mathcal{M}_1(\mathcal{S})$. We may try to construct the limiting process

$$\mu_t[\omega] \equiv \lim_{n\uparrow\infty} \mu^t_{\Lambda_n}[\omega] \tag{2.20}$$

where the limit again can in general be expected only in distribution. Obviously, in our CLT example, this is precisely how we construct the Brownian motion in the invariance principle. We can now of course repeat the construction of the Aizenman-Wehr metastate on the level of processes. To do this, one must make some choices for the topological space one wants to work in. A natural possibility is to consider the space $C\left([0,1],\mathcal{M}_1(\mathcal{S}^\infty)\right)$ of continuous measure valued function equipped with the uniform weak topology[8], i.e. we say that a sequence of its elements λ_i converges to λ, if and only if, for all continuous functions $f : \mathcal{S}^\infty \to I\!\!R$,

$$\lim_{i\to\infty} \sup_{t\in[0,1]} |\lambda_{i,t}(f) - \lambda_t(f)| = 0 \tag{2.21}$$

Since the weak topology is metrizable, so is the uniform weak topology and $C\left([0,1],\mathcal{M}_1(\mathcal{S}^\infty)\right)$ becomes a metric space so we may define the corresponding sigma-algebra generated by the open sets. Taking the tensor product with our old Ω, we can thus introduce the set $\mathcal{M}_1\left(C\left([0,1],\mathcal{M}_1(\mathcal{S}^\infty)\right)\otimes\Omega\right)$ of probability measures on this space tensored with Ω. Then we define the elements

$$\mathcal{K}_n \in \mathcal{M}_1\left(C\left([0,1],\mathcal{M}_1(\mathcal{S}^\infty)\right)\otimes\Omega\right)$$

whose marginals on Ω are $I\!\!P$ and whose conditional measure on $C\left([0,1],\mathcal{M}_1(\mathcal{S}^\infty)\right)$, given $\mathcal{F}$ are the Dirac measure on the measure valued function $\mu_{\Lambda_{[tn]}}[\omega]$, $t \in [0,1]$. Convergence, and even the

[7] The outcome of our construction will depend on the choice of this sequence. Our philosophy here would be to choose a natural sequence of volumes for the problem at hand. In mean field examples this would be $\Lambda_n=\{1,\dots,n\}$, on a lattice one might choose cubes of sidelength n.

[8] Another possibility would be a measure valued version of the space $D([0,1],\mathcal{M}_1(\mathcal{S}))$ of measure valued Càdlàg functions. The choice depends essentially on the properties we expect from the limiting process (i.e. continuous sample paths or not).

existence of limit points for this sequence of measures is now no longer a trivial matter. The problem of the existence of limit points can be circumvented by using a weaker notion of convergence, e.g. that of the convergence of any finite dimensional marginal. Otherwise, some tightness condition is needed [HH], e.g. we must check that for any continuous function f, $\sup_{|s-t|\leq\delta}|\mu^t_{\Lambda_n}(f)-\mu^s_{\Lambda_n}(f)|$ converges to zero in probability, uniformly in N, as $\delta\downarrow 0$.[9]

We can always hope that the limit as n goes to infinity of $\mathcal{K}_n$ exists. *If* the limit, $\mathcal{K}$ exists, we can again consider its conditional distribution given $\mathcal{F}$, and the resulting object is the functional analog of the Aizenman-Wehr metastate. (We feel tempted to call this object the "superstate". Note that the marginal distribution of the superstate "at time $t=1$" is the Aizenman-Wehr metastate, and the law of the empirical distribution of the underlying process is the empirical metastate). The "superstate" contains an enormous amount of information on the asymptotic volume dependence of the random Gibbs measures; on the other hand, its construction in any explicit form is generally hardly feasible.

Finally, we want to stress that the superstate will normally depend on the choice of the basic sequences Λ_n used in its construction. This feature is already present in the empirical metastate. In particular, sequences growing extremely fast will give different results than slowly increasing sequences. On the other hand, the very precise choice of the sequences should not be important. A natural choice would appear to us sequences of cubes of sidelength n, or, in mean field models, simply the sequence of volumes of size n.

Boundary conditions, external fields, conditioning. In the discussion of Newman and Stein, metastates are usually constructed with simple boundary conditions such as periodic or "free" ones. They emphasize the feature of the "selection of the states" by the disorder in a given volume without any bias through boundary conditions or symmetry breaking fields. Our point of view is somewhat different in this respect in that we think that the idea to apply special boundary conditions or, in mean field models, symmetry breaking terms, to improve convergence properties, is still to some extend useful, the aim ideally being to achieve the situation (1). Our only restriction in this is really that our procedure shall have some *predictive power*, that is, it should give information of the approximate form of a finite volume Gibbs state. This excludes any construction involving subsequences via compactness arguments. We thus are interested to know to what extend it is possible to reduce the "choice" of available states for the randomness to select from, to smaller subsets and to classify the minimal possible subsets (which then somehow play the rôle of *extremal states*). In fact, in the examples considered in [K,BGK] it would be possible to reduce the size of

[9] There are pathological examples in which we would not expect such a result to be true. An example is the "highly disordered spin glass model" of Newman and Stein [NS5]. Of course, tightness may also be destroyed by choosing very rapidly growing sequences of volumes Λ_n.

such subsets to one, while in the example of the present paper, we shall see that this is *impossible.* We have to discuss this point carefully.

While in short range lattice models the DLR construction gives a clear framework how the class of infinite volume Gibbs measures is to be defined, in mean field models this situation is somewhat ambiguous and needs discussion.

If the infinite volume Gibbs measure is unique (for given ω), quasi by definition, (1) must hold. So our problems arise from non-uniqueness. Hence the following recipe: modify μ_Λ in such a way that uniqueness holds, while otherwise perturbing it in a minimal way. Two procedures suggest themselves:

(i) Tilting, and

(ii) Conditioning

Tilting consists in the addition of a *symmetry breaking* term to the Hamiltonian whose strength is taken to zero. Mostly, this term is taken *linear* so that it has the natural interpretation of a *magnetic field.* More precisely, define

$$\mu^{\{h\}}_{\Lambda,\epsilon}[\omega](\cdot) \equiv \frac{\mu_\Lambda[\omega]\left(\cdot\, e^{-\beta\epsilon\sum_{i\in\Lambda} h_i\sigma_i}\right)}{\mu_\Lambda[\omega]\left(e^{-\beta\epsilon\sum_{i\in\Lambda} h_i\sigma_i}\right)} \tag{2.22}$$

Here h_i is some sequence of numbers that in general will have to be allowed to depend on ω if anything is to be gained. One may also allow them to depend on Λ explicitly, if so desired. From a physical point of view we might wish to add further conditions, like some locality of the ω-dependence; in principle there should be a way of writing them down in some explicit way. We should stress that tilting by linear functions is not always satisfactory, as some states that one might wish to obtain are lost; an example is the generalized Curie-Weiss model with Hamiltonian $H_N(\sigma) = -\frac{N}{4}[m_N(\sigma)]^4$ at the critical point. There, the free energy has three degenerate absolute minima at $-m^*, 0$, and $+m^*$, and while we might want to think of tree coexisting phases, only the measures centered at $\pm m^*$ can be extracted by the above method. Of course this can be remedied by allowing arbitrary perturbation $h(m)$ with the only condition that $\|h\|_\infty$ tends to zero at the end.

By conditioning we mean always conditioning the macroscopic variables to be in some set $\mathcal{A}$. This appears natural since, in lattice models, extremal measures can always be extracted from arbitrary DLR measures by conditioning on events in the tail sigma fields; the macroscopic variables are measurable with respect to the tail sigma fields. Of course only conditioning on events that do not have too small probability will be reasonable. Without going into too much of a motivating discussion, we will adopt the following conventions. Let $\mathcal{A}$ be an event in the sigma algebra generated by the macroscopic function. Put

$$f_{\Lambda,\beta}(\mathcal{A}) = -\frac{1}{\beta|\Lambda|} \ln \mu_{\Lambda,\beta}[\omega](\mathcal{A}) \tag{2.23}$$

We call $\mathcal{A}$ admissible for conditioning if and only if

$$\lim_{|\Lambda|\uparrow\infty} f_{\Lambda,\beta}[\omega](\mathcal{A}) = 0 \tag{2.24}$$

We call $\mathcal{A}$ minimal if it cannot be decomposed into two admissible subsets. In analogy with (2.22) we then define

$$\mu^{\mathcal{A}}_{\Lambda,\beta}[\omega](\cdot) \equiv \mu_{\Lambda,\beta}[\omega]\left(\cdot|\mathcal{A}\right) \tag{2.25}$$

We define the set of all limiting Gibbs measures to be the set of limit points of measures $\mu^{\mathcal{A}}_{\Lambda,\beta}$ with admissible sets $\mathcal{A}$. Choosing $\mathcal{A}$ minimal, we improve our chances of obtaining convergent sequences and the resulting limits are serious candidates for *extremal* limiting Gibbs measures, but we stress that this is not guaranteed to succeed, as will become manifest in our examples. This will not mean that adding such conditioning is not going to be useful. It is in fact, as it will reduce the disorder in the metastate and may in general allow to construct various *different* metastates in the case of phase transitions. The point to be understood here is that within the general framework outlined above, we should consider two different notions of *uniqueness*:

(a) *Strong uniqueness* meaning that for almost all ω there is only one limit point $\mu_\infty[\omega]$, and

(b) *Weak uniqueness*[10] meaning that there is a unique metastate, in the sense that for any choice of $\mathcal{A}$, the metastate constructed taking the infinite volume limit with the measures $\mu^{\mathcal{A}}_{\Lambda,\epsilon}$ is the same.

In fact, it may happen that the addition of a symmetry breaking term or conditioning does not lead to strong uniqueness. Rather, what may be true is that such a field selects a subset of the states, but to which of them the state at given volume resembles can depend on the volume in a complicated way.

If weak uniqueness does not hold, one has a non-trivial set of metastates.

It is quite clear that a sufficiently general tilting approach is equivalent to the conditioning approach; we prefer for technical reasons to use the conditioning in the present paper. We also note that by dropping condition (2.24) one can enlarge the class of limiting measures obtainable to include *metastable states*, which in many applications, in particular in the context of dynamics, are also relevant.

[10] Maybe the notion of meta-uniqueness would be more appropriate

3. Properties of the induced measures.

In this section we collect a number of results on the distribution of the overlap parameters in the Hopfield model that were obtained in some of our previous papers [BG1,BG2,BG3]. We cite these results mostly from [BG3] where they were stated in the most suitable form for our present purposes and we refer the reader to that paper for the proofs.

We recall some notation. Let $m^*(\beta)$ be the largest solution of the mean field equation $m = \tanh(\beta m)$. Note that $m^*(\beta)$ is strictly positive for all $\beta > 1$, $\lim_{\beta\uparrow\infty} m^*(\beta) = 1$, $\lim_{\beta\downarrow 1} \frac{(m^*(\beta))^2}{3(\beta-1)} = 1$ and $m^*(\beta) = 0$ if $\beta \leq 1$. Denoting by e^μ the μ-th unit vector of the canonical basis of $I\!R^M$ we set, for all $(\mu, s) \in \{-1, 1\} \times \{1, \ldots, M(N)\}$,

$$m^{(\mu,s)} \equiv s m^*(\beta) e^\mu, \tag{3.1}$$

and for any $\rho > 0$ we define the balls

$$B_\rho^{(\mu,s)} \equiv \left\{ x \in I\!R^M \middle| \|x - m^{(\mu,s)}\|_2 \leq \rho \right\} \tag{3.2}$$

For any pair of indices (μ, s) and any $\rho > 0$ we define the conditional measures

$$\mu_{N,\beta,\rho}^{(\mu,s)}[\omega](\mathcal{A}) \equiv \mu_{N,\beta}[\omega](\mathcal{A} \mid B_\rho^{(\mu,s)}), \quad \mathcal{A} \in \mathcal{B}(\{-1,1\}^N) \tag{3.3}$$

and the corresponding induced measures

$$\mathcal{Q}_{N,\beta,\rho}^{(\mu,s)}[\omega](\mathcal{A}) \equiv \mathcal{Q}_{N,\beta}[\omega](\mathcal{A} \mid B_\rho^{(\mu,s)}), \quad \mathcal{A} \in \mathcal{B}(I\!R^{M(N)}) \tag{3.4}$$

The point here is that for $\rho \geq c\frac{\sqrt{\alpha}}{m^*(\beta)}$, the sets $B_\rho^{(\mu,s)}$ are admissible in the sense of the last section.

