Statistical Mechanics of Disordered Systems

Lecture notes

by

Anton Bovier[1,2]

Weierstrass-Institut für Angewandte Analysis und Stochastik

Mohrenstrasse 39, D-10117 Berlin, Germany

May 29, 2001

[1] Research partially supported by the Deutsche Forschungsgemeinschaft in the program "Interacting Random Systems of High Complexity".

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

Learning, undigested by thought, is labour lost.
Thinking, unassisted by learning, is perilous.

K'ung-tsze,
"Lun Yu"

Preface

These notes were compiled for a Concentrated Advanced Course at the University of Copenhagen in the framework of the MaPhySto program. Part of the material was also used for advanced courses at the Technical University of Berlin. These lectures were designed to introduce advanced students with a good background in probability theory but not necessarily in physics to problems in disordered systems of statistical mechanics. This is a difficult, but in my opinion, important task in many respects.

Statistical mechanics is a now more than hundred years old branch of physics that has been remarkably successful. As a subject of mathematics, or more particularly a "branch of probability theory", it is much younger. After early work in the 50's by Kac, Lee and Yang, and Montroll, it essentially started to exist with the early works of Dobrushin in '62 and a first textbook manifestation is the by now classical monograph of Ruelle [Ru1]. This book provides an axiomatic formulation of a subject called the theory of *lattice gases* or equivalently *lattice spin systems*. Further and partly more complete textbooks dealing with the fundamental theory are two books by Preston's [Pr1,Pr2] and Ruelle's monograph in the Encyclopedia of Mathematics [Ru2]. Over the last 30 years this subject has been blooming and there is an enormously rich literature compiled. Nonetheless, considered as a mathematical subject, it is still far from a satisfactory state. Apart from the basic *Gibbsian framework* in the axiomatic formulation of Dobrushin, Lanford and Ruelle, it consists by and large of a number of techniques (correlation inequalities, cluster expansions, Pirogov-Sinai theory, Dobrushin uniqueness, FK-representations, renormalization group, etc.) that work for some models and some sets of parameters. Many of the physically most interesting phenomena (critical exponents, etc.) are well beyond the reach of todays mathematics. The reason for this situation is rather simple: the models people in statistical mechanics are interested in come from physics, and they tend to be extremely varied and hard to analyse. Unlike classical probability theory that by and large evolved around the concept of independent random variables and enriched its scope by slowly adding more complexity, the interesting models in statistical mechanics involve extremely complex random fields and their analysis has to go frequently in a case by case fashion. For this reason, there is not, and maybe cannot be, a reasonable comprehensive textbook that would acquaint the student with the whole of mathematical statistical mechanics. There are some modern texts that cover some of the area. Georgii's text [Ge1] is a comprehensive introduction to the general theory of Gibbs measures; however nothing of the heavy machinery needed to investigate specific models like expansions, Pirogov-Sinai theory, et. is covered. Simon's book [Sim] also emphasizes the more basic aspects of the theory, contains however an introduction to expansion methods. Another useful introductory text is Sinai's small book [Sin], which is quite complementary to the other two with an emphasis on expansions and Pirogov-Sinai theory.

If the situation concerning statistical mechanics is not satisfactory, it is dramatically worse in the now growing field of disordered systems, or random spin systems, to be more specific. An attempt to cover the state of the art in 1984 was made in J. Fröhlich's Les Houches lecture notes [F]. They still can be read with profit today, and, one is afraid to say that many of the most important issues mentioned there remain wide open today. Still, 15 years have passed, and some very important steps have been taken in the meantime. A more recent text is the slim book by Ch. Newman, covering his ETH course [N]. It focuses on two items: percolation methods, and the spin glass problem. It is a very useful book, but neither comprehensive, nor really introductory. A collection of partly review papers is the volume issued by the author and P. Picco in 1998 [BP].

In these notes I try to fill some of these gaps, albeit again from a biased point of view. I will give an introduction to the statistical mechanics of disordered systems that essentially assumes no prior knowledge of statistical mechanics. To that end, the first part of these notes gives a self-contained introduction to the Gibbsian formalism. It emphasizes, as do these notes in general, the fundamental question of infinite volume Gibbs measures and the associated existence and (non)-uniqueness problems. Thus I will present the two basic tools for tackling this problem: the Dobrushin uniqueness condition, and the Peierls argument. This leaves out, regrettably, high and low temperature expansions, and the Pirogov-Sinai theory, but their inclusion would have largely exceeded the scope of these notes. Moreover, I felt that in such an introductory course it might be better to avoid methods that are considered technically very difficult.

The second part of the notes deals with disordered spin systems on the lattice with short range interactions. It begins with a comprehensive introduction to the formalism of *random Gibbs measures* and *metastates*. Then I discuss the extensions and limitations of the two methods introduced in the first section. The bulk of this part is devoted to the random field Ising model and the question how uniqueness and non-uniqueness can be analysed in this case. Again, I will fall short of providing the full renormalization proof of non-uniqueness in three dimensions, since it relies heavily on expansions. I hope to give rather convincing, though incomplete arguments based on concentration of measure ideas. On the other hand, I will give a full proof of uniqueness in $d = 2$ due to Aizenman and Wehr. In the process, we are introduced to applications of correlation inequalities, notably the FKG-inequalities. This covers most of the interesting results in the field, with the notable exception of long-range spin glasses which again would require technically heavy expansions techniques. Another, per se interesting subject I decided to leave out are results on short-range spin glasses. The reason is two-fold: first, there are no solid genuinely low-temperature results available, with the sole exception of some very general results on basic properties of Gibbs measures that follow essentially from ergodic considerations. In that sense they are somewhat orthogonal to the more constructive spirit of these notes. But most importantly, they are rather easily accessible in Newman's ETH-lectures [N].

Thus genuine spin glass models are treated only in the context of mean field theory, and this makes up the rest of these notes. I will basically treat two classes of models: Gaussian processes on the hypercube, and models of the Hopfield type. I will go to great length to explain in all detail the case of the *random energy model* (REM) which will give us an idea how a complete solution of such models could look like. Much less space will be devoted to the Sherrington-Kirkpatrick model and its p-spin counterparts, first since an exposition by M. Talagrand [T2000] is now available, and second because the state of our knowledge is both limited and under active development. Finally I turn to non-Gaussian models, and notably the Hopfield model. Although there exists a rather extensive account in [BP] covering the state of the art in 1997, I will try to give a more accessible account that takes into account some of the more recent developments.

Acknowledgements. I first of all like to thank Ole E. Barndorff-Nielsen and Martin Jacobsen for the invitation to teach this course in Copenhagen in the MAPHYSTO program; without this, it is most unlikely that I would have written these notes. On this occasion I wish to express my profound gratitude to Martin Jacobson for his hospitality at Copenhagen and for the excellent organisation of all the practical matters of the course. My second thanks go to my collaborators, past and present, on the subjects related to these notes: first and foremost to Véronique Gayrard for the long standing and enjoyable collaboration on disordered systems. Next, to all my other collaborators in recent years, A.C.D. van Enter, Ch. Külske, I. Kurkova, M. Löwe, D. Mason, B. Niederhauser, P. Picco, and M. Zahradník, all of whom have contributed to my understanding of the subject. Special thanks are due to Chuck Newman who through his writing and in many discussions has been a source of inspiration over many years. I am very grateful to all those who read preliminary versions and pointed out errors, misprints and omissions, in particular V. Gayrard, B. Gentz, M. Jacobson, I. Kurkova, M. Löwe, Tomász Schreiber, and, especially, A.C.D. van Enter. Last but not least, thanks are due to Michel Talagrand, for bringing so many new ideas and so much enthusiasm into this field.

Berlin, May 2001 Anton Bovier

Contents

1. A survey of the Gibbsian formalism for lattice spin systems

We begin this lecture with a brief survey of the basic formalism of the statistical mechanics of lattice spin systems, or lattice gases. The literature on this subject is well developed and the interested student can find in depth material for further reading in [Ge1,Sim,Sin,Pr1,Pr2,] and the classical monographs by Ruelle [Ru1,Ru2]. A nice short introduction with a somewhat particular aim in view is also given in the first sections of the paper [vEFS]. Here we will be rather sketchy, but try to emphasize some issues and concepts that will become particularly relevant a) for the discussion of disordered systems and b) for the discussion of mean field models (which do not exactly fall into the general framework we develop here).

1.1. Spin systems and Gibbs measures.

The idea of the spin system was born at about 1920 in an attempt to understand the phenomenon of ferromagnetism[3]. At that time it was understood that ferromagnetism should be due to the alignment of the elementary magnetic moments ("spins") of the (iron) atoms that persists even after an external field is turned off. The phenomenon is temperature dependent: if one heats the material, the coherent alignment is lost. It was understood that the magnetic moments should exert an "attractive" ("ferromagnetic") interaction among each others which however is rather short range. The question was then how such a short range interaction could sustain the observed very long range coherent behaviour of the material, and why such an effect should depend on the temperature. In this situation, Lenz had an extraordinarily consequential idea: to invent a model that would simplify the ferromagnetic system to its most rudimentary features and to apply the formalism of statistical mechanics to it. The idea behind this was that the particular phenomenon that was not understood should have to do with the collective behaviour of the many microscopic elements in the system and should be independent of the precise details of these and their

[3] To be historically correct, we should mention that the idea of a spin model goes back to Weiss [We] in 1907 following P. Curie's discovery of the critical Tempeature (Curie temperature) for ferromagnetic order in 1895. This gave rise to what is now called the Curie–Weiss, or "mean-field", model for ferromagnetism. Although this model also had a considerable impact (see Section 3 of these note!), it is deemed more a mathematical toy model compared to the "realistic" Ising spin system.

interaction. Lenz gave the analysis of this model as a subject of a Ph.D. thesis to his student Ernst Ising[4] (1900-1998) who published his findings on the one-dimensional model in 1923 [Is]. Nonwithstanding the fact that he found (correctly) no sign of ferromagnetism and conjectured (wrongly[5]) that the same was true in higher dimensions and that therefore the model could not explain the phenomenon, the model invented by Lenz was destined to become, under the name of "Ising's model" one of the most investigated and successful models in the history of statistical mechanics, and more general to what is now know as the theory of lattice spin systems.

The simplifications proposed by Lenz were dramatic: He assumed the atoms placed on the sites of a regular lattice $\mathbb{Z}^d$ and represented by the simplest possible spin variables taking only the two values ± 1. Spins would interact only if they were at neighbouring sites on the lattice, and this interaction would favour such spins to take on the same values. In addition, there can be an external magnetic field h favouring globally either the plus or the minus-sign. This interaction can be introduced via a *Hamiltonian* function H that assigns to a spin-configuration $\sigma \equiv \{\sigma_i\}_{i\in\mathbb{Z}^d}$ the energy

$$H(\sigma) \equiv - \sum_{\substack{i,j\in\mathbb{Z}^d \\ \|i-j\|_1=1}} \sigma_i\sigma_j - h\sum_{i\in\mathbb{Z}^d}\sigma_i \tag{1.1}$$

Of course this formula makes no sense, as the sums do not converge, and should be given a sensible interpretation. One would immediately like to argue that this problem results from the fact that we are looking at a spin-configuration on an infinite lattice, and that since in nature all magnets consist of a finite, albeit very large, number of atoms, we should always consider finite sets $\Lambda \subset \mathbb{Z}^d$ and spin configurations $\sigma_\Lambda \equiv \{\sigma_i\}_{i\in\Lambda}$ and compute the energy of such a configuration by restricting the sums in (1.1) to run over the set Λ only. This touches on an important fundamental issue of statistical mechanics that we will have occasion to discuss repeatedly in these lectures. It is tempting to formulate this as an (informal) axiom of the approach of statistical mechanics:

---

[4] An account of the life of Ising can be found in [Ko] and is definitely worth reading.

[5] Ising's assertion that his one-dimensional result would hold true equally in higher dimension should serve as a warning against hasty generalizations. In statistical mechanics, new phenomena tend to appear where one would not always expect them.

A system composed of a very large number of degrees of freedom can be well approximated by an infinite system.

We will have to see how to interpret this statement and what its limitations are later. I would ask you to accept this for the moment and take it as an excuse for the otherwise seemingly unreasonable struggle we will enter to describe infinite systems. We will return to this shortly.

The basic axiom of statistical mechanics is now that the (equilibrium) properties of a system shall be described by specifying a probability measure on the space of configurations, in our case $\{-1,+1\}^{\mathbb{Z}^d}$. The particular choice of the probability measure to choose is the subject of the foundations of statistical mechanics and this text is not the place to elaborate on this[6]. We will therefore accept as another axiom that the proper measure to choose is the *Gibbs measure* which formally is given by

$$\mu_\beta(d\sigma) = \frac{1}{Z_\beta} e^{-\beta H(\sigma)} \rho(d\sigma) \tag{1.2}$$

where Z_β is a normalizing constant and ρ is the uniform measure on the configuration space. Again this expression makes no sense as it is written for the infinite system, but would make perfect sense if we replaced $\mathbb{Z}^d$ by a finite set Λ everywhere[7]

We will now see how to obtain a sensible version of (1.2) in the infinite-volume setting. We start with the "a priori" measure ρ that is supposed to describe the non-interacting system. In finite volumes, the uniform measure on the finite space $\{-1,+1\}^\Lambda$ can be seen alternatively as the product Bernoulli measure

$$\rho_\Lambda(\sigma_\Lambda = s_L) = \prod_{i\in\Lambda} \rho_i(\sigma_i = s_i) \tag{1.3}$$

[6] An philosophical discussion on the conceptual basis of the probabilistic foundations of statistical mechanics can be found in the recent book [Gu].

[7] Here we are touching a crucial point. The problem with a finite-volume description is that it appears to be unable to reflect the very phenomenon we want to describe, namely the existence of several phases, i.e. the persistence of magnetized states after the magnetic field has been turned off. The argument was brought forward that a single formula could not possibly describe different physical states at the same time. The question is indeed quite intricate and a full understanding will require to consider the dynamical aspects of the problem. On the level of the equilibrium theory, the issue is however, as we will see, solved precisely and elegantly by the adoption of the infinite-volume axiom.

where $\rho_i(\sigma_i = +1) = \rho_i(\sigma_i = -1) = 1/2$. Now it is of course a standard construction to extend this to infinite-volume. First we make $\{-1,+1\}^{\mathbb{Z}^d}$ into a measure space by equipping it with the product topology of the discrete topology on $\{-1,+1\}$. The corresponding sigma-algebra $\mathcal{F}$ is then just the product sigma-algebra. The measure ρ is then defined by specifying that for all cylinder events $\mathcal{A}_\Lambda$ (i.e. events that for some finite set $\Lambda \subset \mathbb{Z}^d$ depend only on the values of the variables σ_i with $i \in \Lambda$,

$$\rho(\mathcal{A}_\Lambda) = \rho_\Lambda(\mathcal{A}_\Lambda) \tag{1.4}$$

with ρ_Λ defined in (1.3). With this we have set up an a-priori probability space $(\mathcal{S}, \mathcal{F}, \rho)$ describing a system of non-interacting spins. It is worth noting that this set-up is not totally innocent and reflects a certain *physical* attitude towards our problem. Namely, the choice to consider the system as truly infinite and to use the product topology implies that we consider the individual degrees of freedom, or finite collections of them, as the main physical observables whose behaviour is to be measured. While this is rather natural, it should not be forgotten that this has important implications in the interpretation of the infinite-volume results as asymptotic results for large systems that may not in all cases be the most desirable ones[8].

To continue the interpretation of (1.2), one might now be tempted to specify again the measure μ_β by prescribing the finite dimensional marginals, e.g. by specifying that $\mu_{\beta,\Lambda}(d\sigma_\Lambda) = Z_{\beta,\Lambda}^{-1} \exp(-\beta H_\Lambda(\sigma_\Lambda))\rho_\Lambda(d\sigma_\Lambda)$, with $H_\Lambda(\sigma_\Lambda)$ the restriction of (1.1) to the finite volume Λ. The problem with this, however, are the compatibility conditions that are required for such a set measures to specify a measure on $(\mathcal{S}, \mathcal{F})$; Kolmogorov's theorem would require that for $\Lambda \subset \Lambda'$, $\mu_{\beta,\Lambda}(\mathcal{A}_\Lambda) = \mu_{\beta,\Lambda'}(\mathcal{A}_\Lambda)$. While in the case of the non-interacting system, this is trivially checked, this will not hold in the interacting case (Exercise 1.1: Check explicitly that the compatibility conditions do not hold in the case where Λ,Λ' consist of 1 resp. 2 points!). Since there appears no other feasible way how one could specify marginal measures, we need some better

[8] For instance, it might be that one is interested in collections of variables that are composed of enormously many local variables. It may then be that an appropriate description requires intermediate divergent ("mesoscopic") scales in between the "macroscopic" volume and the microscopic degrees of freedom. This would require a slightly different approach to the problem.

idea. Actually, there is not so much choice: if we cannot fix marginals, we should fix conditional distributions. This now seems quite natural from the point of view of the theory of Markov processes, but was only realized in 1968-69 by Roland L. Dobrushin [D1] (and shortly after that by O. Lanford and D. Ruelle [LR]), and is now seen as one the cornerstones of the foundation of modern mathematical statistical mechanics. To understand this construction, we have to return to (1.1) and give an new interpretation to this formal expression. The Hamiltonian should measure the energy of a configuration; this makes no sense in infinite-volume, but what we could ask, is what is the energy of an *infinite-volume* configuration *within a finite-volume* Λ. A natural definition of this quantity is

$$H_\Lambda(\sigma) \equiv - \sum_{\substack{i \vee j \in \Lambda \\ \|i-j\|_1=1}} \sigma_i \sigma_j - h \sum_{i \in \Lambda} \sigma_i \tag{1.5}$$

This differs from the simple restriction of (1.1) to Λ by a term $2\sum_{\substack{i\in\Lambda, j\notin\Lambda \\ \|i-j\|_1=1}} \sigma_i\sigma_j$ which represents the interaction of the spins in Λ with those outside of it; as we see, it actually involves only spins at the boundary of Λ. The notion of finite-volume restriction given by (1.5) has the nice feature that it is compatible under iteration: if $\Lambda' \supset \Lambda$, then

$$\left(H_{\Lambda'}\right)_\Lambda(\sigma) = H_\Lambda(\sigma) \tag{1.6}$$

(1.5) will furnish our standard interpretation of a Hamiltonian function H; we will always consider it as a function from the pairs consisting of finite subsets of $\mathbb{Z}^d$ and configurations in $\mathcal{S}$ to the real numbers that maps $(\Lambda, \sigma) \to H_\Lambda(\sigma)$. This allows to define, for any fixed configuration of spins $\eta \in \mathcal{S}$ and finite subset $\Lambda \subset \mathbb{Z}^d$, a probability measure

$$\mu_\Lambda^\eta(d\sigma_\Lambda) = \frac{1}{Z_{\beta,\Lambda}^\eta} e^{-\beta H_\Lambda((\sigma_\Lambda, \eta_{\Lambda^c}))} \rho_\Lambda(d\sigma_\Lambda) \tag{1.7}$$

(Note: We will change our point of view slightly a bit later on when we formalize this discussion. At the moment it is convenient to consider (1.7) as specifying finite-volume measures).

(1.7) defines a much richer class of measures than just the marginals. The idea now is that these should be the family of conditional probabilities of some measures

μ_β defined on the infinite-volume space. The point is that they satisfy automatically the compatibility conditions required for conditional probabilities (see below), and so have a chance to be conditional probabilities of some infinite-volume measure. Dobrushin's idea was to start from this observation to define the notion of the infinite-volume Gibbs measure, i.e. as the proper definition for the formal expression (1.2):

A probability measure μ_β on $(\mathcal{S}, \mathcal{F})$ is a Gibbs measure for the Hamiltonian H and inverse temperature β, if and only if its conditional distributions (conditioned on configurations in the complement of any finite set Λ) are given by (1.7).

Two immediate questions pose themselves:

(i) Does such a measure exist?

(ii) If it exist, is it uniquely specified?

We will see soon that there is a large class of systems for which existence of such a measure can be shown. That means that Dobrushin's formalism is meaningful and defines a rich theory. The second question makes all the charm of the Gibbsian formalism: There are situations when the infinite-volume measure is not uniquely specified and when *several* infinite-volume measures exist for the same Hamiltonian and the same temperature[9]. This observation will furnish the explanation for strikingly different behaviour of the ferromagnet at high and low temperatures: if $d \geq 2$, the temperature is low, and $h = 0$, there will be measures describing a state with positive magnetization and one with negative magnetization, and the system will have to be in either of them.

Before we continue the investigation of these two questions in the Ising model, we will provide a more general and more formal set up of the preceding discussion.

[9] This could be phrased as saying that the one (meaningless) formula (1.2) defines several (meaningful) Gibbs measures. This resolves the (serious) dispute in the first half of the 20th century on the question whether statistical mechanics could possibly account for phase transitions. See the very amusing citations in the prologue of a recent Thesis [Ue1]

1.2 Regular interactions.

1.2.1. Some topological background.

We will now describe the general framework of spin systems with regular interactions. Our setting will always be lattice systems and our lattice will always be $\mathbb{Z}^d$. Λ will always denote a finite subset of $\mathbb{Z}^d$. Spins will take values in a set $\mathcal{S}_0$ which will always be a complete separable metric space. One could develop the theory in this generality, but to avoid discussions that are not the main concern of the present lectures, I will assume almost always that $\mathcal{S}_0$ is a finite set. We equip $\mathcal{S}_0$ with its sigma-algebra generated by the open sets in the metric topology (resp. the discrete topology in the finite case), $\mathcal{F}_0$, to obtain a measure space $(\mathcal{S}_0, \mathcal{F}_0)$. To complete the description of the single-spin space, we add a probability measure ρ_0, the so-called a-priori distribution of the spin. This gives a single-site probability space $(\mathcal{S}_0, \mathcal{F}_0, \rho_0)$.

As discussed in the previous paragraph, we first want to furnish the setting for infinitely many non-interacting spins. To do this we consider the infinite-product space

$$\mathcal{S} \equiv \mathcal{S}_0^{\mathbb{Z}^d} \tag{1.8}$$

which we turn into a complete separable space by equipping it with the *product topology*. This is done by saying that the open sets are generated by the balls $B_{\epsilon,\Lambda}(\sigma)$ defined as

$$B_{\epsilon,\Lambda}(\sigma) \equiv \left\{ \sigma' \in \mathcal{S} | \max_{i\in\Lambda} |\sigma_i - \sigma'_i| < \epsilon \right\} \tag{1.9}$$

where $\sigma \in \mathcal{S}$, $\Lambda \subset \mathbb{Z}^d$, and $\epsilon \in \mathbb{R}_+$. The product topology of a metric space is metrizable, and $\mathcal{S}$ is a complete separable metric space if $\mathcal{S}_0$ is. The Borel sigma-algebra of $\mathcal{S}$, $\mathcal{F}$, is the product sigma-algebra

$$\mathcal{F} = \mathcal{F}_0^{\mathbb{Z}^d} \tag{1.10}$$

An important fact is *Tychonov's theorem*:

Theorem 1.1: *If $\mathcal{S}_0$ is a compact then the space $\mathcal{S}$ equipped with the product topology is compact.*

We will use the notation $\mathcal{S}_\Lambda \equiv \mathcal{S}_0^\Lambda$ and $\mathcal{F}_\Lambda = \mathcal{F}_0^\Lambda$ for finite-volume configuration space and the sigma-algebra of local events. Note that we identify $\mathcal{F}_\Lambda \subset \mathcal{F}$ with the sub-sigma-algebra of events depending only on the co-ordinates $\sigma_i, i \in \Lambda$. We will call an event that is measurable with respect to $\mathcal{F}_\Lambda$ for some finite Λ a *local* event, or a *cylinder* event. A sequence of volumes $\Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_n \subset \cdots \subset \mathbb{Z}^d$ of volumes with the property that for any finite $\Lambda' \subset \mathbb{Z}^d$, there exists n such that $\Lambda' \subset \Lambda_n$ will be called an *increasing and absorbing* sequence. The corresponding family of sigma-algebras $\mathcal{F}_{\Lambda_n}$ then forms a filtration of the sigma-algebra $\mathcal{F}$. Similarly, $\mathcal{S}_{\Lambda^c} \equiv \mathcal{S}_0^{\mathbb{Z}^d\setminus\Lambda}$ and $\mathcal{F}_{\Lambda^c} \equiv \mathcal{F}_0^{\mathbb{Z}^d\setminus\Lambda}$. A special rôle will be played later by the so-called "tail sigma-algebra" $\mathcal{F}^t \equiv \cap_{\Lambda\subset\mathbb{Z}^d}\mathcal{F}_{\Lambda^c}$. The events in $\mathcal{F}^t$ will be called tail-events or non-local events.

We will refer to various spaces of (real valued) functions on $\mathcal{S}$ in the sequel. In the physical terminology, such functions are sometimes referred to as *observables*. The largest space one usually considers is $B(\mathcal{S},\mathcal{F})$, the space of bounded, measurable functions. (Recall that a function f from a measure space $\mathcal{S}$ into the real numbers is called measurable, if for any Borel set $B \subset \mathcal{B}(\mathbb{R})$, the set $\mathcal{A} \equiv \{\sigma : f(\sigma) \in B\}$ is contained in $\mathcal{F}$).

Correspondingly, we write $B(\mathcal{S},\mathcal{F}_\Lambda)$ for bounded functions measurable with respect to $\mathcal{F}_\Lambda$, i.e. depending only on the values of the spins in Λ. Functions that are in some $B(\mathcal{S},\mathcal{F}_\Lambda)$ are called *local* or *cylinder* functions; we denote their space by

$$B_{\mathrm{loc}}(\mathcal{S}) \equiv \cup_{\Lambda\subset\mathbb{Z}^d} B(\mathcal{S},\mathcal{F}_\Lambda) \tag{1.11}$$

A slight enlargement of the space of local functions are the so-called *quasi-local* functions, $B_{\mathrm{ql}}(\mathcal{S})$; this is the closure of the set of local functions under uniform convergence. Quasi-local functions are characterized by the property that

$$\lim_{\Lambda\uparrow\mathbb{Z}^d} \sup_{\substack{\sigma,\sigma'\in\mathcal{S}\\ \sigma_\Lambda=\sigma'_\Lambda}} |f(\sigma)-f(\sigma')| = 0 \tag{1.12}$$

In the same way one introduces the spaces of continuous, local continuous and quasi-local continuous functions, $C(\mathcal{S})$, $C_{\mathrm{loc}}(\mathcal{S})$, and $C_{\mathrm{ql}}(\mathcal{S})$.

The reader should be warned that in general (i.e. under the hypothesis that $\mathcal{S}_0$ is just a complete separable metric space), neither are all quasi-local functions continuous nor all continuous functions quasi-local (see e.g. [vEFS] for nice examples). However, under stronger hypothesis on $\mathcal{S}_0$, the different spaces acquire relations:

Lemma 1.2:

(i) If $\mathcal{S}_0$ is compact, then $C(\mathcal{S}) = C_{\mathrm{ql}}(\mathcal{S}) \subset B_{\mathrm{ql}}(\mathcal{S})$.

(ii) If $\mathcal{S}_0$ is discrete , then $B_{\mathrm{ql}}(\mathcal{S}) = C_{\mathrm{ql}}(\mathcal{S}) \subset C(\mathcal{S})$.

(iii) If $\mathcal{S}_0$ is finite, then $C(\mathcal{S}) = B_{\mathrm{ql}}(\mathcal{S}) = C_{\mathrm{ql}}(\mathcal{S})$.

Proof. Left as Exercise 2. $\diamondsuit$

Remark. Since we are mostly interested in finite-spin spaces, quasi-locality will be the essential aspect of continuity in the product topology.

We can now turn to the space $\mathcal{M}_1(\mathcal{S}, \mathcal{F})$ of probability measures on $(\mathcal{S}, \mathcal{F})$ and its topological structure. There are several possibilities to equip this space with a topology. The most convenient and commonly used one is that of weak convergence with respect to continuous functions. This topology is generated by the open balls

$$B_{f,\epsilon}(\mu) \equiv \left\{\mu' \in \mathcal{M}_1(\mathcal{S}, \mathcal{F}) \middle| |\mu(f) - \mu(f')| < \epsilon\right\} \tag{1.13}$$

where $f \in \mathbb{C}(\mathcal{S}), \epsilon \in \mathbb{R}_+, \mu \in \mathcal{M}_1(\mathcal{S}, \mathcal{F})$. The main advantage of this topology is that it turns $\mathcal{M}_1(\mathcal{S}, \mathcal{F})$ into a complete separable metric space, and moreover, if $\mathcal{S}_0$ is compact, then $\mathcal{M}_1(\mathcal{S}, \mathcal{F})$ is compact. (Exercise: Prove this using Tychonov's theorem.)

1.2.2. Interactions, local specifications, Gibbs measures.

We can now define a very large class of Hamiltonians for which the Gibbsian theory can be set up. We begin by defining the concept of an interaction.

Definition 1.3: *An interaction is a family $\Phi \equiv \{\Phi_A\}_{A \subset \mathbb{Z}^d}$ where $\Phi_A \in B(\mathcal{S}, \mathcal{F}_A)$. If all $\Phi_A \in C(\mathcal{S}, \mathcal{F}_A)$, the interaction is called continuous.*

An interaction is called regular, if for all $x \in \mathbb{Z}^d$, there exists a constant c, such that

$$\sum_{A \ni x} \|\Phi_A\|_\infty \leq c < \infty \tag{1.14}$$

Remark. What we call 'regular' interaction is called 'absolutely summable' interaction in Georgii's book [Ge1]. In most of the standard literature one finds the stronger condition that

$$|||\Phi||| \equiv \sup_{x \in \mathbb{Z}^d} \sum_{A \ni x} \|\Phi_A\|_\infty < \infty \tag{1.15}$$

With this definition the set of all regular interactions equipped with the norm $|||\cdot|||$ forms a Banach space, $\mathcal{B}_0$, while the weaker condition we use makes the set of regular interactions only into a Fréchet space [Ge1]. In the case of translation invariant interactions, both conditions coincide. However, in the case of random systems the stronger condition (1.15) would introduce some unnatural restrictions on the class of admissible interactions.

Remark. Unbounded interactions occur naturally in two settings: in the case of non-compact state space (e.g. "Gaussian models") or as so called "hard-core" exclusions to describe models in which certain configurations are forbidden (e.g. so called "subshifts of finite type"). While some of such models can be treated quite well, they require special work and we will not discuss them here.

From a regular interaction one now constructs a Hamiltonian by setting, for all finite volumes $\Lambda \subset \mathbb{Z}^d$,

$$H_\Lambda(\sigma) \equiv - \sum_{A \cap \Lambda \neq \emptyset} \Phi_A(\sigma) \tag{1.16}$$

If Φ is in $\mathcal{B}_0$, H_Λ is even guaranteed to satisfy the bound

$$\|H_\Lambda\|_\infty \leq C|\Lambda| \tag{1.17}$$

for some $C < \infty$. Moreover, it is easy to check that H_Λ is a quasi-local function, and if Φ is continuous, even a continuous quasi-local function, for any finite Λ.

The Hamiltonians defined in this way have all the nice properties of the Ising Hamiltonian defined in Section 1.1, and we can proceed to use them to construct Gibbs measures. We begin with the definition of what we will now call *local specifications*:

Definition 1.4: *A* local specification *for Φ is a family of probability kernels $\left\{\mu^{(\cdot)}_{\Lambda,\beta}\right\}_{\Lambda\subset\mathbb{Z}^d}$ such that*

(i) for all Λ and all $\mathcal{A}\in\mathcal{F}$, $\mu^{(\cdot)}_{\Lambda,\beta}(\mathcal{A})$ is a $\mathcal{F}_{\Lambda^c}$-measurable function.

(ii) For any $\eta\in\mathcal{S}$, $\mu^{\eta}_{\Lambda,\beta}$ is a probability measure on $(\mathcal{S},\mathcal{F})$.

(iii) For any pair of volumes Λ,Λ' with $\Lambda\subset\Lambda'$ and any measurable function f

$$\int \mu^{\eta}_{\Lambda',\beta}(d\sigma')\mu^{(\eta_{\Lambda'^c},\sigma'_{\Lambda'})}_{\Lambda,\beta}(d\sigma) f((\sigma_\Lambda,\sigma_{\Lambda'\setminus\Lambda},\eta_{\Lambda'^c})) = \int \mu^{\eta}_{\Lambda',\beta}(d\sigma') f((\sigma_{\Lambda'},\eta_{\Lambda'^c})) \tag{1.18}$$

The most important point is that local specifications satisfy compatibility conditions analogous to conditional expectations. Given a regular interaction, we can now construct local specifications for the Gibbs measures to come.

Lemma 1.5: *If Φ is a regular interaction, then the formula*

$$\int \mu^{\eta}_{\Lambda,\beta}(d\sigma)f(\sigma) \equiv \int \rho_\Lambda(d\sigma_\Lambda)\frac{e^{-\beta H_\Lambda((\sigma_\Lambda,\eta_{\Lambda^c}))}}{Z^{\eta}_{\Lambda,\beta}} f((\sigma_\Lambda,\eta_{\Lambda^c})) \tag{1.19}$$

defines a local specification, called the Gibbs specification for the interaction Φ at inverse temperature β.

Proof. Left as an exercise. The crucial point is that we have (1.6).$\diamondsuit$

We will use a shorthand notation for relations like (1.18) and symbolize this equation by

$$\mu^{(\cdot)}_{\Lambda',\beta}\mu^{(\cdot)}_{\Lambda,\beta} = \mu^{(\cdot)}_{\Lambda',\beta} \tag{1.20}$$

Lemma 1.5 shows that local specifications are “conditional expectations waiting for a measure”; thus nothing is more natural than to define infinite-volume Gibbs measures as follows:

Definition 1.6: *Let $\left\{\mu_{\Lambda,\beta}^{(\cdot)}\right\}$ be a local specification. A measure μ_β is called compatible with this local specification, if and only if, for all $\Lambda \subset \mathbb{Z}^d$ and all $\mathcal{A} \in \mathcal{F}$,*

$$\mu_\beta\left(\mathcal{A} \middle| \mathcal{F}_{\Lambda^c}\right) = \mu_{\Lambda,\beta}^{(\cdot)}(\mathcal{A}), \quad \mu_\beta - \text{a.s.} \tag{1.21}$$

A measure μ_β that is compatible with the local specification for the regular interaction Φ and a priori measure ρ at inverse temperature β is called a Gibbs measure corresponding to Φ and ρ at inverse temperature β.

Remark. Note that by construction Gibbs measure inherit a remarkable property: their conditional distributions given $\mathcal{F}_{\Lambda^c}$ exist for all η, and not only for μ-almost all η, as is usually required of conditional distributions. On the other hand, this observation also hints towards possible generalizations of the constructions beyond the context of regular interactions by weakening the requirement that local specifications be defined for all $\eta \in \mathcal{S}$. The associated concepts of weaker notions of Gibbs measures are currently under active debate, see e.g. [MMR,DS].