It will be extremely useful to introduce the Hubbard-Stratonovich transformed measures $\widetilde{\mathcal{Q}}_{N,\beta}[\omega]$ which are nothing but the convolutions of the induced measures with a gaussian measure of mean zero and variance $1/\beta N$, i.e.

$$\widetilde{\mathcal{Q}}_{N,\beta}[\omega] \equiv \mathcal{Q}_{N,\beta}[\omega] \star \mathcal{N}(0, \frac{1\!\!1}{\beta N}) \tag{3.5}$$

Similarly we define the conditional Hubbard-Stratonovich transformed measures

$$\widetilde{\mathcal{Q}}_{N,\beta,\rho}^{(\mu,s)}[\omega](\mathcal{A}) \equiv \widetilde{\mathcal{Q}}_{N,\beta}[\omega](\mathcal{A} \mid B_\rho^{(\mu,s)}), \quad \mathcal{A} \in \mathcal{B}(I\!R^{M(N)}) \tag{3.6}$$

We will need to consider the Laplace transforms of these measures which we will denote by[10]

$$\mathcal{L}_{N,\beta,\rho}^{(\mu,s)}[\omega](t) \equiv \int e^{(t,x)} d\mathcal{Q}_{N,\beta,\rho}^{(\mu,s)}[\omega](x), \quad t \in I\!R^{M(N)} \tag{3.7}$$

[10] This notation is slightly different from the one used in [BG3].

and

$$\widetilde{\mathcal{L}}^{(\mu,s)}_{N,\beta,\rho}[\omega](t) \equiv \int e^{(t,x)} d\widetilde{\mathcal{Q}}^{(\mu,s)}_{N,\beta,\rho}[\omega](x)\,, \quad t \in I\!R^{M(N)} \tag{3.8}$$

The following is a simple adaptation of Proposition 2.1 of [BG3] to these notations.

Proposition 3.1: *Assume that $\beta > 1$. There exist finite positive constants $c \equiv c(\beta), \tilde{c} \equiv \tilde{c}(\beta), \bar{c} \equiv \bar{c}(\beta)$ such that, with probability one, for all but a finite number of indices N, if ρ satisfies*

$$\frac{1}{2}m^* > \rho > c(\beta)\left\{\tfrac{1}{N^{1/4}} \wedge \sqrt{\alpha}\right\} \tag{3.9}$$

then, for all t with $\frac{\|t\|_2}{\sqrt{N}} < \infty$,

i)

$$\mathcal{L}^{(\mu,s)}_{\beta,N,\rho}[\omega](t)\left(1 - e^{-\tilde{c}M}\right) \le e^{-\frac{1}{2N\beta}\|t\|_2^2}\widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\rho}[\omega](t) \le e^{-\tilde{c}M} + \mathcal{L}^{(\mu,s)}_{\beta,N,\rho}(t)\left(1 + e^{-\tilde{c}M}\right) \tag{3.10}$$

ii) for any $\rho, \bar{\rho}$ satisfying (3.9)

$$\widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\bar{\rho}}[\omega](t)\left(1 - e^{-\bar{c}M}\right) \le \widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\rho}[\omega](t) \le e^{-\bar{c}M} + \widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\bar{\rho}}[\omega](t)\left(1 + e^{-\bar{c}M}\right) \tag{3.11}$$

iii) for any $\rho, \bar{\rho}$ satisfying (3.9)

$$\left|\left(\int d\mathcal{Q}^{(\mu,s)}_{N,\beta,\rho}[\omega](m)m - \int d\widetilde{\mathcal{Q}}^{(\mu,s)}_{N,\beta,\bar{\rho}}[\omega](z)z\ ,t\right)\right| \le \|t\|_2 e^{-\bar{c}M} \tag{3.12}$$

A closely related result that we will need is also an adaptation of estimates from [BG3], i.e. it is obtained combining Lemmata 3.2 and 3.4 of that paper.

Lemma 3.2: *There exists $\gamma_a > 0$, such that for all $\beta > 1$ and $\sqrt{\alpha} < \gamma_a (m^*)^2$, if $c_0\frac{\sqrt{\alpha}}{m^*} < \rho < m^*/\sqrt{2}$ then, with probability one, for all but a finite number of indices N, for all $\mu \in \{1,\dots,M(N)\}$, $s \in \{-1,1\}$, for all $b > 0$ such that $\rho + b < \sqrt{2}m^*$,*

$$1 \le \frac{\mathcal{Q}_{\beta,N}\left(B^{(\mu,s)}_{\rho+b}\right)}{\mathcal{Q}_{\beta,N}\left(B^{(\mu,s)}_{\rho}\right)} \le 1 + e^{-c_2\beta M} \tag{3.13}$$

where $0 < c_2 < \infty$ is a numerical constant.

We finally recall our result on local convexity of the function Φ.

Theorem 3.3: *Assume that $1 < \beta < \infty$. If the parameters α, β, ρ are such that for $\epsilon > 0$,*

$$\begin{aligned}\inf_{\tau}\Big(&\beta(1 - \tanh^2(\beta m^*(1-\tau)))(1 + 3\sqrt{\alpha}) \\ &+ 2\beta\tanh^2(\beta m^*(1-\tau))\Gamma(\alpha, \tau m^*/\rho)\Big) \le 1 - \epsilon\end{aligned} \tag{3.14}$$

Then with probability one for all but a finite number of indices N, $\Phi_{N,\beta}[\omega](m^* e^1 + v)$ is a twice differentiable and strictly convex function of v on the set $\{v : \|v\|_2 \leq \rho\}$, and

$$\lambda_{min}\left(\nabla^2 \Phi_{N,\beta}[\omega](m^* e^1 + v)\right) > \epsilon \tag{3.15}$$

on this set.

Remark: This theorem was first obtained in [BG1], the above form is cited and proven in [BG2]. With ρ chosen as $\rho = c\frac{\sqrt{\alpha}}{m^*}$, the condition (3.14) means (i) For β close to 1: $\frac{\sqrt{\alpha}}{(m^*)^2}$ small and, (ii) For β large: $\alpha \leq c\beta^{-1}$. The condition on α for large β seems unsatisfactory, but one may easily convince oneself that it cannot be substantially improved.

4. Brascamp-Lieb inequalities.

A basic tool of our analysis are the so-called Brascamp-Lieb inequalities [BL]. In fact, we need such inequalities in a slightly different setting than they are presented in the literature, namely for measures with bounded support on some domain $D \subset \mathbb{R}^M$. Our derivation follows the one given in [H] (see also [HS]), and is in this context almost obvious.

Let $D \subset \mathbb{R}^M$ be a bounded connected domain. Let $V \in C^2(D)$ be a twice continuously differentiable function on D, let $\nabla^2 V$ denote its Hessian matrix and assume that, for all $x \in D$, $\nabla^2 V(x) \geq c > 0$ (where we say that a matrix $A > c$, if and only if for all $v \in R^M$, $(v, Av) \geq c(v,v)$). We define the probability measure ν on $(D, \mathcal{B}(D))$ by

$$\nu(dx) \equiv \frac{e^{-NV(x)} d^M x}{\int_D e^{-NV(x)} d^M x} \tag{4.1}$$

Our central result is

Theorem 4.1: *Let ν the probability measure defined above. Assume that $f, g \in C^1(D)$, and assume that (w.r.g.) $\int_D d\nu(x) g(x) = \int_D d\nu(x) f(x) = 0$. Then*

$$\begin{aligned} \left| \int_D d\nu(x) f(x) g(x) \right| \leq & \frac{1}{cN} \int_D d\nu(x) \, \|\nabla f(x)\|_2 \|\nabla g(x)\|_2 \\ & + \frac{1}{cN} \frac{\int_{\partial D} |g(x)| \, \|\nabla f(x)\|_2 \, e^{-NV(x)} d^{M-1} x}{\int_D e^{-NV(x)} d^M x} \end{aligned} \tag{4.2}$$

where $d^{M-1}x$ is the Lebesgue measure on ∂D.

Proof: We consider the Hilbert space $L^2(D, \mathbb{R}^M, \nu)$ of R^M valued functions on D with scalar product $\langle F, G \rangle \equiv \int_D d\nu(x)(F(x), G(x))$. Let ∇ be the gradient operator on D defined with a domain of all bounded C^1-function that vanish on ∂D. Let ∇^* denote its adjoint. Note that $\nabla^* = -e^{NV(x)} \nabla e^{-NV(x)} = -\nabla + N(\nabla V(x))$. One easily verifies by partial integration that on this domain the operator $\nabla\nabla^* \equiv \nabla e^{NV(x)} \nabla e^{-NV(x)} = \nabla^* \nabla + N \nabla^2 V(x)$ is symmetric and $\nabla^* \nabla \geq 0$, so that by our hypothesis, $\nabla\nabla^* \geq cN > 0$. As a consequence, $\nabla\nabla^*$ has a self-adjoint extension whose inverse $(\nabla\nabla^*)^{-1}$ exists on all $L^2(D, \mathbb{R}^M, \nu)$ and is bounded in norm by $(cN)^{-1}$.

As a consequence of the above, for any $f \in C^1(D)$, we can uniquely solve the differential equation

$$\nabla\nabla^* \, \nabla u = \nabla f \tag{4.3}$$

for ∇u. Now note that (4.3) implies that $\nabla^* \nabla u = f + k$, where k is a constant[11]. Hence for real

[11] Observe that this is only true because D is connected. For D consisting of several connected components the theorem is obviously false.

valued f and g as in the statement of the theorem,

$$\begin{aligned}\int_D d\nu(x)\left(\nabla g(x), \nabla u(x)\right) &= \int_D d\nu(x) e^{NV(x)} \,\mathrm{div}\left(e^{-NV(x)} g \nabla u(x)\right) + \int_D d\nu(x) g(x) \nabla^* \nabla u(x) \\ &= \frac{1}{Z}\int_D d^M x \,\mathrm{div}\left(e^{-NV(x)} g \nabla u(x)\right) + \int_D d\nu(x) g(x) f(x)\end{aligned} \tag{4.4}$$

where $Z \equiv \int_D d^M x\, e^{-NV(x)}$. Therefore, taking into account that $\nabla u = (\nabla\nabla^*)^{-1}\nabla f$,

$$\begin{aligned}\left|\int_D d\nu(x) g(x) f(x)\right| &\leq \left|\int_D d\nu(x)\left(\nabla g(x), (\nabla\nabla^*)^{-1}\nabla f(x)\right)\right| \\ &\quad + \frac{1}{Z}\left|\int_D d^M x \,\mathrm{div}\left(e^{-NV(x)} g \nabla u(x)\right)\right| \\ &\leq \frac{1}{cN}\int_D d\nu(x) \|\nabla g(x)\|_2 \|\nabla f(x)\|_2 \\ &\quad + \frac{1}{cNZ}\int_{\partial D} |g(x)|\, \|\nabla f(x)\|_2\, e^{-NV(x)} d^{M-1}x\end{aligned} \tag{4.5}$$

Note that in second term we used the Gauss-Green formula to convert the integral over a divergence into a surface integral. This concludes the proof.$\diamondsuit$

Remark: As is obvious from the proof above and as was pointed out in [H], one can replace the bound on the lowest eigenvalue of the Hessian of V by a bound on the lowest eigenvalue of the operator $\nabla\nabla^*$. So far we have not seen how to get a better bound on this eigenvalue in our situation, but it may well be that this observation can be a clue to an improvement of our results.