Theorem 1.7: *A probability measure μ_β is a Gibbs measure for Φ, ρ, β, if and only if, for all $\Lambda \subset \mathbb{Z}^d$,*

$$\mu_\beta \mu_{\Lambda,\beta}^{(\cdot)} = \mu_\beta \tag{1.22}$$

Proof. Obviously, (1.22) holds if $\mu_{\Lambda,\beta}^{(\cdot)}$ is the conditional probability for μ_β. We only have to show the converse. But for any Λ, and any $\mathcal{A} \in \mathcal{F}$,

$$\mu_\beta(\mathcal{A}) = \mathbb{E}_{\mu_\beta} \mu_\beta\left(\mathcal{A} \middle| \mathcal{F}_{\Lambda^c}\right) \tag{1.23}$$

Inserting this in the right hand side of (1.22) and comparing with the left hand side yields

$$\mu_{\Lambda,\beta}^{(\cdot)}(\mathcal{A}) = \mu_\beta\left(\mathcal{A} \middle| \mathcal{F}_{\Lambda^c}\right), \quad \mu_\beta - \text{a.s.} \tag{1.24}$$

This proves the theorem.$\diamondsuit$

Eq. (1.22) are called the DLR equations after Dobrushin, Lanford and Ruelle, to whom this construction is due. We have now achieved a rigorous definition of what the symbolic expression (1.2) is supposed to mean. Of course this should be completed by an observation saying that such Gibbs measures exist in typical situations. This will turn out to be rather easy.

Theorem 1.8: *Let Φ be a continuous regular interaction and $\mu_{\Lambda,\beta}^{(\cdot)}$ be a corresponding local specification. Let Λ_n be and increasing and absorbing sequence of finite volumes. If for some $\eta \in \mathcal{S}$, the sequence $\mu_{\Lambda_n,\beta}^{\eta}$ of measures converges weakly to some probability measure ν, then ν is a Gibbs measure w.r.t. to Φ, ρ, β.*

Proof. Let f be a continuous function. By hypothesis, we have that

$$\mu_{\Lambda_n,\beta}^{\eta}(f) \to \nu(f), \quad \text{as } n \uparrow \infty \tag{1.25}$$

On the other hand, for all $\Lambda_n \supset \Lambda$,

$$\mu_{\Lambda_n,\beta}^{\eta}\mu_{\Lambda,\beta}^{(\cdot)}(f) = \mu_{\Lambda_n,\beta}^{\eta}(f) \tag{1.26}$$

We would like to assert that $\mu_{\Lambda_n,\beta}^{\eta}\mu_{\Lambda,\beta}^{(\cdot)}(f)$ converges to $\nu\mu_{\Lambda,\beta}^{(\cdot)}(f)$, since this would immediately imply that ν satisfies the DLR equations (1.22) and hence is a Gibbs measure. To be able to make this assertion, we would need to know that $\mu_{\Lambda,\beta}^{(\cdot)}(f)$ is a continuous function. The property of a specification to map continuous functions to continuous functions is called the *Feller property.*

Lemma 1.9: *The local specifications of a continuous regular interaction have the Feller property.*

Proof. We must show that if $\eta_n \to \eta$, then $\mu_{\Lambda,\beta}^{\eta_n}(f) \to \mu_{\Lambda,\beta}^{\eta}(f)$. A rather simple consideration shows that since f is continuous, this property follows if

$$H_\Lambda(\sigma_\Lambda, \eta_{n,\Lambda^c}) \to H_\Lambda(\sigma_\Lambda, \eta_{\Lambda^c}) \tag{1.27}$$

But H_Λ is by assumption a uniformly convergent sum of continuous functions, so it is itself continuous. Then (1.27) is immediate. $\diamondsuit$

The proof of Theorem 1.8 is now obvious.$\diamondsuit$

Remark. Local specifications have even nicer properties than Feller. In particular, they are "quasi-local", in the sense that they map local functions into quasi-local functions. This is expanded on in [vEFS]. Exercise: Prove the quasi-locality of local specifications. This gives also occasion to fill in the details in the proof of Lemma 1.9.

The constructive criterion of Theorem 1.8 gives us now a cheap existence result:

Corollary 1.10: *Assume that $\mathcal{S}_0$ is compact and Φ is regular and continuous. Then there exists at least one Gibbs measure for any $0 \leq \beta < \infty$.*

Proof. By Tychonov's theorem $\mathcal{S}$ is compact. The set of probability measures on a compact space is compact with respect to the weak topology, and so any sequence $\mu^{\eta}_{\Lambda_n,\beta}$ must have convergent subsequences. Any one of them provides a Gibbs measure by Theorem 1.8. $\diamondsuit$

Remark. There are models with non-compact state space for which no Gibbs measure exists.

Theorem 1.8 is of absolutely central importance in the theory of Gibbs measures, since it gives a way how to construct infinite-volume Gibbs measures. Physicists would view this even as the definition of infinite-volume Gibbs measures (and we will have to return to this attitude later when we discuss mean field models). The procedure of taking increasing sequences of finite-volume measures is called the passing to the "thermodynamic limit". It is instructive to compare the physical "approximation" statement contained in the DLR equations and in the weak limit construction. The DLR equations can be interpreted in the sense that if we consider a physical finite system, when we apply "boundary conditions[10] and weigh these with the infinite-volume measure μ_β, then the finite-volume measure within Λ will look exactly like the infinite-volume measure $\mu_{\Lambda,\beta}$. On the other hand, the constructive criterion of Theorem 1.7 means that there are suitable configurations η and suitable volumes Λ, such that if we fix boundary conditions η, the finite-volume measure looks, for large Λ, very much like a infinite-volume Gibbs state ν. Now it is experimentally not very

[10] In the formal discussion we fixed configurations in the entire complement of Λ. Of course for models with short range interactions, like the Ising model, the inside of a volume Λ depends only on the configuration on a layer of width one around Λ. Thus it is physically feasible to emulate the effect of the exterior of Λ by just boundary conditions.

feasible to apply boundary conditions weighted according to some Gibbs measure, while the second alternative seems a bit more realistic. But here difficulties will arise if the dependence on the boundary conditions and on the volumes is too dramatic. There will be no serious problems in simple systems, but we will have occasion to discuss to what extent the Gibbs measures as defined above are the suitable objects to approximately describe real physical situations.

Let us note that there is a different approach that characterizes Gibbs measures in terms of a *variational principle*. Such characterizations always carry a philosophical appeal as they appear to justify the particular choice of Gibbs measures as principle objects of interest. Excellent references are again [Ge1] or [Sim], but also [Isr], and the recent lecture notes by Ch. Pfister [Pf]. Although several important notions linking statistical mechanics, thermodynamics and the theory of large deviations arise in this context, I will not pursue this theme here.

1.3. Structure of Gibbs measures; phase transitions.

In the previous section we have established the concept of infinite-volume Gibbs measures and established the existence of such measures for a large class of systems. The next natural question is to understand the circumstances under which for a given interaction and a given temperature there exists a unique Gibbs measure, and when this is not the case. We have already seen that the possibility that the local specifications might be compatible with several Gibbs measures is precisely providing for the possibility to describe phase transitions in this framework, and therefore this will be the case that we shall be most interested in. Nonetheless, it is important to understand under what conditions one must expect uniqueness. For this reason we start our discussion with some results on uniqueness conditions.

1.3.1. High temperatures. The Dobrushin uniqueness criterion.

In a certain sense one should expect that as a rule a local specification is compatible only with one Gibbs measure. But there are specific interactions (or specific values of the parameters of an interaction) where this rule is violated. However, there are general conditions that preclude this degenerate situation; vaguely, these conditions say that βH is "small"; in this case one can see the Gibbs measure as a weak perturbation of the a priori measure ρ. There are several ways of establishing such conditions. Possibly the most elegant one is due to Dobrushin, and we will not resist the temptation to present it here. Our treatment follows closely that given in Simon's book [Si] where the interested reader may find more material.

Let us introduce the total variation distance of two measures ν, μ by

$$\|\nu - \mu\| \equiv 2 \sup_{\mathcal{A} \in \mathcal{F}} |\nu(\mathcal{A}) - \mu(\mathcal{A})| \tag{1.28}$$

Theorem 1.11: *Let $\mu_{\Lambda,\beta}^{(\cdot)}$ be a local specification. Set, for $x, y \in \mathbb{Z}^d$,*

$$\rho_{x,y} \equiv \frac{1}{2} \sup_{\substack{\eta,\eta' \\ \forall_{z \neq x} \eta_z = \eta'_z}} \left\| \mu_{y,\beta}^{\eta} - \mu_{y,\beta}^{\eta'} \right\| \tag{1.29}$$

If $\sup_{y \in \mathbb{Z}^d} \sum_{x \in \mathbb{Z}^d} \rho_{x,y} < 1$, then the local specification is compatible with at most one Gibbs measure.

Proof. For a continuous function f we define its *variation* at x

$$\delta_x(f) = \sup_{\substack{\eta,\eta' \\ \forall_{z \neq x} \eta_z = \eta'_z}} |f(\eta) - f(\eta')| \tag{1.30}$$

and the *total variation*

$$\Delta(f) \equiv \sum_{x \in \mathbb{Z}^d} \delta_x(f) \tag{1.31}$$

We define the set of functions of finite total variation $\mathcal{T} \equiv \{f \in C(\mathcal{S}) | \Delta(f) < \infty\}$. It is easy to check that this set is a dense subset of $C(\mathcal{S})$. The idea of the proof is

i) Show that Δ is a semi-norm and $\Delta(f) = 0 \Rightarrow f = const.$.

ii) Construct a contraction $\mathbb{T}$ with respect to Δ so that any solution of the DLR equations is $\mathbb{T}$-invariant.

Then, it holds that for any solution of the DLR equations, $\mu(f) = \mu(\mathbb{T}f) = \mu(\mathbb{T}^n f) \to c(f)$, independent of which one we choose. But the value on continuous functions determines μ, so all solutions of the DLR equations are identical.

To simplify notation we drop the reference to β in the course of the proof. Let us first establish (ii). We construct the map $\mathbb{T}$. Let $x_1, x_2, \dots, x_n, \dots$ be an enumeration of all points in $\mathbb{Z}^d$ (this implies that x_n must disappear to infinity as $n \uparrow \infty$). Set

$$\mathbb{T} f \equiv \lim_{n \uparrow \infty} \mu_{x_1}^{(\cdot)} \dots \mu_{x_n}^{(\cdot)}(f) \tag{1.32}$$

For any continuous function, the limit in (1.32) exists in norm. (Exercise! Hint: Check the convergence first on loal functions!). This implies that $\mathbb{T}$ maps continuous functions to continuous functions, which is a crucial property we will use.

It is obvious by construction that if μ satisfies the DLR-equation w.r.t. the specification $\mu_\Lambda^{(\cdot)}$, then

$$\mu(\mathbb{T} f) = \mu(f) \tag{1.33}$$

It remains to show that $\mathbb{T}$ is a contraction w.r.t. Δ, if $\sup_{y \in \mathbb{Z}^d} \sum_{x \in \mathbb{Z}^d} \rho_{x,y} \leq \alpha < 1$. In fact we will show that under this hypothesis, $\Delta(\mathbb{T} f) \leq \alpha \Delta(f)$, for any continuous function f. We look at $\delta_x(\mu_y(f))$. Obviously, $\delta_x(\mu_x(f)) = 0$. Now let $x \neq y$. Then

$$\begin{aligned}
\delta_x(\mu_y(f)) \equiv & \sup_{\substack{\eta, \eta' \\ \forall_{z \neq x} \eta_z = \eta'_z}} \left| \mu_y^\eta(f) - \mu_y^{\eta'}(f) \right| \\
= & \sup_{\substack{\eta, \eta' \\ \forall_{z \neq x} \eta_z = \eta'_z}} \left| \int f(\sigma_y, \eta_{y^c}) \mu_y^\eta(d\sigma_y) - \int f(\sigma_y, \eta'_{y^c}) \mu_y^\eta(d\sigma_y) \right. \\
& \left. + \int f(\sigma_y, \eta'_{y^c}) \left(\mu_y^\eta(d\sigma_y) - \mu_y^{\eta'}(d\sigma_y) \right) \right| \\
\leq & \; \delta_x(f) + \sup_{\substack{\eta, \eta' \\ \forall_{z \neq y} \eta_z = \eta'_z}} |f(\eta) - f(\eta')| \sup_{\substack{\eta, \eta' \\ \forall_{z \neq x} \eta_z = \eta'_z}} \sup_{\mathcal{A} \in \mathcal{F}} \left| \mu_y^\eta(\mathcal{A}) - \mu_y^{\eta'}(\mathcal{A}) \right| \\
= & \; \delta_x(f) + \frac{1}{2} \left\| \mu_y^\eta - \mu_y^{\eta'} \right\| \delta_y(f) \\
= & \; \delta_x(f) + \rho_{x,y} \delta_y(f)
\end{aligned} \tag{1.34}$$

Lemma 1.12: *Under the hypothesis* $\sup_{y \in \mathbb{Z}^D} \sum_{x \in \mathbb{Z}^d} \rho_{x,y} \leq \alpha$, *for all* $n \in \mathbb{N}$,

$$\Delta(\mu_{x_1}^{(\cdot)} \dots \mu_{x_n}^{(\cdot)} f) \leq \alpha \sum_{i=1}^n \delta_{x_i}(f) + \sum_{j \geq n+1} \delta_{x_j}(f) \tag{1.35}$$

Proof. By induction. For $n = 0$ (1.35) is just the definition of Δ. Assume that (1.34) holds for n. Then from (1.33)

$$\begin{aligned}
&\Delta(\mu_{x_1}^{(\cdot)} \dots \mu_{x_n}^{(\cdot)} \mu_{x_{n+1}}^{(\cdot)} f) \leq \alpha \sum_{i=1}^{n} \delta_{x_i}(\mu_{x_{n+1}}^{(\cdot)} f) + \sum_{j \geq n+1} \delta_{x_j}(\mu_{x_{n+1}}^{(\cdot)} f) \\
&\leq \alpha \sum_{i=1}^{n} \left[\delta_{x_i}(f) + \rho_{x_i, x_{n+1}} \delta_{x_{n+1}}(f)\right] \\
&+ \sum_{j \geq n+2} \left[\delta_{x_j}(f) + \rho_{x_j, x_{n+1}} \delta_{x_n+1}(f)\right] \\
&= \alpha \sum_{i=1}^{n} \delta_{x_i}(f) + \sum_{i=1}^{\infty} \rho_{x_i, x_{n+1}} \delta_{x_{n+1}}(f) + \sum_{j \geq n+2} \delta_{x_j}(f) \\
&\leq \alpha \sum_{i=1}^{n+1} \delta_{x_i}(f) + \sum_{j \geq n+2} \delta_{x_j}(f)
\end{aligned} \tag{1.36}$$

so that (1.34) holds for $n+1$. This proves the Lemma.$\diamondsuit$

Passing to the limit $n \uparrow \infty$ yields the desired estimate

$$\Delta(\mathbb{T} f) \leq \alpha \Delta(f) \tag{1.37}$$

It remains to be proven that $\Delta(f) = 0$ implies that $f = const.$. In fact we will show that $\Delta(f) \geq \sup(f) - \inf(f)$. Now, since f is continuous, for any $\epsilon > 0$ there exists a finite Λ and a configurations ω^+, ω^- with $\omega^+_{\Lambda^c} = \omega^-_{\Lambda^c}$ such that

$$\begin{aligned}
\sup(f) &\leq f(\omega^+) + \epsilon, \\
\inf(f) &\geq f(\omega^-) - \epsilon
\end{aligned} \tag{1.38}$$

But, using a simple telescopic expansion,

$$f(\omega^+) - f(\omega^-) \leq \sum_{x \in \Lambda} \delta_x(f) \leq \Delta(f) \tag{1.39}$$

Thus, $\sup(f) - \inf(f) \leq \Delta(f) + 2\epsilon$, for all ϵ which implies the claimed bound. This concludes the proof of the theorem. $\diamondsuit\diamondsuit$

For Gibbs specifications with respect to regular interactions, the uniqueness criterion in Dobrushin's theorem becomes

$$\sup_{x \in \mathbb{Z}^d} \sum_{A \ni x} (|A| - 1) \|\Phi_A(\sigma)\|_\infty < \beta^{-1} \tag{1.40}$$

Thus it always applies if the temperature β^{-1} is sufficiently "high".

Exercise: Compute the bound on the temperature for which Dobrushin's criterion applies in the Ising model (1.1).

The techniques of the Dobrushin uniqueness theorem can be pushed farther to get more information about the unique Gibbs measure; in particular it allows to prove decay of correlations. Since this is not of immediate concern for us, we will not go into this. The interested reader is referred to the very clear exposition in Simon's book [Si].

1.3.2. Low temperatures. The Peierls argument.

Having established a condition for uniqueness, it is natural to seek for situations where uniqueness does not hold. As we mentioned earlier, this possibility was disbelieved for a long time and the solid establishment of the fact that such situations occur in reasonable models like the Ising model was one of the triumphs of statistical mechanics.

Contrary to the very general uniqueness criterion, situations with coexisting Gibbs measures are much more evasive and require a case by case study of the respective interactions. There exist a number of tools to investigate this problem in many situations, the most powerful being what is called the *Pirogov–Sinai* theory, but even in its most recent developments it is far from being able to give a reasonably complete answer for a class of interactions as large as, e.g. the regular interactions[11]. An exposition of this theory in any detail goes beyond the scope of these notes.

The basis of most methods to prove the existence of multiple Gibbs states is the *Peierls argument.* We will explain this in the context it was originally derived, the Ising model, and discuss extensions later.

[11] Of course it would be unreasonable to expect such a theory in any general form to exist.

The basic intuition for the large β (low temperature) behaviour of the Ising model is that the Gibbs measure should in this case strongly favour configurations with minimal H. If $h \neq 0$, one sees that there is a unique such configuration, $\sigma_i = \operatorname{sign}(h)$, whereas for $h = 0$ there are two degenerate minima, $\sigma_i \equiv +1$ and $\sigma_i \equiv -1$. It is a natural idea to characterize a configuration by its deviation from such an optimal one. This leads to the concept of the *contour.* We denote by $< i, j >$ an edge of the lattice $\mathbb{Z}^d$ and by $< ij >^*$ the corresponding *dual plaquette*, i.e. the unique $d-1$ dimensional facet that cuts the edge in the middle. We set

$$\Gamma(\sigma) \equiv \left\{ < ij >^* \middle| \sigma_i \sigma_j = -1 \right\} \tag{1.41}$$

$\Gamma(\sigma)$ forms a surface in $\mathbb{R}^d$. The following properties are immediate from the definition:

Lemma 1.13: *Let Γ be the surface defined above, and let $\partial\Gamma$ denote its $d-2$-dimensional boundary.*

(i) $\partial\Gamma(\sigma) = \emptyset$ for all $\sigma \in \mathcal{S}$. Note that $\Gamma(\sigma)$ may have unbounded connected components.

(ii) Let Γ be a surface in the dual lattice such that $\partial\Gamma = \emptyset$. Then there are exactly two configurations, σ and $-\sigma$, such that $\Gamma(\sigma) = \Gamma(-\sigma) = \Gamma$.

Any G can be decomposed into its connected components γ_i. A connected component γ_i is called a contour. For any σ, any contour γ_i satisfies $\partial\gamma_i(\sigma) = \emptyset$. That is, each contour is either a closed or unbounded surface.

The following theorem goes back to Peierls [Pei]. Its rigorous version is due to Dobrushin [D2] and Griffiths [Gri].

Theorem 1.14: *Let μ_β be a Gibbs measure for the Ising model (1.1) with $h = 0$ and ρ the symmetric product measure defined in (1.3). Assume that $d \geq 2$. Then there is $\beta_d < \infty$ such that $\beta > \beta_d$*

$$\mu_\beta \left[\exists_{\gamma \in \Gamma(\sigma): 0 \in \operatorname{int}\gamma} \right] < \frac{1}{2} \tag{1.42}$$

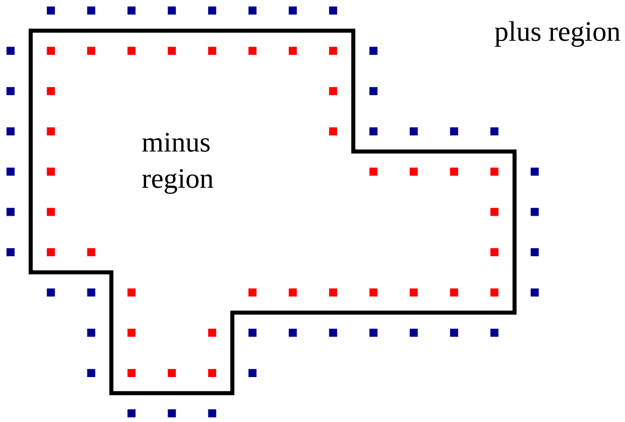
(we write $\gamma \in \Gamma$ for "γ is a connected component of Γ".)

The proof of this theorem is almost immediate from the following

Lemma 1.15: *Let μ_β be a Gibbs measure for the Ising model (1.1) with $h=0$. Let γ be a finite contour. Then*

$$\mu_\beta\left[\gamma \in \Gamma(\sigma)\right] \leq e^{-2\beta|\gamma|} \tag{1.43}$$

Proof. We present the proof again as an application of the DLR construction. Recall that γ is finite and thus closed. We will denote by γ^{in} and γ^{out} the layer of sites in $\mathbb{Z}^d$ adjacent to γ to the interior of γ and the exterior of γ (the interiour and exteriour boundary of the contour).



A contour (solid line) and its interiour (red dots) and exteriour (blue dots) boundary

Apparently we have

$$\mu_\beta\left[\gamma \subset \Gamma(\sigma)\right] \equiv \mu_\beta\left[\sigma_{\gamma^{out}}=+1, \sigma_{\gamma^{in}}=-1\right] + \mu_\beta\left[\sigma_{\gamma^{out}}=-1, \sigma_{\gamma^{in}}=+1\right] \tag{1.44}$$

Now the DLR-equations give

$$\mu_\beta\left[\sigma_{\gamma^{out}}=+1, \sigma_{\gamma^{in}}=-1\right] = \mu_\beta[\sigma_{\gamma^{out}}=+1]\mu^{+1}_{\mathrm{int}\,\gamma,\beta}[\sigma_{\gamma^{in}}=-1] \tag{1.45}$$

But

$$\begin{aligned}
&\mu^{+1}_{\mathrm{int}\,\gamma,\beta}[\sigma_{\gamma^{in}} = -1] \\
&= \frac{\mathbb{E}_{\sigma_{\mathrm{int}\,(\gamma)\setminus\gamma^{in}}}\rho(\sigma_{\gamma^{in}} = -1)e^{-\beta H_{\mathrm{int}\,(\gamma)}(\sigma_{\mathrm{int}\,(\gamma)\setminus\gamma^{in}},-1_{\gamma^{in}},+1_{\gamma^{out}})}}{\mathbb{E}_{\sigma_{\gamma^{in}}}\mathbb{E}_{\sigma_{\mathrm{int}\,(\gamma)\setminus\gamma^{in}}}e^{-\beta H_{\mathrm{int}\,(\gamma)}(\sigma_{\mathrm{int}\,(\gamma)\setminus\gamma^{in}},\sigma_{\gamma^{in}},+1_{\gamma^{out}})}} \\
&= \frac{e^{-\beta|\gamma|}Z^{(-1)}_{\mathrm{int}\,(\gamma)\setminus\gamma^{in}}\rho(\sigma_{\gamma^{in}} = -1)}{\mathbb{E}_{\sigma_{\gamma^{in}}}e^{\beta\sum_{i\in\gamma^{in},j\in\gamma^{out}}\sigma_j}Z^{\sigma_{\gamma^{in}}}_{\mathrm{int}\,(\gamma)\setminus\gamma^{in}}} \\
&\leq e^{-2\beta|\gamma|}\frac{Z^{(-1)}_{\mathrm{int}\,(\gamma)\setminus\gamma^{in}}}{Z^{(+1)}_{\mathrm{int}\,(\gamma)\setminus\gamma^{in}}} \\
&= e^{-2\beta|\gamma|}
\end{aligned} \tag{1.46}$$

Note that in the last line we used the symmetry of H_Λ under the global change $\sigma_i \to -\sigma_i$ to replace the ratio of the two partition functions with spin-flip related boundary conditions by one. In the presence of $h \neq 0$, this would not have been possible. The second term in (1.44) is treated in the same way. Thus (1.43) follows. $\diamondsuit$

Proof of Theorem 1.14 . The proof of the Theorem now follows just by the trivial estimate

$$\mu_\beta\left[\exists_{\gamma\in\Gamma(\sigma):0\in\mathrm{int}\,\gamma}\right] \leq \sum_{\gamma\in\Gamma(\sigma):0\in\mathrm{int}\,\gamma}\mu_\beta[\gamma\in\Gamma(\sigma)] \tag{1.47}$$

and by (roughly) counting the number of contours of area k that enclose the origin:

$$\#\{\gamma : 0 \in \mathrm{int}\gamma, |\gamma| = k\} \leq C_d^k k^{d/(d-1)} \tag{1.48}$$

where C_d is some dimension dependent constant (e.g. it is immediate to see that $C_2 \leq 3$).Thus

$$\mu_\beta\left[\exists_{\gamma\in\Gamma(\sigma):0\in\mathrm{int}\gamma}\right] \leq \sum_{k=2d}^{\infty}k^{d/(d-1)}e^{-k(2\beta-\ln C_d)} \tag{1.49}$$

so choosing β a little larger than $\frac{1}{2}\ln C_d$ we get the claimed estimate. $\diamondsuit$

Theorem 1.14 brings us very close to showing the existence of at least two Gibbs states. Intuitively, it implies that, with probability greater than 1/2, the spin at the origin has the same sign as "the spins at infinity" which in turn could be plus one or minus one. Most importantly, the spin in the origin is correlated to those at infinity, establishing the existence of long-rang correlation. Making these somewhat vague statements precise will give occasion to learn a little more about properties of Gibbs states.

Notice first that Theorem 1.14 does *not* imply that there are no *infinite* contours with positive probability. We will show, however, that μ_β can be decomposed into Gibbs measures containing infinite contours with probability zero and one, respectively.

This leads us to the definition of the important concept of extremal Gibbs measures or *"pure states"*.

By the characterization of Gibbs measures through the DLR-equations it is obvious that with any two Gibbs measures μ_β, μ'_β for the same local specification, their convex combinations $p\mu_\beta + (1-p)\mu'_\beta$, $p \in [0,1]$, are also Gibbs measures. Thus the set of Gibbs measures for a local specification forms a closed convex set. One calls the extremal points of this set *extremal Gibbs measures* or *pure states*[12]

The following gives an important characterization of extremal Gibbs measures.

Proposition 1.16: *A Gibbs measure μ_β is extremal if and only if it is trivial on the tail sigma-field $\mathcal{F}^t$, i.e. if for all $\mathcal{A} \in \mathcal{F}^t$, $\mu_\beta(\mathcal{A}) \in \{0,1\}$.*

To prove this lemma, we need two important observations:

The first says that a Gibbs measure for a given specification is characterized by its value on the tail sigma-field.

Proposition 1.17: *Let μ_β and ν_β be two Gibbs measures for the same specification. If for all $\mathcal{A} \in \mathcal{F}^t$, $\nu_\beta(\mathcal{A}) = \mu_\beta(\mathcal{A})$, then $\nu_\beta = \mu_\beta$.*

Proof. Again we use the DLR equations. Let f be any local function. Since for any

[12] The name pure state is sometimes reserved to translation invariant extremal Gibbs measures.

Λ,

$$\begin{aligned} \mu_\beta(f) &= \mu_\beta\left(\mu_{\beta,\Lambda}^{(\cdot)}(f)\right) \\ \nu_\beta(f) &= \nu_\beta\left(\mu_{\beta,\Lambda}^{(\cdot)}(f)\right) \end{aligned} \tag{1.50}$$

the lemma follows if $\lim_{\Lambda\uparrow\mathbb{Z}^d}\mu_{\beta,\Lambda}^{(\cdot)}(f)$ is measurable with respect to $\mathcal{F}^t$. But by definition, $\mu_{\beta,\Lambda}^{(\cdot)}(f)$ is measurable with respect to $\mathcal{F}_{\Lambda^c}$, and so $\lim_{\Lambda\uparrow\mathbb{Z}^d}\mu_{\beta,\Lambda}^{(\cdot)}(f)$ is measurable with respect to $\cap_{\Lambda\uparrow\mathbb{Z}^d}\mathcal{F}_{\Lambda^c}$, i.e. $\mathcal{F}^t$. $\diamondsuit$

The second observation is

Lemma 1.18: *Let μ be a Gibbs measure, and $\mathcal{A}\in\mathcal{F}^t$ with $\mu(\mathcal{A})>0$. Then the conditioned measure $\mu(\cdot|\mathcal{A})$ is also a Gibbs measure for the same specification.*

Proof. We again consider a local function f. Then

$$\begin{aligned} \mu(f|\mathcal{A}) &\equiv \frac{\mu(f\mathbb{1}_\mathcal{A})}{\mu(\mathcal{A})} \\ &= \frac{\mu\mu_\Lambda^{(\cdot)}(f\mathbb{1}_\mathcal{A})}{\mu(\mathcal{A})} = \frac{\mu\mathbb{1}_\mathcal{A}\mu_\Lambda^{(\cdot)}(f)}{\mu(\mathcal{A})} \\ &= \mu(\mu_\Lambda^{(\cdot)}(f)|\mathcal{A}) \end{aligned} \tag{1.51}$$

for any Λ, so $\mu(\cdot|\mathcal{A})$ satisfies the DLR-equations.$\diamondsuit$

Now we can prove proposition 1.16: Assume that μ is trivial on the tail field and $\mu = p\mu' + (1-p)\mu''$, for $p\in(0,1)$. Then for any $\mathcal{A}\in\mathcal{F}^t$, by Lemma 1.17,

$$p\mu'(\mathcal{A}) + (1-p)\mu''(\mathcal{A}) \in \{0,1\} \tag{1.52}$$

But this can only hold if $\mu'(\mathcal{A}) = \mu''(\mathcal{A}) \in \{0,1\}$, so, in particular, $\nu=\mu$.

To prove the converse, assume that μ is not trivial on the tail field. Then then there exists $\mathcal{A}\in\mathcal{F}^t$ with $\mu(\mathcal{A}) = p\in(0,1)$. So, by Lemma 1.18

$$\mu = p\mu(\cdot|\mathcal{A}) + (1-p)\mu(\cdot|\mathcal{A}^c) \tag{1.53}$$

and, by Lemma 1.18, $\mu(\cdot|\mathcal{A})$ and $\mu(\cdot|\mathcal{A}^c)$ are Gibbs measures, so μ is not extremal. This concludes the proof of the proposition.$\diamondsuit$

Tail field triviality is equivalent to a certain uniform decay of correlations which is a common alternative characterization of extremal Gibbs measures:

Corollary 1.19: *A Gibbs measure μ is trivial on the tail sigma-field if and only if, for all $\mathcal{A} \in \mathcal{F}$,*

$$\lim_{\Lambda \uparrow \mathbb{Z}^d} \sup_{\mathcal{B} \in \mathcal{F}_{\Lambda^c}} |\mu(\mathcal{A} \cup \mathcal{B}) - \mu(\mathcal{A})\mu(\mathcal{B})| = 0 \tag{1.54}$$

We can now return to the investigation of the phase structure of the Ising model. We define the event $\mathcal{U} \equiv \{\Gamma(\sigma)$ contains no infinite contour$\}$. Clearly this is a tail event. Therefore, by Lemma 1.18, if μ is any Gibbs measure, then $\mu(\cdot|\mathcal{U})$ is also a Gibbs measure, provided $\mu(\mathcal{U}) > 0$. But it is easy to see that such μ exist: Take the local specifications with boundary conditions either $\eta \equiv +1$ or $\eta \equiv -1$. They are supported on $\mathcal{U}$, and so any weak limit $\mu^{\pm}$ of these sequences satisfies $\mu^{\pm}(\mathcal{U}) = 1$.

Theorem 1.20: *Consider the Ising model for parameters where the conclusion of Theorem 1.14 hold. Then there exist (at least) two extremal Gibbs measures μ_β^+ and μ_β^- satisfying $\mu^+(\sigma_0) = -\mu^-(\sigma_0) > 0$.*

Proof. We know that there exists a Gibbs measure with $\mu_\beta(\mathcal{U}) = 1$. Now on $\mathcal{U}$, the set of points $x \in \mathbb{Z}^d$ that is not surrounded by a contour (the *exterior of the contour*) is connected and the spin configuration on this set is constant either $+1$ or -1. Clearly the value of the spin on the exterior is a function of the tail sigma-algebra, so if μ_β is extremal it takes either one or the other value with probability one. Let us denote these measures by $\mu_\beta^{\pm}$. Then

$$\mu_\beta^+(\sigma_0 = -1) = \mu_\beta^+\left(\exists_{\gamma \in \Gamma(\sigma)} : 0 \in \operatorname{int} \gamma\right) < \frac{1}{2} \tag{1.55}$$

which implies the theorem.$\diamondsuit$

On a qualitative level we have now solved Ising's problem: The Ising model in dimension two and more has a unique Gibbs state with decaying correlations at high temperatures, while at low temperature there are at least two extremal one that exhibit spontaneous magnetization. Thus the observed phenomenon of a phase transition in ferromagnets is reproduced by this simple system with short range interaction.

I have said earlier that the Peierls argument is the basis of most proofs of the existence of multiple Gibbs states. This is true in the sense that whenever one will prove such a fact, one will want to introduce some notion of contours that characterizes a locally unlikely configuration; one will then want to conclude that "typical" configurations do no contain large regions where configurations are atypical, and finally one will want to use that there are several choices for configurations not containing large undesirable regions. What is lacking then is an argument showing that these "good" regions are equally likely; on a more technical level, this corresponds to being able to pass from the one-but-last line in (1.46) to the last one. In the Ising model we were helped by the spin flip symmetry of the problem. This should be considered accidental, as should be the fact that the ratio of the two partition function appearing in (1.45) equals one. In fact, they equal one because the parameter h was chosen equal to zero. In a situation without symmetry one should expect that there will be some value of h (or other parameters of the model) for which the ratio of the partition function is close enough to one for all γ. This is a subtle issue and at the heart of what is called the *Pirogov–Sinai theory* [PS,Z1,Z2] which in rather general situations allows to establish criteria for the existence of phase-coexistence in lattice spin models. I will not be able to cover this theory in these lectures.

2. Disordered Systems. Generalities and Lattice models.

We will now begin our discussion of disordered systems. In this first part, we will mainly concentrate on disordered systems that are in some sense perturbations of translation invariant systems. From a physical point of view, it is obviously essential to understand the effect of such perturbations, since the hypothesis of perfect lattice symmetry can hardly be expected to be verified in a system consisting of 10^{23} constituents. Impurities and imperfections are omnipresent, and for the results of statistical mechanics to be applicable, they must show some robustness against perturbations. The basic questions are thus: what properties of the translation invariant system are only mildly affected by what types of weak perturbations?

The perturbations we will study here are to model "impurities" and other poorly controlled effects. We will assume them to be "spatially random". In most cases, very simple assumptions on the type of randomness are made, and as these introduce already serious difficulties, no attempts towards great generality are at present reasonable. In fact, the status of the field is that of a few, not overly well understood, simple examples. Nonetheless, it has emerged that a proper probabilistic setup is useful, and we will to some degree insist on this point.

2.1.Random Gibbs measures and metastates.

We will now give a general definition of disordered lattice spin systems. This will not be as general as possible, as we allow disorder only in the interactions, but not in the lattice structure or the spin spaces. Thus as in Section I we consider a lattice $\mathbb{Z}^d$, a single site spin space $(\mathcal{S}_0, \mathcal{F}_0, \nu_0)$ and the corresponding a priori product space $(\mathcal{S}, \mathcal{F}, \nu)$. As a new ingredient, we add a (rich enough) probability space $(\Omega, \mathcal{B}, \mathbb{P})$ where Ω will always be assumed to be a Polish space. On this probability space we construct a *random interaction* as follows:

Definition 2.1: *A random interaction Φ is a family $\{\Phi_A\}_{A\subset\mathbb{Z}^d}$ of random variables on $(\Omega, \mathcal{B}, \mathbb{P})$ taking values in $B(\mathcal{S}, \mathcal{F}_A)$, i.e. measurable maps $\Phi_A : \Omega \ni \omega \to \Phi_A[\omega] \in B(\mathcal{S}, \mathcal{F}_A)$. A random interaction is called regular, if, for $\mathbb{P}$-almost all ω, for any*

$x \in \mathbb{Z}^d$, there exists a finite constant $c[\omega]$, such that

$$\sum_{A \ni x} \|\Phi_A[\omega]\|_\infty \leq c[\omega] < \infty \tag{2.1}$$

A regular random interaction is called continuous if for each $A \subset \Lambda$, Φ_A is jointly continuous in the variables η and ω.

In the present section we discuss only regular random interactions. Some of the most interesting physical systems do correspond to irregular random interactions. In particular, many real spin glasses have a non-absolutely summable interaction, called the RKKY-interaction. See [FZ1-2,Z] for some rigorous results.

Remark. In most examples of interest one assumes that the random interaction has the property that Φ_A and Φ_B are *independent*, if $A \cap B = \emptyset$.

Given a random interaction, it is straightforward to define random finite-volume Hamiltonians $H_\Lambda[\omega]$, as in the deterministic case. Note that for regular random interactions, H_Λ is a random variable taking values in the space $\mathcal{B}_{ql}(\mathcal{S})$, i.e. the mapping $\omega \to H_\Lambda[\omega]$ is measurable. If moreover the Φ_A are continuous functions of ω, then the local Hamiltonians are also continuous functions of ω.

Random local specifications $\mu_{\Lambda,\beta}^{(\cdot)}[\omega]$ are again defined in complete analogy to the deterministic case, i.e.

$$\mu_{\Lambda,\beta}^{(\eta)}[\omega](d\sigma) \equiv \frac{1}{Z_{\beta,\Lambda}^{\eta}[\omega]} e^{-\beta H_\Lambda[\omega](\sigma_\Lambda,\eta_{\Lambda^c})} \rho_\Lambda(d\sigma_\Lambda)\delta_{\eta_{\Lambda^c}}(d\sigma_{\Lambda^c}) \tag{2.2}$$

The important point is that the maps $\omega \to \mu_{\Lambda,\beta}^{(\cdot)}[\omega]$ are again measurable in all appropriate senses. In particular:

Lemma 2.2: *Let Φ be a regular random interaction. Then*

(i) for all $\Lambda \subset \mathbb{Z}^d$ and $\mathcal{A} \in \mathcal{F}$, $\mu_{\beta,\Lambda}^{(\cdot)}(\mathcal{A})$ is measurable function w.r.t. the product sigma-algebra $\mathcal{F}_{\Lambda^c} \times \mathcal{B}$.

(ii) For $\mathbb{P}$-almost all ω, for all $\eta \in \mathcal{S}$, $\mu_{\Lambda,\beta}^{(\eta)}[\omega](d\sigma)$ is a probability measure on $\mathcal{S}$.

(iii) For almost all ω, the family $\left\{\mu_{\beta,\Lambda}^{(\cdot)}[\omega]\right\}_{\Lambda \subset \mathbb{Z}^d}$ is a local specification for the interaction $\Phi[\omega]$ and inverse temperature β.

(iv) If Φ is a continuous regular random interactions, then for any finite Λ, $\mu^{\eta}_{\beta,\Lambda}[\omega]$ is jointly continuous in η and ω.

We now feel ready to define random infinite-volume Gibbs measures. The following is surely reasonable:

Definition 2.3: *A measurable map $\mu_\beta : \Omega \to \mathcal{M}_1(\mathcal{S}, \mathcal{F})$ is called a random Gibbs measure for the regular random interaction Φ at inverse temperature β, if, for almost all ω, $\mu_\beta[\omega]$ is compatible with the local specification $\left\{\mu^{(\cdot)}_{\beta,\Lambda}[\omega]\right\}_{\Lambda\subset\mathbb{Z}^d}$ for this interaction.*

The natural question concerns the existence of such random Gibbs measures. One would expect that again for compact state space, the same argument as in the deterministic situation should work. Now it is indeed obvious that for almost all ω, any sequence $\mu^{\eta}_{\beta,\Lambda_n}[\omega]$ taken along an increasing and absorbing sequence of volumes possesses limit points, and therefore, there exist convergent subsequences $\Lambda_{n[\omega]}$ such that $\mu^{\eta}_{\beta,\Lambda_{n[\omega]}}[\omega]$ converges and the limit is a Gibbs measure for the interaction $\Phi[\omega]$. The non-trivial issue provoked by the fact that the subsequence $\Lambda_n[\omega]$ must in general depend on the realization of the disorder is, whether the measures obtained by this construction depend on ω in a *measurable* way?

This question may first sound like some irrelevant mathematical sophistication, and indeed this problem was mostly disregarded in the literature. To my knowledge this problem was first discussed in a paper by van Enter and Griffiths [vEG] and studied in more detail by Aizenman and Wehr [AW1], but it is the merit of Ch. Newman and D. Stein [NS1,NS2,N] to have brought the intrinsic *physical relevance* of this issue to light. Needless to say the issue arises only when limits along *deterministic* subsequences cannot be constructed, and this could be feared mainly in very strongly disordered systems such as spin-glasses that we will discuss in later sections.

In more pragmatic terms, the construction of infinite-volume Gibbs measures via limits along random subsequences can be criticised by its lack of actual approximative power. An infinite-volume Gibbs measure is supposed to approximate reasonably a very large system under controlled conditions. If however this approximation is only valid for certain very special finite volumes that depend on the specific realization of the disorder, while for other volumes the system is described by other measures, knowledge of just what are the infinite-volume measures is surely not enough, if nothing is known about the relevant subsequences.

As far as proving the existence of random Gibbs measures is concerned, there is a rather simple way out of the random subsequence problem. This goes by extending the local specifications to probability measures $K^{\eta}_{\beta,\Lambda}$ on $\Omega \times \mathcal{S}$ in such a way that the marginal distribution of $K^{\eta}_{\beta,\Lambda}$ on Ω is simply $\mathbb{P}$, while the conditional distribution, given $\mathcal{B}$, is $\mu^{(\eta)}_{\beta,\Lambda}[\omega]$.

Theorem 2.4: *Let Φ be a continuous regular random interaction. Let $K^{(\cdot)}_{\beta,\Lambda}$ be the corresponding measure defined as above. Then*

(i) If for some increasing and absorbing sequence, Λ_n, and some $\eta \in \mathcal{S}$ the weak limit $\lim_{n\uparrow\infty} K^{\eta}_{\beta,\Lambda_n} \equiv K^{\eta}_{\beta}$ exists, then its conditional distribution $K^{\eta}_{\beta}(\cdot|\mathcal{B}\times\mathcal{S})$ given $\mathcal{B}$ is a random Gibbs measure for the interaction Φ.

(ii) If $\mathcal{S}$ is compact, then there exist increasing and absorbing sequences Λ_n such that the hypothesis of (i) is satisfied.

Proof. The proof of this theorem is rather instructive. Let $f \in C(\mathcal{S},\mathcal{F})$ be a continuous function. We must show that

$$K^{\eta}_{\beta}(f|\mathcal{B}\times\mathcal{S})[\omega] = K^{\eta}_{\beta}(\mu^{(\cdot)}_{\beta,\Lambda}[\omega](f)|\mathcal{B}\times\mathcal{S})[\omega] \tag{2.3}$$

Let $\mathcal{B}_k$, $k \in \mathbb{N}$ be a filtration of the sigma-algebra $\mathcal{B}$ where $\mathcal{B}_k$ is generated by the interaction potentials Φ_A with $A \subset \Lambda_k$ with Λ_k some increasing and absorbing sequence of volumes. Note that

$$K^{\eta}_{\beta}(f|\mathcal{B}\times\mathcal{S})[\omega] \equiv \lim_{k\uparrow\infty}\lim_{n\uparrow\infty} K^{\eta}_{\beta,\Lambda_n}(f|\mathcal{B}_k\times\mathcal{S})[\omega] \tag{2.4}$$

Let us denote by $\mathcal{B}_k[\omega]$ the set of all $\omega' \in \Omega$ that have the same projection to $\mathcal{B}_k$ as ω, more formally

$$\mathcal{B}_k[\omega] \equiv \left\{\omega' \in \Omega \,\middle|\, \forall_{\mathcal{A}\in\mathcal{B}_k:\omega\in\mathcal{A}} : \omega' \in \mathcal{A}\right\} \tag{2.5}$$

But for any fixed Λ and n large enough,

$$\begin{aligned}\int_{\mathcal{B}_k[\omega]} \mathbb{P}[d\omega']\mu^{(\eta)}_{\beta,\Lambda_n}[\omega'](f) &= \int_{\mathcal{B}_k[\omega]} \mathbb{P}[d\omega']\mu^{(\eta)}_{\beta,\Lambda_n}[\omega'] \left(\mu^{(\cdot)}_{\beta,\Lambda}[\omega'](f)\right) \\ &= \int_{\mathcal{B}_k[\omega]} \mathbb{P}[d\omega']\mu^{(\eta)}_{\beta,\Lambda_n}[\omega'] \left(\mu^{(\cdot)}_{\beta,\Lambda}[\omega](f)\right) \\ &+ \int_{\mathcal{B}_k[\omega]} \mathbb{P}[d\omega']\mu^{(\eta)}_{\beta,\Lambda_n}[\omega'] \left(\mu^{(\cdot)}_{\beta,\Lambda}[\omega'](f) - \mu^{(\cdot)}_{\beta,\Lambda}[\omega](f)\right)\end{aligned} \tag{2.6}$$

The first term in the last expression converges to $K^{\eta}_{\beta}(\mu^{(\eta)}_{\beta,\Lambda}[\omega](f)|\mathcal{B}\times\mathcal{S})[\omega]$, while for the last we observe that due to the continuity of the local specifications in ω, uniformly in n,

$$\begin{aligned}&\left|\int_{\mathcal{B}_k[\omega]} \mathbb{P}[d\omega']\mu^{(\eta)}_{\beta,\Lambda_n}[\omega'] \left(\mu^{(\eta)}_{\beta,\Lambda}[\omega'](f) - \mu^{(\eta)}_{\beta,\Lambda}[\omega](f)\right)\right| \\ &\leq \sup_{\omega'\in\mathcal{B}_k[\omega]} \sup_{\eta\in\mathcal{S}} \left|\mu^{(\eta)}_{\beta,\Lambda}[\omega'](f) - \mu^{(\eta)}_{\beta,\Lambda}[\omega](f)\right| \downarrow 0,\end{aligned} \tag{2.7}$$

as $k\uparrow\infty$. This proves the Theorem.$\diamondsuit$

Theorem 2.4 appears to solve our problems concerning the proper Gibbsian set-up for random systems. We understand what a random infinite-volume Gibbs measure is and we can prove their existence in reasonable generality. Moreover, there is a constructive procedure that allows us to obtain such measures through a procedure of taking infinite-volume limits. However, upon closer inspection, the construction is not quite as satisfactory as it seems. The unsatisfactory point lies actually hidden in equation (2.4) that tells us what conditioning on $\mathcal{B}$ actually amounts to. In all the examples of interest, the space Ω will itself be some infinite product space $\Omega=\Omega_0^{\mathbb{Z}^d}$, and will be equipped with the product topology. The filtration $\mathcal{B}_k$ will then consist of the Borel-field of $\Omega_0^{\Lambda_k}$ for some increasing and absorbing sequence of finite volumes Λ_k. That is, the measures $K^{\eta}_{\beta}(\cdot|\mathcal{B}_k\times\mathcal{S})$ are actually averages of Gibbs measures over the values of the random interactions outside a finite region Λ_k, and so their limit still contains an averaging over the realization of the disorder "at infinity". This manifests itself in the fact that the measures $K^{\eta}_{\beta}(\cdot|\mathcal{B}\times\mathcal{S})$ will often be mixed states. In particular, this state will not actually describe the result of the observations of one sample of the material at given conditions, but rather the average over many samples that have been prepared to look alike locally. This is clearly not a very physical situation.

While we have come to understand that it may not be realistic to construct a state that predicts the outcome of observations on a single (infinite) sample, it would already be more satisfactory to obtain a probability distribution for these predictions (i.e. a random probability measure) rather than just a mean prediction (and average over probability measures). This led Aizenman and Wehr [AW1] and more emphatically Newman and Stein [NS1] to an extension of the preceding construction to a measure-valued setting. That is, rather than to consider measures on the space $\Omega \times \mathcal{S}$, they introduced measures $\mathcal{K}^{\eta}_{\beta,\Lambda}$ on the space $\Omega \times \mathcal{M}_1(\mathcal{S})$, defined in such a way that the marginal distribution of $\mathcal{K}^{\eta}_{\beta,\Lambda}$ on Ω is again $\mathbb{P}$, while the conditional distribution, given $\mathcal{B}$, is $\delta_{\mu^{(\eta)}_{\beta,\Lambda}[\omega]}$, the Dirac-measure concentrated on the corresponding local specification. We will introduce the symbolic notation

$$\mathcal{K}^{\eta}_{\beta,\Lambda} \equiv \mathbb{P} \times \delta_{\mu^{(\eta)}_{\beta,\Lambda}[\omega]} \tag{2.8}$$

One has the following analogue of Theorem 2.4:

Theorem 2.5: *Let Φ be a continuous regular random interaction. Let $K^{(\cdot)}_{\beta,\Lambda}$ be the corresponding measure defined as above. Then*

(i) If for some increasing and absorbing sequence, Λ_n, and some $\eta \in \mathcal{S}$ the weak limit $\lim_{n\uparrow\infty} \mathcal{K}^{\eta}_{\beta,\Lambda_n} \equiv \mathcal{K}^{\eta}_{\beta}$ exists, then its conditional distribution $\mathcal{K}^{\eta}_{\beta}(\cdot|\mathcal{B} \times \mathcal{S})$ given $\mathcal{B}$ is a probability distribution on $\mathcal{M}_1(\mathcal{S})$ that, for almost all ω, gives full measure to the set of infinite-volume Gibbs measures corresponding to the interaction $\Phi[\omega]$ at inverse temperature β. Moreover,

$$K^{\eta}_{\beta}(\cdot|\mathcal{B} \times \mathcal{S}) = \mathcal{K}^{\eta}_{\beta}(\mu|\mathcal{B} \times \mathcal{S}) \tag{2.9}$$

(ii) If $\mathcal{S}$ is compact, then there exist increasing and absorbing sequences Λ_n such that the hypothesis of (i) is satisfied for any η.

Remark. The conditional measure

$$\kappa^{\eta}_{\beta} \equiv \mathcal{K}^{\eta}_{\beta}(\cdot|\mathcal{B} \times \mathcal{S}) \tag{2.10}$$

is called the Aizenman-Wehr *metastate* (following the suggestion of Newman and Stein [NS1]).

Proof. A proof of this theorem can be found in [N]. Here I will give a simple proof following [AW1]. Note that the assertion (i) will follow if for any bounded continuous function $f : \mathcal{S} \to \mathbb{R}$, and any finite $\Lambda \subset \mathbb{Z}^d$, we can show that

$$\mathbb{E} \int \mathcal{K}_\beta^\eta(d\mu|\mathcal{B} \times \mathcal{S})(\omega) \left| \mu(f) - \mu\left(\mu_{\beta,\Lambda}^{(\cdot)}[\omega](f)\right) \right| = 0 \tag{2.11}$$

But the left hand side clearly equals

$$\int \mathcal{K}_\beta^\eta(d\mu, d\omega) \left| \mu(f) - \mu\left(\mu_{\beta,\Lambda}^\eta[\omega](f)\right) \right| \tag{2.12}$$

Now $\mu(g)$ is trivially a continuous function of μ if g is continuous. By Lemma 1.9, $\mu_{\beta,\Lambda}^\eta[\omega](f)$ is continuous in f whenever $\Phi[\omega]$ is regular and continuous, i.e. for almost all ω. Thus, both $\mu(f)$ and $\mu\left(\mu_{\beta,\Lambda}^\eta[\omega](f)\right)$ are continuous in μ, and hence the integrand in (2.11) is a bounded continuous function of μ and ω. But then, by definition, the left-hand side of (2.12) is given by the limit

$$\begin{aligned} &\lim_{n\uparrow\infty} \int \mathcal{K}_{\beta,\Lambda_n}^\eta(d\mu, d\omega) \left| \mu(f) - \mu\left(\mu_{\beta,\Lambda}^\eta[\omega](f)\right) \right| \\ &= \lim_{n\uparrow\infty} \mathbb{E} \left| \mu_{\beta,\Lambda_n}^\eta[\omega](f) - \mu_{\beta,\Lambda_n}^\eta[\omega]\left(\mu_{\beta,\Lambda}^\eta[\omega](f)\right) \right| \end{aligned} \tag{2.13}$$

But the first term in the last line is equal to zero as soon as n is so large that $\Lambda \subset \Lambda_n$ which implies that (2.11) holds. Assertion (ii) follows by compactness. $\diamondsuit$

At this stage the reader may rightly hold his breath and ask the question whether all this abstract formalism is really necessary, or whether in reasonable situations, we will not get away without all of this? To answer this question, we need to look at specific results, and above all, at examples.

2.2. Remarks on uniqueness conditions.

As in the case of deterministic interactions, having established existence of Gibbs states, the next basic question is that of uniqueness. As in the deterministic case, uniqueness conditions can be formulated in a rather general setting that amounts to say that the interaction is "weak" enough. Indeed, Theorem 1.10 can of course be applied directly for any given realization of the disorder. However, a simple application of such a criterion will not capture any of the particularities of a disordered system, and will therefore give no or just bad answers in most interesting examples. The simple reason for this lies in the fact that the criterion of Theorem 1.10 is formulated in terms of a supremum over $y \in \mathbb{Z}^d$; in a translation invariant situation, this is not bothersome. However, in a random system, we will often find that while for most points y the condition will be satisfied, there may exist rare random points where it is false. Extensions of Dobrushin's criteria have been developed by van den Berg and Maes [vdBM] and are further developed by Gielis [Gi]. I will not pursue them here. Uniqueness for weak interactions in the class of regular interactions can be proven with the help of cluster expansion techniques rather easily. The best results in this direction are due to A. Klein and Masooman [KM], although the basic ideas go back to Berretti [Be].

It should be pointed out that the really interesting problems in high-temperature disordered systems concern the case on non-regular interactions. The most interesting work here is that of Fröhlich and Zegarlinski [FZ1,FZ2] who prove uniqueness (in a weak sense) for mean zero square integrable interactions in the Ising case.

2.3. Phase transitions.

Of course, the most interesting questions in disordered systems concern again the case of non-uniqueness of Gibbs measures, i.e. phase transitions. Already in the case of deterministic models we have seen that there is no real general theory for the classification of the extremal Gibbs states in the low-temperature regime; in the case of disordered systems the situation is considerably worse. Basically, one should distinguish between two situations:

(1) Random perturbations of a deterministic model (whose phase structure is known).

(2) Fully disordered models.

Of course, this distinction is a bit vague. Nonetheless, we say that we are in situation (1) if we can represent the Hamiltonian in the form

$$H[\omega](\sigma) = H^{(0)}(\sigma) + \epsilon H^{(1)}[\omega](\sigma) \tag{2.14}$$

where $H^{(0)}$ is a non-random Hamiltonian (corresponding to a regular interaction) and $H^{(1)}$ is a random Hamiltonian corresponding to a regular random interaction. ϵ plays the rôle of a small parameter. The main question is then whether the phase diagram of H is a continuous deformation of that of $H^{(0)}$ or not. In particular, if $H^{(0)}$ has a first order phase transition in some parameter, will the same be true for H if ϵ is small enough?

There are situations when this question can be answered rather easily; they occur when the different extremal states of $H^{(0)}$ are related by a symmetry group and if for any realisation of the disorder, this symmetry is respected by the random perturbation $H^{(1)}[\omega]$. The classical example of this situation is the

Dilute Ising Model: The Hamiltonian of this model is given by

$$H[\omega](\sigma) = -\sum_{|i-j|=1} J_{ij}[\omega]\sigma_i\sigma_j \tag{2.15}$$

where J_{ij} are i.i.d. random variables taking the values[13] 0 and 1 with probability ϵ and $1-\epsilon$, respectively. It is very easy to see that the Peierls argument (Theorem 1.14) applies with just minor modifications, as was observed in [ARS].

Theorem 2.6: *Let μ_β be a Gibbs measure for the dilute Ising model defined by (2.15) and assume that $d \geq 2$. Then there exists $\epsilon_0 > 0$, such that for all $\epsilon \leq \epsilon_0$, there exists $\beta = \beta(\epsilon) < \infty$, such that for $\beta \geq \beta(\epsilon)$,*

$$\mathbb{P}\left[\mu_\beta[\omega]\left[\exists_{\gamma\in\Gamma(\sigma):0\in\mathrm{int}\gamma}\right] < \frac{1}{2}\right] > 0 \tag{2.16}$$

Proof. Define the random contour energy $E(\gamma)$ by

$$E(\gamma) \equiv \sum_{<ij>^*\in\gamma} J_{ij} \tag{2.17}$$

[13] The precise distribution of the J_{ij} plays of course no rôle for the arguments that follow; it is enough to have $\mathbb{E}J_{ij} = J_0 > 0$, and $\mathrm{var}(J_{ij}) \ll J_0$.

Repeating the proof of Lemma 1.15 mutatis mutandis, one gets immediately the analog

Lemma 2.7: *Let μ_β be a Gibbs measure for the dilute Ising model defined by (2.15). Let γ be a finite contour. Then*

$$\mu_\beta[\omega]\left[\gamma \in \Gamma(\sigma)\right] \leq e^{-2\beta E[\omega](\gamma)} \tag{2.18}$$

But by the law of large numbers, for large γ, $E(\gamma)$ will tend to be proportional to $|\gamma|$; indeed we have that

$$\mathbb{P}\left[E(\gamma) = x|\gamma|\right] = \binom{|\gamma|}{x|\gamma|}(1-\epsilon)^{x|\gamma|}\epsilon^{(1-x)|\gamma|} \tag{2.19}$$

(for $x|\gamma|$ integer). Now define the event

$$\mathcal{A} \equiv \left\{\exists_{\gamma:0\in\text{int}\gamma} : E(\gamma) < |\gamma|/2\right\} \tag{2.20}$$

Then by (2.19), it is an easy exercise to show that

$$\begin{aligned}\mathbb{P}[\mathcal{A}] &\leq \sum_{\gamma:0\in\text{int}\gamma} \mathbb{P}\left[E(\gamma) < |\gamma|/2\right] \\ &\leq \sum_{\gamma:0\in\text{int}\gamma} e^{-|\gamma|\left(\frac{1}{2}\ln\frac{1}{2\epsilon} - \frac{1}{2}\ln 2\right)} \\ &\leq \sum_{k=2d}^{\infty} C_d^k e^{-k\left(\frac{1}{2}\ln\frac{1}{2\epsilon} - \frac{1}{2}\ln 2\right)} \\ &\leq C_d \epsilon^d\end{aligned} \tag{2.21}$$

if ϵ is sufficiently small so that the sum converges. But if $\omega \in \mathcal{A}^c$,

$$\mu_\beta[\exists_{\gamma\in\Gamma(\sigma):0\in\text{int}\gamma}] \leq \sum_{\gamma:0\in\text{int}\gamma} \mu_\beta[\omega]\left[\gamma \in \Gamma(\sigma)\right] \leq \sum_{\gamma:0\in\text{int}\gamma} e^{-\beta|\gamma|} \tag{2.22}$$

which is smaller than $1/2$ if β is large enough. Thus for such β, the event considered in (2.16) holds with probability at least $1 - C_d\epsilon^d$. Of course the $1/2$ in the definition of $\mathcal{A}$ can be replaced by an ϵ-dependent value to improve on the admissible values of β. This proves the theorem. $\diamondsuit$

From Theorem 2.6 one can now deduce the existence of at least two distinct random Gibbs states in the dilute Ising model with weak dilution.

Corollary 2.8: *In the dilute Ising model, for any $d \geq 2$, there exists $\epsilon_0 > 0$, such that for all $\epsilon \leq \epsilon_0$, there exists $\beta(\epsilon) > 0$, such that for all $\beta \geq \beta(\epsilon)$, with probability one, there exist at least two extremal random Gibbs states.*

Proof. Theorem 2.6 implies, by the arguments put forward in Section 1.3.2, that there exists at least two extremal Gibbs states with positive probability. However, the number of extremal Gibbs measures for a given random interaction (with sufficient ergodic properties which are trivially satisfied here) is an almost sure constant([N], Proposition 4.4.).$\diamondsuit$

Of course the Peierls approach indicated here is not giving optimal results (but has the advantage of clarity and robustness). It is known that the maximal ϵ for which $\beta(\epsilon)$ is finite is the critical percolation probability for bond percolation. This has been proven first by Georgii [Ge1] in $d = 2$ and in more generality in [ACCN]. The latter paper also obtains precise results on the dependence of $\beta(\epsilon)$ on ϵ. These results are all based on some profound facts from percolation theory which is a subject we will not exploit in these notes.

The situation when the random perturbation respects the symmetries of the unperturbed interaction for any realization of the disorder must be considered exceptional. In general, the perturbation $H^{(1)}[\omega]$ will break all symmetries of the model for typical ω and thus will render the Peierls argument useless. The simplest example of such models is the *random field Ising model*, whose Hamiltonian is

$$H[\omega](\sigma) \equiv - \sum_{|i-j|=1} \sigma_i \sigma_j - \epsilon \sum_i h_i[\omega] \sigma_i \tag{2.23}$$

with h_i a family of i.i.d. random variables (say of mean zero and variance one). In the 1980's the question of in which dimensions this model would or would not exhibit a first order phase transition was among the most vividly discussed issues in the theory of disordered systems, both on the level of theoretical and experimental physics. The problem was solved at the end of the decade in two rigorous papers by Bricmont and

Kupiainen [BK] (who proved the existence of a phase transition in $d \geq 3$ for small ϵ) and Aizenman and Wehr [AW1] (who showed the uniqueness of the Gibbs state in $d = 2$ for all temperatures).

Most of the heat of the debate arose from the fact that two competing heuristic theories making different predictions on the critical value of the dimension for which a phase transition would occur co-existed. There appeared to be no other way but a rigorous mathematical proof to allow people to chose between the different heuristic considerations[14]. This little piece of history is rather instructive and should warn people against too firm trust in speculative heuristic theories. Disordered systems, more than others, tend to elude common intuition.

The issue of the RFIM being one of the first occurrences where profound probabilistic thinking has entered the field, I will devote the following section to the analysis of this model (although we will fall short from giving the full proof of the results in [BK]).

2.4. The random field Ising model.

2.4.1. The Imry–Ma argument.

The earliest attempt to address the question of the phase transition in the RFIM goes back to Imry and Ma [IM] and is nothing but an attempt to extend the beautiful and simple Peierls argument to a situation with symmetry breaking randomness. Let us recall that the Peierls argument in its essence relies on the observation that in order to deform one ground state (+) in the interior of a contour γ to another ground state (-) costs a "surface energy" $2|\gamma|$, while by symmetry, the "bulk energies" of the two ground states are the same. Since the number of contours of a given length L is only of order C^L, the Boltzmann-factors $e^{-2\beta L}$ suppress such deformations sufficiently to make their existence unlikely, if β is large enough. What goes wrong with the argument in the RFIM is the fact that the bulk energies of the two ground states are

[14] It is interesting to recall that three previous papers [Ch,FFS,I] that reached the same conclusions as [BK] but failed to give fully rigorous proofs ([Ch] and [FFS] considered a somewhat artificial modified model and the proof in [I] worked only a zero temperature) were not considered sufficiently convincing evidence to close the debate.

no longer the same. Indeed, if all σ_i in $\text{int}\,\gamma$ take the value $+1$, then the random field term gives a contribution

$$E_{bulk}(\gamma) = +\epsilon \sum_{i \in \text{int}\,\gamma} h_i[\omega] \tag{2.24}$$

while it is equal to minus the same quantity if all σ_i equal -1. Thus deforming the plus state to the minus state within γ produces, in addition to the surface energy term a bulk term of order $2\epsilon \sum_{i\in\text{int}\,\gamma} h_i[\omega]$ which can take on any sign. Note that even when the random fields h_i are uniformly bounded, this contribution is bounded uniformly only by $2\epsilon|\text{int}\,\gamma|$ in absolute value and thus can be considerably bigger in modulus than the surface term, no matter how small ϵ is, if $|\gamma|$ is sufficiently large. Now Imry and Ma argued that the uniform bound on $E_{bulk}(\gamma)$ should not be the relevant quantity to consider. Rather, they argued, the "typical" value of $E_{bulk}(\gamma)$ would be much smaller, namely, by the central limit theorem,

$$E_{bulk}(\gamma) \sim \pm\epsilon\sqrt{|\text{int}\,\gamma|} \tag{2.25}$$

Since by the classical isoperimetric inequalities $|\text{int}\,\gamma| \leq c|\gamma|^{\frac{d}{d-1}}$, this means that the typical value of the bulk energy is only $E_{bulk}(\gamma) \sim \pm\epsilon|\gamma|^{\frac{d}{2(d-1)}}$, which is small compared to $|\gamma|$ if $d > 2$, while otherwise it is comparable or even larger. This very simple consideration led Imry and Ma to the (*correct!!*) prediction that the RFIM undergoes a phase transition in $d \geq 3$ and does not in $d \leq 2$.

Admittedly, that was a rather sketchy argument, and anyone would be excused for not trusting it. We will thus distance us a bit from Imry and Ma and try to repeat their reasoning in a somewhat more precise way. What we would obviously want to do is to reprove something like Theorem 2.6. When trying to re-run the proof of Lemma 1.15, all works as before until the last line of (1.46). One obtains instead the two bounds

$$\begin{aligned} \mu^{+1}_{\text{int}\,\gamma,\beta}\left[\sigma_{\gamma^{in}} = -1\right] &\leq e^{-2\beta|\gamma|}\frac{Z^{-1}_{\text{int}\,\gamma\setminus\gamma^{in},\beta}}{Z^{+1}_{\text{int}\,\gamma\setminus\gamma^{in},\beta}} \\ \mu^{-1}_{\text{int}\,\gamma,\beta}\left[\sigma_{\gamma^{in}} = +1\right] &\leq e^{-2\beta|\gamma|}\frac{Z^{+1}_{\text{int}\,\gamma\setminus\gamma^{in},\beta}}{Z^{-1}_{\text{int}\,\gamma\setminus\gamma^{in},\beta}} \end{aligned} \tag{2.26}$$

whence the analogue of Lemma 1.15 becomes

Lemma 2.9: *In random field Ising model, for any Gibbs state μ_β,*

$$\mu_\beta\left[\gamma \in \Gamma(\sigma)\right] \leq \exp\left(-2\beta|\gamma| + \left|\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} - \ln Z^{-1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta}\right|\right) \tag{2.27}$$

Arrived at this point, one may be tempted to loose all hope when facing the difference of the logarithm of the two partition functions, and one may not even see any good reason how to arrive at Imry and Ma's assertion on the "typical value" of this bulk term[15] However, the situation is much better than could be feared. We will see that it is indeed easy to prove the following Lemma:

Lemma 2.10: *Assume that the random fields have a symmetric distribution*[16] *and are bounded*[17] *(i.e. $|h_i| \leq 1$). Then there is a constant $C < \infty$ such that for any $z \geq 0$,*

$$\mathbb{P}\left[\left|\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} - \ln Z^{-1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta}\right| > z\right] \leq C\exp\left(-\frac{z^2}{\epsilon^2\beta^2 C|\mathrm{int}\,\gamma|}\right) \tag{2.28}$$

Proof. The key to the proof are what are called *concentration of measure* inequalities. Note that by symmetry of the distribution of h, the two partition functions we consider have, as random variables, the same distribution. In particular,

$$\mathbb{E}\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} = \mathbb{E}\ln Z^{-1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} \tag{2.29}$$

---

[15] Historically, this has indeed been considered to be the truly difficult part of the problem. [Ch] and [FFS] gave a solution of the problem where this difference was ad hoc replaced by the sum over the random fields within $\mathrm{int}\,\gamma$. As we will see, the real difficulty of the problem lies, however, elsewhere.

[16] This assumption appears necessary even for the result; otherwise the phase coexistence point could be shifted to some finite value of the external magnetic field.

[17] We make this assumption for convenience; as a matter of fact essentially the same result holds if we only assume that the h_i have finite exponential moments.

Therefore,

$$
\begin{aligned}
&\mathbb{P}\left[\left|\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} - \ln Z^{-1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta}\right| > z\right]\\
&\leq \mathbb{P}\left[\left|\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} - \mathbb{E}\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta}\right| + \left|\mathbb{E}\ln Z^{-1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} - \ln Z^{-1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta}\right| > z\right]\\
&\leq 2\mathbb{P}\left[\left|\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} - \mathbb{E}\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta}\right| > z/2\right]
\end{aligned}
\tag{2.30}
$$

The point is now that $\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta}$ is a function of the independent random variables h_i, with $i \in \mathrm{int}\,\gamma\backslash\gamma^{in}$; there exists an extensive theory about the fluctuation properties of such functions that has been developed over the last three decades, with important culminating results produced in the last few years by M. Talagrand. Very roughly, these theorems imply that in many situations, a Lipshitz continuous function of i.i.d. random variables has fluctuations that are not bigger than those of a corresponding linear function. We will use the following Theorem, due to M. Talagrand, whose proof can be found in [T1]:

Theorem 2.11: *Let $f : [-1,1]^N \to \mathbb{R}$ be a function whose level sets are convex. Assume that f is Lipshitz with uniform constant C_{Lip}, i.e. for any $X, Y \in [-1,1]^N$,*

$$
|f(X) - f(Y)| \leq C_{Lip}\|X - Y\|_2 \tag{2.31}
$$

Then, if $X_1, \dots X_N$ are i.i.d. random variables taking values in $[-1,1]$, and $Z = f(X_1, \dots, X_N)$, if $\mathbb{M}_Z$ is a median[18] of Z

$$
\mathbb{P}\left[|Z - \mathbb{M}_Z| \geq z\right] \leq 4\exp\left(-\frac{z^2}{16C_{Lip}^2}\right) \tag{2.32}
$$

Remark. In most applications, and in particular when C_{Lip} is small compared to z^2, one can replace the median in (2.32) by the expectation $\mathbb{E}Z$ without harm (Exercise!).

Remark. The theme of concentration of measure will recur in these notes. Physical quantities satisfying such inequalities are often also called "self-averaging".