The typical situation where we want to use Theorem 4.1 is the following: Suppose we are given a measure like (4.1) but not on D, but on some bigger domain. We may be able to establish the lower bound on $\nabla^2 V$ not everywhere, but only on the smaller domain D, but such that the measure is essentially concentrated on D anyhow. It is then likely that we can also estimate away the boundary term in (4.2), either because $V(x)$ will be large on ∂D, or because ∂D will be very small (or both). We then have essentially the Brascamp-Lieb inequalities at our disposal.

We mention the following corollary which shows that the Brascamp-Lieb inequalities give rise to concentration inequalities under certain conditions.

Corollary 4.2: *Let ν be as in Lemma 4.3. Assume that $f \in C^1(D)$ and that moreover $V_t(x) \equiv V(x) - tf(x)/N$ for $t \in [0,1]$ is still strictly convex and $\lambda_{min}(\nabla^2 V_t) \geq c' > 0$. Then*

$$\begin{aligned} 0 \leq \ln\int_D d\nu(x) e^{f(x)} - \int_D d\nu(x) f(x) &\leq \frac{1}{2c'N}\sup_{t\in[0,1]}\int_D d\nu_t(x)\|\nabla f\|_2^2 \\ + \sup_{t\in[0,1]}\frac{1}{c'N}\frac{\int_{\partial D}|g(x)|\,\|\nabla f(x)\|_2\, e^{-NV_t(x)} d^{M-1}x}{\int_D e^{-NV_t(x)} d^M x}\end{aligned} \tag{4.6}$$

where ν_t is the corresponding measure with V replaced by V_t.

Proof: Note that

$$\begin{aligned}
\ln I\!E_V e^f &= I\!E_V f + \int_0^1 ds \int_0^s ds' \frac{I\!E_V \left[e^{s'f} \left(f - \frac{I\!E_V e^{s'f} f}{I\!E_V e^{s'f}} \right)^2 \right]}{I\!E_V e^{s'f}} \\
&= I\!E_V f + \int_0^1 ds \int_0^s ds' I\!E_{V_{s'}} \left(f - I\!E_{V_{s'}} f \right)^2
\end{aligned} \tag{4.7}$$

where by assumption $V_s(x)$ has the same properties as V itself. Thus using (4.2) gives (4.7).$\diamondsuit$

Remark: We would like to note that a concentration estimate like Corollary 4.2 can also be derived under slightly different hypothesis on f using logarithmic Sobolev inequalities (see [Le]) which hold under the same hypothesis as Theorem 4.1, and which in fact can be derived as a special case using $f = h^2$ and $g = \ln h^2$ in Theorem 4.1.

In the situations where we will apply the Brascamp-Lieb inequalities, the correction terms due to the finite domain D will be totally irrelevant. This follows from the following simple observation.

Lemma 4.3: *Let B_ρ denote the ball of radius ρ centered at the origin. Assume that for all $x \in D$, $d \geq \nabla^2 V(x) \geq c > 0$. If x^* denotes the unique minimum of V, assume that $\|x^*\|_2 \leq \rho/2$. Then there exists a constant $K < \infty$ (depending only on c and d) such that if $\rho \geq K\sqrt{M/N}$, then for N large enough*

$$\frac{\int_{\partial D} e^{-NV(x)} d^{M-1}x}{\int_D e^{-NV(x)} d^M x} \leq e^{-\rho^2 N/K} \tag{4.8}$$

The proof of this lemma is elementary and will be left to the reader.

5. The convergence of the Gibbs measures.

After these preliminaries we can now come to the central part of the paper, namely the study of the marginal distributions of the Gibbs measures $\mu^{(\mu,s)}_{N,\beta,\rho}$. Without loss of generality it suffices to consider the case $(\mu,s)=(1,1)$, of course. Let us fix $I\subset I\!N$ arbitrary but finite. We assume that $\Lambda\supset I$, and for notational simplicity we put $|\Lambda|=N+|I|$. We are interested in the probabilities

$$\mu^{(1,1)}_{\Lambda,\beta,\rho}[\omega]\left(\{\sigma_I=s_I\}\right)\equiv\frac{I\!E_{\sigma_{\Lambda\setminus I}}e^{\frac12\beta|\Lambda|\left\|m_\Lambda(s_I,\sigma_{\Lambda\setminus I})\right\|_2^2}1\!\!I_{\{m_\Lambda(s_I,\sigma_{\Lambda\setminus I})\in B^{(1,1)}_\rho\}}}{I\!E_{\sigma_I}I\!E_{\sigma_{\Lambda\setminus I}}e^{\frac12\beta|\Lambda|\left\|m_\Lambda(\sigma_I,\sigma_{\Lambda\setminus I})\right\|_2^2}1\!\!I_{\{m_\Lambda(s_I,\sigma_{\Lambda\setminus I})\in B^{(1,1)}_\rho\}}}\tag{5.1}$$

Note that $\|m_I(\sigma)\|_2\leq\sqrt{M}$. Now we can write

$$m_\Lambda(\sigma)=\frac{N}{|\Lambda|}m_{\Lambda\setminus I}(\sigma)+\frac{|I|}{|\Lambda|}m_I(\sigma)\tag{5.2}$$

Then

$$\begin{aligned}1\!\!I_{\{m_\Lambda(s_I,\sigma_{\Lambda\setminus I})\in B^{(1,1)}_\rho\}}&\leq1\!\!I_{\{m_{\Lambda\setminus I}(\sigma)\in B^{(1,1)}_{\rho_+}\}}\\1\!\!I_{\{m_\Lambda(s_I,\sigma_{\Lambda\setminus I})\in B^{(1,1)}_\rho\}}&\geq1\!\!I_{\{m_{\Lambda\setminus I}(\sigma)\in B^{(1,1)}_{\rho_-}\}}\end{aligned}\tag{5.3}$$

where $\rho_\pm\equiv\rho\pm\frac{\sqrt{M}|I|}{N}$. Setting $\beta'\equiv\frac{N}{|\Lambda|}\beta$, this allows us to write

$$\begin{aligned}\mu^{(1,1)}_{\Lambda,\beta,\rho}[\omega]\left(\{\sigma_I=s_I\}\right)&\leq\frac{\int_{B^{(1,1)}_{\rho_+}}d\mathcal{Q}_{\Lambda\setminus I,\beta'}(m)e^{\beta'|I|(m_I(s_I),m)}\,e^{\beta\frac{|I|^2}{2|\Lambda|}\|m_I(s_I)\|_2^2}}{2^{|I|}I\!E_{\sigma_I}\int_{B^{(1,1)}_{\rho_-}}d\mathcal{Q}_{\Lambda\setminus I,\beta'}(m)e^{\beta'|I|(m_I(\sigma_I),m)}\,e^{\beta\frac{|I|^2}{2|\Lambda|}\|m_I(\sigma_I)\|_2^2}}\\&\quad\times\frac{\int_{B^{(1,1)}_{\rho_-}}d\mathcal{Q}_{\Lambda\setminus I,\beta'}(m)}{\int_{B^{(1,1)}_{\rho_+}}d\mathcal{Q}_{\Lambda\setminus I,\beta'}(m)}\\&\leq\frac{\mathcal{L}_{\Lambda/I,\beta,\rho_+}[\omega](\beta'|I|m_I(s_I))\,e^{\beta\frac{|I|^2}{2|\Lambda|}\|m_I(s_I)\|_2^2}}{2^{|I|}I\!E_{\sigma_I}\mathcal{L}_{\Lambda/I,\beta,\rho_-}[\omega](\beta'|I|m_I(\sigma_I))\,e^{\beta\frac{|I|^2}{2|\Lambda|}\|m_I(\sigma_I)\|_2^2}}\frac{\mathcal{Q}_{\Lambda\setminus I,\beta'}\left(B^{(1,1)}_{\rho_+}\right)}{\mathcal{Q}_{\Lambda\setminus I,\beta'}\left(B^{(1,1)}_{\rho_-}\right)}\end{aligned}\tag{5.4}$$

and

$$\begin{aligned}\mu^{(1,1)}_{\Lambda,\beta,\rho}[\omega]\left(\{\sigma_I=s_I\}\right)&\geq\frac{\int_{B^{(1,1)}_{\rho_-}}d\mathcal{Q}_{\Lambda\setminus I,\beta'}(m)e^{\beta'|I|(m_I(s_I),m)}\,e^{\beta\frac{|I|^2}{2|\Lambda|}\|m_I(s_I)\|_2^2}}{2^{|I|}I\!E_{\sigma_I}\int_{B^{(1,1)}_{\rho_+}}d\mathcal{Q}_{\Lambda\setminus I,\beta'}(m)e^{\beta'|I|(m_I(\sigma_I),m)}\,e^{\beta\frac{|I|^2}{2|\Lambda|}\|m_I(\sigma_I)\|_2^2}}\\&\quad\times\frac{\mathcal{Q}_{\Lambda\setminus I,\beta'}\left(B^{(1,1)}_{\rho_-}\right)}{\mathcal{Q}_{\Lambda\setminus I,\beta'}\left(B^{(1,1)}_{\rho_+}\right)}\\&=\frac{\mathcal{L}_{\Lambda/I,\beta,\rho_-}[\omega](\beta'|I|m_I(s_I))\,e^{\beta\frac{|I|^2}{2|\Lambda|}\|m_I(s_I)\|_2^2}}{2^{|I|}I\!E_{\sigma_I}\mathcal{L}_{\Lambda/I,\beta,\rho_+}[\omega](\beta'|I|m_I(\sigma_I))\,e^{\beta\frac{|I|^2}{2|\Lambda|}\|m_I(\sigma_I)\|_2^2}}\frac{\mathcal{Q}_{\Lambda\setminus I,\beta'}\left(B^{(1,1)}_{\rho_-}\right)}{\mathcal{Q}_{\Lambda\setminus I,\beta'}\left(B^{(1,1)}_{\rho_+}\right)}\end{aligned}\tag{5.5}$$

Now the term $\frac{|I|^2}{N}\|m_I(s)\|_2^2$ is, up to a constant that is independent of the s_i, irrelevantly small. More precisely, we have that

Lemma 5.1: *There exist $\infty > C, c > 0$ such that for all I, M, and for all $x > 0$,*

$$
\begin{aligned}
&I\!P\left[\sup_{\sigma_I\in\{-1,1\}^I} \frac{|I|^2}{N}\left|\|m_I(s)\|_2^2 - \frac{M|I|}{N}\right| \geq \frac{|I|M}{N}\left(\sqrt{\frac{|I|}{N}} + x\right)\right]\\
&\leq C\exp\left(-cM\left(\sqrt{1+x}-1\right)^2\right)
\end{aligned}
\tag{5.6}
$$

Proof: This Lemma is a direct consequence of estimates on the norm of the random matrices obtained, e.g. in Theorem 4.1 of [BG6].$\diamondsuit$

Together with Proposition 3.1 and Lemma 3.2, we can now extract the desired representation for our probabilities.