[18] A median of a random variable Z is any number such that $\mathbb{P}[Z \geq \mathbb{M}_Z] \geq 1/2$ and $\mathbb{P}[Z \leq \mathbb{M}_Z] \geq 1/2$.

Now it is quite an easy matter to verify that $\ln Z$ is always a Lipshitz continuous function. In fact,

$$\begin{aligned}
&\left|\ln Z^{+1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta}[\omega] - \ln Z^{+1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta}[\omega']\right| \\
&\leq \sup_{\omega''}\left|\sum_{i\in\text{int}\,\gamma\backslash\gamma^{in}} (h_i[\omega]-h_i[\omega'])\frac{\partial \ln Z^{+1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta}}{\partial h_i}[\omega'']\right| \\
&\leq \epsilon\beta \sup_{i\in\text{int}\,\gamma\backslash\gamma^{in}} |\mu_{\text{int}\,\gamma\backslash\gamma^{in},\beta}(\sigma_i)| \sum_{i\in\text{int}\,\gamma\backslash\gamma^{in}} |h_i[\omega]-h_i[\omega']| \\
&\leq \epsilon\beta\sqrt{|\text{int}\,\gamma|}\|h_{\text{int}\,\gamma}[\omega]-h_{\text{int}\,\gamma}[\omega']\|_2
\end{aligned} \tag{2.33}$$

where we used in the last step that the expectation of σ_i is of course bounded by one together with the Cauchy–Schwarz inequality. Since moreover $\ln Z$ is a convex function of h, one can apply now Theorem 2.11 to arrive at (2.28).

Lemma 2.10 implies indeed that for typical γ,

$$\mu_\beta\left[\gamma\in\Gamma(\sigma)\right] \leq e^{-2\beta|\gamma|+\epsilon\beta\sqrt{|\text{int}\,\gamma|}} \tag{2.34}$$

However, the immediate attempt to prove the analog of Theorem 2.6 fails, since the typical γ are not what really matters. Namely, to conclude such a result, we would have to show that

$$\mathbb{P}\left[\exists_{\gamma:\text{int}\,\gamma\ni 0}\left|\ln Z^{+1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta} - \ln Z^{-1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta}\right| > \beta|\gamma|\right]$$

is small (for small ϵ). Now

$$\begin{aligned}
&\mathbb{P}\left[\exists_{\gamma:\text{int}\,\gamma\ni 0}\left|\ln Z^{+1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta} - \ln Z^{-1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta}\right| > \beta|\gamma|\right] \\
&\leq \sum_{\gamma:\text{int}\,\gamma\ni 0} \mathbb{P}\left[\left|\ln Z^{+1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta} - \ln Z^{-1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta}\right| > \beta|\gamma|\right] \\
&\leq \sum_{\gamma:\text{int}\,\gamma\ni 0} e^{-\frac{|\gamma|^2}{C\epsilon^2|\text{int}\,\gamma|}}
\end{aligned} \tag{2.35}$$

But $\frac{|\gamma|^2}{|\text{int}\,\gamma|}$ can be as small (and is for many γ) as $|\gamma|^{(d-2)/(d-1)}$, and since the number of γ's of given length is of order $C^{|\gamma|}$, the last sum in (2.35) diverges.

Some reflection shows that it is the first inequality in (2.35) that spoiled the estimates. This step would be reasonable if the partition functions for different γ were more or less independent. However, if γ and γ' are very similar, it is clear that this is not the case. A more careful analysis should exploit this fact and hopefully lead to a better bound. Such situations are quite common in probability theory, and in principle there are well-known techniques that go under the name of *chaining* to systematically improve estimates like (2.35). This was done in the papers [Ch] and [FFS], however in a model where $\ln Z^{+1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta} - \ln Z^{-1}_{\text{int}\,\gamma\backslash\gamma^{in},\beta}$ is ad hoc replaced by $\beta\sum_{i\in\text{int}\,\gamma\backslash\gamma^{in}} h_i$ (the so-called "no contours within contours" approximation). In fact, they prove the following:

Theorem 2.12: [FFS]*Assume in the RFIM that there is a finite positive constant C such that for all $\Lambda,\Lambda'\subset\mathbb{Z}^d$*

$$\mathbb{P}\left[\left|\ln Z^{+1}_{\Lambda,\beta} - \ln Z^{+1}_{\Lambda',\beta} - \mathbb{E}[\ln Z^{+1}_{\Lambda,\beta} - \ln Z^{+1}_{\Lambda',\beta}]\right| \geq z\right] \leq \exp\left(-\frac{z^2}{C\epsilon^2\beta^2|\Lambda\Delta\Lambda'|}\right) \quad (2.36)$$

where $\Lambda\Delta\Lambda'$ denotes the symmetric difference of the two sets Λ and Λ'. Then, if $d\geq 3$, there exists $\epsilon_0>0$, $\beta_0<\infty$, such that for all $\epsilon\leq\epsilon_0$ and $\beta\geq\beta_0$, for $\mathbb{P}$ almost all $\omega\in\Omega$, there exist at least two extremal infinite-volume Gibbs states μ^+_β, and μ^-_β.

Remark. There are good reasons to believe that (2.36) holds, but in spite of multiple efforts, I have not been able to find an easy proof. On a somewhat heuristic level, the argument is that the difference appearing in (2.36) should depend very little on the random variables that appear in the intersection of Λ and Λ'. More precisely, when computing the Lipshitz norm we get instead of (2.33)

$$\begin{aligned}
&\left|\ln Z^{+1}_{\Lambda,\beta}[\omega] - \ln Z^{+1}_{\Lambda,\beta}[\omega'] - \ln Z^{+1}_{\Lambda',\beta}[\omega] + \ln Z^{+1}_{\Lambda',\beta}[\omega']\right| \\
&\leq \sup_{\omega''}\left|\sum_{i\in\Lambda\backslash(\Lambda\cap\Lambda')}(h_i[\omega]-h_i[\omega'])\frac{\partial\ln Z^{+1}_{\Lambda,\beta}}{\partial h_i}[\omega'']\right| \\
&\quad+\left|\sum_{i\in\Lambda'\backslash(\Lambda\cap\Lambda')}(h_i[\omega]-h_i[\omega'])\frac{\partial\ln Z^{+1}_{\Lambda',\beta}}{\partial h_i}[\omega'']\right| +
\end{aligned}$$

$$
\begin{aligned}
+ & \quad \left| \sum_{i\in\Lambda\cap\Lambda'} (h_i[\omega]-h_i[\omega']) \left(\frac{\partial \ln Z^{+1}_{\Lambda,\beta}}{\partial h_i}[\omega''] - \frac{\partial \ln Z^{+1}_{\Lambda',\beta}}{\partial h_i}[\omega''] \right) \right| \\
& \le \epsilon\beta \left| \sum_{i\in\Lambda\Delta\Lambda'} |h_i[\omega]-h_i[\omega']| \right| \\
& \quad + \epsilon\beta \left| \sum_{i\in\Lambda\cap\Lambda'} (h_i[\omega]-h_i[\omega']) \left(\mu^+_{\beta,\Lambda}[\omega''](\sigma_i) - \mu^+_{\beta,\Lambda'}[\omega''](\sigma_i) \right) \right| \\
& \le \epsilon\beta\sqrt{|\Lambda\Delta\Lambda'|}\|h_{\Lambda\Delta\Lambda'}[\omega]-h_{\Lambda\Delta\Lambda'}[\omega']\|_2 \\
& \quad + \epsilon\beta \sqrt{\sum_{i\in\Lambda\cap\Lambda'} \left(\mu^+_{\beta,\Lambda}[\omega''](\sigma_i) - \mu^+_{\beta,\Lambda'}[\omega''](\sigma_i) \right)^2} \|h_{\Lambda\cap\Lambda'}[\omega]-h_{\Lambda\cap\Lambda'}[\omega']\|_2
\end{aligned}
\tag{2.37}
$$

It is natural to expect that the expectation of σ_i with respect to the two measures will be essentially the same for i well in the intersection of Λ and Λ', so that it should be possible to bound the coefficient in the last line by $\sqrt{|\Lambda\Delta\Lambda'|}$. But to do this requires estimates that we currently do not dispose of. The reader should realize that this argument is more convincing and robust than the one given in [FFS]; they argue that (2.36) will hold *if* the $\mu^+_{\beta,\Lambda}[\omega](\sigma_i)$ depend 'weakly' on ω which is essentially what we are out to prove anyway. On the other hand, smallness of (2.37) can even be expected to hold if the expectation of σ_i depends strongly on the disorder.

Even though we still do not know how to prove that the hypothesis of Theorem 2.12 is satisfied, I find it instructive to give some details of the proof.

Proof. To simplify notation, let us set

$$
F_\gamma \equiv \ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} - \mathbb{E}\ln Z^{+1}_{\mathrm{int}\,\gamma\backslash\gamma^{in},\beta} - \tag{2.38}
$$

The idea behind chaining arguments is to define a sequence of sets Γ_k, $\kappa \in \mathbb{N}$ of 'coarse grained' contours and maps $\gamma_k : \Gamma_0 \to \Gamma_k$, for $k > 0$ where Γ_0 is the original set of contours Now write for k to be chosen

$$
F_\gamma = F_{\gamma_k(\gamma)} + (F_{\gamma_{k-1}(\gamma)} - F_{\gamma_k(\gamma)}) + (F_{\gamma_{k-2}(\gamma)} - F_{\gamma_{k-1}(\gamma)}) + \cdots + (F_\gamma - F_{\gamma_1(\gamma)}) \tag{2.39}
$$

Then we can write

$$
\begin{aligned}
\mathbb{P}\left[\sup_{\gamma:|\gamma|=n} F_\gamma > z\right] \leq \sum_{\ell=1}^{k(n)} \mathbb{P}\left[\sup_{\gamma:|\gamma|=n} F_{\gamma_\ell(\gamma)} - F_{\gamma_{\ell-1}(\gamma)} > z_\ell\right] \\
+ \mathbb{P}\left[\sup_{\gamma:|\gamma|=n} F_{\gamma_k(\gamma)} > z_{k+1}\right]
\end{aligned}
\tag{2.40}
$$

for any choice of $k = k(n)$ and sequences z_ℓ with $\sum_{\ell=1}^{k+1} z_\ell \leq z$. Now to estimate the individual probabilities, we just need to count the number $A_{\ell,n}$ of image points in Γ_ℓ that are reached when mapping all the γ occurring in the sup (i.e. those of length n and encircling the origin) and use the assumption to get the obvious estimate

$$
\begin{aligned}
\mathbb{P}\left[\sup_{\gamma:|\gamma|=n} F_\gamma > z\right] \leq \sum_{\ell=1}^{k(n)} A_{\ell-1,n} A_{\ell,n} \exp\left(-\frac{z_\ell^2}{C\epsilon^2\beta^2 \inf_\gamma |\mathrm{int}\,\gamma_\ell(\gamma)\Delta \mathrm{int}\,\gamma_{\ell-1}(\gamma)|}\right) \\
+ A_{k,n} \exp\left(-\frac{z_{k+1}^2}{C\epsilon^2\beta^2 \inf_\gamma |\mathrm{int}\,\gamma_k(\gamma)|}\right)
\end{aligned}
\tag{2.41}
$$

We must now make a choice for the sets Γ_k. For this we cover the lattice $\mathbb{Z}^d$ with squares of side-length 2^k centered at the coarser lattice $(2^k\mathbb{Z})^d$. The set Γ_k will then be nothing but collections of such squares. Next we need to define the maps γ_k. This is done as follows: Let $V_k(\gamma)$ be the set of all cubes, c, of side-length 2^k from the covering introduced above such that

$$
|c \cap \mathrm{int}\,\gamma| \geq \frac{1}{2} 2^{dk} \tag{2.42}
$$

Then let $\gamma_k(\gamma) \equiv \partial V_k(\gamma)$ be the set of boundary cubes of $V_k(\gamma)$. Note that the images $\gamma_k(\gamma)$ are in general not connected, but one verifies that the number of connected components cannot exceed $const.|\gamma|2^{-(d-1)(k-1)}$, and the maximal distance between any of the connected components is less than $|\gamma|$. This leads easily to the estimate that

$$
A_{\ell,n} \leq \exp\left(\frac{C\ell n}{2^{(d-1)\ell}}\right) \tag{2.43}
$$

On the other hand, one readily sees that

$$
|\mathrm{int}\,\gamma_\ell(\gamma) \Delta \mathrm{int}\,\gamma_{\ell-1}(\gamma)| \leq |\gamma| 2^\ell \tag{2.44}
$$

for whatever γ. Finally, one chooses $k(n)$ so that for any γ of length n that encircles the origin, the image $\gamma_{k(n)}$ is empty (i.e. $k(n) \sim \ln n$). Inserting these estimates into (2.42), one concludes that for small enough ϵ the sum in (2.42) is bounded by

$$\sum_{\ell=1}^{k(n)} \exp\left(-\frac{z_\ell^2}{C\beta^2\epsilon^2 2^\ell n} + \frac{C(\ell-1)n}{2^{(d-1)(\ell-1)}} + \frac{C(\ell)n}{2^{(d-1)(\ell)}}\right) \tag{2.45}$$

This allows us to choose for instance $z_\ell = c\epsilon\beta n\ell^{-2}$ to get an overall bound of order

$$\mathbb{P}\left[\sup_{\gamma:|\gamma|=n} F_\gamma > c\epsilon\beta n\right] \leq e^{-cn} \tag{2.46}$$

and hence

$$\mathbb{P}\left[\sup_{\gamma:\mathrm{int}\,\gamma\ni 0} F_\gamma > c\epsilon\beta|\gamma|\right] \leq e^{-c} \tag{2.47}$$

But from here follows the conclusion of Theorem 2.6 and hence the existence of two Gibbs states. $\diamondsuit\diamondsuit$

The only existing full proof of the existence of a phase transition in the RFIM is due to Bricmont and Kupiainen [BK] and requires much harder work. In that approach a detailed analysis of the Gibbs measure for a given realization of the disorder is performed using what is called a *real space renormalization group* method. In this approach a sequence of image measures of the Gibbs measure under a map that maps spin configurations to local averages on blocks (not unlike the blocks considered in the preceding proof) is analyzed, and the properties of this sequence of measures are used to deduce the desired information about the Gibbs measures. In this procedure, the detailed properties of the random fields configurations are analyzed carefully on a sequence of length scales. Technically, this method relies on very carefully performed (partial) cluster expansions, which makes its implementation very difficult and involved. For this reason I cannot present any details in these notes. There is no really good pedagogical exposition of this work, and the reading of the original paper [BK] is a rather demanding task. For readers that are not frightened by a little more complication in the model, I recommend to look at the paper [BoKu1] by myself and Ch. Külske, where the same method is applied in the more complicated random solid-on-solid (SOS) model. In that paper an attempt towards a pedagogical presentation was made, albeit with only partial success. An exposition of the renormalization group method without cluster expansions is given in [BoKu2,BoKu3] in the context of hierarchical SOS models.

2.4.2. Absence of phase transitions: The Aizenman–Wehr method.

We have seen that in $d \geq 3$ the random energy that can be gained by flipping spins locally cannot compensate, if the disorder is weak and the temperature low, the surface energy produced in doing so. On the other hand, in dimension $d \leq 2$, the Imry–Ma argument predicts that the random bulk energies outweigh surface terms, and this should imply that the particular realization of the random fields determine locally the orientation of the spins, and the effects of boundary conditions are not felt in the interior of the system, implying a unique (random) Gibbs state. This argument was made rigorous by Aizenman and Wehr [AW1,AW2] and this required a number of clever and interesting ideas. Roughly, the proof is based on the following reasoning: Consider a volume Λ and two boundary conditions, say all spins plus and all spins minus. Then the difference between the corresponding free energies $f_{\beta,\Lambda^{\pm}} \equiv \ln Z^{\pm}_{\beta,\Lambda}$ must always be bounded by $const.|\partial\Lambda|$ (just introduce a contour right next to the boundary and proceed as in the Peierls argument). Now get a lower bound on the random fluctuations of that free energy; from the upper bound (2.28) one might guess that these can be of order $C(\beta)\sqrt{|\Lambda|}$. If this is so, there is a dilemma: by symmetry, the difference between the two free energies must be as big as the random part, and this implies that $C(\beta)\sqrt{|\Lambda|} \leq const.|\partial\Lambda|$. In $d \leq 2$, this implies that $C(\beta) = 0$. But $C(\beta)$ will be seen to be linked to an *order parameter*, here the *magnetization*, and its vanishing will imply the uniqueness of Gibbs state. To make this rough argument precise requires, however, a rather delicate procedure. In what follows I will give the proof of Aizenman and Wehr only for the special case of the RFIM (in fact for any system where FKG inequalities hold).

A key technical idea in [AW1] is to carry out the sketch of the argument above in such a way that it gives directly information about infinite-volume states. This will allow the use of ergodicity arguments and this in turn will force us to investigate some covariance properties of random Gibbs measures.

To do this, we will equip our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with some extra structure. First, we define the action T of the translation group $\mathbb{Z}^d$ on Ω. We will assume that $\mathbb{P}$ is invariant under this action and that the dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is stationary and ergodic. In the random field case, we have of course that $(h_{x_1}[T_y\omega], \dots, h_{x_n}[T_y\omega]) \equiv (h_{x_1+y}[\omega], \dots, h_{x_n+y}[\omega])$, and the assumptions of stationarity and ergodicity are trivially satisfied if the h_i are i.i.d..

Next we will assume that Ω is equipped with an affine structure, i.e. we will set $(h_{x_1}[\omega+\omega'], \dots, h_{x_n}[\omega+\omega']) \equiv (h_{x_1}[\omega]+h_{x_1}[\omega'], \dots, h_{x_n}[\omega]+h_{x_n}[\omega'])$. We will introduce a subset $\Omega_0 \subset \Omega$ of random fields that differ from zero only in some finite set[19], i.e.

$$\Omega_0 \equiv \left\{\delta\omega \in \Omega : \exists \Lambda \subset \mathbb{Z}^d, \text{finite}, \forall y \notin \Lambda, h_y[\delta\omega] = 0\right\} \tag{2.48}$$

We will use the convention to denote elements of Ω_0 by $\delta\omega$.

Definition 2.13: *A random Gibbs measure μ_β is called* covariant *if*

(i) For all $x \in \mathbb{Z}^d$, and any continuous function f,

$$\mu_\beta[\omega](T_{-x}f) = \mu_\beta[T_x\omega](f), \quad a.s. \tag{2.49}$$

and

(ii) for all $\delta\omega \in \Omega_0$, for almost all ω and all bounded, continuous f,

$$\mu_\beta[\omega+\delta\omega](f) = \frac{\mu_\beta[\omega]\left(f e^{-\beta(H[\omega+\delta\omega]-H[\omega])}\right)}{\mu_\beta[\omega]\left(e^{-\beta(H[\omega+\delta\omega]-H[\omega])}\right)} \tag{2.50}$$

(Note that $H[\omega+\delta\omega]-H[\omega]$ is a finite sum, i.e. if $\delta\omega$ is supported on Λ, then $H[\omega+\delta\omega](\omega)-H[\omega](\sigma) = \sum_{x\in\Lambda}\sigma_x h_x[\delta\omega]$).

The properties of covariant random Gibbs measures look rather natural, but their verification is in general elusive (recall that even the construction of random Gibbs measures was problematic). Essentially, one can verify the hypothesis for Gibbs measures constructed as limits over arbitrary sequences of volumes of local specifications with translation invariant boundary conditions. It is a priori far from clear

[19] Never mind that this set will usually have measure zero.

that this is possible. The only cases occur if convergence can be assured by strong monotonicity arguments. Since this is a frequently occuring topic, we will devote a short excursion to it.

2.4.2.1. The FKG inequalities and monotonicity.

A large number of results in statistical mechanics are based on so-called *correlation inequalities*. They reflect certain structural properties of the interactions, and most useful equalities are proven for *ferromagnetic* interactions. The FKG-inequalities, named after Fortuin, Kasteleyn, and Ginibre [FKG] are amongst the most useful ones.

Definition 2.14: *Let the single-spin space $\mathcal{S}$ be a linearly ordered set. We say that a probability measure μ on $\mathcal{S}_\Lambda$ for a finite $\Lambda \subset \mathbb{Z}^d$* satisfies the FKG inequalities *or* is positively correlated, *if for all bounded $\mathcal{F}_\Lambda$-measurable functions f, g that are non-decreasing with respect to the partial order on $\mathcal{S}_\Lambda$ induced by the ordering of $\mathcal{S}$, it holds that*

$$\mu(fg) \geq \mu(f)\mu(g) \tag{2.51}$$

Theorem 2.15: [FKG]*Assume that the cardinality of $\mathcal{S}_0$ is 2, and the interaction is nearest neighbour and attractive. Then any finite-volume Gibbs measure for this interaction satisfies the FKG inequalities.*

Proof. *A survey of various proofs can be found in [dHK].*$\diamondsuit$

We will now show how the FKG inequalities can be used to prove useful facts about the Gibbs measures.

Lemma 2.16: *Let $\mu^\eta_{\beta,\Lambda}$ local specifications for Gibbs measure that satisfies the FKG inequalities. Denote by $+$ the spin configuration $\eta_i = +1, \forall_{i\in\mathbb{Z}^d}$. Then*

(i) For any $\Lambda \subset \mathbb{Z}^d$ and any $\eta \in \mathcal{S}$, and any increasing function $f : \mathcal{S}_\Lambda \to \mathbb{R}$,

$$\mu^+_{\beta,\Lambda}(f) \geq \mu^\eta_{\beta,\Lambda}(f) \tag{2.52}$$

(ii) For any $\Lambda_2 \supset \Lambda_1$, and any increasing function $f : \mathcal{S}_{\Lambda_1} \to \mathbb{R}$,

$$\mu^+_{\beta,\Lambda_2}(f) \leq \mu^+_{\beta,\Lambda_1}(f) \tag{2.53}$$

Proof. For the proof we only consider the case where $\mathcal{S}_0 = \{-1, 1\}$. We first prove (i). Let $i \in \Lambda^c$, and consider η_i as an element of $[-1, 1]$. We will show that $\mu^\eta_{\beta,\Lambda}(f)$ is increasing in η_i. If this is true, (2.52) is immediate. Now compute

$$\frac{\partial}{\partial \eta_i} \mu^\eta_{\beta,\Lambda}(f) = \sum_{j\in\Lambda} \beta J_{ij} \left(\mu^\eta_{\beta,\Lambda}(\sigma_j f) - \mu^\eta_{\beta,\Lambda}(\sigma_j) \mu^\eta_{\beta,\Lambda}(f) \right) \tag{2.54}$$

Now since all J_{ij} are positive, and since σ_j is an increasing function, by the FKG inequalities, the right hand side of (2.54) is non-negative and (i) is proven.

To prove (ii), consider $\mu^+_{\beta,\Lambda_2}(\mathbb{I}_{+1_{\Lambda_2\setminus\Lambda_1}} f)$. By FKG,

$$\begin{aligned} \mu^+_{\beta,\Lambda_2}(\mathbb{I}_{+1_{\Lambda_2\setminus\Lambda_1}} f) &\geq \mu^+_{\beta,\Lambda_2}(\mathbb{I}_{+1_{\Lambda_2\setminus\Lambda_1}}) \mu^+_{\beta,\Lambda_2}(f) \\ &= \exp\left(\beta \sum_{\substack{i,j\in\Lambda_1^c \\ i\vee i\in\Lambda_2\setminus\Lambda_1}} J_{ij} - \beta\epsilon \sum_{i\in\Lambda_2\setminus\Lambda_1} h_i \right) \frac{Z^+_{\beta,\Lambda_1}}{Z^+_{\beta,\Lambda_2}} \mu^+_{\beta,\Lambda_2}(f) \end{aligned} \tag{2.55}$$

where the second equality uses the DLR equations. On the other hand, applying the DLR equations directly on the left hand side of (2.55), we get

$$\mu^+_{\beta,\Lambda_2}(\mathbb{I}_{+1_{\Lambda_2\setminus\Lambda_1}} f) = \mu^+_{\beta,\Lambda_1}(f) \exp\left(\beta \sum_{\substack{i,j\in\Lambda_1^c \\ i\vee i\in\Lambda_2\setminus\Lambda_1}} J_{ij} - \beta\epsilon \sum_{i\in\Lambda_2\setminus\Lambda_1} h_i \right) \frac{Z^+_{\beta,\Lambda_1}}{Z^+_{\beta,\Lambda_2}} \tag{2.56}$$

and combining both observations we have (ii).$\diamondsuit$

An immediate corollary of this theorem is

Corollary 2.17: *Under the hypothesis of 2.16,*

(i) For any sequence of increasing and absorbing sequence of volumes $\Lambda_n \subset \mathbb{Z}^d$, the limit

$$\lim_{n\uparrow\infty} \mu^+_{\beta,\Lambda_n} \equiv \mu^+_\beta \tag{2.57}$$

exists and is independent of the particular sequence. Moreover,

(ii) The Gibbs measure μ_β^+ is extremal.

(iii) Similarly, the limit

$$\lim_{n\uparrow\infty} \mu_{\beta,\Lambda_n}^- \equiv \mu_\beta^- \tag{2.58}$$

also exists, is independent of the sequence Λ_n and is an extremal Gibbs measure. Moreover,

(iv) for all Gibbs measures for the same interaction and temperature, and any increasing bounded continuous function f,

$$\mu_\beta^-(f) \leq \mu_\beta(f) \leq \mu_\beta^+(f) \tag{2.59}$$

Proof. Note that compactness and monotonicity (2.53) implies that for all increasing bounded continuous functions, for any sequence Λ_n of increasing absorbing sequences the limit $\mu_{\beta,\Lambda_n}^+(f)$ exists. Now let Λ_n, Λ_n' be two such sequences. Since both sequences are absorbing, it follows that there exist infinite sequences n_k and n_k' such that for all $k \in \mathbb{N}$, $\Lambda_{n_k} \subset \Lambda'_{n_k'} \subset \Lambda_{n_{k+1}}$. But this implies that

$$\lim_{n\uparrow\infty} \mu_{\beta,\Lambda_n}^+(f) = \lim_{k\uparrow\infty} \mu_{\beta,\Lambda_{n_k}}^+(f) \geq \lim_{k\uparrow\infty} \mu_{\beta,\Lambda'_{n_k}}^+(f) = \lim_{n\uparrow\infty} \mu_{\beta,\Lambda'_n}^+(f) \tag{2.60}$$

and

$$\lim_{n\uparrow\infty} \mu_{\beta,\Lambda_n}^+(f) = \lim_{k\uparrow\infty} \mu_{\beta,\Lambda_{n_{k+1}}}^+(f) \leq \lim_{k\uparrow\infty} \mu_{\beta,\Lambda'_{n_k}}^+(f) = \lim_{n\uparrow\infty} \mu_{\beta,\Lambda'_n}^+(f) \tag{2.61}$$

and so

$$\lim_{n\uparrow\infty} \mu_{\beta,\Lambda_n}^+(f) = \lim_{n\uparrow\infty} \mu_{\beta,\Lambda'_n}^+(f) \tag{2.62}$$

Thus all possible limit points of $\mu_{\beta,\Lambda}^+$ coincide on the set of increasing bounded continuous functions. But then by standard approximation arguments, the limits coincide on all bounded continuous functions implying that in fact the limiting measures exist and are independent of the subsequences chosen. This proves (i). Assume now that

μ_β^+ was not extremal. Then there exist two distinct Gibbs measures μ and ν such that $\mu_\beta^+ = \alpha\mu_\beta + (1-\alpha)\nu_\beta$ with $\alpha > 0$. In particular, for f increasing,

$$\mu_\beta^+(f) = \alpha\mu_\beta(f) + (1-\alpha)\nu_\beta(f) \tag{2.63}$$

Now by (2.52) and the DLR-equations, for any local increasing function f, for all Λ large enough so that f is $\mathcal{F}_\Lambda$-measurable, for any Gibbs measure ν_β

$$\nu_\beta(f) = \nu_\beta\left(\mu_{\beta,\Lambda}^{\cdot}(f)\right) \leq \mu_{\beta,\Lambda}^+(f) \tag{2.64}$$

Since $\mu_{\beta,\Lambda}^+$ converges to μ_β^+, this implies that

$$\nu_\beta(f) \leq \mu_\beta^+(f) \tag{2.65}$$

Thus (2.63) can only hold if both $\mu_\beta(f)$ and $\nu_\beta(f)$ are equal to $\mu_\beta^+(f)$. But then, by the same argument as before, we conclude $\mu_\beta = \nu_\beta = \mu_\beta^+$, contradicting the assumption that μ_β and ν_β are different. This proves (ii). (iii) is obvious by repeating all arguments with decreasing functions which also yields the complementing version of (2.65) which implies (iv). $\diamondsuit$

We can now apply these results to the random field Ising model.

Theorem 2.18: *Consider the random field Ising model (2.25) with h_i a stationary and ergodic random field. Then there exist two covariant random Gibbs measures μ_β^+ and μ_β^- that satisfy*

(i) For almost all ω,

$$\mu_\beta^\pm[\omega] = \lim_{\Lambda\uparrow\mathbb{Z}^d} \mu_{\beta,\Lambda}^\pm[\omega] \tag{2.66}$$

(ii) Suppose that for some β, $\mu_\beta^+ = \mu_\beta^-$. Then for this value of β, the Gibbs measure for the RFIM model is unique for almost all ω.

Proof. We observe first that due to Corollary 2.17, the functions $\omega \to \mu_\beta^\pm$ are measurable, since they are limits of measurable functions. It remains to check the covariance properties. Property (ii) follows immediately from the fact that $\mu_\beta^\pm$ can be represented as a limit of local specifications, and that the formula (2.51) holds trivially for local specifications with Λ large enough to contain the support of $\delta\omega$. Property (i) on the contrary requires the independence of the limit from the chosen sequence Λ_n. Indeed we have

$$\mu^+_{\beta,\Lambda}[\omega](T_{-x}f) = \mu^+_{\beta,\Lambda+x}[T_x\omega](f) \tag{2.67}$$

which implies by Corollary 2.17 that $\mu^+_\beta[\omega](T_{-x}f) = \mu^+_\beta[T_x\omega](f)$ almost surely, as desired. The second assertion of the theorem follows directly from (iv) of Corollary 2.17. $\diamondsuit$

Remark. It is remarkably hard to prove the translation covariance property in the absence of strong result like the FKG inequalities. In fact there are two difficulties. The first is that of the measurability of the limits which we have already discussed above. This can be resolved by the introduction of metastates, and it was precisely in this context that Aizenman and Wehr first applied this idea. The second problem is that without comparison results between local specifications in different volumes, the relative shift between the function and the volume implicit in (2.67) cannot be removed. The general way out of this problem is to construct Gibbs states with *periodic* boundary conditions (i.e. one chooses instead of Λ a torus, i.e. $(\mathbb{Z} \bmod n)^d$). In that case, one may recover the translation covariance of the limit from translation covariance of the finite-volume measures under the automorphisms of the torus. From the point of view of the general theory as we have presented it so far, this is of course somewhat unsatisfactory. For this reason we have prefered to restrict our exposition of the Aizenman–Wehr method to a smaller class of models and refer the reader to the original articles for the more general results.

2.4.2.2. Order parameters and generating functions.

We can conclude from the preceding subsection that (due to FKG) we will have a unique Gibbs state for almost all ω, provided we have a unique covariant random Gibbs state. Moreover, it is easy to see that uniqueness will follow from the vanishing of a so-called *order parameter* which in the present case is nothing but the total magnetization. If μ is a Gibbs measure, we set

$$m^\mu[\omega] \equiv \lim_{\Lambda\uparrow\infty} \frac{1}{|\Lambda|} \sum_{i\in\Lambda} \mu[\omega](\sigma_i) \tag{2.68}$$

provided the limit exists. We will also use the abusive notations $m_\beta^\pm = m^{\mu_\beta^\pm}$.

Some simple facts follow from covariance and FKG:

Lemma 2.19: *Suppose that μ is a covariant Gibbs state. Then the total magnetization $m^\mu[\omega]$ exists for almost all ω and is almost surely independent of ω.*

Proof. By the covariance of μ,

$$m^\mu[\omega] = \lim_{\Lambda\uparrow\infty} \frac{1}{|\Lambda|} \sum_{i\in\Lambda} \mu[T_{-i}\omega](\sigma_0) \tag{2.69}$$

But $\mu_\beta[\omega](\sigma_0)$ is a bounded measurable function of ω. Since we assumed that $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is stationary and ergodic, it follows from the ergodic theorem (see e.g. Appendix A3 of [Ge1] for a good exposition and proofs) that the limit exists almost surely and is given by

$$m^\mu = \mathbb{E}\,\mu(\sigma_0) \tag{2.70}$$

Lemma 2.20: *In the random field Ising model,*

$$m^+ - m^- = 0 \Leftrightarrow \mu_\beta^+ = \mu_\beta^- \tag{2.71}$$

Proof. (2.70) implies that almost surely

$$0 = m^+ - m^- = \mathbb{E}(\mu_\beta^+(\sigma_i) - \mu_\beta^-(\sigma_i)) \tag{2.72}$$

and so, since $\mu_\beta^+(\sigma_i) - \mu_\beta^-(\sigma_i) \geq 0$, and there are only countably many sites i, almost surely for all $i \in \mathbb{Z}^d$, $\mu_\beta^+(\sigma_i) - \mu_\beta^-(\sigma_i) = 0$.

Interestingly enough, in the presence of FKG inequalities, this fact implies that the two measures μ_β^+ and μ_β^- coincide. This result is due to Lebowitz and Martin-Löf [LML] and Ruelle [Ru3]. We give a proof in the Ising case following Preston [Pr1]. It is based on the following simple lemma:

Lemma 2.21: *Consider a model with Ising spins for which the FKG inequalities hold. Then for any finite sets $A, B \subset \Lambda$,*

$$\begin{aligned} &\mu_\beta^+(\sigma_{A\cup B} = +1) - \mu_\beta^-(\sigma_{A\cup B} = +1) \\ &\leq \mu_\beta^+(\sigma_A = +1) - \mu_\beta^-(\sigma_A = +1) + \mu_\beta^+(\sigma_B = +1) - \mu_\beta^-(\sigma_B = +1) \end{aligned} \tag{2.73}$$

(where $\sigma_A = +1$ is shorthand for $\forall_{i\in A}\sigma_i = +1$).