Lemma 5.2: *For all $\beta > 1$ and $\sqrt{\alpha} < \gamma_a(m^*)^2$, if $c_0\frac{\sqrt{\alpha}}{m^*} < \rho < m^*/\sqrt{2}$ then, with probability one, for all but a finite number of indices N, for all $\mu \in \{1,\dots,M(N)\}$, $s \in \{-1,1\}$,*

(i)

$$
\begin{aligned}
\mu^{(1,1)}_{\Lambda,\beta,\rho}[\omega]\left(\{\sigma_I = s_I\}\right) = &\frac{\mathcal{L}^{(1,1)}_{\Lambda/I,\beta,\rho}[\omega](\beta'|I|m_I(s_I))}{2^{|I|}I\!E_{\sigma_I}\mathcal{L}^{(1,1)}_{\Lambda/I,\beta,\rho}[\omega](\beta'|I|m_I(\sigma_I))}\\
&+O(N^{-1/4})
\end{aligned}
\tag{5.7}
$$

and alternatively

(ii)

$$
\begin{aligned}
\mu^{(1,1)}_{\Lambda,\beta,\rho}[\omega]\left(\{\sigma_I = s_I\}\right) = &\frac{\widetilde{\mathcal{L}}^{(1,1)}_{\Lambda/I,\beta,\rho}[\omega](\beta'|I|m_I(s_I))}{2^{|I|}I\!E_{\sigma_I}\widetilde{\mathcal{L}}^{(1,1)}_{\Lambda/I,\beta,\rho}[\omega](\beta'|I|m_I(\sigma_I))}\\
&+O\left(e^{-O(M)}\right)
\end{aligned}
\tag{5.8}
$$

We leave the details of the proof to the reader. We see that the computation of the marginal distribution of the Gibbs measures requires nothing but the computation of the Laplace transforms of the induced measures or its Hubbard-Stratonovich transform at the random points $t = \sum_{i\in I} s_i\xi_i$. Alternatively, these can be seen as the Laplace transforms of the distribution of the random variables (ξ_i, m).

Now it is physically very natural that the law of the random variables (ξ_i, m) should determine the Gibbs measures completely. The point is that in a mean field model, the distribution of the spins in a finite set I is determined entirely in terms of the effective mean fields produced by the rest of the system that act on the spins σ_i. These fields are precisely the (ξ_i, m). In a "normal" mean field situation, the mean fields are constant almost surely with respect to the Gibbs measure. In the Hopfield model with subextensively many patterns, this will also be true, as m will be concentrated near one of the values $m^* e^\mu$ (see [BGP1]). In that case (ξ_i, m) will depend only in a local and very explicit form on the disorder, and the Gibbs measures will inherit this property. In a more general situation, the local mean fields may have a more complicated distribution, in particular they may not be constant under the Gibbs measure, and the question is how to determine this. The approach of the *cavity method* (see e.g. [MPV]) as carried out by Talagrand [T1] consists in deriving this distribution by induction over the volume. [PST] also followed this approach, using however the assumption of "self-averaging" of the order parameter to control errors. Our approach consists in using the detailed knowledge obtained on the measures $\widetilde{\mathcal{Q}}$, and in particular the local convexity to determine a priori the form of the distribution; induction will then only be used to determine the remaining few parameters.

Let us begin with some general preparatory steps which will not yet require special properties of our measures. To simplify the notation, we we introduce the following abbreviations:

We write $\mathbb{E}_{\Phi_N}$ for the expectation with respect to the measures $\widetilde{\mathcal{Q}}_{\Lambda\setminus I,\beta,h}[\omega]$ conditioned on B_ρ and we set $\bar{Z} \equiv Z - \mathbb{E}_{\Phi_N} Z$. We will write $\mathbb{E}_{\xi_I}$ for the expectation with respect to the family of random variables ξ_i^μ, $i \in I$, $\mu = 1, \dots, M$.

The first step in the computation of our Laplace transform consists in centering, i.e. we write

$$\mathbb{E}_{\Phi_N} e^{\sum_{i\in I} \beta s_i (\xi_i, Z)} = e^{\sum_{i\in I} \beta s_i (\xi_i, \mathbb{E}_{\Phi_N} Z)} \mathbb{E}_{\Phi_N} e^{\sum_{i\in I} \beta s_i (\xi_i, \bar{Z})} \tag{5.9}$$

While the first factor will be entirely responsible for the for the distribution of the spins, our main efforts have to go into controlling the second. To do this we will use heavily the fact, established first in [BG1], that on $B_\rho^{(1,1)}$ the function Φ is convex with probability close to one. This allows us to exploit the Brascamp-Lieb inequalities in the form given in Section 3. The advantage of this procedure is that it allows us to identify immediately the leading terms and to get a priori estimates on the errors. This is to be contrasted to the much more involved procedure of Talagrand [T1] who controls the errors by induction.

General Assumption: For the remainder of this paper we will always assume that the parameters α and β of our model are such that the hypotheses of Proposition 3.1 and Theorem 3.3 are satisfied. All lemmata, propositions and theorem are valid under this provision only.

Lemma 5.3: *Under our general assumption,*

(i)

$$\mathbb{E}_{\xi_I} \mathbb{E}_{\Phi_N} e^{\sum_{i\in I} \beta s_i (\xi_i, \bar{Z})} = e^{\frac{\beta^2}{2} \sum_{i\in I} s_i^2 \mathbb{E}_{\Phi_N} \|\bar{Z}\|_2^2} \times e^{O(1/(\epsilon N))} \tag{5.10}$$

(ii) There is a finite constant C such that

$$\mathbb{E}_{\xi_I}\left[\ln\left(\frac{\mathbb{E}_{\Phi_N} e^{\sum_{i\in I}\beta s_i(\xi_i,\bar{Z})}}{\mathbb{E}_{\xi_I}\mathbb{E}_{\Phi_N} e^{\sum_{i\in I}\beta s_i(\xi_i,\bar{Z})}}\right)\right]^2 \leq \frac{C}{N} \tag{5.11}$$

Remark: The immediate consequence of this lemma is the observation that the family of random variables $\left\{(\xi_i,\bar{Z})\right\}_{i\in I}$ is asymptotically close to a family of i.i.d. centered gaussian random variables with variance $U_N \equiv \mathbb{E}_{\Phi_N}\|\bar{Z}\|_2^2$. U_N will be seen to be one of the essential parameters that we will need to control by induction. Note that for the moment, we cannot say whether the law of the $(\xi_i,\bar{Z})$ converges in any sense, as it is not a priori clear whether U_N will converge as $N\uparrow\infty$, although this would be a natural guess. Note that as far as the computation of the marginal probabilities of the Gibbs measures is concerned, this question is, however, completely irrelevant, in as far as this term is an even function of the s_i.

Remark: It follows from Lemma 5.3 that

$$\ln \mathbb{E}_{\Phi_N}\exp\left(\sum_{i\in I}\beta s_i(\xi_i,\bar{Z})\right) = \frac{\beta^2}{2}|I|\mathbb{E}_{\Phi_N}\|\bar{Z}\|_2^2 + O\left(\tfrac{1}{\epsilon N}\right) + R_N \tag{5.12}$$

where

$$\mathbb{E}_{\xi_I} R_N^2 \leq \frac{C}{N} \tag{5.13}$$

Proof: The proof of this Lemma relies heavily on the use of the Brascamp-Lieb inequalities, Theorem 4.1, which are applicable due to our assumptions and Theorem 3.3. It was given in [BG1] for I being a single site, and we repeat the main steps. First note that

$$\begin{aligned}
\mathbb{E}_{\xi_I}\mathbb{E}_{\Phi_N} e^{\sum_{i\in I}\beta s_i(\xi_i,\bar{Z})} &\leq \mathbb{E}_{\Phi_N} e^{\frac{\beta^2}{2}\sum_{i\in I} s_i^2\|\bar{Z}\|_2^2}\\
\mathbb{E}_{\xi_I}\mathbb{E}_{\Phi_N} e^{\sum_{i\in I}\beta s_i(\xi_i,\bar{Z})} &\geq \mathbb{E}_{\Phi_N} e^{\frac{\beta^2}{2}\sum_{i\in I} s_i^2\|\bar{Z}\|_2^2 - \frac{\beta^4}{4}\sum_{i\in I} s_i^4\|\bar{Z}\|_4^4}
\end{aligned} \tag{5.14}$$

Note first that if the smallest eigenvalue of $\nabla^2\Phi \geq \epsilon$, then the Brascamp-Lieb inequalities Theorem 4.1 yield

$$\mathbb{E}_{\Phi_N}\|\bar{Z}\|_2^2 \leq \frac{M}{\epsilon N} + O(e^{-\rho^2 N/K}) \tag{5.15}$$

and by iterated application

$$\mathbb{E}_{\Phi_N}\|\bar{Z}\|_4^4 \leq 4\frac{M}{\epsilon^2 N^2} + O(e^{-\rho^2 N/K}) \tag{5.16}$$

In the bounds (5.14) we now use Corollary 4.2 with f given by $\beta^2|I|/2\|\bar{Z}\|_2^2$, respectively by $\beta^2|I|/2\|\bar{Z}\|_2^2 - \beta^4|I|/4\|\bar{Z}\|_4^4$ to first move the expectation into the exponent, and then (5.15) and

(5.16) (applied to the slightly modified measures $\mathbb{E}_{\Phi_{N-tf/N}}$, which still retain the same convexity properties) to the terms in the exponent. This gives (5.10).

By very similar computations one shows first that

$$\mathbb{E}\left(\mathbb{E}_{\Phi_N} e^{\sum_{i\in I}\beta s_i(\xi_i,\bar{Z})} - \mathbb{E}_{\xi_I}\mathbb{E}_{\Phi_N} e^{\sum_{i\in I}\beta s_i(\xi_i,\bar{Z})}\right) \leq \frac{C}{N} \tag{5.17}$$

Moreover, using again Corollary 4.2, one obtains that (on the subspace $\bar{\Omega}$ where convexity holds)

$$e^{-\beta^2|I|/2\frac{\alpha}{\epsilon}} \leq \mathbb{E}_{\Phi_N} e^{\sum_{i\in I}\beta s_i(\xi_i,\bar{Z})} e^{-\beta^2|I|/2\frac{\alpha}{\epsilon}} \tag{5.18}$$

These bounds, together with the obvious Lipshitz continuity of the logarithm away from zero yield (5.11). $\diamondsuit$

Remark: The above proof follows ideas of the proof of Lemma 4.1 on [T1]. The main difference is that the systematic use of the Brascamp-Lieb inequalities that allows us to avoid the appearance of uncontrolled error terms.

We now turn to the mean values of the random variables $(\xi_i, \mathbb{E}_{\Phi_N} Z)$. These are obviously random variables with mean value zero and variance $\|\mathbb{E}_{\Phi_N} Z\|_2$. Moreover, the variables $(\xi_i, \mathbb{E}_{\Phi_N} Z)$ and $(\xi_j, \mathbb{E}_{\Phi_N} Z)$ are uncorrelated for $i \neq j$. Now $\mathbb{E}_{\Phi_N} Z$ has one macroscopic component, namely the first one, while all others are expected to be small. It is thus natural to expect that these variables will actually converge to a sum of a Bernoulli variable $\xi_i^1 \mathbb{E}_{\Phi_N} Z_1$ plus independent gaussians with variance $T_N \equiv \sum_{\mu=2}^M [\mathbb{E}_{\Phi_N} Z_\mu]^2$, but it is far from trivial to prove this. It requires in particular at least to show that T_N converges.

We will first prove the following proposition:

Proposition 5.4: *In addition to our general assumption, assume that* $\liminf_{N\uparrow\infty} N^{1/4} T_N = +\infty$, *a.s.. For* $i \in I$, *set* $X_i(N) \equiv \frac{1}{\sqrt{T_N}} \sum_{\mu=2} \xi_i^\mu \mathbb{E}_{\Phi_N} Z_\mu$. *Then this family converges to a family of i.i.d. standard normal random variables.*

Remark: The assumption on the divergence of $N^{1/4} T_N$ is harmless. We will see later that it is certainly verified provided $\liminf_{N\uparrow\infty} N^{1/8} \mathbb{E} T_N = +\infty$. Recall that our final goal is to approximate (in law) $\sum_{\mu=2}^M \xi_i^\mu \mathbb{E}_{\Phi_N} Z_\mu$ by $\sqrt{T_N} g_i$, where g_i is gaussian. So if $T_N \leq N^{-1/4}$, then $\sum_{\mu=2}^M \xi_i^\mu \mathbb{E}_{\Phi_N} Z_\mu$ is close to zero (in law) anyway, as is $\sqrt{T_N} g_i$, and no harm is done if we exchange the two. We will see that this situation only arises in fact if M/N tends to zero rapidly, in which case all this machinery is not needed.