Proof. Notice the set-equality

$$\mathbb{1}_{\sigma_A=+1\wedge\sigma_B=+1} = \mathbb{1}_{\sigma_A=+1} + \mathbb{1}_{\sigma_B=+1} - \mathbb{1}_{\sigma_A=+1\vee\sigma_B=+1} \tag{2.74}$$

This implies that

$$\begin{aligned}\mu_\beta^+(\sigma_{A\cup B} = +1) - \mu_\beta^-(\sigma_{A\cup B} = +1) &= \mu_\beta^+(\sigma_A = +1) - \mu_\beta^-(\sigma_A = +1)\\ + \mu_\beta^+(\sigma_B = +1) - \mu_\beta^-(\sigma_B = +1) &+ \mu_\beta^-(\sigma_A = +1 \vee \sigma_B = +1) - \mu_\beta^+(\sigma_A = +1 \vee \sigma_B = +1)\end{aligned} \tag{2.75}$$

But $\{\sigma_A = +1 \vee \sigma_B = +1\}$ is an increasing event, and so by (2.59),

$$\mu_\beta^-(\sigma_A = +1 \vee \sigma_B = +1) - \mu_\beta^+(\sigma_A = +1 \vee \sigma_B = +1) \leq 0 \tag{2.76}$$

This implies the assertion of the lemma.$\diamondsuit$

Now in the Ising model, all local functions can expressed in terms of the indicator functions $\mathbb{1}_{\sigma_A=+1}$, for finite $A \subset \Lambda$. But by repeated application of Lemma 2.21, we get

$$0 \leq \mu_\beta^+(\sigma_A = +1) - \mu_\beta^-(\sigma_A = +1) \leq \sum_{i\in A} \mu_\beta^+(\sigma_i = +1) - \mu_\beta^-(\sigma_i = +1) \tag{2.77}$$

Therefore, if for all i, $\mu_\beta^+(\sigma_i = +1) = \mu_\beta^-(\sigma_i = +1)$, it follows indeed that $\mu_\beta^+ = \mu_\beta^-$. This concludes the proof of Lemma 2.20.$\diamondsuit$

The (macroscopic) functions m^μ are called *order parameters* because their values allow to decide (in this model) on the uniqueness respectively co-existence of phases. One can generalize this notion to other models, and one may set up a general theory that is able to produce rather interesting abstract results (see [Ge1]). Recall that after all, extremal Gibbs measures are characterized by their values on the tail-sigma-field, i.e. by their values on macroscopic functions. The general philosophy would thus be to identify a (hopefully) finite set of macroscopic functions whose values suffice to characterize all possible Gibbs states of the system. We will not enter this subject here, but will have occasion to return to the notion of order parameters quite extensively in our discussion of mean field models.

The order parameters introduced above can be computed as derivatives of certain *generating functions.* We set

$$G_\Lambda^\mu \equiv \frac{1}{\beta} \ln \mu \left(e^{\beta \sum_{i\in\Lambda} h_i \sigma_i} \right) \tag{2.78}$$

Note that if μ is a covariant Gibbs state, then

$$G_\Lambda^{\mu[\omega]} = -\frac{1}{\beta} \ln \mu[\omega - \omega_\Lambda] \left(e^{-\beta \sum_{i\in\Lambda} h_i \sigma_i} \right) \tag{2.79}$$

Here $\omega_\Lambda \in \Omega_0$ is such that $h_i[\omega_\Lambda] = h_i[\omega]$, if $i \in \Lambda$, and $h_i[\omega_\Lambda] = 0$, if $i \notin \Lambda$. Therefore, for $i \in \Lambda$,

$$\frac{\partial}{\partial h_i} G_\Lambda^{\mu[\omega]} = \frac{\mu[\omega - \omega_\Lambda] \left(\sigma_i e^{-\beta \sum_{i\in\Lambda} h_i \sigma_i} \right)}{\mu[\omega - \omega_\Lambda] \left(e^{-\beta \sum_{i\in\Lambda} h_i \sigma_i} \right)} = \mu[\omega](\sigma_i) \tag{2.80}$$

where the first equality follows from the fact that $\mu[\omega - \omega_\Lambda]$ is $\mathcal{F}_{\Lambda^c}$-measurable and the second follows using (2.50). In particular, we get that

$$\mathbb{E} \frac{\partial}{\partial h_i} G_\Lambda^{\mu^\pm} = m^\pm \tag{2.81}$$

Let us now introduce the function

$$F_{\beta,\Lambda} \equiv \mathbb{E} \left[G_\Lambda^{\mu^+} - G_\Lambda^{\mu^-} | \mathcal{F}_\Lambda \right] \tag{2.82}$$

Clearly $\mathbb{E} \frac{\partial}{\partial h_0} F_\Lambda = m^+ - m^-$, and our purpose will be to prove that this quantity must be zero. The important point is the following a-priori upper bound:

Lemma 2.22: *For any temperature and any volume Λ,*

$$|F_\Lambda| \leq 2|\partial\Lambda| \tag{2.83}$$

Proof. The first step in the proof makes use of (2.79) to express F_Λ in terms of measures that do not depend on the disorder within Λ anymore. Namely,

$$\begin{aligned} F_\Lambda =& \beta^{-1} \mathbb{E} \left[\ln \frac{\mu_\beta^+[\omega] \left(e^{\beta \sum_{i\in\Lambda} h_i \sigma_i} \right)}{\mu_\beta^-[\omega] \left(e^{\beta \sum_{i\in\Lambda} h_i \sigma_i} \right)} | \mathcal{B}_\Lambda \right] \\ =& \mathbb{E} \left[\ln \frac{\mu_\beta^-[\omega - \omega_\Lambda] \left(e^{-\beta \sum_{i\in\Lambda} h_i \sigma_i} \right)}{\mu_\beta^+[\omega - \omega_\Lambda] \left(e^{-\beta \sum_{i\in\Lambda} h_i \sigma_i} \right)} | \mathcal{B}_\Lambda \right] \end{aligned} \tag{2.84}$$

Next we use the spin-flip symmetry which assures that $\mu_\beta^+[\omega] = \mu_\beta^-[-\omega]$ together with the fact that the distribution of the h_i is symmetric:

$$
\begin{aligned}
&\mathbb{E}\left[\ln \frac{\mu_\beta^-[\omega-\omega_\Lambda]\left(e^{-\beta\sum_{i\in\Lambda}h_i\sigma_i}\right)}{\mu_\beta^+[\omega-\omega_\Lambda]\left(e^{-\beta\sum_{i\in\Lambda}h_i\sigma_i}\right)}|\mathcal{B}_\Lambda\right]\\
&=\mathbb{E}\left[\ln \frac{\mu_\beta^+[-(\omega-\omega_\Lambda)]\left(e^{+\beta\sum_{i\in\Lambda}h_i\sigma_i}\right)}{\mu_\beta^+[\omega-\omega_\Lambda]\left(e^{-\beta\sum_{i\in\Lambda}h_i\sigma_i}\right)}|\mathcal{B}_\Lambda\right]\\
&=\mathbb{E}\left[\ln \frac{\mu_\beta^+[\omega-\omega_\Lambda]\left(e^{+\beta\sum_{i\in\Lambda}h_i\sigma_i}\right)}{\mu_\beta^+[\omega-\omega_\Lambda]\left(e^{-\beta\sum_{i\in\Lambda}h_i\sigma_i}\right)}|\mathcal{B}_\Lambda\right]
\end{aligned}
\tag{2.85}
$$

We are left with the ratio of two expectations with respect to the same measure. Here we use the DLR equations to compare them:

$$
\begin{aligned}
&\mu_\beta^+[\omega-\omega_\Lambda]\left(e^{+\beta\sum_{i\in\Lambda}h_i\sigma_i}\right) = \sum_{\sigma_{\Lambda^c}}\mu_\beta^+[\omega-\omega_\Lambda](\sigma_{\Lambda^c})\mu_{\beta,\Lambda}^{\sigma_{\Lambda^c}}\left(e^{\beta\sum_{i\in\Lambda}h_i\sigma_i}\right)\\
&=\sum_{\sigma_{\Lambda^c}}\mu_\beta^+[\omega-\omega_\Lambda](\sigma_{\Lambda^c})\frac{1}{Z_{\beta,\Lambda^c}^{\sigma_{\Lambda^c}}}\sum_{\sigma_\Lambda}e^{\beta\sum_{\substack{i,j\in\Lambda\\|i-j|=1}}\sigma_i\sigma_j+\sum_{\substack{i\in\Lambda,j\in\Lambda^c\\|i-j|=1}}\sigma_i\sigma_j+\sum_{i\in\Lambda}h_i\sigma_i}\\
&=\sum_{\sigma_{\Lambda^c}}\mu_\beta^+[\omega-\omega_\Lambda](\sigma_{\Lambda^c})\frac{1}{Z_{\beta,\Lambda^c}^{\sigma_{\Lambda^c}}}\sum_{\sigma_\Lambda}e^{\beta\sum_{\substack{i,j\in\Lambda\\|i-j|=1}}\sigma_i\sigma_j-\sum_{\substack{i\in\Lambda,j\in\Lambda^c\\|i-j|=1}}\sigma_i\sigma_j-\sum_{i\in\Lambda}h_i\sigma_i}\\
&\leq\sum_{\sigma_{\Lambda^c}}\mu_\beta^+[\omega-\omega_\Lambda](\sigma_{\Lambda^c})\frac{1}{Z_{\beta,\Lambda^c}^{\sigma_{\Lambda^c}}}\sum_{\sigma_\Lambda}e^{\beta\sum_{\substack{i,j\in\Lambda\\|i-j|=1}}\sigma_i\sigma_j+\sum_{\substack{i\in\Lambda,j\in\Lambda^c\\|i-j|=1}}\sigma_i\sigma_j-\sum_{i\in\Lambda}h_i\sigma_i}e^{2\beta|\partial\Lambda|}\\
&=\mu_\beta^+[\omega-\omega_\Lambda]\left(e^{-\beta\sum_{i\in\Lambda}h_i\sigma_i}\right)e^{2\beta|\partial\Lambda|}
\end{aligned}
\tag{2.86}
$$

Inserting this bound into (2.85) gives the desired estimate immediately.$\diamondsuit$

Next we proof a lower bound on the fluctuations of F_Λ, or more precisely on its Laplace transform. Namely:

Lemma 2.23: *Assume that for some $\epsilon > 0$, the distribution of the random fields h satisfies $\mathbb{E}|h|^{2+\epsilon} < \infty$. Then, for any $t \in \mathbb{R}$, we have that*

$$
\liminf_{\Lambda=[-L,L]^d;L\uparrow\infty}\mathbb{E}e^{tF_\Lambda/\sqrt{|\Lambda|}} \geq e^{\frac{t^2b^2}{2}}
\tag{2.87}
$$

where

$$b^2 \geq \mathbb{E}\left[\mathbb{E}\left[F_\Lambda|\mathcal{B}_0\right]^2\right] \tag{2.88}$$

Remark. It is easy to see that Lemmata 2.22 and 2.23 contradict each other in $d \leq 2$ *unless* $b = 0$. On the other hand we will see that $b = 0$ implies $m^+ = m^-$ and thus uniqueness of the Gibbs state.

Proof. The proof of this lemma uses a decomposition of F_Λ as a martingale-difference sequence. That is, we order all the points $i \in \Lambda$ and denote by $\mathcal{B}_{\Lambda,i}$ the sigma-algebra generated by the variables $\{h_j\}_{j\in\Lambda:j\leq i}$. Then we have trivially that

$$F_\Lambda = \sum_{i=1}^{|\Lambda|} \left(\mathbb{E}\left[F_\Lambda|\mathcal{B}_{\Lambda,i}\right] - \mathbb{E}\left[F_\Lambda|\mathcal{B}_{\Lambda,i-1}\right]\right) \equiv \sum_{i=1}^{|\Lambda|} Y_i \tag{2.89}$$

(note that $\mathbb{E}F_\Lambda = 0$!). Using this, we can represent the generating function as

$$\begin{aligned} \mathbb{E}e^{tF_\Lambda} &= \mathbb{E}\prod_{i=1}^{|\Lambda|} e^{tY_i} \\ &= \mathbb{E}\left[\mathbb{E}\left[\dots\mathbb{E}\left[\mathbb{E}\left[e^{tY_{|\Lambda|}}|\mathcal{B}_{\Lambda,|\Lambda|-1}\right]e^{tY_{|\Lambda|-1}}|\mathcal{B}_{\Lambda,|\Lambda|-2}\right]\dots e^{tY_2}|\mathcal{B}_{\Lambda,1}\right]e^{tY_1}\right] \end{aligned} \tag{2.90}$$

We want to work up the conditional expectation from the inside out. For this we need a bound from below for any of the terms $\mathbb{E}\left[e^{tY_i}|\mathcal{B}_{\Lambda,i-1}\right]$. To do this we will use the elementary observation (to be found in [AW1], Lemma A.2.2) that there exists a continuous functions $g(a)$ with $g(a) \downarrow 0$ as $a \downarrow 0$, such that for all real and all x and $a \geq 0$, $e^x \geq 1 + x + \frac{1}{2}(1-g(a))x^2 \mathbb{1}_{|x|\leq a}$. Since moreover for all $|x| \leq a$, $e^{x^2 e^{-a^2}} \leq 1 + x^2$, it follows that if $\mathbb{E}X = 0$, then for $f(a) = 1-(1-g(a))e^{-a^2}$,

$$\mathbb{E}e^X \geq e^{\frac{1}{2}(1-f(a))\mathbb{E}[X^2\mathbb{1}_{|X|\leq a}]} \tag{2.91}$$

Using this estimate we see that

$$\mathbb{E}\left[e^{tY_i}|\mathcal{B}_{\Lambda,i-1}\right] \geq \exp\left(\frac{t^2}{2}(1-f(a))\mathbb{E}\left[Y_i^2\mathbb{1}_{t|Y_i|\leq a}|\mathcal{B}_{\Lambda,i-1}\right]\right) \tag{2.92}$$

But this implies that (we switch to the desired normalization)

$$1 \leq \mathbb{E}e^{tF_\Lambda/\sqrt{\Lambda}-\frac{t^2}{2|\Lambda|}(1-f(a))\sum_{i=1}^{|\Lambda|}\mathbb{E}\left[Y_i^2\mathbb{1}_{t|Y_i|\leq a\sqrt{|\Lambda|}}|\mathcal{B}_{\Lambda,i-1}\right]} \tag{2.93}$$

Now by the Hölder inequality, for any p,q with $1/p+1/q=1$,

$$1 \leq \left(\mathbb{E}e^{qtF_\Lambda/\sqrt{\Lambda}}\right)^{1/q} \left(\mathbb{E}e^{-p\frac{t^2}{2|\Lambda|}(1-f(a))\sum_{i=1}^{|\Lambda|}\mathbb{E}\left[Y_i^2 \mathbb{1}_{t|Y_i|\leq a\sqrt{|\Lambda|}}|\mathcal{B}_{\Lambda,i-1}\right]}\right)^{1/p} \tag{2.94}$$

Now if the term

$$V_\Lambda(a) \equiv |\Lambda|^{-1}\sum_{i=1}^{|\Lambda|}\mathbb{E}\left[Y_i^2 \mathbb{1}_{t|Y_i|\leq a\sqrt{|\Lambda|}}|\mathcal{B}_{\Lambda,i-1}\right] \tag{2.95}$$

in the exponent in the last factor converges to a constant C, independent of $a>0$, in probability (as $|\Lambda|\uparrow\infty$), then for all $a>0$,

$$\limsup_{\Lambda\uparrow\mathbb{Z}^d}\left(\mathbb{E}e^{-p\frac{t^2}{2|\Lambda|}(1-f(a))\sum_{i=1}^{|\Lambda|}\mathbb{E}\left[Y_i^2 \mathbb{1}_{t|Y_i|\leq a\sqrt{|\Lambda|}}|\mathcal{B}_{\Lambda,i-1}\right]}\right)^{1/p} \geq e^{-C(1-f(a))t^2} \tag{2.96}$$

independent of p and therefore, using first (2.94) and letting finally q go to one and a to zero,

$$\mathbb{E}e^{tF_\Lambda/\sqrt{\Lambda}} \geq e^{t^2 C} \tag{2.97}$$

Thus we are left with controlling and identifying the limit of $V_\Lambda(a)$. This will be done by a clever use of the ergodic theorem. Note that the summands in (2.95) depend on Λ

For this we introduce new sigma-algebras $\mathcal{B}_i^{\leq}$, generated by the random variables h_j with $j\leq i$, where $\leq$ refers to the lexicographical ordering. Define

$$W_i \equiv \mathbb{E}\left[G_\Lambda^{\mu^+} - G_\Lambda^{\mu^-}|\mathcal{B}_i^{\leq}\right] - \mathbb{E}\left[G_\Lambda^{\mu^+} - G_\Lambda^{\mu^-}|\mathcal{B}_{i-1}^{\leq}\right] \tag{2.98}$$

Using (2.80) one may indeed show that for all i in Λ, W_i is independent of Λ (the proof goes by using (2.80) to represent G_Λ^μ in terms of integrals over $\mu(\sigma_i)$, which is independent of Λ). On the other hand, we have the obvious relation that

$$Y_i = \mathbb{E}[W_i|\mathcal{B}_\Lambda] \tag{2.99}$$

We use this first to show that the indicator function in the conditional expectation can be removed, i.e. for all $\epsilon>0$,

$$\lim_{\Lambda\uparrow\infty}\mathbb{P}\left[|\Lambda|^{-1}\sum_{i=1}^{|\Lambda|}\mathbb{E}\left[Y_i^2 \mathbb{1}_{t|Y_i|>a\sqrt{|\Lambda|}}|\mathcal{B}_{\Lambda,i-1}\right]>\epsilon\right]=0 \tag{2.100}$$

To see this, compute the expectation of the left-hand side in the probability, and use the Hölder inequality to get

$$\begin{aligned}\mathbb{E}|\Lambda|^{-1}\sum_{i=1}^{|\Lambda|}\mathbb{E}\left[Y_i^2 \mathbb{1}_{t|Y_i|>a\sqrt{|\Lambda|}}|\mathcal{B}_{\Lambda,i-1}\right] &= |\Lambda|^{-1}\sum_{i=1}^{|\Lambda|}\mathbb{E}\left[Y_i^2 \mathbb{1}_{t|Y_i|>a\sqrt{|\Lambda|}}\right] \\ \leq |\Lambda|^{-1}\sum_{i=1}^{|\Lambda|}\left(\mathbb{E}Y_i^{2q}\right)^{1/q}\left(\mathbb{P}\left[|Y_i|>a\sqrt{|\Lambda|}/t\right]\right)^{1/p}\end{aligned} \tag{2.101}$$

with $1/p+1/q=1$. Now using Jensen's inequality and (2.99), we see that for any $q>1$ $\mathbb{E}Y_i^{2q}\leq\mathbb{E}W_0^{2q}$ However, using e.g. (2.80), it is easy to see that $|W_0|\leq C|h_0|$, so that if the $2q$-th moment of h is finite, then $\mathbb{E}W_0^{2q}<\infty$. Using the Chebyshev inequality and the same argument as before, we also conclude that $\mathbb{P}\left[|Y_i|>a\sqrt{|\Lambda|}/t\right]\leq\frac{t^2\mathbb{E}W_0^2}{a^2|\Lambda|}$ which tends to zero as $\Lambda\uparrow\infty$. We see that (2.101) tends to zero whenever $p<\infty$, for any $a>0$. By Chebyshev's inequality, this allows in turn to conclude (2.100)[20].

Next observe that W_i is shift covariant, i.e.

$$W_i[\omega]=W_0[T_{-i}\omega] \tag{2.102}$$

Therefore, by the ergodic theorem, we can conclude that

$$\lim_{\Lambda\uparrow\mathbb{Z}^d}|\Lambda|^{-1}\sum_{i\in\Lambda}\mathbb{E}\left[W_i^2|\mathcal{B}_{i-1}^{\leq}\right]=\mathbb{E}W_0^2,\quad \text{in Prob.} \tag{2.103}$$

Now we will be done if we can show that

$$E\left[Y_i^2|\mathcal{B}_{\Lambda,i}\right]-E\left[W_i^2|\mathcal{B}_i^{\leq}\right] \tag{2.104}$$

goes to zero as Λ goes to infinity, in probability. This follows by estimating the expectation of the square of (2.104), some simple estimates using the Cauchy-Schwartz inequality and the fact that for any square integrable function f, $\mathbb{E}[(f-\mathbb{E}[f|\mathcal{B}_\Lambda])^2]$ tends to zero as Λ approaches $\mathbb{Z}^d$.

To arrive at the final conclusion, note that

$$\mathbb{E}W_0^2\geq\mathbb{E}[(\mathbb{E}[W_0|\mathcal{B}_0])^2] \tag{2.105}$$

[20] In [AW1], only two moments are required for h. However, the proof given there is in error as it pretends that the function $x^2\mathbb{1}_{|x|>a}$ is convex which is manifestly not the case.

(where $\mathcal{B}_0$ is the sigma-algebra generated by the single variable h_0), and $\mathbb{E}[W_0|\mathcal{B}_0] = \mathbb{E}[F_\Lambda|\mathcal{B}_0]$. $\diamondsuit$

Finally we observe that by (2.80),

$$\frac{\partial}{\partial h_0}\mathbb{E}[F_\Lambda|\mathcal{B}_0] = \mathbb{E}\left[(\mu_\beta^+(\sigma_0) - \mu_\beta^-(\sigma_0))|\mathcal{B}_0\right] \tag{2.106}$$

Let us denote by $f(h) = \mathbb{E}[F_\Lambda|\mathcal{B}_0]$ (where $h = h_0[\omega]$). Since $1 \geq f'(h) \geq 0$ for all h, $\mathbb{E}f^2 = 0$ implies that $f(h) = 0$ (for $\mathbb{P}$-almost all points) on the support of the distribution of h. But then $f'(h)$ must also vanish on the convex hull of the support of the distribution of h (except of course if the distribution is concentrated on a single point). Therefore, barring that case $\mathbb{E}[F_\Lambda|\mathcal{B}_0] = 0 \Rightarrow m^+ - m^- = 0$.

Collecting our results we can now prove the following

Theorem 2.24: [**AW1**] *In the random field Ising model with i.i.d. random fields whose distribution is not concentrated on a single point and possesses at least $2+\epsilon$ finite moments, for some $\epsilon > 0$, if $d \leq 2$, there exists a unique infinite-volume Gibbs state.*

Proof. Lemma 2.22 implies that for any Λ, $Ee^{tF_{\beta,\Lambda}} \leq e^{t|\partial\Lambda|}$. Combining this with Lemma 2.23, we deduce that if $d \leq 2$, necessarily $b = 0$. But from what was just shown, this implies $m^+ = m^-$, which in turn implies uniqueness of the Gibbs state.$\diamondsuit$

With this result we conclude our discussion of the random field Ising model. We may stress that Theorem 2.24 is in some sense a soft result that gives uniqueness without saying anything more precise about the properties of the Gibbs state. Clearly, there are many interesting questions that could still be asked. How does the Gibbs state at high temperatures distinguish itself from the one at low temperatures, or how does the low temperature Gibbs state look like in dependence on the strength of the random fields? It is clear that for very low temperatures and very large ϵ, the Gibbs state will be concentrated near configurations $\sigma_i = \operatorname{sign} h_i$. For small ϵ, on the contrary, a more complicated behaviour is expected. Results of this type are available in $d = 1$ [BRZ], but much less is known in $d = 2$ [AW2].

The natural continuation of this section would be to start to discuss models with bond disorder in situations where the simple Peierls argument (Theorem 2.6) does not apply. This is the case in the so-called *Edwards–Anderson model*, whose Hamiltonian is identical to that of the dilute Ising model (2.15), except that the random bond variables J_{ij} take both positive and negative values with roughly the same probability; e.g. one may chose them to be symmetric Rademacher variables (i.e. taking values ± 1 with equal probability), or centered Gaussian random variables. Such models are called *spin glasses* and are expected to exhibit most remarkable properties. However, exceedingly little is known about their low temperature phases, and even in the physics literature various speculative theories compete. I will therefor approach the subject of spin glasses only via so-called *mean-field models*, which will be the subject of the rest of these notes. Readers interested to learn more about the Edwards–Anderson model should turn to Newman's book [N] or the review papers [NS1,NS2] by Newman and Stein.

3. Mean-field models 1: Gaussian processes.

In the previous chapters we have seen that with considerable work it is possible to study some simple aspects of random perturbations of the Ising model. On the other hand, models that are genuinely random and promise to bring some new features to light are at the moment not seriously accessible to rigorous analysis. Therefore we now turn our attention to simplified models that promise more explicit solutions. A class of such models is generally called "mean-field models".

3.1. What are disordered mean-field models?

The common feature of mean-field models is that the spatial structure of the lattice $\mathbb{Z}^d$ is abandoned in favour of a simpler setting where sites are indexed by the natural numbers and all spins are supposed to interact with each other irrespective of their 'distance'. The prototype of all such models is the *Curie–Weiss model* of ferromagnetism which is nothing but the mean-field trivialization of the Ising model. Let us briefly describe this model: We consider the single-site state space $\mathcal{S}_0 = \{-1,+1\}$. Let $\Lambda_N \equiv \{1,2,\dots,N\} \subset \mathbb{N}$, and define the state spaces $\mathcal{S}_N \equiv \mathcal{S}_{\Lambda_N}$ and $\mathcal{S} = \mathcal{S}_0^{\mathbb{N}}$. We would like to define a ferromagnetic interaction of equal strength between all sites i,j. Of course this is not possible within the context of regular interactions. Thus, we will have to leave the Gibbsian setting introduced in Section 1 and return to a more naive description of finite volume systems. Define

$$H_{N,h}(\sigma) \equiv -\frac{1}{2N} \sum_{i,j\in\Lambda_N\times\Lambda_N} \sigma_i\sigma_j - \sum_{i\in\Lambda_N} h\sigma_i \tag{3.1}$$

Note that there are no terms in (3.1) that correspond to an interaction with the complement of Λ_N, and it is easy to see that as long as we maintain the decision to ignore distance between sites, there is no reasonable way of writing such an interaction (we will understand later that the magnetic field term in the Hamiltonian will take over this role to some extent). Note also that the Hamiltonian contains an explicit coefficient $1/(2N)$ that depends on the volume. In particular, there is no way of thinking of H_N as the restriction of some infinite-volume Hamiltonian to finite volume; the alert reader will immediately worry that this will also prevent us from constructing infinite-volume Gibbs measures by prescribing local specifications.

While from a formal point of view all this looks very worrying, there are good reasons not to worry and to continue in a naive way. Namely, we can define (finite-volume) measures

$$\mu_{N,\beta,h}(f) \equiv \int \nu_{\Lambda_N}(d\sigma_{\Lambda_N})\nu^{h'}(d\sigma_{\Lambda_N^c})\frac{e^{-\beta H_{N,h}(\sigma)}}{Z_{N,\beta,h}}f(\sigma) \tag{3.2}$$

where ν^h is the product measure with marginal $\nu(\sigma_i = \pm 1) = \frac{e^{\pm\beta h}}{2\cosh(\beta h)}$, and the choice $h' = h + m^*(\beta,h)$, where $m^*(\beta,h)$ solves $m = \tanh\beta(h+m)$ looks particularly attractive[21]. Of course these measures are not local specifications, but we may nonetheless look for limit points, or even limits, as $N\uparrow\infty$. Such objects may be interpreted in good faith as "infinite-volume Gibbs measures", even though they lack some of the nice properties of genuine Gibbs measures. In compensation, we will see that we are rewarded – at least in many cases – by the possibility to actually compute things explicitly, that is to compute the actual limit measures in a more or less explicit form.

We have seen that in lattice systems 'boundary conditions' played an important rôle in the construction of Gibbs states. In particular, the use of suitable boundary conditions does allow, often even in disordered models (see e.g. the RFIM), to construct sequences of finite-volume measures that *converge* to a given extremal Gibbs state. In mean-field models, such a construction is not immediately available. However, there is an often used alternative that is some way emulates the effect of an external configurations, namely an *external field*. This just means that one constructs for a given Hamiltonian Gibbs measures $\mu_{N,\beta,\{h\}}$ depending on an entire family of parameters $\{h_i\}_{i=1}^N$ that are thought of as external fields acting on the spin σ_i. One may then consider the set of Gibbs measures as the set of limit points of all these sequences when first N tends to infinity and then all h_i tend to zero[22].

We will at this point not linger much around deterministic mean-field models such as the Curie–Weiss model and its generalizations. The method of choice for their

[21] The definition given here is however not the only possible one. It is in fact irrelevant which measure is put outside Λ_N.

[22] One can, and sometimes does, apply this procedure also in the case of lattice models. It offers the possibility to construct finite-volume measures on tori, thus allowing to obtain finite-volume measures with explicit translation in- or covariance. See for instance [AW].

analysis is the method of *large deviations*, and an excellent exposition of this method together with applications to mean-field models of this type may be found in the textbook by Ellis [E].

What should be the natural class of *disordered* mean-field models to study? This question requires some discussion and we will see that there are at least two natural classes that propose themselves. To understand this point, we must consider a second form of writing the Hamiltonian of the Curie–Weiss model. Namely, note that

$$\begin{aligned} H_{N,h}(\sigma) = & -\frac{N}{2}\left(\frac{1}{N}\sum_{i\in\Lambda_N}\sigma_i\right)^2 - hN\frac{1}{N}\sum_{i\in\Lambda_N}\sigma_i \\ & = -N\left(\frac{(m_N(\sigma))^2}{2} + hm_N(\sigma)\right) \end{aligned} \tag{3.3}$$

where

$$m_N(\sigma) \equiv \frac{1}{N}\sum_{i\in\Lambda_N}\sigma_i \tag{3.4}$$

is the *empirical magnetisation*. Thus, the Hamiltonian of the Curie–Weiss model is a function of a *macroscopic variable*. This fact is indeed instrumental for the treatment of the model with large deviation methods, and, more generally, is responsible for the fact that it is rather easily solvable. One may thus consider as proper mean-field models only models that share this property, i.e. whose Hamiltonian is a function of macroscopic variables[23]. One can then seek to introduce macroscopic variables that are in turn dependent on some quenched disorder. The simplest natural example of this kind would be the *random field Curie–Weiss model*. Its Hamiltonian is

$$\begin{aligned} H_N[\omega](\sigma) \equiv & -\frac{1}{2N}\sum_{i,j\in\Lambda_N\times\Lambda_N}\sigma_i\sigma_j - \sum_{i\in\Lambda_N}h_i[\omega]\sigma_i \\ & -\frac{N}{2}(m_N(\sigma))^2 - Nn_N[\omega](\sigma) \end{aligned} \tag{3.5}$$

where

$$n_N[\omega](\sigma) \equiv -\frac{1}{N}\sum_{i\in\Lambda_N}h_i[\omega]\sigma_i \tag{3.6}$$

[23] It is in fact natural from the point of view of large deviations to extend this notion slightly and to consider as macroscopic variables also measure valued functions, such as the empirical distribution. This is of relevance when richer single-spin spaces than $\{-1,+1\}$ are considered.

is a second, random, macroscopic variable. We will discuss more complicated and more interesting generalizations along this line in the context of the so-called *Hopfield models* in Part 4 of these lectures.

While the mean-field analogue of the random field model fits naturally in this class, models with random pair interactions are less naturally incorporated. In fact, the naive analog of the Curie–Weiss Hamiltonian with random couplings would be

$$H_N[\omega](\sigma) = -\frac{1}{2N} \sum_{i,j \in \Lambda_N \times \Lambda_N} J_{ij}[\omega]\sigma_i\sigma_j \tag{3.7}$$

for, say, J_{ij} some family of i.i.d. random variables. This Hamiltonian cannot be written as a function of some macroscopic variables. The properties of this model depend strongly on the choice of random variables J_{ij}. The main interest in this model concerns the case when the random couplings have mean zero. In this case, we will see shortly that the normalization factor N^{-1} is actually inappropriate and must be replaced by $N^{-1/2}$ to obtain an interesting model. This is easily understood by the following argument. We certainly want the energy to be an extensive quantity, i.e. to be of the order of the volume of the system. This means that for typical realisations of the disorder, at least for some spin configurations, we must demand that $H_N(\sigma) = CN$, for some $C > 0$. Thus we want to estimate $\mathbb{P}[\max_\sigma H_N(\sigma) \geq CN]$. But the most simple estimate shows that, with the Hamiltonian defined as in (3.7),

$$\begin{aligned} \mathbb{P}[\max_\sigma H_N(\sigma) \geq CN] &\leq \sum_{\sigma \in \mathcal{S}_N} \mathbb{P}[H_N(\sigma) \geq CN] \\ &= \sum_{\sigma \in \mathcal{S}_N} \inf_{t \geq 0} e^{-tCN} \mathbb{E} e^{t \frac{1}{2N} \sum_{i,j \in \Lambda_N \times \Lambda_N} J_{ij}[\omega]\sigma_i\sigma_j} \\ &= \sum_{\sigma \in \mathcal{S}_N} \inf_{t \geq 0} e^{-tCN} \prod_{i,j \in \Lambda_N \times \Lambda_N} \mathbb{E} e^{t \frac{1}{2N} J_{ij}[\omega]\sigma_i\sigma_j} \end{aligned} \tag{3.8}$$

where we assumed that the exponential moments of J_{ij} exist. A standard estimate then shows that for some constant c, $\mathbb{E} e^{t \frac{1}{2N} J_{ij}[\omega]\sigma_i\sigma_j} \leq e^{c \frac{t^2}{2N^2}}$, and so

$$\mathbb{P}[\max_\sigma H_N(\sigma) \geq CN] \leq 2^N \inf_{t \geq 0} e^{-tCN} e^{ct^2/2} \leq 2^N e^{-\frac{C^2N^2}{2c}} \tag{3.9}$$

which tends to zero with N. Thus, our Hamiltonian is never of order N, but at best of order $\sqrt{N}$. The proper Hamiltonian for what is called the *Sherrington–Kirkpatrick model* (or short SK-model), is thus

$$H_N^{SK} \equiv -\frac{1}{\sqrt{2N}} \sum_{i,j \in \Lambda_N \times \Lambda_N} J_{ij} \sigma_i \sigma_j \tag{3.10}$$

where the random variables $J_{ij} = J_{ji}$ are i.i.d. for $i \leq j$ with mean zero (or at most $J_0 N^{-1/2}$) and variance normalized to one for $i \neq j$ and to two for $i = j$[24]. In its original, and mostly considered form, the distribution is moreover taken to be Gaussian. Note that $\sum_{ij} |N^{-1/2} J_{ij} \sigma_i \sigma_j| \sim N^{3/2}$, and that thus competing signs play a major role.

This model was introduced by Sherrington and Kirkpatrick in 1976 [SK] as an attempt to furnish a simple, solvable mean-field model for the then rather newly discovered class of materials called spin-glasses. However, it turned out that the innocent looking modifications made to create a spin glass model that looks similar to the Curie–Weiss model had rather thoroughly destroyed the simplifying properties that made the latter so easily solvable, and that a model with an enormously complex structure, which on a mathematically rigorous level remains today largely un-understood, had been invented. Using highly innovative ideas based on ad hoc mathematical structures, Parisi (see [MPV]) produced in the mid-eighties a heuristic framework that explained the properties of the model. Only the simplest predictions of this theory have been established with considerable effort, on a rigorous basis over the last years.