Proof: To prove such a result requires essentially to show that $\mathbb{E}_{\Phi_N} Z_\mu$ for all $\mu \geq 2$ tend to zero as $N \uparrow \infty$. We note first that by symmetry, for all $\mu \geq 2$, $\mathbb{E}\mathbb{E}_{\Phi_N} Z_\mu = \mathbb{E}\mathbb{E}_{\Phi_N} Z_2$. On the other hand,

$$\sum_{\mu=2}^{M}[\mathbb{E}\mathbb{E}_{\Phi_N} Z_\mu]^2 \le \mathbb{E}\sum_{\mu=2}^{M}[\mathbb{E}_{\Phi_N} Z_\mu]^2 \le \rho^2 \tag{5.19}$$

so that $|\mathbb{E}\mathbb{E}_{\Phi_N} Z_\mu| \le \rho M^{-1/2}$.

To derive from this a probabilistic bound on $\mathbb{E}_{\Phi_N} Z_\mu$ itself we will use concentration of measure estimates. To do so we need the following lemma:

Lemma 5.5: *Assume that $f(x)$ is a random function defined on some open neighborhood $U \subset \mathbb{R}$. Assume that f verifies for all $x \in U$ that for all $0 \le r \le 1$,*

$$\mathbb{P}\left[|f(x) - \mathbb{E}f(x)| > r\right] \le c\exp\left(-\frac{Nr^2}{c}\right) \tag{5.20}$$

and that, at least with probability $1-p$, $|f'(x)| \le C$, $|f''(x)| \le C < \infty$ both hold uniformly in U. Then, for any $0 < \zeta \le 1/2$, and for any $0 < \delta < N^{\zeta/2}$,

$$\mathbb{P}\left[|f'(x) - \mathbb{E}f'(x)| > \delta N^{-\zeta/2}\right] \le \frac{32C^2}{\delta^2}N^{\zeta}\exp\left(-\frac{\delta^4 N^{1-2\zeta}}{256c}\right) + p \tag{5.21}$$

Proof: Let us assume that $|U| \le 1$. We may first assume that the boundedness conditions for the derivatives of f hold uniformly; by standard arguments one shows that if they only hold with probability $1-p$, the effect is nothing more than the final summand p in (5.21). The first step in the proof consists in showing that (5.20) together with the boundedness of the derivative of f implies that $f(x) - \mathbb{E}f(x)$ is uniformly small. To see this introduce a grid of spacing ϵ, i.e. let $U_\epsilon = U \cap \epsilon\mathbb{Z}$. Clearly

$$\begin{aligned}
&\mathbb{P}\left[\sup_{x\in U}|f(x) - \mathbb{E}f(x)| > r\right] \\
&\le \mathbb{P}\left[\sup_{x\in U_\epsilon}|f(x) - \mathbb{E}f(x)| \right.\\
&\qquad \left. + \sup_{x,y:|x-y|\le\epsilon}|f(x) - f(y)| + |\mathbb{E}f(x) - \mathbb{E}f(y)| > r\right] \\
&\le \mathbb{P}\left[\sup_{x\in U_\epsilon}|f(x) - \mathbb{E}f(x)| > r - 2C\epsilon\right] \\
&\le \epsilon^{-1}\mathbb{P}\left[|f(x) - \mathbb{E}f(x)| > r - 2C\epsilon\right]
\end{aligned} \tag{5.22}$$

If we choose $\epsilon = \frac{r}{4C}$, this yields

$$\mathbb{P}\left[\sup_{x\in U}|f(x) - \mathbb{E}f(x)| > r\right] \le \frac{4C}{r}\exp\left(-\frac{Nr^2}{4c}\right) \tag{5.23}$$

Next we show that *if* $\sup_{x\in U}|f(x)-g(x)|\leq r$ for two functions f,g with bounded second derivative, then

$$|f'(x)-g'(x)|\leq\sqrt{8Cr} \tag{5.24}$$

For notice that

$$\left|\frac{1}{\epsilon}[f(x+\epsilon)-f(x)]-f'(x)\right|\leq\frac{\epsilon}{2}\sup_{x\leq y\leq x+\epsilon}f''(y)\leq C\frac{\epsilon}{2} \tag{5.25}$$

so that

$$\begin{aligned}|f'(x)-g'(x)|&\leq\frac{1}{\epsilon}|f(x+\epsilon)-g(x+\epsilon)-f(x)+g(x)|+C\epsilon\\&\leq\frac{2r}{\epsilon}+C\epsilon\end{aligned} \tag{5.26}$$

Choosing the optimal $\epsilon=\sqrt{2r/C}$ gives (5.24). It suffices to combine (5.24) with (5.23) to get

$$I\!P\left[|f'(x)-I\!Ef'(x)|>\sqrt{8rC}\right]\leq\frac{4C}{r}\exp\left(-\frac{Nr^2}{4c}\right) \tag{5.27}$$

Setting $r=\frac{\delta^2}{CN^\zeta}$, we arrive at (5.21). $\diamondsuit$

We will now use Lemma 5.5 to control $I\!E_{\Phi_N}Z_\mu$. We define

$$f(x)=\frac{1}{\beta N}\ln\int_{B_\rho^{(1,1)}}d^Mze^{\beta Nxz_\mu}e^{-\beta N\Phi_{\beta,N,M}(z)} \tag{5.28}$$

and denote by $I\!E_{\Phi_N,x}$ the corresponding modified expectation. As has by now been shown many times [T1,BG1], $f(x)$ verifies (5.20). Moreover, $f'(x)=I\!E_{\Phi_N,x}Z_\mu$ and

$$f''(x)=\beta NI\!E_{\Phi_N,x}\left(Z_\mu-I\!E_{\Phi_N,x}Z_\mu\right)^2 \tag{5.29}$$

Of course the addition of the linear term to Φ does not change its second derivative, so that we can apply the Brascamp-Lieb inequalities also to the measure $I\!E_{\Phi_N,x}$. This shows that

$$I\!E_{\Phi_N,x}\left(Z_\mu-I\!E_{\Phi_N,x}Z_\mu\right)^2\leq\frac{1}{\epsilon N\beta} \tag{5.30}$$

which means that $f(x)$ has a second derivative bounded by $c=\frac{1}{\epsilon}$.

This gives the

Corollary 5.6: *There are finite positive constants c,C such that, for any $0<\zeta\leq\frac{1}{2}$, for any μ,*

$$I\!P\left[|I\!E_{\Phi_N}Z_\mu-I\!EI\!E_{\Phi_N}Z_\mu|\geq N^{-\zeta/2}\right]\leq CN^\zeta\exp\left(-\frac{N^{1-2\zeta}}{c}\right) \tag{5.31}$$

We are now ready to conclude the proof of our proposition. We may choose e.g. $\zeta = 1/4$ and denote by Ω_N the subset of Ω where, for all μ, $|I\!\!E_{\Phi_N} Z_\mu - I\!\!E I\!\!E_{\Phi_N} Z_\mu| \leq N^{-1/8}$. Then $I\!\!P[\Omega_N^c] \leq O\left(e^{-N^{1/2}}\right)$.

We will prove the proposition by showing convergence of the characteristic function to that of product standard normal distributions, i.e. we show that for any $t \in I\!\!R^I$, $I\!\!E \prod_{j\in I} e^{it_j X_j(N)}$ converges to $\prod_{j\in I} e^{-\frac{1}{2}t_j^2}$. We have

$$
\begin{aligned}
I\!\!E \prod_{j\in I} e^{it_j X_j(N)} &= I\!\!E_{\xi_{I^c}} \left[1\!\!1_{\Omega_N} I\!\!E_{\xi_I} e^{i\sum_{j\in I} t_j X_j(N)} + 1\!\!1_{\Omega_N^c} I\!\!E_{\xi_I} e^{i\sum_{j\in I} t_j X_j(N)} \right] \\
&= I\!\!E_{\xi_{I^c}} \left[1\!\!1_{\Omega_N} \prod_{\mu\geq 2} \prod_{j\in I} \cos\left(\frac{t_j}{\sqrt{T_N}} I\!\!E_{\Phi_N} Z_\mu \right) \right] + O\left(e^{-N^{1/2}}\right)
\end{aligned}
\tag{5.32}
$$

Thus the second term tends to zero rapidly and can be forgotten. On the other hand, on Ω_N,

$$
\sum_{\mu=2}^{M} (I\!\!E_{\Phi_N} Z_\mu)^4 \leq N^{-1/4} \sum_{\mu=2}^{M} (I\!\!E_{\Phi_N} Z_\mu)^2 \leq N^{-1/4} T_N \tag{5.33}
$$

Moreover, for any finite t_j, for N large enough, $\left| \frac{t_j}{\sqrt{T_N}} I\!\!E_{\Phi_N} Z_\mu \right| \leq 1$. Thus, using that $|\ln\cos x - x^2/2| \leq cx^4$ for $|x| \leq 1$, and that

$$
\begin{aligned}
&I\!\!E_{\xi_{I^c}} 1\!\!1_{\Omega_N} I\!\!E_\eta e^{i\sum_{j\in I} t_j X_j(N)} \\
&\leq e^{-\sum_{j\in I} t_j^2/2} \sup_{\Omega_N} \left[\prod_{j\in I} \exp\left(c \frac{t_j^4 N^{-1/4}}{T_N} \right) \right] I\!\!P_\xi(\Omega_N)
\end{aligned}
\tag{5.34}
$$

Clearly, the right hand side converges to $e^{-\sum_{j\in I} t_j^2/2}$, provided only that $N^{1/4} T_N \uparrow \infty$. Since this was assumed, the Proposition is proven. $\diamondsuit$

We now control the convergence of our Laplace transform except for the three parameters $m_1(N) \equiv I\!\!E_{\Phi_N} Z_1$, $T_N \equiv \sum_{\mu=2}^{M} \left[I\!\!E_{\Phi_N} Z_\mu\right]^2$ and $U_N \equiv I\!\!E_{\Phi_N} \|\bar{Z}\|_2^2$. What we have to show is that these quantities converge almost surely and that the limits satisfy the equations of the replica symmetric solution of Amit, Gutfreund and Sompolinsky [AGS].

While the issue of convergence is crucial, the technical intricacies of its proof are largely disconnected to the question of the convergence of the Gibbs measures. We will therefore assume for the moment that these quantities do converge to some limits and draw the conclusions for the Gibbs measures from the results of this section under this assumption (which will later be proven to hold).

Indeed, collecting from Lemma 5.3 (see the remark following that lemma) and Proposition 5.4, we can write

$$
\mu_{\Lambda,\beta,\rho}^{(1,1)}[\omega](\{\sigma_I = s_I\}) = \frac{e^{\beta'_N \sum_{i\in I} s_i \left[m_1(N)\xi_i^1 + X_i(N)\sqrt{T_N}\right] + R_N(s_I)}}{2^I I\!\!E_{\sigma_I} e^{\beta'_N \sum_{i\in I} \sigma_i \left[m_1(N)\xi_i^1 + X_i(N)\sqrt{T_N}\right] + R_N(\sigma_I)}} \tag{5.35}
$$

where

$$\begin{aligned}
&\beta'_N \to \beta \\
&R_N(s_I) \to 0 \qquad \text{in Probability} \\
&X_i(N) \to g_i \qquad \text{in law} \\
&T_N \to \alpha r \qquad \text{a.s.} \\
&m_1(N) \to m_1 \qquad \text{a.s.}
\end{aligned}$$

for some numbers r, m_1 and there $\{g_i\}_{i\in I\!N}$ is a family of i.i.d. standard gaussian random variables.

Putting this together we get that

Proposition 5.7: *In addition to our general assumptions, assume that $T_N \to \alpha r$, a.s. and $m_1(N) \to m_1$, a.s. Then, for any finite $I \subset I\!N$*

$$\mu^{(1,1)}_{\Lambda,\beta,\rho}\left(\{\sigma_I = s_I\}\right) \to \prod_{i\in I} \frac{e^{\beta s_i\left[m_1\bar\xi^1_i + g_i\sqrt{\alpha r}\right]}}{2\cosh\left(\beta\sigma_i\left[m_1\bar\xi^1_i + g_i\sqrt{\alpha r}\right]\right)} \tag{5.36}$$

where the convergence holds in law with respect to the measure $I\!P$, and $\{g_i\}_{\in\in I\!N}$ is a family of i.i.d. standard normal random variables and $\{\bar\xi^1_i\}_{i\in I\!N}$ are independent Bernoulli random variables, independent of the g_i and having the same distribution as the variables ξ^1_i.

To arrive at the convergence in law of the random Gibbs measures, it is enough to show that (5.36) holds jointly for any finite family of cylinder sets, $\{\sigma_i = s_i, \forall_{i\in I_k}\}, I_k \subset I\!N$, $k = 1,\dots,\ell$ (C.f. [Ka], Theorem 4.2). But this is easily seen to hold from the same arguments. Therefore, denoting by $\mu^{(1,1)}_{\infty,\beta}$ the random measure

$$\mu^{(1,1)}_{\infty,\beta}[\omega](\sigma) \equiv \prod_{i\in I\!N} \frac{e^{\beta\sigma_i[m_1\xi^1_i[\omega] + \sqrt{\alpha r}g_i[\omega]]}}{2\cosh\left(\beta[m_1\xi^1_i[\omega] + \sqrt{\alpha r}g_i[\omega]]\right)} \tag{5.37}$$

we have

Theorem 5.8: *Under the assumptions of Proposition 5.7, and with the same notation,*

$$\mu^{(1,1)}_{\Lambda,\beta,\rho} \to \mu^{(1,1)}_{\infty,\beta}, \quad \textit{in law, as } \Lambda \uparrow \infty \quad , \tag{5.38}$$

This result can easily be extended to the language of metastates. The following Theorem gives an explicit representation of the Aizenman-Wehr metastate in our situation:

Theorem 5.9: *Let $\kappa_\beta(\cdot)[\omega]$ denote the Aizenman-Wehr metastate. Under the hypothesis of Proposition 5.7, for almost all ω, for any continuous function $F : I\!R^k \to I\!R$, and cylinder functions*

f_i on $\{-1,1\}^{I_i}$, $i=1,\dots,k$, one has

$$\begin{aligned}\int_{\mathcal{M}_1(\mathcal{S}_\infty)} &\kappa_\beta(d\mu)[\omega] F\left(\mu(f_1),\dots,\mu(f_k)\right)\\ =\int \prod_{i\in I} d\mathcal{N}(g_i) F&\Biggl(\mathbb{E}_{s_{I_1}} f_i(s_{I_1}) \prod_{i\in I_1} \frac{e^{\beta\left[\sqrt{\alpha r} g_i + m_1 \xi_i^1[\omega]\right]}}{2\cosh\left(\sqrt{\alpha r} g_i + m_1\xi_i^1[\omega]\right)},\dots\\ &\dots,\mathbb{E}_{s_{I_k}} f_k(s_{I_k}) \prod_{i\in I_k} \frac{e^{\beta\left[\sqrt{\alpha r} g_i + m_1 \xi_i^1[\omega]\right]}}{2\cosh\left(\sqrt{\alpha r} g_i + m_1\xi_i^1[\omega]\right)}\Biggr)\end{aligned} \tag{5.39}$$

where $\mathcal{N}$ denotes the standard normal distribution.

Remark: Modulo the convergence assumptions, that will be shown to hold in the next section, Theorem 5.9 is the precise statement of Theorem 1.1. Note that the only difference from Theorem 5.8 is that the variables ξ_i^1 that appear here on the right hand side are now the same as those on the left hand side.

Proof: This theorem is proven just as Theorem 5.8, except that the "almost sure version" of the central limit theorem, Proposition 5.4, which in turn is proven just as Lemma 2.1, is used. The details are left to the reader.$\diamondsuit$

Remark: Our conditions on the parameters α and β place us in the regime where, according to [AGS] the "replica symmetry" is expected to hold. This is in nice agreement with the remark in [NS4] where replica symmetry is linked to the fact that the metastate is concentrated on product measures.

Remark: One would be tempted to exploit also the other notions of "metastate" explained in Section 2. We see that the key to these constructions would be an invariance principle associated to the central limit theorem given in Proposition 5.4. However, there are a number of difficulties that so far have prevented us from proving such a result. We would have to study the random process

$$X_i^t(N) \equiv \sum_{\mu=2}^{M(tN)} \xi_i^\mu \mathbb{E}_{\Phi_{tN}} Z_\mu \tag{5.40}$$

(suitably interpolated for t that are not integer multiples of $1/N$). If this process was to converge to Brownian motion, its increments should converge to independent Gaussians with suitable variance. But

$$\begin{aligned}X_i^t(N) - X_i^s(N) = &\sum_{\mu=M(sN)}^{M(tN)} \xi_i^\mu \mathbb{E}_{\Phi_{tN}} Z_\mu\\ &+\sum_{\mu=2}^{M(sN)} \xi_i^\mu \left(\mathbb{E}_{\Phi_{tN}} Z_\mu - \mathbb{E}_{\Phi_{sN}} Z_\mu\right)\end{aligned} \tag{5.41}$$

The first term on the right indeed has the desired properties, as is not too hard to check, but the second term is hard to control.

To get some idea of the nature of this process, we recall from [BG1,BG2] that $\mathbb{E}_{\Phi_N} Z$ is approximately given by $c(\beta)\frac{1}{N}\sum_{j\in\Lambda\setminus I}\xi_j$ (in the sense that the ℓ_2 distance between the two vectors is of order $\sqrt{\alpha}$ at most). Let us for simplicity consider only the case $I=\{0\}$. If we replace $\mathbb{E}_{\Phi_N} Z$ by this approximation, we are led to study the process

$$Y^t(N) \equiv \frac{1}{t}\sum_{\mu=2}^{\alpha tN} \xi_0^\mu \frac{1}{N}\sum_{i=1}^{tN} \xi_i^\mu \tag{5.42}$$

for $tN, \alpha tN$ integer and linearly interpolated otherwise.

Proposition 5.10: *The sequence of processes $Y^t(N)$ defined by (5.42) converges weakly to the gaussian process $t^{-1}B_{\alpha t^2}$, where B_s is a standard Brownian motion.*

Proof: Notice that $\xi_0^\mu \xi_i^\mu$ has the same distribution as ξ_i^μ, and therefore $Y^t(N)$ has the same distribution as

$$\widetilde{Y}^t(N) \equiv \frac{1}{tN}\sum_{\mu=2}^{\alpha tN}\sum_{i=1}^{tN} \xi_i^\mu \tag{5.43}$$

for which the convergence to $B_{\alpha t^2}$ follows immediately from Donsker's theorem. $\diamondsuit$

At present we do not see how to extend this result to the real process of interest, but at least we can expect that some process of this type will emerge.

As a final remark we investigate what would happen if we adopted the "standard" notion of limiting Gibbs measures as weak limit points along possibly random subsequences. The answer is the following

Proposition 5.10: *Under the assumptions of Proposition 5.7, for any finite $I\subset \mathbb{N}$, for any $x\in \mathbb{R}^I$, for $\mathbb{P}$-almost all ω, there exist sequences $N_k[\omega]$ tending to infinity such that for any $s_I\in\{-1,1\}^I$*

$$\begin{aligned}&\lim_{k\uparrow\infty}\mu^{(1,1)}_{N_k,\beta}[\omega](\{\sigma_I=s_I\})\\ &\quad=\prod_{i\in I}\frac{e^{\beta s_i[m_1\xi_i^1[\omega]+\sqrt{\alpha r}x_i]}}{2\cosh(\beta[m_1\xi_i^1[\omega]+\sqrt{\alpha r}x_i])}\end{aligned}\tag{5.44}$$

Proof: To simplify the notation we will write the proof only for the case $i=\{0\}$. The general case differs only in notation. It is clear that we must show that for almost all ω there exist subsequences $N_k[\omega]$ such that $X_0(N_k)[\omega]$ converges to x, for any chosen value x. Since by assumption T_N converges almost surely to αr, it is actually enough to show that the variables $Y_k\equiv\sqrt{T_{N_k}}X_0(N_k)$ converge to x. But this follows from the following lemma:

Lemma 5.11: *Define* $Y_k \equiv \sqrt{T_{N_k}} X_0(N_k)$. *For any* $x \in I\!R^I$ *and any* $\epsilon > 0$,

$$I\!P\left[Y_k \in (x_0 - \epsilon, x_0 + \epsilon) \ i.o.\right] = 1 \tag{5.45}$$

Proof: Let us denote by $\mathcal{F}_\xi$ the sigma algebra generated by the random variables $\xi_i^\mu, \mu \in I\!N, i \geq 1$. Note that

$$I\!P\left[X_k \in (x_0 - \epsilon, x_0 + \epsilon) \text{ i.o.}\right] = I\!E\left(I\!P\left[X_k \in (x_0 - \epsilon, x_0 + \epsilon) \text{ i.o.} \mid \mathcal{F}_\xi\right]\right) \tag{5.46}$$

so that it is enough to prove that for almost all ω, $I\!P\left[X_k \in (x_0 - \epsilon, x_0 + \epsilon) \text{ i.o.} \mid \mathcal{F}_\xi\right] = 1$.

Let us define the random variables

$$\widetilde{Y}_k \equiv \sum_{\mu=M(N_{k-1})+1}^{M(N_k)} \xi_0^\mu I\!E_{\Phi_{N_k}} Z_\mu \tag{5.47}$$

Note first that

$$I\!E\left(Y_k - \widetilde{Y}_k\right)^2 = I\!E \sum_{\mu=2}^{M(N_{k-1})} \left(I\!E_{\Phi_{N_k}} Z_\mu\right)^2 \leq M(N_{k-1}) I\!E\left(I\!E_{\Phi_{N_k}} Z_2\right)^2 \leq \rho^2 \frac{N_{k-1}}{N_k} \tag{5.48}$$

Thus, if N_k is chosen such that $\sum_{k=1}^\infty \frac{N_{k-1}}{N_k} < \infty$, by the first Borel-Cantelli lemma,

$$\lim_{k\uparrow\infty} (Y_k - \widetilde{Y}_k) = 0 \quad \text{a.s.} \tag{5.49}$$

On the other hand, the random variables $\widetilde{Y}_k$ are conditionally independent, given $\mathcal{F}_\xi$. Therefore, by the second Borel-Cantelli lemma

$$I\!P\left[X_k \in (x_0 - \epsilon, x_0 + \epsilon) \text{ i.o.} \mid \mathcal{F}_\xi\right] = 1 \tag{5.50}$$

if

$$\sum_{k=1}^\infty I\!P\left[X_k \in (x_0 - \epsilon, x_0 + \epsilon) \mid \mathcal{F}_\xi\right] = \infty \tag{5.51}$$

But for almost all ω, $\widetilde{Y}_k$ conditioned on $\mathcal{F}_\xi$ converges to a gaussian of variance αr (the proof is identical to that of Proposition 5.3), so that for almost all ω, as $k \uparrow \infty$

$$I\!P\left[X_k \in (x_0 - \epsilon, x_0 + \epsilon) \mid \mathcal{F}_\xi\right] \to \frac{1}{\sqrt{2\pi\alpha r}} \int_{x-\epsilon}^{x+\epsilon} dy e^{-\frac{y^2}{2\alpha r}} > 0 \tag{5.52}$$

which implies (5.51) and hence (5.50). Putting this together with (5.49) concludes the proof of the lemma, and of the proposition. $\diamondsuit$

Some remarks concerning the implications of this proposition are in place. First, it shows that if the standard definition of limiting Gibbs measures as weak limit points is adapted, then we have discovered that in the Hopfield model all product measures on $\{-1,1\}^{\mathbb{N}}$ are extremal Gibbs states. Such a statement contains some information, but it is clearly not useful as information on the approximate nature of a finite volume state. This confirms our discussion in Section 2 on the necessity to use a metastate formalism.