We will introduce a somewhat different point of view on the SK-model that allows us to put it in a wider context where simpler models can be found, and some of the reasons for its complexity are more easily understood. This point of view consists in regarding the Hamiltonian (3.10) as a *Gaussian random process* indexed by the set $\mathcal{S}_N$, i.e. by the N-dimensional hypercube. We will restrict our attention to the case when the J_{ij} are centered Gaussian random variables. In this case, $H_N(\sigma)$ is in fact

[24] This choice is for notational convenience. Of course the self-couplings J_{ii} have no physical relevance whatsoever.

a centered Gaussian random process which is fully characterized by its covariance function

$$
\begin{aligned}
r_N(\sigma,\sigma') \equiv &\mathrm{cov}(H_N(\sigma),H_N(\sigma')) = \frac{1}{2N}\sum_{1\leq i,j,l,k\leq N} \mathbb{E}J_{ij}\,J_{kl}\sigma_i\sigma_j\sigma'_k\sigma'_l \\
=&\frac{1}{N}\sum_{1\leq i,j\leq N}\sigma_i\sigma'_i\sigma_j\sigma'_j = N\left(N^{-1}\sum_{i=1}\sigma_i\sigma'_i\right)^2 = NR_N(\sigma,\sigma')^2
\end{aligned} \tag{3.11}
$$

where $R_N(\sigma,\sigma')$ is usually called the *overlap* between the two configurations σ and σ'. It is useful to recall that the overlap is closely related to the *Hamming distance* $d_{HAM}(\sigma,\sigma') \equiv \#(i\leq N : \sigma_i \neq \sigma'_i)$, namely $R_N(\sigma,\sigma') = (1-2N^{-1}d_{HAM}(\sigma,\sigma'))$.

Seen this way, the SK model is a particular example of a class of models whose Hamiltonian are centered Gaussian random process on the hypercube with covariance depending only on $R_N(\sigma,\sigma')$,

$$
\mathrm{cov}(H_N(\sigma),H_N(\sigma')) = Nf(R_N(\sigma,\sigma')) \tag{3.12}
$$

normalized such that $f(1)=1$. A class of particular examples considered in the literature are the so-called p-spin SK models, which are obtained by choosing $f(x) = |x|^p$. They enjoy the property that they may be represented in a form similar to the SK Hamiltonian, except that the two-spin interaction must be replaced by a p-spin one:

$$
H_N^{p-SK}(\sigma) = \frac{-1}{\sqrt{N^{p-1}}}\sum_{1\leq i_1,\dots,i_p\leq N} J_{i_1\dots i_p}\sigma_{i_1}\dots\sigma_{i_p} \tag{3.13}
$$

with $J_{i_1,\dots,i_p}$ i.i.d. standard normal random variables[25]. We will see that the faster decay of the covariance with $p>2$ entails considerable simplifications, but as yet is not enough to get fully to grips with the model. As we will see later, the difficulties in studying the statistical mechanics of these models is closely linked to the understanding of the extremal properties of the corresponding random process. While Gaussian processes have been heavily analyzed in the mathematical literature (see

[25] Sometimes the terms there some indices in the sum (3.13) coincide are ommitted. Although this has an effect for instance on the fluctuations of the free energy (see [BKL]), for our present purposes this is not relevant and we choose the form with the simplest expression for the covariance.

e.g. [LT,A]), it turns out that still not quite enough is known. On the other hand, the heuristic results obtained in the physics literature suggest facts about extremes that go far beyond the existing mathematical results in the field. This is one reason why this particular field of mean-field spin-glass models has considerable intrinsic interest for mathematics.

The class of Gaussian models we have just introduced has, fortunately, an extreme member that turns out to be so simple that full answers can be provided with full rigour. It corresponds to the case when the Gaussian process is just an i.i.d. field. The corresponding model is known as the *random energy model* or *REM* and we present a full analysis in the next section.

3.2. The REM.

The random energy model, introduced by Derrida [D1,D2] can be considered as the ultimate toy model of a disordered system. In this model, rather little is left of the structure of interacting spins, but we will still be able to gain a lot of insight into the peculiarities of disordered systems by studying this simple system. For earlier rigorous results on the REM see e.g. [Ei,OP,GMP,DW].

The REM is a model with state space $\mathcal{S}_N = \{-1,+1\}^N$. For fixed N, the Hamiltonian is given by

$$H_N(\sigma) = -\sqrt{N} X_\sigma \tag{3.14}$$

where X_σ, is a family of 2^N i.i.d. centered random variables. In its original form, these variables are taken to be standard normal random variables, but in this case generalizations pose no problems. We will, however, only consider the Gaussian case here.

3.2.1. The free energy.

Before turning to the question of Gibbs measures, we turn to the simpler question of analysing in some detail the partition function. In this model, the partition function is of course just the sum of i.i.d. random variables, i.e.

$$Z_{\beta,N} \equiv 2^{-N} \sum_{\sigma \in \mathcal{S}_N} e^{\beta\sqrt{N} X_\sigma} \tag{3.15}$$

One usually asks first for the exponential asymptotics of this quantity, i.e. one introduces the *free energy*,

$$F_{\beta,N} \equiv -\frac{1}{N} \ln Z_{\beta,N} \tag{3.16}$$

and tries to find its limit as $N \uparrow \infty$. A remark is in place here. While in lattice models with regular interactions the almost sure *existence* of a non-random limit follows from general principles (e.g. sub-additivity arguments), in mean-field models this is in general not the case. For example, it is unknown whether in the SK model such a limit always exists.

In our simple model we expect of course to be able to compute this limit exactly. In fact, the first guess would be that a *law of large numbers* might hold, implying that $Z_{\beta,N} \sim \mathbb{E} Z_{\beta,N}$, and hence

$$\lim_{N\uparrow\infty} F_{\beta,N} = \lim_{N\uparrow\infty} -\frac{1}{N} \ln \mathbb{E} Z_{\beta,N} = -\frac{\beta^2}{2}, \quad \text{a.s.} \tag{3.17}$$

It turns out that this is indeed true, but only for small enough values of β, and this can be linked precisely to a critical value for the breakdown of the law of large numbers. The analysis of this problem will allow us to compute the free energy exactly.

Theorem 3.1: *In the REM,*

$$\lim_{N\uparrow\infty} \mathbb{E} F_{\beta,N} = \begin{cases} -\frac{\beta^2}{2}, & \text{for } \beta \leq \beta_c \\ -\frac{\beta_c^2}{2} - (\beta-\beta_c)\beta_c, & \text{for } \beta \geq \beta_c \end{cases} \tag{3.18}$$

where $\beta_c = \sqrt{2\ln 2}$.

Proof. We will not give the most efficient proof of this result, but one that introduces useful ideas that can be applied to other models. It uses what we call the method of truncated second moments which was introduced by M. Talagrand [T0,T2,T3]. The REM furnishes a particularly simple setting to explain how this works.

We will first derive a lower bound for $\mathbb{E}F_{\beta,N}$. Note first that by Jensen's inequality, $\mathbb{E}\ln Z \leq \ln \mathbb{E}Z$, and thus

$$\mathbb{E}F_{\beta,N} \geq -\frac{\beta^2}{2} \tag{3.19}$$

On the other hand we have that

$$-\mathbb{E}\frac{d}{d\beta}F_{\beta,N} = N^{-1/2}\mathbb{E}\frac{\mathbb{E}_\sigma X_\sigma e^{\beta\sqrt{N}X_\sigma}}{Z_{\beta,N}} \leq N^{-1/2}\mathbb{E}\max_{\sigma\in\mathcal{S}_N} X_\sigma \leq \beta\sqrt{2\ln 2}(1+C/N) \tag{3.20}$$

for some constant C. (Exercise: Prove the last inequality!). Moreover, since $\frac{d^2}{d\beta^2}F_{\beta,N} \leq 0$ (Exercise!), we may combine (3.19) and (3.20) to deduce that

$$\mathbb{E}F_{\beta,N} \geq \sup_{\beta_0\geq 0}\begin{cases} -\frac{\beta^2}{2}, & \text{for } \beta\leq\beta_0 \\ -\frac{\beta_0^2}{2}-(\beta-\beta_0)\sqrt{2\ln 2}(1+C/N), & \text{for } \beta\geq\beta_0 \end{cases} \tag{3.21}$$

It is easy to see that the supremum is realized (ignore the C/N correction) for $\beta_0 = \sqrt{2\ln 2}$. This shows that the right hand side of (3.18) is a lower bound.

It remains to show the corresponding upper bound. The basic idea behind this approach is to obtain a variance estimate on the partition function[26] . Naively, one would compute

$$\begin{aligned}\mathbb{E}Z_{\beta,N}^2 &= \mathbb{E}_\sigma\mathbb{E}_{\sigma'}\mathbb{E}e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})} \\ &= 2^{-2N}\left(\sum_{\sigma\neq\sigma'}e^{N\beta^2}+\sum_\sigma e^{2N\beta^2}\right) \\ &= e^{N\beta^2}\left[(1-2^{-N})+2^{-N}e^{N\beta^2}\right]\end{aligned} \tag{3.22}$$

where all we used is that for $\sigma\neq\sigma'$ X_σ and $X_{\sigma'}$ are independent. Now we see that the second term in the square brackets is exponentially small if and only if $\beta^2 < \ln 2$.

[26]This idea can be traced to Aizenman, Lebowitz, and Ruelle [ALR], and later Comets and Neveu [CN] who used it in the proofs of a central limit theorem for the free energy.

For such values of β we have that

$$
\begin{aligned}
\mathbb{P}\left[\left|\ln \frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}}\right| > \epsilon N\right] &= \mathbb{P}\left[\frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} < e^{-\epsilon N} \text{ or } \frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} > e^{\epsilon N}\right] \\
&\leq \mathbb{P}\left[\left(\frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} - 1\right)^2 > \left(1 - e^{-\epsilon N}\right)^2\right] \\
&\leq \frac{\mathbb{E}Z^2_{\beta,N}/(\mathbb{E}Z_{\beta,N})^2 - 1}{(1 - e^{-\epsilon N})^2} \\
&\leq \frac{2^{-N} + 2^{-N} e^{N\beta^2}}{(1 - e^{-\epsilon N})^2}
\end{aligned}
\tag{3.23}
$$

which is more than enough to get (3.17). But of course this does not correspond to the critical value of β claimed in the proposition! Some reflection shows that the point here is that when computing $\mathbb{E}e^{\beta\sqrt{N}2X_\sigma}$, the dominant contribution comes from the part of the distribution of X_σ where $X_\sigma \sim 2\beta\sqrt{N}$, whereas in the evaluation of $\mathbb{E}Z_{\beta,N}$ the values of X_σ where $X_\sigma \sim \beta\sqrt{N}$ give the dominant contribution. Thus one is led to the realization that it is not the second moment of Z one should control, but rather that of a truncated version of Z, namely, for $c \geq 0$,

$$
\widetilde{Z}_{\beta,N}(c) \equiv \mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma} \mathbb{1}_{X_\sigma < c\sqrt{N}} \tag{3.24}
$$

An elementary computation using the standard bound, for $u > 0$,

$$
\frac{1}{\sqrt{2\pi}u} e^{-u^2/2}\left(1 - 2u^{-2}\right) \leq \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-x^2/2} dx \leq \frac{1}{\sqrt{2\pi}u} e^{-u^2/2} \tag{3.25}
$$

shows that

$$
\mathbb{E}\widetilde{Z}_{\beta,N}(c) = \begin{cases} e^{\beta^2 N}\left(1 - \frac{e^{-N(-\beta)^2/2}}{\sqrt{2\pi N}(c-\beta)}(1 + O(1/N))\right), & \text{if } \beta < c \\ \frac{1+O(1/N)}{\sqrt{2\pi N}(\beta-c)} e^{N\beta c - \frac{Nc^2}{2}}, & \text{if } \beta > c \end{cases} \tag{3.26}
$$

Note that (3.26) shows that this truncation essentially does not influence the mean partition function, if $\beta < c$.

But now compute the mean of the square of the truncated partition function:

$$
\mathbb{E}\widetilde{Z}^2_{\beta,N}(c) = (1 - 2^{-N})[\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2 + 2^{-N}\mathbb{E}e^{\beta\sqrt{N}2X_\sigma} \mathbb{1}_{X_\sigma < c\sqrt{N}} \tag{3.27)}
$$

where the second term satisfies (we do not mention the irrelevant $O(1/N)$ error term anymore)

$$2^{-N}\mathbb{E}\, e^{2\beta\sqrt{N}X_\sigma}\mathbb{1}_{X_\sigma<c\sqrt{N}} \leq \begin{cases} 2^{-N}e^{2\beta^2 N}, & \text{if } 2\beta < c \\ 2^{-N}\frac{e^{2c\beta N - \frac{c^2 N}{2}}}{(2\beta-c)\sqrt{2\pi N}}, & \text{otherwise,} \end{cases} \tag{3.28}$$

Combined with (3.26) this implies

$$\frac{2^{-N}\mathbb{E}\, e^{2\beta\sqrt{N}X_\sigma}\mathbb{1}_{X_\sigma<c\sqrt{N}}}{\left(\mathbb{E}\,\widetilde{Z}_{N,\beta}\right)^2} \leq \begin{cases} e^{-N(\ln 2-\beta^2)}, & \text{if } \beta < \frac{c}{2}, \\ \frac{e^{-N(c-\beta)^2 - N(\ln 2 - \frac{c^2}{2})}}{(2\beta-c)\sqrt{N}}, & \text{if } \frac{c}{2} < \beta < c, \\ e^{(c^2/2-\ln 2)N}\sqrt{2\pi N}\frac{(\beta-c)^2}{2\beta-c}, & \text{if } \beta > c \end{cases} \tag{3.29}$$

Therefore, for all $c < \sqrt{2\ln 2}$, and all $\beta \neq c$,

$$\mathbb{E}\left[\frac{\widetilde{Z}_{\beta,N}(c) - \mathbb{E}\widetilde{Z}_{\beta,N}(c)}{\mathbb{E}\widetilde{Z}_{\beta,N}(c)}\right]^2 \leq e^{-Ng(c,\beta)} \tag{3.30}$$

with $g(c,\beta) > 0$. Thus Chebyshev's inequality implies that

$$\mathbb{P}\left[|\widetilde{Z}_{\beta,N}(c) - \mathbb{E}\widetilde{Z}_{\beta,N}(c)| > \delta\mathbb{E}\widetilde{Z}_{\beta,N}(c)\right] \leq \delta^{-2}e^{-Ng(c,\beta)} \tag{3.31}$$

which implies in particular that

$$\lim_{N\uparrow\infty}\frac{1}{N}\mathbb{E}\ln\widetilde{Z}_{\beta,N}(c) = \lim_{N\uparrow\infty}\frac{1}{N}\ln\mathbb{E}\widetilde{Z}_{\beta,N}(c) \tag{3.32}$$

for all $c < \sqrt{2\ln 2} = \beta_c$. But this implies that for all $c < \beta_c$

$$\lim_{N\uparrow\infty}\frac{1}{N}\ln\mathbb{E}Z_{\beta,N} \geq \lim_{N\uparrow\infty}\frac{1}{N}\ln\mathbb{E}\widetilde{Z}_{\beta,N}(c) = \begin{cases} \frac{\beta^2}{2}, & \text{for } \beta < c \\ \frac{c^2}{2} + c(\beta - c), & \text{for } \beta > c \end{cases} \tag{3.33}$$

which converges to minus the right hand side of (3.18) as $c \uparrow \beta_c$. This proves the theorem.$\diamondsuit$

3.2.2. Fluctuations and limit theorems.

In the previous section we went to some length in computing the limit of the free energy. However, computing the free energy is not quite enough to get a full understanding of a model, and in particular the Gibbs states. The REM offers the advantage that we can go much further, as we will see now. The limit of the free energy has been seen to be a non-random quantity. A question of central importance for further progress is to understand on what level and how the randomness shows up in the corrections to the limiting behaviour. This is the subject of the present section. We will see later that the knowledge gained here is then sufficient to derive all information about the Gibbs measures we would want. The results of this section were obtained in [BKL]. For earlier rigorous results see also [Ei,OP,GMP].

Let us first state the results:

Theorem 3.2: *The partition function of the REM has the following fluctuations:*

(i) If $\beta < \sqrt{\ln 2/2}$, *then*

$$e^{\frac{N}{2}(\ln 2-\beta^2)} \ln \frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.34}$$

(ii) If $\beta = \sqrt{\ln 2/2}$, *then*

$$\sqrt{2}e^{\frac{N}{2}(\ln 2-\beta^2)} \ln \frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.35}$$

(iii) Let $\alpha \equiv \beta/\sqrt{2\ln 2}$. *If* $\sqrt{\ln 2/2} < \beta < \sqrt{2\ln 2}$, *then*

$$e^{\frac{N}{2}(\sqrt{2\ln 2}-\beta)^2+\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]} \ln \frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} \xrightarrow{\mathcal{D}} \int_{-\infty}^{\infty} e^{\alpha z}(\mathcal{P}(dz)-e^{-z}dz), \tag{3.36}$$

where $\mathcal{P}$ *denotes the Poisson point process*[27] *on* $\mathbb{R}$ *with intensity measure* $e^{-x}dx$.

(iv) If $\beta = \sqrt{2\ln 2}$, *then*

$$e^{\frac{1}{2}[\ln(N\ln 2)+\ln 4\pi]}\left(\frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}}-\frac{1}{2}+\frac{\ln(N\ln 2)+\ln 4\pi}{4\sqrt{\pi N\ln 2}}\right) \xrightarrow{\mathcal{D}} \int_{-\infty}^{0} e^{z}(\mathcal{P}(dz)-e^{-z}dz)+\int_0^{\infty} e^{z}\mathcal{P}(dz). \tag{3.37}$$

[27] For a thorough exposition on point processes and their connection to extreme value theory, see in particular [Re].

(v) If $\beta > \sqrt{2\ln 2}$, then

$$e^{-N[\beta\sqrt{2\ln 2}-\ln 2]+\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]} Z_{\beta,N} \xrightarrow{\mathcal{D}} \int_{-\infty}^{\infty} e^{\alpha z}\mathcal{P}(dz) \tag{3.38}$$

and

$$\ln Z_{\beta,N} - \mathbb{E}\ln Z_{\beta,N} \xrightarrow{\mathcal{D}} \ln \int_{-\infty}^{\infty} e^{\alpha z}\mathcal{P}(dz) - \mathbb{E}\ln \int_{-\infty}^{\infty} e^{\alpha z}\mathcal{P}(dz). \tag{3.39}$$

Remark. Note that expressions like $\int_{-\infty}^{0} e^{z}(\mathcal{P}(dz) - e^{-z}dz)$ are always understood as $\lim_{y\downarrow-\infty}\int_{y}^{0} e^{z}(\mathcal{P}(dz) - e^{-z}dz)$. We will see that all the functionals of the Poisson point process appearing are almost surely finite random variables.

We will now prove these results step by step. The first thing one might guess about fluctuations in the REM is that they should stem from a central limit theorem for the partition function, that is, after all, a sum of i.i.d. random variables. The theorem shows that this can be true only for quite small values of β, much smaller even than the critical value.

Proof. We first prove (i). Here the result follows from the standard CLT for triangular arrays. Let us first write

$$\ln \frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} = \ln\Big(1 + \frac{Z_{\beta,N} - \mathbb{E}Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}}\Big). \tag{3.40}$$

We will show that the second term in the logarithm properly normalized will converge to a normal random variable. To see this, write

$$\frac{Z_{\beta,N} - \mathbb{E}Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} = \sum_{\sigma\in\mathcal{S}_N} e^{-N(\ln 2+\beta^2/2)}\left(e^{\beta\sqrt{N}X_\sigma} - e^{N\beta^2/2}\right) \equiv \sum_{\sigma\in\mathcal{S}_N}\mathcal{Y}_N(\sigma). \tag{3.41}$$

Note that $\mathbb{E}\mathcal{Y}_N(\sigma) = 0$ and $\mathbb{E}\mathcal{Y}_N^2(\sigma) = e^{-N(2\ln 2-\beta^2)}[1 - e^{-N\beta^2}]$ and thus

$$\mathbb{E}\Big(\frac{Z_{\beta,N} - \mathbb{E}Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}}\Big)^2 = e^{-N(\ln 2-\beta^2)}[1 - e^{-N\beta^2}]. \tag{3.42}$$

Therefore we can write

$$\frac{Z_{\beta,N} - \mathbb{E}Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} = e^{-\frac{N}{2}(\ln 2-\beta^2)}\sqrt{1 - e^{-N\beta^2}}\,\frac{1}{2^{N/2}}\sum_{\sigma\in\mathcal{S}_N}\widetilde{\mathcal{Y}}_N(\sigma), \tag{3.43}$$

where $\widetilde{\mathcal{Y}}_N(\sigma) = e^{\frac{N}{2}(2\ln 2-\beta^2)}[1-e^{-N\beta^2}]^{-1/2}\mathcal{Y}_N(\sigma)$ has mean zero and variance one. By the CLT for triangular arrays (see [Shi]), it follows readily that

$$\frac{1}{2^{N/2}}\sum_{\sigma\in\mathcal{S}_N}\widetilde{\mathcal{Y}}_N(\sigma)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1) \tag{3.44}$$

if the Lindeberg condition holds, that is, if for any $\epsilon>0$,

$$\lim_{N\uparrow\infty}\mathbb{E}\widetilde{\mathcal{Y}}_N^2(\sigma)\mathbb{1}_{\{|\widetilde{\mathcal{Y}}_N(\sigma)|\geq\epsilon 2^{N/2}\}}=0. \tag{3.45}$$

But

$$\begin{aligned}\mathbb{E}\widetilde{\mathcal{Y}}_N^2(\sigma)\mathbb{1}_{\{|\widetilde{\mathcal{Y}}_N(\sigma)|\geq\epsilon 2^{N/2}\}}&=\frac{1}{\sqrt{2\pi}(1-e^{-N\beta^2})}e^{-2N\beta^2}\int\limits_{\sqrt{N}(\frac{\ln 2}{2\beta}+\beta)+\frac{\ln\epsilon}{\sqrt{N}\beta}+o(\frac{1}{\sqrt{N}})}^{\infty}e^{2\sqrt{N}\beta z-\frac{z^2}{2}}dz+o(1)\\&=\frac{1}{\sqrt{2\pi}(1-e^{-N\beta^2})}\int\limits_{\sqrt{N}(\frac{\ln 2}{2\beta}-\beta)+\frac{\ln\epsilon}{\sqrt{N}\beta}+o(\frac{1}{\sqrt{N}})}^{\infty}e^{-\frac{z^2}{2}}dz+o(1).\end{aligned} \tag{3.46}$$

It is easy to check that the latter integral converges to zero if and only if $\beta^2<\ln 2/2$. Using now the fact that $e^x=1+x+o(x)$ as $x\to 0$, it is now a trivial matter to deduce the assertion of the proposition. $\diamondsuit$

Since the Lindeberg condition clearly fails for $2\beta^2\geq\ln 2$, it is clear that we cannot expect a simple CLT beyond this regime. Such a failure of a CLT is always a problem related to "heavy tails", and results from the fact that extremal events begin to influence the fluctuations of the sum. It appears therefore reasonable to separate from the sum the terms where X_σ is anomalously large. For Gaussian r.v.'s it is well known that the right scale of separation is given by $u_N(x)$ defined by

$$2^N\int\limits_{u_N(x)}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-z^2/2}=e^{-x} \tag{3.47}$$

which (for $x>-\ln N/\ln 2$) is equal to (see e.g. [LLR])

$$u_N(x)=\sqrt{2N\ln 2}+\frac{x}{\sqrt{2N\ln 2}}-\frac{\ln(N\ln 2)+\ln 4\pi}{2\sqrt{2N\ln 2}}+o(1/\sqrt{N}), \tag{3.48}$$

$x \in \mathbb{R}$ is a parameter. The key to most of what follows relies on the famous result on the convergence of the extreme value process to a Poisson point process. Let us now introduce the point process on $\mathbb{R}$ given by

$$\mathcal{P}_N \equiv \sum_{\sigma \in \mathcal{S}_N} \delta_{u_N^{-1}(X_\sigma)}. \tag{3.49}$$

A classical result from the theory of extreme order statistics (see e.g. [LLR]) asserts that **Theorem 3.3:** *The point process $\mathcal{P}_N$ converges weakly to a Poisson point process on $\mathbb{R}$ with intensity measure $e^{-x}dx$.*

Let us now define

$$Z^x_{N,\beta} \equiv \mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma} \mathbb{1}_{\{X_\sigma \leq u_N(x)\}}. \tag{3.50}$$

We may write

$$\frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} = 1 + \frac{\mathbb{Z}^x_{\beta,N} - \mathbb{E}Z^x_{\beta,N}}{\mathbb{E}Z_{\beta,N}} + \frac{Z_{\beta,N} - Z^x_{\beta,N} - \mathbb{E}(Z_{\beta,N} - Z^x_{\beta,N})}{\mathbb{E}Z_{\beta,N}} \tag{3.51}$$

Let us first consider the last summand. We introduce the random variable

$$\mathcal{W}_N(x) = \frac{Z_{\beta,N} - Z^x_{\beta,N}}{\mathbb{E}Z_{\beta,N}} = e^{-N(\ln 2 + \beta^2/2)} \sum_{\sigma \in \mathcal{S}_N} e^{\beta\sqrt{N}X_\sigma} \mathbb{1}_{\{X_\sigma > u_N(x)\}} \tag{3.52}$$

It will be convenient to rewrite this as (we ignore the sub-leading corrections to $u_N(x)$ and only keep the explicit representation (3.48))

$$\begin{aligned} \mathcal{W}_N(x) &= e^{-N(\ln 2 + \beta^2/2)} \sum_{\sigma \in \mathcal{S}_N} e^{\beta\sqrt{N}u_N(u_N^{-1}(X_\sigma))} \mathbb{1}_{\{u_N^{-1}(X_\sigma) > x\}} \\ &= e^{-N(\ln 2 + \beta^2/2)} e^{\beta N\sqrt{2\ln 2} - \beta\frac{\ln(N\ln 2) + \ln 4\pi}{2\sqrt{2\ln 2}}} \sum_{\sigma \in \mathcal{S}_N} e^{\frac{\beta}{\sqrt{2\ln 2}}u_N^{-1}(X_\sigma)} \mathbb{1}_{\{u_N^{-1}(X_\sigma) > x\}}. \end{aligned} \tag{3.53}$$

Clearly, the weak convergence of $\mathcal{P}_N$ to $\mathcal{P}$ implies convergence in law of the right hand side of (3.53), provided that $e^{\alpha x}$ is integrable on $[x, \infty)$ w.r.t. the Poisson process with intensity e^{-x}. This is, in fact, never a problem: the Poisson point process has almost surely support on a finite set, and therefore $e^{\alpha x}$ is always a.s. integrable. Note, however, that for $\beta \geq \sqrt{2\ln 2}$ the mean of the integral is infinite, indicating the passage to the low temperature regime. Note also that the variance of the integral is finite exactly if $\alpha < 1/2$, i.e. $\beta^2 < \ln 2/2$, i.e. when the CLT holds. On the other hand, the mean of the integral diverges if $x \downarrow \infty$; note that at minus infinity the points of the Poisson point process accumulate, and there is no finite support argument as before that would assure the existence of the integral if x is taken to $-\infty$. The following lemma provides the first step in the proof of part (iii) of Theorem 3.2:

Lemma 3.4: *Let $\mathcal{W}_N(x), \alpha$ be defined as above, and let $\mathcal{P}$ be the Poisson point process with intensity measure $e^{-z}dz$. Then*

$$e^{\frac{N}{2}(\sqrt{2\ln 2}-\beta)^2+\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]}\mathcal{W}_N(x) \xrightarrow{\mathcal{D}} \int_x^\infty e^{\alpha z}\mathcal{P}(dz). \tag{3.54}$$

Remark. Note that the mean of the right hand side is finite if and only if $\beta < \sqrt{2\ln 2}$. Thus only in that case does this lemma also allows us to deal with the centered variable appearing in (3.51).

We now need to turn to the remaining term,

$$\frac{Z^x_{\beta,N} - \mathbb{E}Z^x_{\beta,N}}{\mathbb{E}Z_{\beta,N}} = \frac{\mathcal{V}_N(x)}{\mathbb{E}Z_{\beta,N}}, \tag{3.55}$$

where

$$\mathcal{V}_N(x) \equiv Z^x_{\beta,N} - \mathbb{E}Z^x_{\beta,N}. \tag{3.56}$$

One might first hope that this term upon proper scaling would converge to a Gaussian; however, one can easily check that this is not the case (the Lindeberg condition will not be verified). However, it will not be hard to compute all moments of this term:

Lemma 3.5: *Let $\mathcal{V}_N(x)$ be defined by (3.56). Then for $\alpha > 1/2$ and any integer $k \geq 2$*

$$\begin{aligned}\lim_{N\uparrow+\infty} &\frac{\mathbb{E}[\mathcal{V}_N(x)]^k}{\left[2^{-N}e^{N\beta\sqrt{2\ln 2}-\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]}\right]^k}\\ &= \sum_{i=1}^k \frac{1}{i!} \sum_{\substack{\ell_1\geq 2,\ldots,\ell_i\geq 2\\ \sum_j \ell_j=k}} \frac{k!}{\ell_1!\ldots\ell_i!}\frac{e^{(k\alpha-i)x}}{(\ell_1\alpha-1)\ldots(\ell_i\alpha-1)}\end{aligned} \tag{3.57}$$

For $\alpha = 1/2$, we have for k even

$$\lim_{N\uparrow+\infty} \frac{\mathbb{E}[\mathcal{V}_N(x)]^k}{\left[2^{-N}e^{N\beta\sqrt{2\ln 2}}\right]^k} = \frac{k!}{(k/2)!\,2^k} = \frac{(k-1)!!}{2^{k/2}} \tag{3.58}$$

and for k odd

$$\lim_{N\uparrow+\infty} \frac{\mathbb{E}[\mathcal{V}_N(x)]^k}{\left[2^{-N}e^{N\beta\sqrt{2\ln 2}}\right]^k} = 0 \tag{3.59}$$

(which are the moments of the normal distribution with variance $1/2$).

Proof. This is a pure computation. Set $T_N(\sigma) \equiv e^{\beta\sqrt{N}X_\sigma}\mathbb{1}_{\{X_\sigma \le u_N(x)\}}$. Note that for $\beta < \sqrt{2\ln 2}$

$$\mathbb{E}T_N(\sigma) = \int_{-\infty}^{u_N(x)} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}+\beta\sqrt{N}z} = e^{N\beta^2/2}\Big(1 - \int_{u_N(x)-\beta\sqrt{N}}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\Big) \sim e^{\beta^2 N/2}. \tag{3.60}$$

while for $\beta > \sqrt{2\ln 2}$ and all $k \ge 1$, and for $\beta > \sqrt{\ln 2/2}$ and for $k \ge 2$,

$$\begin{aligned} \mathbb{E}[T_N(\sigma)]^k &= \int_{-\infty}^{u_N(x)} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}+k\beta\sqrt{N}z} = e^{Nk^2\beta^2/2} \int_{-\infty}^{u_N(x)-k\beta\sqrt{N}} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \\ &\sim e^{Nk^2\beta^2/2} \frac{e^{-(u_N(x)-k\beta\sqrt{N})^2/2}}{\sqrt{2\pi}(k\beta\sqrt{N} - u_N(x))} \sim \frac{2^{-N}e^{-x}}{k\alpha - 1} e^{k[\beta\sqrt{2\ln 2}N + \alpha x - \frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]]}. \end{aligned} \tag{3.61}$$

Formula (3.61) is also valid for $\beta = \sqrt{2\ln 2}$ with $k > 1$ and for $\beta = \sqrt{\ln 2/2}$ with $k > 2$. It is easy to see from the computations above that for $\beta = \sqrt{2\ln 2}$ with $k = 1$ and also for $\beta = \sqrt{\ln 2/2}$ with $k = 2$ we have

$$\mathbb{E}[T_N(\sigma)]^k \sim \frac{e^{k^2\beta^2N/2}}{2} = \frac{2^{-N}e^{-x}}{2} e^{k[\beta\sqrt{2\ln 2}N + \alpha x]}. \tag{3.62}$$

We set $\widetilde{T}_N(\sigma) \equiv 2^{-N}T_N(\sigma)$; by (3.61) we get for $\beta > \sqrt{\ln 2/2}$ with $k \ge 2$ and also for $\beta > \sqrt{2\ln 2}$ with $k \ge 1$

$$\mathbb{E}[\widetilde{T}_N(\sigma)]^k = \frac{2^{-N}e^{-x}}{k\alpha - 1} e^{k[\beta\sqrt{2\ln 2}N - \ln 2 + \alpha x - \frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]]}. \tag{3.63}$$

This formula is also true for $\beta = \sqrt{\ln 2/2}$, $k > 2$ and $\beta = \sqrt{2\ln 2}$, $k > 1$. For $\beta = \sqrt{2\ln 2}$ and $k = 1$ and also for $\beta = \sqrt{\ln 2/2}$ and $k = 2$ by (3.62)

$$\mathbb{E}[\widetilde{T}_N(\sigma)]^k = \frac{2^{-N}e^{-x}}{2} e^{k[\beta\sqrt{2\ln 2}N - \ln 2 + \alpha x]}. \tag{3.64}$$

Now

$$\begin{aligned}\mathbb{E}\,[\mathcal{V}_N(x)]^k &= \mathbb{E}\Big(\sum_{\sigma\in\mathcal{S}_N}[\widetilde{T}_N(\sigma) - \mathbb{E}\widetilde{T}_N(\sigma)]\Big)^k = \sum_{\sigma_1,\dots,\sigma_k\in\mathcal{S}_N} \mathbb{E}\prod_{i=1}^k[\widetilde{T}_N(\sigma_i) - \mathbb{E}\widetilde{T}_N(\sigma_i)] \\ &= \sum_{i=1}^k \sum_{\substack{\ell_1,\dots,\ell_i\geq 2\\ \sum_j \ell_j = k}} \frac{k!}{\ell_1!\dots\ell_i!}\binom{2^N}{i}\mathbb{E}[\widetilde{T}_N(\sigma) - \mathbb{E}\widetilde{T}_N(\sigma)]^{\ell_1}\dots\mathbb{E}[\widetilde{T}_N(\sigma) - \mathbb{E}\widetilde{T}_N(\sigma)]^{\ell_i}.\end{aligned} \tag{3.65}$$

Note finally that for $l \geq 2$ and $\beta \geq \sqrt{\ln 2/2}$

$$\mathbb{E}\big[\widetilde{T}_N(\sigma) - \mathbb{E}\widetilde{T}_N(\sigma)\big]^\ell = \sum_{j=1}^\ell (-1)^j\binom{\ell}{j}\mathbb{E}\widetilde{T}_N(\sigma)^{\ell-j}[\mathbb{E}\widetilde{T}_N(\sigma)]^j \sim \mathbb{E}\widetilde{T}_N(\sigma)^\ell. \tag{3.66}$$

In fact, if $\sqrt{\ln 2/2} \leq \beta < \sqrt{2\ln 2}$, $l \geq 2$, $j \geq 1$, $j \neq l-1, l$, then by (3.60) and (3.63), (3.64)

$$\frac{\mathbb{E}[T_N^{l-j}(\sigma)][\mathbb{E}T_N(\sigma)]^j}{\mathbb{E}[T_N^l(\sigma)]} = e^{Nj(\beta^2/2 - \beta\sqrt{2\ln 2})}O\left(N^{\alpha j/2}\right) \tag{3.67}$$

is exponentially small. For $l \geq 2$, $j = l-1, l$

$$\frac{\mathbb{E}[T_N^{l-j}(\sigma)][\mathbb{E}T_N(\sigma)]^j}{\mathbb{E}[T_N^l(\sigma)]} = e^{Nl(\beta^2/2 - \beta\sqrt{2\ln 2}) + N\ln 2}O\left(N^{\alpha l/2}\right) \leq e^{-N\ln 2}O\left(N^{l\alpha/2}\right) \tag{3.68}$$

For $\beta \geq \sqrt{2\ln 2}$, $l \geq 2$ and $j \geq 1$ by (3.63) and (3.64)

$$\frac{\mathbb{E}[T_N^{l-j}(\sigma)][\mathbb{E}T_N(\sigma)]^j}{\mathbb{E}[T_N^l(\sigma)]} = O(2^{-Nj}). \tag{3.69}$$

Thus for $l \geq 2$ and $\beta > \sqrt{\ln 2/2}$ and also for $l \geq 3$ and $\beta = \sqrt{\ln 2/2}$

$$\mathbb{E}\big[\widetilde{T}_N(\sigma) - \mathbb{E}\widetilde{T}_N(\sigma)\big]^\ell \sim \frac{2^{-N}e^{-x}}{k\alpha - 1}\big[2^{-N}e^{N\beta\sqrt{2\ln 2}}e^{\alpha x}e^{-\frac{\alpha}{2}[\ln(N\ln 2) + \ln 4\pi]}\big]^\ell. \tag{3.70}$$

Inserting this result into (3.65) gives the assertion of the lemma (3.57).