Second, one may ask whether conditioning or the application of external fields of vanishing strength as discussed in Section 2 can improve the convergence behaviour of our measures. The answer appears obviously to be no. Contrary to a situation where a symmetry is present whose breaking biases the system to choose one of the possible states, the application of an arbitrarily weak field cannot alter anything.

Third, we note that the total set of limiting Gibbs measures does not depend on the conditioning on the ball $B_\rho^{(1,1)}$, while the metastate obtained does depend on it. Thus the conditioning allows us to construct two metastates corresponding to each of the stored patterns. These metastates are in a sense extremal, since they are concentrated on the set of extremal (i.e. product) measures of our system. Without conditioning one can construct other metastates (which however we cannot control explicitly in our situation).

6. Induction and the replica symmetric solution

We now conclude our analysis by showing that the quantities $U_N \equiv \mathbb{E}_{\Phi_N}\|\bar{Z}\|_2^2$, $m_1(N) \equiv \mathbb{E}_{\Phi_N} Z_1$ and $T_N \equiv \sum_{\mu=2}^{M} [\mathbb{E}_{\Phi_N} Z_\mu]^2$ actually do converge almost surely under our general assumptions. The proof consist of two steps: First we show that these quantities are self-averaging and then the convergence of their mean values is proven by induction. We will assume throughout this section that the parameters α and β are such that local convexity holds. We stress that this section is entirely based on ideas of Talagrand [T1] and Pastur, Shcherbina and Tirozzi [PST] and is mainly added for the convenience of the reader.

Thus our first result will be:

Proposition 6.1: *Let A_N denote any of the three quantities U_N, $m_1(N)$ or T_N. Then there are finite positive constants c, C such that, for any $0 < \zeta \leq \frac{1}{2}$,*

$$\mathbb{P}\left[|A_N - \mathbb{E}A_N| \geq N^{-\zeta/2}\right] \leq CN^{\zeta} \exp\left(-\frac{N^{1-2\zeta}}{c}\right) \tag{6.1}$$

Proof: The proofs of these three statements are all very similar to that of Corollary 5.6. Indeed, for $m_1(N)$, (6.1) is a special case of that corollary. In the two other cases, we just need to define the appropriate analogues of the 'generating function' f from (5.28). They are

$$g(x) \equiv \frac{1}{\beta N} \ln \mathbb{E}_{\Phi_N} \mathbb{E}'_{\Phi_N} e^{\beta N x (\bar{Z}, \bar{Z}')} \tag{6.2}$$

in the case of T_N and

$$\tilde{g}(x) \equiv \frac{1}{\beta N} \ln \mathbb{E}_{\Phi_N} \mathbb{E}'_{\Phi_N} e^{\beta N x \|\bar{Z}\|_2^2} \tag{6.3}$$

The proof then proceeds as in that of Corollary 6.6. We refrain from giving the details. $\diamondsuit$

We now turn to the induction part of the proof and derive a recursion relation for the three quantities above. In the sequel it will be convenient to introduce a site 0 that will replace the set I and to set $\xi_0 = \eta$. Let us define

$$u_N(\tau) \equiv \ln \mathbb{E}_{\Phi_N} e^{\beta\tau(\eta, Z)} \tag{6.4}$$

We also set $v_N(\tau) \equiv \tau\beta(\eta, \mathbb{E}_{\Phi_N} Z)$ and $w_N(\tau) \equiv u_N(\tau) - v_N(\tau)$. In the sequel we will need the following auxiliary result

Lemma 6.2: *Under our general assumptions*

(i) $\frac{1}{\beta\sqrt{T_N}} \frac{d}{d\tau} v_N(\tau)$ *converges weakly to a standard gaussian random variable.*

(ii) $\left|\frac{d}{d\tau}w_N(\tau) - \tau\beta^2 I\!E I\!E_{\Phi_N}\|\bar{Z}\|_2^2\right|$ *converges to zero in probability.*

Proof: (i) is obvious from Proposition 5.4 and the definition of $v_N(\tau)$. To prove (ii), note that $w_N(\tau)$ is convex and $\frac{d^2}{d\tau^2}w_N(\tau) \leq \frac{\beta\alpha}{\epsilon}$. Thus, *if* $\text{var}\,(w_N(\tau)) \leq \frac{C}{\sqrt{N}}$, then $\text{var}\left(\frac{d}{d\tau}w_N(\tau)\right) \leq \frac{C'}{N^{1/4}}$ by a standard result similar in spirit to Lemma 5.5 (see e.g. [T2], Proposition 5.4). On the other hand, $|I\!E w_N(\tau) - \frac{\tau^2\beta^2}{2} I\!E I\!E_{\Phi_N}\|\bar{Z}\|_2^2| \leq \frac{K}{\sqrt{N}}$, by Lemma 5.3, which, together with the boundedness of the second derivative of $w_N(\tau)$ implies that $|\frac{d}{d\tau}I\!E w_N(\tau) - \tau\beta^2 I\!E I\!E_{\Phi_N}\|\bar{Z}\|_2^2| \downarrow 0$. This means that $\text{var}\,(w_N(\tau)) \leq \frac{C}{\sqrt{N}}$ implies the lemma. Since we already know from G.11ter) that $I\!E R_N^2 \leq \frac{K}{N}$, it is enough to prove $\text{var}\left(I\!E_{\Phi_N}\|\bar{Z}\|_2^2\right) \leq \frac{C}{\sqrt{N}}$. This follows just as the corresponding concentration estimate for U_N. $\diamondsuit$

We are now ready to start the induction procedure. We will place ourselves on a subspace $\widetilde{\Omega} \subset \Omega$ where for all but finitely many N $|U_N - I\!E U_N| \leq N^{-1/4}$, $|T_N - I\!E T_N| \leq N^{-1/4}$, etc. This subspace has probability one by our estimates.

Let us note that by (iii) of Proposition 3.1, $I\!E_{\Phi_N} Z_\mu$ and $\int dQ^{(1,1)}_{N,\beta,\rho}(m)m_\mu$ differ only by an exponentially small term. Thus

$$I\!E_{\Phi_N} Z_\mu = \frac{1}{N}\sum_{i=1}\xi_i^\mu \int \mu^{(1,1)}_{N,\beta,\rho}(d\sigma)\sigma_i + O\left(e^{-cM}\right) \tag{6.5}$$

and, by symmetry,

$$I\!E I\!E_{\Phi_{N+1}}(Z_\mu) = I\!E\eta^\mu \int \mu^{(1,1)}_{N+1,\beta,\rho}(d\sigma)\sigma_0 + O\left(e^{-cM}\right) \tag{6.6}$$

Using Lemma 5.2 and the definition of u_N, this gives

$$I\!E I\!E_{\Phi_{N+1}}(Z_\mu) = I\!E\eta^\mu \frac{e^{u_N(1)} - e^{u_N(-1)}}{e^{u_N(1)} + e^{u_N(-1)}} + O\left(e^{-cM}\right) \tag{6.7}$$

where to be precise one should note that the left and right hand side are computed at temperatures β and $\beta' = \frac{N}{N}\beta$, respectively, and that the value of M is equal to $M(N+1)$ on both sides; that is, both sides correspond to slightly different values of α and β, but we will see that this causes no problems.

Using our concentration results and Lemma 5.3 this gives

$$I\!E I\!E_{\Phi_{N+1}}(Z_\mu) = I\!E\eta^\mu \tanh\left(\beta(\eta^1 I\!E m_1(N) + \sqrt{I\!E T_N}X_0(N))\right) + O(N^{-1/4}) \tag{6.8}$$

Using further Proposition 5.4 we get a first recursion for $m_1(N)$:

$$m_1(N+1) = \int d\mathcal{N}(g)\tanh\left(\beta(I\!E m_1(N) + \sqrt{I\!E T_N}g)\right) + o(1) \tag{6.9}$$

Remark: The error term in (6.9) can be sharpened to $O(N^{-1/4})$ by using instead of Lemma 5.3 a trick, attributed to Trotter, that we learned from Talagrand's paper [T1] (see the proof of Proposition 6.3 in that paper).

We need of course a recursion for T_N as well. From here on there is no great difference from the procedure in [PST], except that the N-dependences have to be kept track of carefully. This was outlined in [BG4] and we repeat the steps for the convenience of the reader. To simplify the notation, we ignore all the $O(N^{-1/4})$ error terms and put them back in the end only. Also, the remarks concerning β and α made above apply throughout.

Note that $T_N = \|I\!\!E_{\Phi_N} Z\|_2^2 - (I\!\!E_{\Phi_N} Z_1)^2$ and

$$
\begin{aligned}
I\!\!E\|I\!\!E_{\Phi_{N+1}} Z\|_2^2 &= \sum_{\mu=1}^{M} I\!\!E \left(\frac{1}{N+1} \sum_{i=0}^{N} \xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i) \right)^2 \\
&= \frac{M}{N+1} I\!\!E \left(\mu_{\beta,N+1,M}^{(1,1)}(\sigma_0) \right)^2 \\
&+ \sum_{\mu=1}^{M} I\!\!E \xi_0^\mu \mu_{\beta,N+1,M}^{(1,1)}(\sigma_0) \left(\frac{1}{N+1} \sum_{i=1}^{N} \xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i) \right)
\end{aligned}
\tag{6.10}
$$

Using Lemma 5.2 as in the step leading to (6.7), we get for the first term in (6.10)

$$
I\!\!E \left(\mu_{\beta,N+1,M}^{(1,1)}(\sigma_0) \right)^2 = I\!\!E \tanh^2 \left(\beta(\eta_1 I\!\!E_{\Phi_N} Z_1 + \sqrt{I\!\!E T_N}) \right) \equiv I\!\!E Q_N \tag{6.11}
$$

For the second term, we use the identity from [PST]

$$
\begin{aligned}
\sum_{\mu=1}^{M} \xi_0^\mu \left(\frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i) \right) &= \frac{\sum_{\sigma_0} I\!\!E_{\Phi_N}(\xi_0, X) e^{\beta\sigma_0(\xi_0,X)}}{\sum_{\sigma_0} I\!\!E_{\Phi_N} e^{\beta\sigma_0(\xi_0,X)}} \\
=& \beta^{-1} \frac{\sum_{\tau=\pm1} u_N{}'(\tau) e^{u_N(\tau)}}{\sum_{\tau=\pm1} e^{u_N(\tau)}}
\end{aligned}
\tag{6.12}
$$

Together with Lemma 6.2 one concludes that in law up to small errors

$$
\begin{aligned}
\sum_{\mu=1}^{M} \xi_0^\mu \left(\frac{1}{N+1} \sum_{i=1}^{N} \xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i) \right) &= \xi_0^1 I\!\!E_{\Phi_N} Z_1 + \sqrt{I\!\!E T_N} X_N \\
+ \beta I\!\!E_{\Phi_N} \|\bar{Z}\|_2^2 \tanh\beta &\left(\xi_0^1 I\!\!E_{\Phi_N} Z_1 + \sqrt{I\!\!E T_N} X_N \right)
\end{aligned}
\tag{6.13}
$$

and so

$$
\begin{aligned}
I\!\!E\|I\!\!E_{\Phi_{N+1}} Z\|_2^2 = \alpha I\!\!E Q_N + I\!\!E \Big[&\tanh\beta \left(\xi_0^1 I\!\!E_{\Phi_N} Z_1 + \sqrt{I\!\!E T_N} X_N \right) \\
&\times \left[\xi_0^1 I\!\!E_{\Phi_N} Z_1 + \sqrt{I\!\!E T_N} X_N \right] \Big] \\
&+ \beta I\!\!E I\!\!E_{\Phi_N} \|\bar{Z}\|_2^2 \tanh^2\beta \left(\xi_0^1 I\!\!E_{\Phi_N} Z_1 + \sqrt{I\!\!E T_N} X_N \right)
\end{aligned}
\tag{6.14}
$$

Using the self-averaging properties of $I\!E_{\Phi_N}\|\bar{Z}\|_2^2$, the last term is of course essentially equal to

$$\beta I\!E I\!E_{\Phi_N}\|\bar{Z}\|_2^2 I\!E Q_N \tag{6.15}$$

The appearance of $I\!E_{\Phi_N}\|\bar{Z}\|_2^2$ is disturbing, as it introduces a new quantity into the system. Fortunately, it is the last one. The point is that proceeding as above, we can show that

$$\begin{aligned} I\!E I\!E_{\Phi_{N+1}}\|Z\|_2^2 =&\alpha + I\!E\Bigg[\tanh\beta\left(\xi^1_{N+1} I\!E_{\Phi_N} Z_1 + \sqrt{I\!E T_N} X_N\right) \\ &\times\left[\xi_0^1 I\!E_{\Phi_N} Z_1 + \sqrt{I\!E T_N} X_N\right]\Bigg] + \beta I\!E I\!E_{\Phi_N}\|\bar{Z}\|_2^2 I\!E Q_N \end{aligned} \tag{6.16}$$

so that setting $U_N \equiv I\!E_{\Phi_N}\|\bar{Z}\|_2^2$, we get, subtracting (6.14) from (6.16), the simple recursion

$$I\!E U_{N+1} = \alpha(1 - I\!E Q_N) + \beta(1 - I\!E Q_N) I\!E U_N \tag{6.17}$$

From this we get (since all quantities considered are self-averaging, we drop the $I\!E$ to simplify the notation), setting $m_1(N) \equiv I\!E_{\Phi_N} Z_1$,

$$\begin{aligned} T_{N+1} =& -(m_1(N+1))^2 + \alpha Q_N + \beta U_N Q_N \\ &+ \int d\mathcal{N}(g)[m_1(N) + \sqrt{T_N} g] \tanh\beta(m_1(N) + \sqrt{T_N} g) \\ &= m_1(N+1)(m_1(N) - m_1(N+1)) + \beta U_N Q_N + \beta T_N(1 - Q_N) + \alpha Q_N \end{aligned} \tag{6.18}$$

where we used integration by parts. The complete system of recursion relations can thus be written as

$$\begin{aligned} m_1(N+1) =& \int d\mathcal{N}(g)\tanh\beta\left(m_1(N) + \sqrt{T_N} g\right) + O(N^{-1/4}) \\ T_{N+1} =& m_1(N-1)(m_1(N) - m_1(N+1)) + \beta U_N Q_N + \beta T_N(1 - Q_N) + \alpha Q_N + O(N^{-1/4}) \\ U_{N+1} =& \alpha(1 - Q_N) + \beta(1 - Q_N) U_N + O(N^{-1/4}) \\ Q_{N+1} =& \int d\mathcal{N}(g)\tanh^2\beta\left(m_1(N) + \sqrt{T_N} g\right) + O(N^{-1/4}) \end{aligned} \tag{6.19}$$

If the solutions to this system of equations converges, than the limits $r = \lim_{N\uparrow\infty} T_N/\alpha$, $q = \lim_{N\uparrow\infty} Q_N$ and $m_1 = \lim_{N\uparrow\infty} m_1(N)$ ($u \equiv \lim_{N\uparrow\infty} U_N$ can be eliminated) must satisfy the equations

$$m_1 = \int d\mathcal{N}(g)\tanh(\beta(m_1 + \sqrt{\alpha r} g)) \tag{6.20}$$

$$q = \int d\mathcal{N}(g)\tanh^2(\beta(m_1 + \sqrt{\alpha r} g)) \tag{6.21}$$

$$r = \frac{q}{(1 - \beta + \beta q)^2} \tag{6.22}$$

which are the equations for the replica symmetric solution of the Hopfield model found by Amit et al. [AGS].

In principle one might think that to prove convergence it is enough to study the stability of the dynamical system above without the error terms. However, this is not quite true. Note that the parameters β and α of the quantities on the two sides of the equation differ slightly (although this is suppressed in the notation). In particular, if we iterate too often, α will tend to zero. The way out of this difficulty was proposed by Talagrand [T1]. We will briefly explain his idea. In a simplified notation, we are in the following situation: We have a sequence $X_n(p)$ of functions depending on a parameter p. There is an explicit sequence p_n, satisfying $|p_{n+1}-p_n| \leq c/n$ and a functions F_p such that

$$X_{n+1}(p_{n+1}) = F_{p_n}(X_n(p_n)) + O(n^{-1/4}) \tag{6.23}$$

In this setting, we have the following lemma.

Lemma 6.3: *Assume that there exist a domain D containing a single fixed point $X^*(p)$ of F_p. Assume that $F_p(X)$ is Lipshitz continuous as a function of X, Lipshitz continuous as a function of p uniformly for $X \in D$ and that for all $X \in D$, $F_p^n(X) \to X^*(p)$. Assume we know that for all n large enough, $X_n(p) \in D$. Then*

$$\lim_{n\uparrow\infty} X_n(p) = X^*(p) \tag{6.24}$$

Proof: Let us choose a integer valued monotone increasing function $k(n)$ such that $k(n) \uparrow \infty$ as n goes to infinity. Assume e.g. $k(n) \leq \ln n$. We will show that

$$\lim_{n\uparrow\infty} X_{n+k(n)}(p) = X^*(p) \tag{6.25}$$

To see this, note first that $|p_{n+k(n)} - p_n| \leq \frac{k(n)}{n}$. By (6.23), we have that using the Lipshitz properties of F

$$X_{n+k(n)}(p) = F_p^{k(n)}(X_n(p_n)) + O(n^{-1/4}) \tag{6.26}$$

where we choose p_n such that $p_{n+k(n)} = p$. Now since $X_n(p_n) \in D$, $\left|F_p^{k(n)}(X_n(p_n) - X^*(p)\right| \downarrow 0$ as n and thus $k(n)$ goes to infinity, so that (6.26) implies (6.25). But (6.25) for any slowly diverging function $k(n)$ implies the convergence of $X_n(p)$, as claimed. $\diamondsuit$

This lemma can be applied to the recurrence (6.18). The main point to check is whether the corresponding F_β attracts a domain in which the parameters $m_1(N), T_N, U_N, Q_N$ are a priori located due tho the support properties of the measure $\widetilde{\mathcal{Q}}^{(1,1)}_{N,\beta,\rho}$. This stability analysis was carried out (for an equivalent system) by Talagrand and answered to the affirmative. We do not want to repeat this tedious, but in principle elementary computation here.

We would like to make, however, some remarks. It is clear that if we consider conditional measures, then we can always force the parameters $m_1(N), R_N, U_N, Q_N$ to be in some domain. Thus, in principle, we could first study the fixpoints of (6.18), determine their domains of attraction and then define corresponding conditional Gibbs measures. However, these measures may then be metastable. Also, of course, at least in our derivation, do we need to verify the local convexity in the corresponding domains since this was used in the derivation of the equations (6.18).

References

[AGS] D.J. Amit, H. Gutfreund and H. Sompolinsky, "Statistical mechanics of neural networks near saturation", Ann. Phys. **173**, 30-67 (1987).

[AW] M. Aizenman, and J. Wehr, "Rounding effects on quenched randomness on first-order phase transitions", Commun. Math. Phys. **130**, 489 (1990). 643-664 (1993).

[BG1] A. Bovier and V. Gayrard, "The retrieval phase of the Hopfield model, A rigorous analysis of the overlap distribution", Prob. Theor. Rel. Fields **107**, 61-98 (1997).

[BG2] A. Bovier and V. Gayrard, "The Hopfield model as a generalized random mean field model", in "Mathematics of spin glasses and neural networks", A. Bovier and P. Picco, Eds., Progress in Probablity, Birkhäuser, Boston, (1997).

[BG3] A. Bovier and V. Gayrard, "An almost sure central limit theorem for the Hopfield model", to appear in Markov Proc. Rel. Fields (1997).

[BGP1] A. Bovier, V. Gayrard, and P. Picco, "Gibbs states of the Hopfield model in the regime of perfect memory", Prob. Theor. Rel. Fields **100**, 329-363 (1994).

[BGP2] A. Bovier, V. Gayrard, and P. Picco, "Gibbs states of the Hopfield model with extensively many patterns", J. Stat. Phys. **79**, 395-414 (1995).

[BGK] A. Bovier, V. Gayrard, and Ch. Külske, in preparation.

[BK] A. Bovier and Ch. Külske, A rigorous renormalization group method for interfaces in random media, Rev. Math. Phys. **6**, 413-496 (1994).

[BL] H.J. Brascamp and E.H. Lieb, "On extensions of the Brunn-Minkowski and Pékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation", J. Funct. Anal. **22**, 366-389 (1976).

[H] B. Helffer, "Recent results and open problems on Schrödinger operators, Laplace integrals, and transfer operators in large dimension" (1996).

[HH] P. Hall and C.C. Heyde, "Martingale limit theory and its applications", Academic Press, New York (1980).

[Ho] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", Proc. Natl. Acad. Sci. USA **79**, 2554-2558 (1982).

[HS] B. Helffer and J. Sjöstrand, "On the correlation for Kac-like models in the convex case", J. Stat. Phys. **74**, 349-409 (1994).

[Ka] O. Kallenberg, "Random measures", Academic Press, New York (1983).

[K] Ch. Külske, "Metastates in disordered mean field models: random field and Hopfield models", to appear in J. Stat. Phys. (1997).

[MPV] M. Mézard, G. Parisi, and M.A. Virasoro, "Spin-glass theory and beyond", World Scientific, Singapore (1988).

[N] Ch. Newman, "Topics in disordered systems", Birkhäuser, Boston (1997).

[NS1] Ch.M. Newman and D.L. Stein, "Multiple states and the thermodynamic limits in short ranged Ising spin glass models", Phys. Rev. **B 72**, 973-982 (1992).

[NS2] Ch.M. Newman and D.L. Stein, "Non-mean-field behaviour in realistic spin glasses", Phys. Rev. Lett. **76**, 515-518 (1996).

[NS3] Ch.M. Newman and D.L. Stein, "Spatial inhomogeneity and thermodynamic chaos", Phys. Rev. Lett. **76**, 4821-4824 (1996).

[NS4] 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).

[NS5] C.M. Newman and D.L. Stein, "Ground state structure in a highly disordered spin glass model", J. Stat. Phys. **82**, 1113-1132 (1996).

[PST] L. Pastur, M. Shcherbina, and B. Tirozzi, "The replica symmetric solution without the replica trick for the Hopfield model", J. Stat. Phys. **74**, 1161-1183 (1994).

[T1] M. Talagrand, "Rigorous results for the Hopfield model with many patterns", preprint 1996, to appear in Probab. Theor. Rel. Fields.

[T2] M. Talagrand, "The Sherrington-Kirkpatrick model: A challenge for mathematicians", preprint 1996, to appear in Prob. Theor. Rel. Fields.