For $\beta = \sqrt{\ln 2/2}$ and $l = 2$ by (3.64) we have

$$\mathbb{E}\big[\widetilde{T}_N(\sigma) - \mathbb{E}\widetilde{T}_N(\sigma)\big]^2 \sim \frac{2^{-N}e^{-x}}{2}\big[2^{-N}e^{N\beta\sqrt{2\ln 2}}e^{\alpha x}\big]^2. \tag{3.71}$$

Inserting this formula into (3.65) we see, that the term with $l_1, \dots, l_i = 2$, $i = k/2$ brings the main contribution to the sum, and all others are of smaller order, because of the polynomial terms $e^{-l\frac{\alpha}{2}\ln(N\ln 2)}$ in (3.70). This implies (3.58) and (3.59) and the lemma is proved. $\diamondsuit$

Remark. One sees that if we let $x \downarrow -\infty$, and rescale properly, the corresponding moments converge to that of a centered Gaussian r.v. This could alternatively be seen by checking that the Lindeberg condition holds for the truncated variables provided $x \leq -2\ln\ln 2^N$.

A standard consequence of Lemma 3.5 is the weak convergence of the normalized version of $\mathcal{V}_N(x)$:

Corollary 3.6: *For $\sqrt{\ln 2/2} < \beta$,*

$$e^{\frac{N}{2}(\sqrt{2\ln 2}-\beta)^2+\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]}\frac{\mathcal{V}_N(x)}{\mathbb{E}Z_{\beta,N}} \xrightarrow{\mathcal{D}} \mathcal{V}(x,\alpha), \tag{3.72}$$

where $\mathcal{V}(x,\alpha)$ is the random variable with mean zero and kth moments given by the right hand side of (3.57). For $\beta = \sqrt{\ln 2/2}$

$$\sqrt{2}e^{\frac{N}{2}(\sqrt{2\ln 2}-\beta)^2}\frac{\mathcal{V}_N(x)}{\mathbb{E}Z_{\beta,N}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.73}$$

Equation (3.73) together with Lemma 3.4 imply that (ii) of Theorem 3.2 holds. The next proposition will imply (iii) of Theorem 3.2.

Proposition 3.7: *Let $\sqrt{\ln 2/2} < \beta < \sqrt{2\ln 2}$. Then for $x \in \mathbb{R}$ chosen arbitrarily,*

$$e^{\frac{N}{2}(\sqrt{2\ln 2}-\beta)^2+\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]}\ln\frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} \xrightarrow{\mathcal{D}} \mathcal{V}(x,\alpha) + \int_x^\infty e^{\alpha z}\mathcal{P}(dz) - \int_x^\infty e^{\alpha z}e^{-z}dz, \tag{3.74}$$

where $\mathcal{V}(x,\alpha)$ and $\mathcal{P}$ are independent random variables.

Proof. (3.74) would be immediate from Lemma 3.4 and Corollary 3.6, if $\mathcal{W}_N(x)$ and $\mathcal{V}_N(x)$ were independent. However, while this is not true, they are not far from independent. To see this, note that if we condition on the number of variables X_σ, $n_N(x)$, that exceed $u_N(x)$, the decomposition in (3.51) is independent. On the other hand, one readily verifies that Corollary 3.6 also holds under the conditional law $\mathbb{P}[\cdot|n_N(x)=n]$, for any finite n, with the same right hand side $\mathcal{V}(x,\alpha)$. But this implies that the limit can be written as the sum of two independent random variables, as desired. $\diamondsuit$

Since for $\beta^2 > \ln 2/2$, $\alpha > 1/2$, one sees that $\mathbb{E}\mathcal{V}(x,\alpha)^2 = e^{x(2\alpha-1)}/(2\alpha-1)$ tends to zero as $x \downarrow -\infty$. Therefore we see that

$$\mathcal{V}(x,\alpha) \stackrel{\mathcal{D}}{=} \lim_{y\uparrow+\infty} \int_{-y}^{x} e^{\alpha z}\mathcal{P}(dz) - \int_{-y}^{x} e^{\alpha z} e^{-z} dz \tag{3.75}$$

which means that we can give sense to the Poisson integral $\int_{-\infty}^{\infty} e^{\alpha z}(\mathcal{P}(dz) - e^{-z}dz)$ From here we get (iii) of Theorem 3.2.

We now turn to the proof of parts (iv) an (v) of Theorem 3.2. We will see that the computations above almost suffice to conclude the low temperature case as well. With the notations from above, we write

$$Z_{\beta,N} = Z^x_{\beta,N} + (Z_{\beta,N} - Z^x_{\beta,N}) \tag{3.76}$$

Clearly for $\beta \geq \sqrt{2\ln 2}$

$$Z_{\beta,N} - Z^x_{\beta,N} = e^{N\left[\beta\sqrt{2\ln 2}-\ln 2\right]-\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]} \sum_{\sigma\in\mathcal{S}_N} \mathbb{I}_{\{u_N^{-1}(\sigma)>x\}} e^{\alpha u_N^{-1}(X_\sigma)} \tag{3.77}$$

so that for any $x \in \mathbb{R}$,

$$(Z_{\beta,N} - Z^x_{\beta,N}) e^{-N\left[\beta\sqrt{2\ln 2}-\ln 2\right]+\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]} \stackrel{\mathcal{D}}{\to} \int_x^{\infty} e^{\alpha z}\mathcal{P}(dz). \tag{3.78}$$

Now write

$$Z^x_{\beta,N} = \mathbb{E}Z^x_{\beta,N}\Big(1 + \frac{Z^x_{\beta,N} - \mathbb{E}Z^x_{\beta,N}}{\mathbb{E}Z^x_{\beta,N}}\Big). \tag{3.79}$$

Let us first treat the case $\beta > \sqrt{2\ln 2}$. By (3.61) we have

$$\mathbb{E}Z^x_{\beta,N} \sim \frac{2^{-N}e^{-x}}{\alpha-1} e^{\beta\sqrt{2\ln 2}N+\alpha x-\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]}. \tag{3.80}$$

Thus

$$e^{-N\left[\beta\sqrt{2\ln 2}-\ln 2\right]+\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]} Z^x_{\beta,N} = \frac{e^{x(\alpha-1)}}{\alpha-1}\Big(1+\frac{Z^x_{\beta,N}-\mathbb{E}Z^x_{\beta,N}}{\mathbb{E}Z^x_{\beta,N}}\Big)(1+o(1)). \tag{3.81}$$

Using Lemma 3.5 we see that now $\frac{Z^x_{\beta,N}-\mathbb{E}Z^x_{\beta,N}}{\mathbb{E}Z^x_{\beta,N}}\frac{e^{x(\alpha-1)}}{\alpha-1}$ converges in distribution to a random variable with moments given by the right hand side of (3.57). Moreover, as $x \downarrow -\infty$, this variable converges to zero in probability. Since the same is true for the prefactor, the assertion of the theorem is now immediate.

Let us consider now the case $\beta = \sqrt{2\ln 2}$. Proceeding as in (3.61),

$$\mathbb{E}Z^0_{\beta,N} = \frac{2^N}{\sqrt{2\pi}} \int_{-\infty}^{u_N(0)-\sqrt{2N\ln 2}} e^{-z^2/2}dz = 2^N\Big(\frac{1}{2} - \frac{\ln(N\ln 2)+\ln 4\pi}{4\sqrt{N\pi\ln 2}} + O\Big(\frac{(\ln N)^2}{N}\Big)\Big). \tag{3.82}$$

We use the decomposition

$$Z_{\beta,N} = Z_{\beta,N} - Z^0_{\beta,N} + \mathbb{E}Z^0_{\beta,N} + (Z^0_{\beta,N} - \mathbb{E}Z^0_{\beta,N}). \tag{3.83}$$

By (3.82), $\mathbb{E}Z^0_{\beta,N}/\mathbb{E}Z_{\beta,N} \sim 1/2$. By (3.53), we see easily that

$$\frac{Z_{\beta,N}-Z^0_{\beta,N}}{\mathbb{E}Z_{\beta,N}} = \mathcal{W}_N(x) \to 0 \quad \text{a.s.} \tag{3.84}$$

even though $\mathbb{E}\mathcal{W}_N(0) = 1/2$! Thus the more precise statement consists in saying that

$$e^{\frac{1}{2}[\ln(N\ln 2)+\ln 4\pi]}\mathcal{W}_N(0) \xrightarrow{\mathcal{D}} \int_0^\infty e^z\mathcal{P}(dz). \tag{3.85}$$

Note that of course the limiting variable has infinite mean, but is a.s. finite. Finally, by Corollary 3.6,

$$e^{\frac{1}{2}[\ln(N\ln 2)+\ln 4\pi]}\frac{Z^0_{\beta,N}-\mathbb{E}Z^0_{\beta,N}}{\mathbb{E}Z_{\beta,N}} \xrightarrow{\mathcal{D}} \mathcal{V}(0,1) \tag{3.86}$$

The same arguments as those given after Proposition 3.7 allow us to identify $\mathcal{V}(0,1)$ with the centered Poisson integral $\int_{-\infty}^{0} e^z \left(\mathcal{P}(dz) - e^{-z}dz\right)$. This implies (3.38). (3.39) is an immediate corollary. This concludes the proof of Theorem 3.2.$\diamond\diamond$

3.2.3. The Gibbs measure.

With our preparation on the fluctuations of the free energy, we have accumulated enough understanding about the partition function that we can deal with the Gibbs measures. Clearly, there are a number of ways of trying to describe the asymptotics of the Gibbs measures. Recalling the general discussion on random Gibbs measures from Part 2, it should be clear that we are seeking a result on the convergence in distribution of random measures. To be able to state such a results, we have to introduce a topology on the spin configuration state that makes it uniformly compact. The usual topology we used to do this was the product topology, and this clearly would be an option here. However, given what we already know about the partition function, this topology does not appear ideally adapted to give adequate information. Recall that at low temperatures, the partition function was dominated by a 'few' spin configurations with exceptionally large energy. This is a feature that should remain visible in a limit theorem. A nice way to do this consists in mapping the hypercube to the interval $[-1,1]$ via

$$\mathcal{S}_N \ni \sigma \to r_N(\sigma) \equiv \sum_{i=1}^{N} \sigma_i 2^{-i} \in [-1,1] \tag{3.87}$$

Define the pure point measure $\tilde{\mu}_{\beta,N}$ on $[-1,1]$ by

$$\tilde{\mu}_{\beta,N} \equiv \sum_{\sigma\in\mathcal{S}_N} \delta_{r_N(\sigma)} \mu_{\beta,N}(\sigma) \tag{3.88}$$

Our results will be expressed in terms of the convergence of these measures. It will be understood in the sequal that the space of measures on $[-1,1]$ is equipped with the topology of weak convergence, and all convergence results hold with respect to this topology.

As the diligent reader will have expected, in the high temperature phase the limit is the same as for $\beta = 0$, namely

Theorem 3.8: *If $\beta \leq \sqrt{2\ln 2}$, then*

$$\tilde{\mu}_{\beta,N} \to \frac{1}{2}\lambda, \quad a.s. \tag{3.89}$$

where λ denotes the Lebesgue measure on $[-1,1]$.

Proof. Note that we have to prove that for any finite collection of intervals $I_1,\dots,I_k \subset [-1,1]$, the family of random variables $\{\tilde{\mu}_{\beta,N}(I_1),\dots,\tilde{\mu}_{\beta,N}(I_k)$ converges jointly almost surely to $\frac{1}{2}|I_1|,\dots,\frac{1}{2}|I_k|$. But by construction these random vectors are independent, so that this will follow automatically, if we can prove the result in the case $k=1$. Our strategy is to get first very sharp estimates for a family of special intervals.

In the sequel we will always assume that $N \geq n$. We will denote by Π_n the canonical projection from $\mathcal{S}_N$ to $\mathcal{S}_n$. To simplify notation, we will often write $\sigma_n \equiv \Pi_n\sigma$ when no confusion can arise. For $\sigma \in \mathcal{S}_N$, set

$$a_n(\sigma) \equiv r_n(\Pi_n\sigma) \tag{3.90}$$

and

$$I_n(\sigma) \equiv [a_n(\sigma) - 2^{-n}, a_n(\sigma) + 2^{-n}) \tag{3.91}$$

Note that the union of all these intervals forms a disjoint covering of $[-1,1)$. Obviously, these intervals are constructed in such a way that

$$\tilde{\mu}_{\beta,N}(I_n(\sigma)) = \mu_{\beta,N}\left(\{\sigma' \in \mathcal{S}_N : \Pi_n(\sigma') = \Pi_n(\sigma)\}\right) \tag{3.92}$$

The first step in the proof consists in showing that the masses of all the intervals $I_n(\sigma)$ are remarkably well approximated by their uniform mass.

Lemma 3.9: *Set $\beta' \equiv \sqrt{\frac{N}{N-n}}\beta$. For any $\sigma \in \mathcal{S}_n$,*

(i) If $\beta' \leq \sqrt{\frac{\ln 2}{2}}$,

$$|\tilde{\mu}_{\beta,N}(I_n(\sigma)) - 2^{-n}| \leq 2^{-n}e^{-(N-n)(\ln 2 - \beta'^2)}Y_{N-n} \tag{3.93}$$

where Y_N has bounded variance, as $N \uparrow \infty$.

(ii) If $\sqrt{\frac{\ln 2}{2}} < \beta' < \sqrt{2\ln 2}$,

$$|\tilde\mu_{\beta,N}(I_n(\sigma)) - 2^{-n}| \leq 2^{-n} e^{-(N-n)(\sqrt{2\ln 2}-\beta')^2/2 - \alpha \ln(N-n)/2} Y_{N-n} \tag{3.94}$$

where Y_N *is a random variable with bounded mean modulus.*

(iii) If $\beta = \sqrt{2\ln 2}$, *then, for any* n *fixed,*

$$|\tilde\mu_{\beta,N}(I_n(\sigma)) - 2^{-n}| \to 0 \quad \textit{in probability} \tag{3.95}$$

Remark. Note that in the sub-critical case, the results imply convergence to the uniform product mesure on $\mathcal{S}$ in a *very strong sense*. In particular, the base-size of the cylinders considered (i.e. n) can grow proportionally to N, *even if almost sure convergence uniformly for all cylinders is required!* This is unusually good. However, one should not be deceived by this fact: even though seen from the cylinder masses the Gibbs measures look like the uniform measure, seen from the point of view of individual spin configurations the picture is quite different. In fact, the measure concentrates on an *exponentially* small fraction of the full hypercube, namely those $O(\exp(N(\ln 2 - \beta^2/2)))$ vertices that have energy $\sim \beta N$ (Exercise!). It is just the fact that this set is still exponentially large, as long as $\beta < \sqrt{2\ln 2}$, and is very uniformly dispersed over $\mathcal{S}_N$, that produces this somewhat paradoxical effect. The rather weak result in the critical case is not artificial. In fact it is not true that almost sure convergence will hold. This follows e.g. from Theorem 1 in [GMP]. One should of course anticipate some signature of the phase transition at the critical point.

Proof. The proof of this lemma is a simple application of the first three points in Theorem 3.2. Just note that the partial partition functions

$$Z_{\beta,N}(\sigma_n) \equiv \mathbb{E}_{\sigma'} e^{\beta\sqrt{N}X_{\sigma'}} \mathbb{1}_{\Pi_n(\sigma')=\sigma_n} \tag{3.96}$$

are independent and have the same distribution as $2^{-n} Z_{\beta',N-n}$. But

$$\tilde\mu_{\beta,N}(I_n(\sigma_n)) = \frac{Z_{\beta,N}(\sigma_n)}{[Z_{\beta,N} - Z_{\beta,N}(\sigma_n)] + Z_{\beta,N}(\sigma_n)} \tag{3.97}$$

Note that $Z_{\beta,N}(\sigma_n)$ and $[Z_{\beta,N} - Z_{\beta,N}(\sigma_n)]$ are independent. It should now be obvious how to conclude the proof with the help of Theorem 3.2. $\diamondsuit$

Once we have the excellent approximation of the measure on all of the intervals $I_n(\sigma)$, almost sure convergence of the measure in the weak topology is a simple consequence. Of course, this is just a coarse version of the finer results we have, and much more precise information on the quality of approximation can be inferred from Lemma 3.9. But since the high-temperature phase is not our prime concern, we will not go further in this direction.

Somehow much more interesting is the behaviour of the measure at low temperatures that we will discuss now. Let us introduce the Poisson point process $\mathcal{R}$ on the strip $[-1,1]\times\mathbb{R}$ with intensity measure $\frac{1}{2}dy \times e^{-x}dx$. If (Y_k, X_k) denote the atoms of this process, define a new point process $\mathcal{W}_\alpha$ on $[-1,1]\times(0,1]$ whose atoms are (Y_k, w_k), where

$$w_k \equiv \frac{e^{\alpha X_k}}{\int \mathcal{R}(dy,dx)e^{\alpha x}} \tag{3.98}$$

for $\alpha > 1$. With this notation we have that

Theorem 3.10: *If $\beta > \sqrt{2\ln 2}$, with $\alpha = \beta/\sqrt{2\ln 2}$,*

$$\tilde{\mu}_{\beta,N} \xrightarrow{\mathcal{D}} \tilde{\mu}_\beta \equiv \int_{[-1,1]\times(0,1]} \mathcal{W}_\alpha(dy,dw)\delta_y w \tag{3.99}$$

Proof. With $u_N(x)$ defined in (3.48), we define the point process $\mathcal{R}_N$ on $[-1,1]\times\mathbb{R}$ by

$$\mathcal{R}_N \equiv \sum_{\sigma\in\mathcal{S}_N} \delta_{(r_N(\sigma),u_N^{-1}(X_\sigma))} \tag{3.100}$$

A standard result of extreme value theory (see [LLR], Theorem 5.7.2) is easily adapted to yield that

$$\mathcal{R}_N \xrightarrow{\mathcal{D}} \mathcal{R}, \quad \text{as } N\uparrow\infty \tag{3.101}$$

where the convergence is in the sense of weak convergence on the space of sigma-finite measures endowed with the (metrizable) topology of vague convergence. Note that

$$\mu_{\beta,N}(\sigma) = \frac{e^{\alpha u_N^{-1}(X_\sigma)}}{\sum_\sigma e^{\alpha u_N^{-1}(X_\sigma)}} = \frac{e^{\alpha u_N^{-1}(X_\sigma)}}{\int \mathcal{R}_N(dy,dx)e^{\alpha x}} \tag{3.102}$$

Since $\int \mathcal{R}_N(dy,dx)e^{\alpha x} < \infty$ a.s., we can define the point process

$$\mathcal{W}_N \equiv \sum_{\sigma\in\mathcal{S}_N} \delta_{\left(r_N(\sigma), \frac{\exp(\alpha u_N^{-1}(X_\sigma))}{\int \mathcal{R}_N(dy,dx)\exp(\alpha x)}\right)} \tag{3.103}$$

on $[-1,1]\times(0,1]$. Then

$$\tilde{\mu}_{\beta,N} = \int \mathcal{W}_N(dy,dw)\delta_y w \tag{3.104}$$

The only non-trivial point in the convergence proof is to show that the the contribution to the partition functions in the denominator from atoms with $u_N(X_\sigma) < x$ vanishes as $x \downarrow -\infty$. But this is precisely what we have shown to be the case in the proof of part (v) of Theorem 3.2. Standard arguments then imply that first $\mathcal{W}_N \stackrel{\mathcal{D}}{\to} \mathcal{W}$, and consequently, (3.99). $\diamondsuit$

Remark. Note that Theorem 3.10 contains in particular the convergence of the Gibbs measure in the product topology on $\mathcal{S}_N$, since cylinders correspond to certain subintervals of $[-1,1]$. On the other hand, it implies that the point process of weights $\sum_{\sigma\in\mathcal{S}_N}\delta_{\mu_{\beta,N}(\sigma)}$ converges in law to the marginal of $\mathcal{W}_N$ on $(0,1]$ which is the process introduced by Ruelle [Ru4]. The formulation of Theorem 3.10 is moreover very much in the spirit of the metastate approach to random Gibbs measures. The limiting measure is a measure on a continuous space, and each point measure on this set may appear as "pure state". The "metastate", i.e. the law of the random measure $\tilde{\mu}_\beta$ is a probability distribution concentrated on the countable convex combinations of pure states randomly chosen by a Poisson point process from an uncountable collection, while the coefficients of the convex combination are again random and selected via another point process. The only aspect of metastates that is missing here is that we have not "conditioned on the disorder". The point is, however, that there is no natural filtration of the disorder space compatible with, say, the product topology, and thus in this model we have no natural urge to "fix the disorder locally"; note that it is possible to represent the i.i.d. family X_σ as a sum of "local" couplings, i.e. let J_I, for any $I\subset\mathbb{N}$ be i.i.d. standard normal variables. Then we can represent $X_\sigma = 2^{-N/2}\sum_{I\subset\{1,\dots,N\}}\sigma_I J_J$; obviously these variables become independent of any of the J_I, with I fixed, so that conditioning on them would not change the metastate.

Let us discuss the properties of the limiting process $\tilde{\mu}_\beta$. It is not hard to see that with probability one, the support of $\tilde{\mu}_\beta$ is the entire interval $[-1,1]$. On the other hand, its mass is concentrated on a countable set, i.e. the measure is pure point. To see this, consider the rectangle $A_\epsilon \equiv (\ln \epsilon, \infty) \times [-1,1]$. Clearly, the process $\mathcal{R}$ restricted to this set has finite total intensity given by ϵ^{-1}. i.e. the number total number of atoms in that set is a Poissonian random variable with parameter ϵ^{-1}. Now if we remove the projection of these finitely many random points from $[-1,1]$, we will show that the total mass that remains goes to zero with ϵ. Clearly, the remaining mass is given by

$$\int_{[-1,1]\times(-\infty,\ln\epsilon)} \mathcal{R}(dy,dx)\frac{e^{\alpha x}}{\int \mathcal{P}(dx')e^{\alpha x'}} = \int_{-\infty}^{\ln\epsilon} \mathcal{P}(dx)\frac{e^{\alpha x}}{\int \mathcal{P}(dx')e^{\alpha x'}} \tag{3.105}$$

We want to get a lower bound in probability on the denominator. The simplest possible bound is obtained by estimating the probability of the integral by the contribution of the largest atom which of course follows the double-exponential distribution. Thus

$$\mathbb{P}\left[\int \mathcal{P}(dx)e^{\alpha x} \leq Z\right] \leq e^{-e^{-\ln Z/\alpha}} = e^{-Z^{-\frac{1}{\alpha}}} \tag{3.106}$$

Setting $\Omega_Z \equiv \{\mathcal{P} : \int \mathcal{P}(dx)e^{\alpha x} \leq Z\}$, we conclude that, for $\alpha > 1$,

$$\begin{aligned}
\mathbb{P}\left[\int_{-\infty}^{\ln\epsilon} \mathcal{P}(dx)\frac{e^{\alpha x}}{\int \mathcal{P}(dx')e^{\alpha x'}} > \gamma\right] &\leq \mathbb{P}\left[\int_{-\infty}^{\ln\epsilon} \mathcal{P}(dx)\frac{e^{\alpha x}}{\int \mathcal{P}(dx')e^{\alpha x'}} > \gamma,\, \Omega_Z^c\right] + \mathbb{P}[\Omega_Z] \\
&\leq \mathbb{P}\left[\int_{-\infty}^{\ln\epsilon} \mathcal{P}(dx)e^{\alpha x} > \gamma Z,\, \Omega_Z^c\right] + \mathbb{P}[\Omega_Z] \\
&\leq \mathbb{P}\left[\int_{-\infty}^{\ln\epsilon} \mathcal{P}(dx)e^{\alpha x} > \gamma Z\right] + \mathbb{P}[\Omega_Z] \\
&\leq \frac{\mathbb{E}\int_{-\infty}^{\ln\epsilon} \mathcal{P}(dx)e^{\alpha x}}{\gamma} + \mathbb{P}[\Omega_Z] \\
&\leq \frac{\epsilon^{\alpha-1}}{(\alpha-1)\gamma Z} + e^{-Z^{-\frac{1}{\alpha}}}
\end{aligned} \tag{3.107}$$

Obviously, for any positive γ it is possible to choose Z as a function of ϵ in such a way that the right hand side tends to zero. But this implies that with probability one, all of the mass of the measure $\tilde{\mu}_\beta$ is carried by a countable set, implying that $\tilde{\mu}_\beta$ is pure point.

So we see that the phase transition in the REM expresses itself via a change of the properties of the infinite volume Gibbs measure mapped to the interval from Lebesgue measure at high temperatures to a random dense pure pure point measure at low temperatures.

3.2.4. The replica overlap.

While the random measure description of the phase transition in the REM yields, in our view, the most complete and elegant description of the thermodynamic limit of this model, it is common in the physics literature on spin glasses to look at coarser indicators of a phase transition that are reminiscent of the "order parameters" we have discussed before. In a situation where no particular reference configuration exists, a natural possibility is to compare two independent copies of spin configurations drawn from the same Gibbs distribution to each other. To make this precise, recall the function $R_N : \mathcal{S}_N \times \mathcal{S}_N \to [-1,1]$ defined in 3.1. We are interested in the probability distribution of $R_N(\sigma,\sigma')$ under the product measure $\mu_{\beta,N} \otimes \mu_{\beta,N}$, i.e. define a probability measure, $f_{\beta,N}$, on $[-1,1]$ by

$$f_{\beta,N}[\omega](dz) \equiv \mu_{\beta,N}[\omega] \otimes \mu_{\beta,N}[\omega] \left(R_N(\sigma,\sigma') \in dz\right) \tag{3.108}$$

As we will see later, the analysis of the replica overlap will become a crucial tool for studying the Gibbs measures of more complicated models. The following exposition is intended to give a first introduction to this approach.

Theorem 3.11:

(i) For all $\beta < \sqrt{2\ln 2}$

$$\lim_{N\uparrow\infty} f_{\beta,N} = \delta_0, \quad a.s. \tag{3.109}$$

(ii) For all $\beta > \sqrt{2\ln 2}$

$$f_{\beta,N} \xrightarrow{\mathcal{D}} \delta_0 \left(1 - \int \mathcal{W}(dy,dw)w^2\right) + \delta_1 \int \mathcal{W}(dy,dw)w^2 \tag{3.110}$$

Proof. We will write for any $I \subset [-1,1]$

$$f_{\beta,N}(I) = Z_{\beta,N}^{-2} \mathbb{E}_\sigma \mathbb{E}_{\sigma'} \sum_{\substack{t \in I \\ R_N(\sigma,\sigma')=t}} e^{\beta\sqrt{N}(X_\sigma + X_{\sigma'})} \tag{3.111}$$

First of all, the denominator is bounded from below by $[\widetilde{Z}_{\beta,N}(c)]^2$, and by (3.31), with probability of order $\delta^{-2}\exp(-Ng(c,\beta))$, this in turn is larger than $(1-\delta)^2[\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2$. Now let first $\beta < \sqrt{2\ln 2}$. Assume first that $I \subset (0,1) \cup [-1,0)$. We conclude that

$$\begin{aligned} \mathbb{E}f_{\beta,N}(I) &\leq \frac{1}{(1-\delta)^2} \mathbb{E}_\sigma \mathbb{E}_{\sigma'} \sum_{\substack{t \in I \\ R_N(\sigma,\sigma')=t}} 1 + \delta^{-2} e^{-g(c,\beta)N} \\ &= \frac{1}{\sqrt{2\pi N}} \frac{1}{(1-\delta)^2} \sum_{t \in I} \frac{2e^{-N\phi(t)}}{1-t^2} + \delta^{-2} e^{-g(c,\beta)N} \end{aligned} \tag{3.112}$$

for any $\beta < c < \sqrt{2\ln 2}$, where $\phi : [-1,1] \to \mathbb{R}$ denotes the Cramèr entropy function

$$\phi(t) = \frac{(1+t)}{2}\ln(1+t) + \frac{(1-t)}{2}\ln(1-t) \tag{3.113}$$

Here we used of course that, firstly, if $(1-t)N = 2\ell$, $\ell = 0, \dots, N$, then

$$\mathbb{E}_\sigma \mathbb{E}_{\sigma'} \mathbb{1}_{R_N(\sigma,\sigma')=t} = 2^{-N} \binom{N}{\ell} \tag{3.114}$$

and, secondly, Stirling's approximation which implies that

$$\binom{N}{\ell} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{N}{\ell(N-\ell)}} \frac{N^N}{\ell^\ell (N-\ell)^{N-\ell}} (1 + o(1)) \tag{3.115}$$

valid if $\ell \sim xN$ with $x \in (0,1)$. Under our assumptions on I, we see immediately from this representation that the right hand side of (3.112) is clearly exponentially small in N. If $1 \in I$, the additional term coming from $t = 1$ is precisely the term that we have estimated in (3.29), so that again this gives an exponentially small contribution. This shows that the measure $f_{\beta,N}$ concentrates asymptotically on the point 0. This proves (3.109).

Now let $\beta > \sqrt{2\ln 2}$. Here we use the sharper truncations introduced in 3.2.2. Note first that for any interval I

$$\left| f_{\beta,N}(I) - Z_{\beta,N}^{-2}\mathbb{E}_\sigma\mathbb{E}_{\sigma'} \sum_{\substack{t\in I\\ R_N(\sigma,\sigma')=t}} \mathbb{I}_{X_\sigma,X_{\sigma'}\geq u_N(x)} e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})} \right| \leq \frac{2Z_{\beta,N}^x}{Z_{\beta,N}} \tag{3.116}$$

We have already seen in the proof of Theorem 3.2 (see (3.81)) that the right hand side of (3.116) tends zero in probability as first $N\uparrow\infty$ and then $x\downarrow -\infty$. On the other hand, for $t\neq 1$

$$\begin{aligned} &\mathbb{P}\left[\exists_{\sigma,\sigma,:R_N(\sigma,\sigma')=t}\, X_\sigma > u_N(x)\wedge X'_\sigma > u_N(x)\right] \\ &\leq \mathbb{E}_\sigma\,\mathbb{I}_{R_N(\sigma,\sigma')=t}\, 2^{-2N}\mathbb{P}\left[X_\sigma > u_N(x)\right]^2 = \frac{2}{\sqrt{2\pi N}\sqrt{1-t^2}} e^{-\phi(t)N}e^{2x} \end{aligned} \tag{3.117}$$

by the definition of $u_N(x)$ (see (3.47)). This implies again that any interval $I\subset(0,1)\cup[-1,0)$ will have zero mass. To conclude the proof it will be enough to compute $f_{\beta,N}(1)$. Clearly

$$f_{\beta,N}(1) = \frac{2^{-N}\mathbb{Z}_{2\beta,N}}{Z_{\beta,N}^2} \tag{3.118}$$

By Theorem 3.2, (v), one sees easily that

$$f_{\beta,N}(1) \stackrel{\mathcal{D}}{\to} \frac{\int e^{2\alpha z}\mathcal{P}(dz)}{\left(\int e^{\alpha z}\mathcal{P}(dz)\right)^2} \tag{3.119}$$

Expressing the left hand side of (3.119) in terms of the point process $\mathcal{W}_\alpha$ defined in (3.98) yields the expression for the mass of the atom at 1; since the only other atom is at zero the full results follows from the fact that $f_{\beta,N}$ is a probability measure. This concludes the proof. $\diamondsuit$

3.2.5. Multi-overlaps and Ghirlanda–Guerra identities.

The distribution of the replica overlap apparently does not contain all the information on the Gibbs state we have acquired so far. It will be instructive to look at the joint distribution of k independent copies of the spin variables. Interestingly, there are some very strong general principles that allow to relate all multiple overlaps in terms of the two-spin overlaps. These are to some extent rather general in the context of Gaussian processes and it will be instructive to look at them in this simple context. These identities have been known in the physics literature and a more rigorous analysis is given in a paper by Girlanda and Guerra [GG], even though some of the claims in that paper are at best confusing. Equivalent relations were in fact derived somewha earlier by Aizenman and Contucci [AC].The importance of these relations has been underlined by Talagrand [T4]. Let us begin with the simplest instance of these relations.

Proposition 3.12: *For any value of* β,

$$\mathbb{E}\frac{d}{d\beta}F_{\beta,N} = -\beta(1-\mathbb{E}f_{\beta,N}(1)) \tag{3.120}$$

Proof. Obviously,

$$\mathbb{E}\frac{d}{d\beta}F_{\beta,N} = -N^{-1}\mathbb{E}\frac{\mathbb{E}_\sigma\sqrt{N}X_\sigma e^{\beta\sqrt{N}X_\sigma}}{\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}} \tag{3.121}$$

Now if X is standard normal variable, and g any function of at most polynomial growth, then

$$\mathbb{E}[Xg(X)] = \mathbb{E}g'(X) \tag{3.122}$$

Using this identity in the right hand side of (3.121) with respect to the average over X_σ, we get immediately that

$$\begin{aligned}\mathbb{E}\frac{\mathbb{E}_\sigma\sqrt{N}X_\sigma e^{\beta\sqrt{N}X_\sigma}}{\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}} &= N\beta\mathbb{E}\left(1-\frac{2^{-N}\mathbb{E}_\sigma e^{2\beta\sqrt{N}X_\sigma}}{(\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma})^2}\right)\\ &= N\beta\mathbb{E}\left(1-\mu_{\beta,N}^{\otimes 2}\left(\mathbb{1}_{\sigma_1=\sigma_2}\right)\right)\end{aligned} \tag{3.123}$$

which is obviously the claim of the lemma.$\diamondsuit$

In exactly the same way one can prove the following generalisation:

Lemma 3.13: *Let* $h:\mathcal{S}_N^n\to\mathbb{R}$ *be any bounded function of* n *spins. Then*

$$\begin{aligned}&\frac{1}{\sqrt{N}}\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(X_{\sigma^k}h(\sigma^1,\ldots,\sigma^n)\right)\\ &=\beta\mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\ldots,\sigma^n)\left(\sum_{l=1}^n\mathbb{1}_{\sigma^k=\sigma^l}-n\mathbb{1}_{\sigma^k=\sigma^{n+1}}\right)\right)\end{aligned} \tag{3.124}$$

Proof. Left as an exercise.$\diamondsuit$

The strength of Lemma 3.13 comes out when combined with a factorization result that in turn is a consequence of self-averaging.

Lemma 3.14: *Let h be as in the previous lemma. For all but possibly a countable number of values of β,*

$$\lim_{N\uparrow\infty}\frac{1}{\sqrt{N}}\left|\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(X_{\sigma^k}h(\sigma^1,\dots,\sigma^n)\right)-\mathbb{E}\mu_{\beta,N}\left(X_{\sigma^k}\right)\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(h(\sigma^1,\dots,\sigma^n)\right)\right|=0 \tag{3.125}$$

Proof. Let us write

$$\begin{aligned}&\left(\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(X_{\sigma^k}h(\sigma^1,\dots,\sigma^n)\right)-\mathbb{E}\mu_{\beta,N}\left(X_{\sigma^k}\right)\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(h(\sigma^1,\dots,\sigma^n)\right)\right)^2\\&=\left(\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(\left(X_{\sigma^k}-\mathbb{E}\mu_{\beta,N}^{\otimes n}X_{\sigma^k}\right)h(\sigma^1,\dots,\sigma^n)\right)\right)^2\\&\leq\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(X_{\sigma^k}-\mathbb{E}\mu_{\beta,N}^{\otimes n}X_{\sigma^k}\right)^2\mathbb{E}\mu_{\beta,N}^{\otimes n}\left(h(\sigma^1,\dots,\sigma^n)\right)^2\end{aligned} \tag{3.126}$$

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

$$\begin{aligned}&\mathbb{E}\left(\mu_{\beta,N}(X_\sigma^2)-[\mu_{\beta,N}(X_\sigma)]^2\right)+\mathbb{E}\left(\mu_{\beta,N}(X_\sigma)-\mathbb{E}\mu_{\beta,N}(X_\sigma)\right)^2\\&=-\beta^{-2}\mathbb{E}\frac{d^2}{d\beta^2}F_{\beta,N}+N\beta^{-2}\mathbb{E}\left(\frac{d}{d\beta}F_{\beta,N}-\mathbb{E}\frac{d}{d\beta}F_{\beta,N}\right)^2\end{aligned} \tag{3.127}$$

We know that $F_{\beta,N}$ converges as $N\uparrow\infty$ and that the limit is infinitely differentiable for all $\beta\geq 0$, except at $\beta=\sqrt{2\ln 2}$; moreover, $-F_{\beta,N}$ is convex in β. Then standard results of convex analysis imply that

$$\limsup_{N\uparrow\infty}\left(-\mathbb{E}\frac{d^2}{d\beta^2}F_{\beta,N}\right)=-\frac{d^2}{d\beta^2}\lim_{N\uparrow\infty}\mathbb{E}F_{\beta,N} \tag{3.128}$$

which is finite for all $\beta\neq\sqrt{2\ln 2}$. Thus, the first term in (3.127) will vanish when divided by N. To see that the coefficient of N of the second term gives a vanishing contribution, we use the general fact that if the variance of family of a convex (or concave) functions tends to zero, then the same is true for its derivative, except possibly on a countable set of values of their argument. In Theorem 3.2 we have more than established that the variance of $F_{\beta,N}$ tends to zero, and hence the result of the Lemma is proven. $\diamondsuit$

If we combine Proposition 3.12, Lemma 3.13, and Lemma 3.14 we arrive immediately at

Proposition 3.15:*For all but a countable set of values β, for any bounded function $h : \mathcal{S}_N^n \to \mathbb{R}$,*

$$\begin{aligned}\lim_{N\uparrow\infty}\Bigg| & \mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\dots,\sigma^n)\mathbb{I}_{\sigma^k=\sigma^{n+1}}\right) \\ & -\frac{1}{n}\mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\dots,\sigma^n)\left(\sum_{l\neq k}^n \mathbb{I}_{\sigma^l=\sigma^k}+\mathbb{E}\mu_{\beta,N}^{\otimes 2}(\mathbb{I}_{\sigma_1=\sigma_2})\right)\right)\Bigg| = 0\end{aligned} \tag{3.129}$$

Together with the fact that the product Gibbs measures are concentrated only on the sets where the overlaps take values 0 and 1, (3.129) permits to compute the distribution of all higher overlaps in terms of the two-replica overlap. E.g., if we put

$$A_n \equiv \lim_{N\uparrow\infty}\mathbb{E}\mu_{\beta,N}^{\otimes n}(\mathbb{I}_{\sigma^1=\sigma^2=\dots=\sigma^n}) \tag{3.130}$$

then (3.129) with $h = \mathbb{I}_{\sigma^1=\sigma^2=\dots=\sigma^n}$ provides the recursion

$$\begin{aligned} A_{n+1} &= \frac{n-1}{n}A_n + \frac{1}{n}A_nA_2 = A_n\left(1-\frac{1-A_2}{n}\right) \\ &= \prod_{k=2}^n\left(1-\frac{1-A_2}{k}\right)A_2 \\ &= \frac{\Gamma(n+A_2)}{\Gamma(n+1)\Gamma(A_2)}\end{aligned} \tag{3.131}$$

Note that we can use alternatively Theorem 3.8 to compute, for the non-trivial case $\beta > \sqrt{2\ln 2}$,

$$\lim_{N\uparrow\infty}\mu_{\beta,N}^{\otimes 2}(\mathbb{I}_{\sigma^1=\sigma^2=\dots=\sigma^n}) = \int \mathcal{W}(dy,dw)w^n \tag{3.132}$$

so that (3.131) implies a formula for the mean of the n-th moments of $\mathcal{W}$,

$$\mathbb{E} \int \mathcal{W}(dy, dw) w^n = \frac{\Gamma(n + A_2)}{\Gamma(n+1)\Gamma(A_2)} \tag{3.133}$$

where $A_2 = \mathbb{E} \int \mathcal{W}(dy, dw) w^2$. This result has been obtained by a direct computation by Ruelle ([Ru], Corollary 2.2), but its derivation via the Ghirlanda–Guerra identities shows a way to approach this problem in a different manner that has the potential to give results in more complicated situations.[28]

3.2.6. Final remarks.

We have seen that the random energy model, although rather trivial and artificial looking when considered as a model of interacting spins, exhibits a rather rich and interesting structure which shows by explicit computation characteristic features of a strongly disordered model. In particular, in contrast to the models we have discussed before, we have seen the appearance of a truely random limiting Gibbs measure, showing that the concept of the metastate introduced in Section 2 is actually meaningful, and unavoidable if the full asymptotic properties of the Gibbs states are to be described adequately. In the remainder of these notes we will have this picture in mind when studying more complicated models.

3.3. Correlated Gaussian processes: p-spin models.

Having seen how much can be found out about the i.i.d. case, we will now return to the more general setting of correlated Gaussian processes introduced in Section 3.1. Physicists have studied such model as early as the REM [D1], and soon a full solution of the p-spin model with $p \geq 3$ based on the heuristic "replica-method" was given in [GM]. In the context of this method it was found that some considerable simplifications take place once $p > 2$, in particular if the temperature is not too small. In 1998, M. Talagrand discovered that these simplifications also allow to obtain substantial rigorous results on the low-temperature phase of the model which

[28] More generally, one may dervive recursion formulas for more general moments of Ruelle's process that show that the identies (3.129) determine completely the process of Ruelle in terms of the two-overlap distribution function.

are as yet unavailable in the case $p = 2$. While I will not review all of Talagrand's results here (many are not even published at the time these notes are written), I will show that using some rather simple estimates it is possible to gain substantial insight in the structure of the Gibbs states of this model. Further material can be found in Talagrand's lecture notes [T4].

Why should there be such a remarkable difference between the case $p = 2$ and the cases $p \geq 3$? We will see that this can be understood as the result of a competition between correlation and entropy. If we consider the number of configurations σ' at Hamming distance r from a given configuration, then clearly this number increases like $\exp(N(\ln 2 - \phi(1-r)))$ (where ϕ is the Cramèr entropy defined in (3.113)) which, for $r \sim 1$ behaves like $\exp(N(\ln 2 - (1-r)^2)$. On the other hand, the correlation between two such spin-configurations behaves like $(1-r)^p$. It will turn out that this will imply that for $p = 2$ correlations over the entire hypercube are important, whereas for larger p, spin-configurations that are at distances close to 1 from each other can be regarded as essentially uncorrelated. This will mean that in the latter case, on the large scale the system behaves much more like the REM, even though this is far from trivial to prove.

In the REM we have seen that essentially all the results can be derived rather easily from the basic Theorem 3.3 which is a classical result from extreme value theory. Naively, one would expect that there should be comparable results covering the correlated case. As it turns out, this is not the case. Even though we are in the relatively convenient and widely investigated setting of Gaussian processes, classical results and methods fall short from giving anything like Theorem 3.3. The main obstacle is that one is unable to determine the precise value of $\lim_{N\uparrow\infty} N^{-1} \sup_{\sigma\in\mathcal{S}_N} H_N(\sigma)$, worse, one is not even able to prove the existence of this limit! Before entering into the "statistical mechanics" approach, it will be rather instructive to learn to appreciate this point by trying out what standard methods could reveal about this problem.

3.3.1. Classical estimates on extremes.

We consider the Hamiltonians defined in (3.13). To work with normalized Gaussian processes and to emphasise the analogy with the REM, we introduce

$$\mathcal{H}_\sigma \equiv -\frac{1}{\sqrt{N}} H_N^{p-SK}(\sigma) \tag{3.134}$$

Classical tools for estimating maxima of Gaussian processes are comparisons with simpler processes for which such estimates are more readily available. The most basic of these comparison methods rely on *Slepian's lemma*:

Lemma 3.16: [Sl]*Let X_σ, Y_σ are standardized Gaussian processes on some set $\mathcal{S}_N$. Assume that for any $\sigma, \tau \in \mathcal{S}_N$,*

$$\mathbb{E}X_\sigma X_\tau \leq \mathbb{E}Y_\sigma Y_\tau \tag{3.135}$$

Then, for any $x \in \mathbb{R}$,

$$\mathbb{P}\left[\sup_{\sigma\in\mathcal{S}_N} X_\sigma > x\right] \geq \mathbb{P}\left[\sup_{\sigma\in\mathcal{S}_N} Y_\sigma > x\right] \tag{3.136}$$

In particular,

$$\mathbb{E}\sup_{\sigma\in\mathcal{S}_N} X_\sigma \geq \mathbb{E}\sup_{\sigma\in\mathcal{S}_N} Y_\sigma \tag{3.137}$$

An immediate and otherwise trivial consequence of Slepian's Lemma is the fact that the ground state energies of the p-spin models are monotone in p and dominated by that of the REM. Naturally one would like to get a lower bound.

Proposition 3.17: *With probability one, for all but finitely many N,*

$$\lim_{N\uparrow\infty} \sup_\sigma \frac{|H_N^{p-SK}(\sigma)|}{N} \geq \sqrt{2\ln 2}(1-c_p) \tag{3.138}$$

where for large p, $c_p \sim p2^{-p}$.

Proof. We will now give a proof based on a standard Gaussian comparison lemma, originally due to Slepian [Sl] that can be found e.g. in [LLR] (Theorem 4.2.1.).

The idea is to see to what extent the independent r.v. X_σ of the REM are reasonable approximations of the dependent $\mathcal{H}_\sigma$ of the SK models, in the sense that their extremes are comparable. To this end one would like to compare the probabilities $\mathbb{P}\left[\forall\sigma, \mathcal{H}_\sigma \leq u_N\right]$ and $\mathbb{P}\left[\forall\sigma, X_\sigma \leq u_N\right] = \Phi(u_N)^{2^N}$. The Normal Comparison Lemma states the following:

Lemma 3.18: [LLR] *Let Z_i be a family of standard normal variables with covariance matrix R. Then for any real u,*

$$\left|\mathbb{P}\left[Z_1 \leq u, \ldots, Z_n \leq u\right] - \Phi(u)^n\right| \leq \frac{1}{2\pi} \sum_{1\leq i<j\leq n} \frac{|R_{ij}|}{1-R_{ij}^2} \exp\left(-\frac{u^2}{1+|R_{ij}|}\right) \tag{3.139}$$

Applied to our problem, a first attempt would yield the bound (due to the symmetry, we identify σ and $-\sigma$)

$$\begin{aligned}
&\left|\mathbb{P}\left[\forall\sigma, \mathcal{H}_\sigma \leq u\right] - \Phi(u)^{2^N}\right| \leq \frac{1}{2\pi} \sum_{\sigma\neq\sigma'} \frac{|R(\sigma,\sigma')^p|}{1-R(\sigma,\sigma')^{2p}} \exp\left(-\frac{u^2}{1+|R(\sigma,\sigma')|^p}\right)\\
&= \frac{1}{2\pi} \sum_{m\in\mathcal{M}_N\setminus\{-1,1\}} \sum_{\sigma\neq\sigma'} \mathbb{1}_{\{R(\sigma,\sigma')=m\}} \frac{|m|^p}{1-m^{2p}} e^{-\frac{u^2}{1+|m|^p}}\\
&= \frac{1}{2\pi} \sum_{m\in\mathcal{M}_N\setminus\{-1,1\}} \frac{|m|^p}{1-m^{2p}} e^{-\frac{u^2}{1+|m|^p}} 2^{2N} \frac{2e^{-N\phi(m)}}{\sqrt{2\pi N(1-m^2)}}
\end{aligned} \tag{3.140}$$

where

$$\mathcal{M}_N \equiv \{-1, -1+2/N, \ldots, 1-2/N, 1\} \tag{3.141}$$

is the set of possible values the "overlap" $R(\sigma,\sigma')$ can take on. Now we would need to have the right-hand side tend to zero for values of u at which the maximum of 2^N standard Gaussians is taken, i.e. where $\Phi(u)^{2^N}$ is between zero and one. Now it is a well-known fact that this means that u has to be very close to $u_N \equiv \sqrt{N 2\ln 2}$. With this value inserted, (3.141) yields indeed

$$\begin{aligned}
&\left|\mathbb{P}\left[\forall\sigma, \mathcal{H}_\sigma \leq u_N\right] - \Phi(u_N)^{2^N}\right|\\
&\leq \frac{1}{2\pi} \sum_{m\in\mathcal{M}_N\setminus\{-1,1\}} \frac{2}{\sqrt{2\pi N(1-m^2)}} \frac{|m|^p}{1-m^{2p}} e^{-N\left(\phi(m)-\frac{m^p 2\ln 2}{1+|m|^p}\right)}
\end{aligned} \tag{3.142}$$

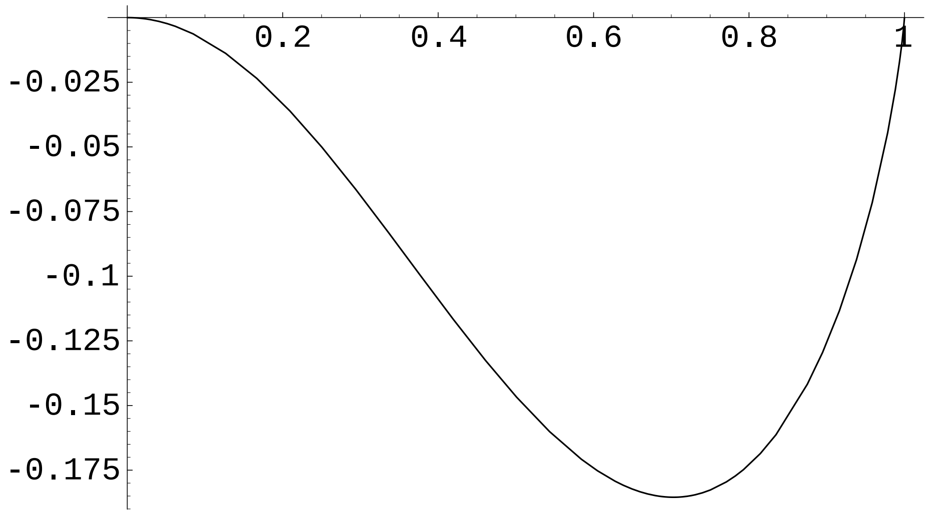
One now sees immediately the crucial difference between the cases $p = 2$ and $p > 2$. In the former, due to the fact that $\phi(m) \sim m^2/2$ for small m, spin-configurations with overlap close to zero give a large contribution to this sum. On the other hand, if $p > 3$, the contribution from a neighbourhood of $m = 0$ is of the order $N^{-(p-1)/2}$ only. One might at first hope that the larger values of m also give no large contribution, but closer inspection shows that this is not the case. Indeed, for all p, there is a (shrinking with N) region near $|m| = 1$ which gives a non-vanishing contribution to the sum, and thus it will not be true that the left hand side of (3.142) will be small for any p. This is in fact very reasonable, since it is clear that there should be some effect coming from the strong correlation of spin-configurations that are very similar.

In the sequel we need the following properties of the functions $\phi(m) - \frac{m^p 2\ln 2}{1+|m|^p}$

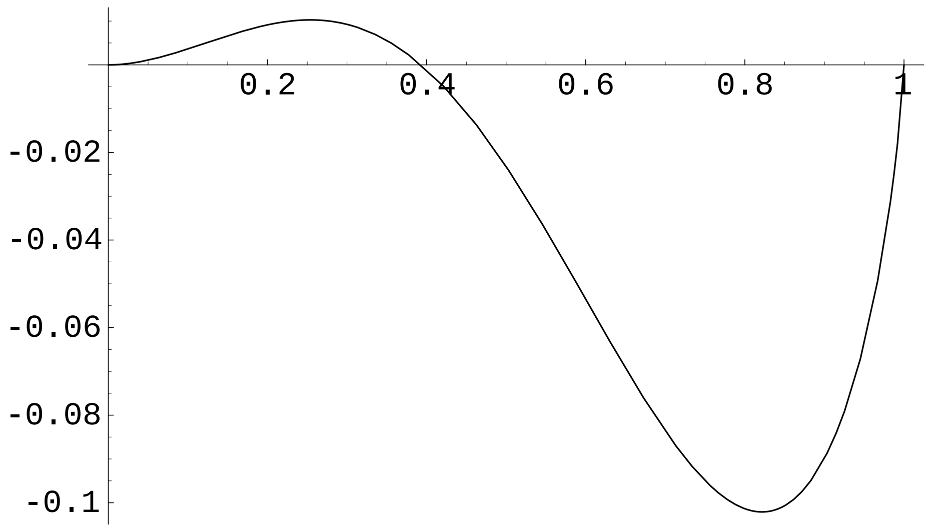
Lemma 3.19: *For any $p \geq 3$, there exist $m_p > 0$, such that for all $m < m_p$,*

$$\psi_p(m) \equiv \phi(m) - \frac{m^p 2\ln 2}{1+|m|^p} > 0 \tag{3.143}$$

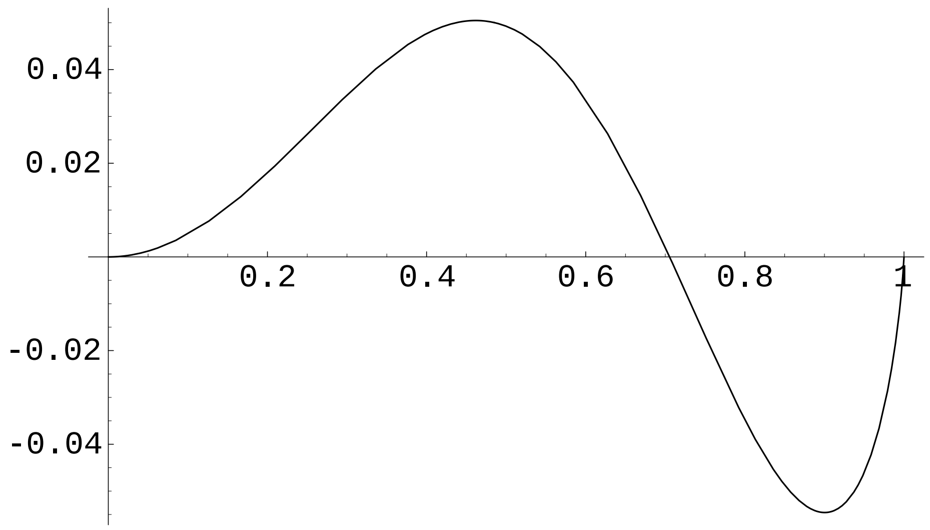
Moreover, m_p can be chosen in such a way that $m_p \uparrow 1$ as $p \uparrow \infty$. For p large, $m_p = 1 - 2^{-p}$.



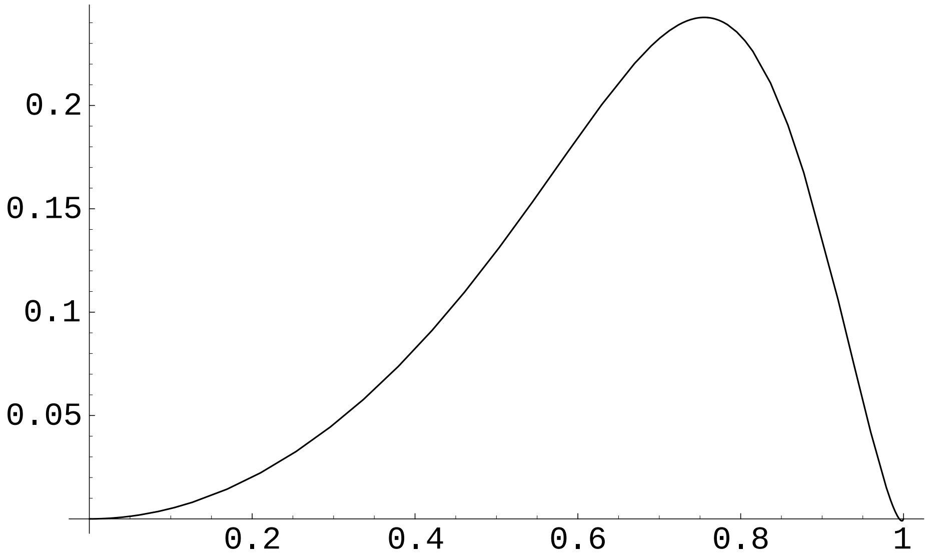
$\psi_2(x)$



$\psi_3(x)$



$\psi_4(x)$



$\psi_{10}(x)$

Based on this lemma, our strategy is now simple: Find a set $\mathcal{K}_p \subset \{-1,1\}^N$ of cardinality K_p as large as possible such that for any $\sigma, \sigma' \in \mathcal{K}_p$, $|R(\sigma,\sigma')| \leq m_p$. Suppose we can find such a set with $K_p = 2^{N(1-\gamma_p)}$. Then on the one hand

$$\sup_{\sigma\in\{-1,1\}^N} \mathcal{H}_\sigma \geq \sup_{\sigma\in\mathcal{K}_p} \mathcal{H}_\sigma \tag{3.144}$$

while if the $\mathcal{H}_\sigma$ with $\sigma \in \mathcal{K}_p$ behaved like independent Gaussians, their maximum would be around $\sqrt{N2(1-\gamma_p)\ln 2}$. The Normal Comparison Lemma applied to these variables on this level then yields

$$\begin{aligned}
&\left|\mathbb{P}\left[\forall\sigma\in\mathcal{K}_p, \mathcal{H}_\sigma \leq \sqrt{N2(1-\gamma_p)\ln 2}\right] - \Phi\left(\sqrt{N2(1-\gamma_p)\ln 2}\right)^{2^{N(1-\gamma_p)}}\right| \\
&\leq \frac{1}{2\pi} \sum_{\sigma\neq\sigma'\in\mathcal{K}_p} \frac{|R(\sigma,\sigma')^p|}{1-R(\sigma,\sigma')^{2p}} \exp\left(-\frac{N(1-\gamma_p)2\ln 2}{1+|R(\sigma,\sigma')|^p}\right) \\
&= \sum_{m=-m_p}^{m_p} \frac{1}{2\pi} \sum_{\sigma\neq\sigma'\in\mathcal{K}_p} \mathbb{1}_{\{R(\sigma,\sigma')=m\}} \frac{|m|^p}{1-m^{2p}} e^{-\frac{N(1-\gamma_p)2\ln 2}{1+|m|^p}}
\end{aligned} \tag{3.145}$$

Assuming a homogeneous distribution of the set $\mathcal{K}_p$, one can reasonably assume that

$$\sum_{\sigma\neq\sigma'\in\mathcal{K}_p} \mathbb{1}_{\{R(\sigma,\sigma')=m\}} \approx 2^{2N(1-\gamma_p)} \frac{2}{2\pi N(1-m^2)} e^{-N\phi(m)} \tag{3.146}$$

in which case the right-hand side of (3.145) would simplify to

$$\begin{aligned}&\frac{1}{2\pi}\sum_{m=-m_p}^{m_p}\frac{m^p}{1-m^{2p}}e^{-N\left(\phi(m)-\frac{(1-\gamma_p)2\ln 2|m|^p}{1+|m|^p}\right)}\\&\leq C_p N^{-p/2}\end{aligned}\tag{3.147}$$

and this would prove that the maximum of the $\mathcal{H}_\sigma$ is indeed of the order of $\sqrt{N2(1-\gamma_p)\ln 2}$.

It remains to establish that we can find sets $\mathcal{K}_p$ with the desired properties. The following construction may look quite ridiculous and is possibly too complicated, but due to the author's lack of insight into the geometry of the hypercube in high dimensions, it is the only one he can provide.

Consider K i.i.d. random elements $\sigma_1,\sigma_2,\dots,\sigma_K$ in $\{-1,1\}^N$, distributed according to the uniform distribution on $\{-1,1\}^N$. We want to estimate the size of the largest subset $\widetilde{\mathcal{K}}\subset\{1,\dots,K\}$ with the property that for all $\sigma,\sigma'\in\widetilde{\mathcal{K}}$, $R(\sigma,\sigma')<m_p$.

Lemma 3.20: *Let $\sigma_1,\sigma_2,\dots,\sigma_K$ be random variables as described above. Then*

$$\mathbb{E}|\widetilde{\mathcal{K}}|\geq K(1-K\exp(-N\phi(m_p)))\tag{3.148}$$

Moreover,

$$\mathbb{P}\left[|\widetilde{\mathcal{K}}|\geq K\big(1-2K\exp(-N\phi(m_p))\big)\right]\geq\frac{1}{4}\tag{3.149}$$

Proof. Note that

$$\begin{aligned}K-|\widetilde{\mathcal{K}}|&\leq\left|\left\{i\in 1,\dots,K\big|\exists j\neq i\,\text{s.t.}|R(\sigma_i,\sigma_j)|\geq m_p\right\}\right|\\&\leq\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{I}_{\{|R(\sigma_i,\sigma_j)|\geq m_p\}}\end{aligned}\tag{3.150}$$

Now

$$\mathbb{E}_\sigma\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{I}_{\{|R(\sigma_i,\sigma_j)|\geq m_p\}}=K^2\mathbb{P}_\sigma\left[|R(\sigma_i,\sigma_j)|\geq m_p\right]\leq K^2e^{-N\phi(m_p)}\tag{3.151}$$

and from this (3.148) is follows. Similarly, one verifies easily that

$$\mathbb{E}_\sigma \left(\sum_{i=1}^{K} \sum_{j\neq 1} \mathbb{1}_{\{|R(\sigma_i,\sigma_j)|\geq m_p\}} \right)^2 \leq K^4 e^{-2N\phi(m_p)} + 6K^3 e^{-2N\phi(m_p)} + 2K^2 e^{-N\phi(m_p)} \tag{3.152}$$

From this one deduces (3.149) via the *Paley-Zygmund inequality*, which states that for any positive random variable X and $0 \leq q \leq 1$,

$$\mathbb{P}[X \geq q\mathbb{E}X] \geq (1-q)^2 \frac{\mathbb{E}X^2}{(\mathbb{E}X)^2} \tag{3.153}$$

$\diamondsuit$

Lemma 3.19 tells us that if we choose K such that $K \exp(-N\phi(m_p))$ is small, then almost all σ drawn will be isolated from the others as desired. Since $\phi(m)$ converges to $\ln 2$ as m tends to one, it is clear that we can find γ_p tending to zero as p tends to infinity such that this holds with $K = 2^{N(1-\gamma_p)}$. An asymptotic analysis of the function $\psi(x)$ for large p yields the estimates claimed in the theorem. $\diamondsuit\diamondsuit$

Having seen how we can prove a lower bound on the ground state energy and thus the free energy of our model, it is tempting to also get an upper bound that improves on the trivial REM bound.

Proposition 3.21: *In the p-spin SK model, there exists $c_p > 0$ such that almost surely for all but finitely many values of N,*

$$\sup_{\sigma\in\mathcal{S}_N} \frac{H_N^{p-SK}(\sigma)}{N} \leq \sqrt{2\ln 2}(1-c_p) \tag{3.154}$$

Moreover, for p large,

$$c_p \geq 2^{-\alpha_p^2 p} \frac{p}{(2-\alpha)^2} \tag{3.155}$$

with $\alpha_p = 1 + \frac{1}{p\ln 2} + \sqrt{1 + \frac{1}{(p\ln 2)^2}}$ tending to 2 as $p \uparrow \infty$.

We will set in this section $\mathbb{E}\mathcal{H}_\sigma\mathcal{H}_{\sigma'} = f(R(\sigma,\sigma'))$ where in the application to the p-spin SK model, f will be the p-th power. We will assume that f is an increasing function.

First we establish the existence of a ρ-covering of the set $\mathcal{S}_N = \{-1,1\}^N$.

Lemma 3.22: *Let $1 > \rho > 0$. Then there exists a set $K_N(\rho) \subset \mathcal{S}_N$ with the following properties:*

(i) For any $\sigma \in \mathcal{S}_N$, $\operatorname{dist}(\sigma, K_N(\rho)) \leq \rho$.

(ii) $N^{-1} \ln |K_N(\rho)| \leq \phi(1-\rho) + o(1)$.

Proof. Again we will establish the existence of such sets by showing that a random set of the cardinality given in (ii) will have property (i) with probability close to 1. Thus let $K = \{k_1, \dots, k_{|K|}\}$ be a set of i.i.d. random variables taking values in $\mathcal{S}_N$ drawn from the uniform distribution on $\mathcal{S}_N$. Let $\mathbb{P}$ denote their law. Then

$$\begin{aligned}
\mathbb{P}\left[\exists_{\sigma\in\mathcal{S}_N} \operatorname{dist}(\sigma, K) > \rho\right] &\leq 2^N \mathbb{P}\left[\operatorname{dist}(\sigma, K) > \rho\right] \\
&\leq 2^N \mathbb{P}\left[\forall_{k\in K} \operatorname{dist}(\sigma, k) > \rho\right] \\
&\leq 2^N \left(\mathbb{P}\left[\operatorname{dist}(\sigma, k) > \rho\right]\right)^{|K|} \\
&\leq 2^N \left(1 - \mathbb{P}\left[\operatorname{dist}(\sigma, k) \leq \rho\right]\right)^{|K|} \\
&\leq 2^N e^{-|K|\mathbb{P}[\operatorname{dist}(\sigma,k)\leq\rho]}
\end{aligned} \tag{3.156}$$

This probability is exponentially small, as soon as for some $\epsilon > 0$,

$$|K| \geq \frac{\ln 2N(1+\epsilon)}{\mathbb{P}\left[\operatorname{dist}(\sigma, k) \leq \rho\right]} \tag{3.157}$$

Since $P\left[\operatorname{dist}(\sigma, k) \leq \rho\right] \sim e^{-N\phi(1-\rho)}$, the claimed result follows.$\diamondsuit$

Let now a set $K_N(\rho)$ be given. We introduce the map $k : \mathcal{S}_N \to K_N(\rho)$ that maps any configuration σ to the point $k(\sigma) \in K_N(\rho)$ that is closest to σ. We will use the following obvious relation, valid for all $1 > \rho > 0$ and all x:

$$\mathbb{P}\left[\sup_{\sigma\in\mathcal{S}_N} \mathcal{H}_\sigma > x\right] \leq \sum_{k\in K_N(\rho)} \mathbb{P}\left[\sup_{\sigma\in\mathcal{S}_N, k(\sigma)=k} \mathcal{H}_\sigma > x\right] \tag{3.158}$$

Now

$$\mathbb{P}\left[\sup_{\sigma\in\mathcal{S}_N, k(\sigma)=k} \mathcal{H}_\sigma > x\right] = \sum_{\ell=-\infty}^{\infty} \mathbb{P}\left[\mathcal{H}_k \in [\ell, \ell+1), \sup_{\sigma\in\mathcal{S}_N, k(\sigma)=k} \mathcal{H}_\sigma > x\right] \tag{3.159}$$

Moreover, it is clearly true that

$$\begin{aligned}&\mathbb{P}\left[\mathcal{H}_k \in [\ell, \ell+1), \sup_{\sigma \in \mathcal{S}_N, k(\sigma)=k} \mathcal{H}_\sigma > x\right] \\ &\leq \min\left(\mathbb{P}\left[\mathcal{H}_k \in [\ell, \ell+1)\right], \sum_{\sigma : k(\sigma)=k} \mathbb{P}\left[\mathcal{H}_k \in [\ell, \ell+1), \mathcal{H}_\sigma > x\right]\right)\end{aligned} \tag{3.160}$$

The idea is now that for ℓ "large" the maximum of the process $\mathcal{H}_\sigma$ can be bounded by the first term, while for ℓ too small, the joint probability is reduced due to the strong correlation between $\mathcal{H}_\sigma$ and $\mathcal{H}_k$. Indeed we have for $\ell \geq 0$

$$\mathbb{P}\left[\mathcal{H}_k \in [\ell, \ell+1)\right] \leq 2e^{-\ell^2/2} \tag{3.161}$$

while

$$\begin{aligned}&\mathbb{P}\left[\mathcal{H}_k \in [\ell, \ell+1), \mathcal{H}_\sigma > x\right] \\ &\quad = \frac{1}{2\pi} \frac{1}{\sqrt{1-(f(R(\sigma,k)))^2}} \int_\ell^{\ell+1} dy \int_x^\infty dz e^{-\frac{y^2+z^2-2f(R(\sigma,k))zy}{2(1-(f(R(\sigma,k)))^2)}}\end{aligned} \tag{3.162}$$

Now

$$\begin{aligned}\frac{y^2+z^2-2f(R(\sigma,k))zy}{1-(f(R(\sigma,k)))^2} &= z^2 + \frac{z^2(f(R(\sigma,k)))^2+y^2-2f(R(\sigma,k))xy}{(1-(f(R(\sigma,k)))^2)} \\ &z^2 + \frac{(zf(R(\sigma,k))-y)^2}{(1-(f(R(\sigma,k)))^2)}\end{aligned} \tag{3.163}$$

Now note that by construction, if $k(\sigma) = k$, then $f(R(\sigma,k)) \geq f(1-\rho)$. It is clear that the second argument in the minimum in (3.160) has a chance to realize the minimum only if x is substantially larger than ℓ. In fact we will bound this minimum by its second argument only if $f(1-\rho)x - (\ell+1) \equiv \delta_\ell \geq 0$. In that case, on the domain of integration in (3.162), the exponent in the integrand is bounded by

$$\frac{1}{2}\left(x^2 + \frac{\delta_\ell^2}{1-(f(1-\rho))^2}\right)$$

(in fact the integral is bounded by replacing the $f(R(\sigma,k))$ by $f(1-\rho)$ everywhere) and so

$$\mathbb{P}\left[\mathcal{H}_k \in [\ell, \ell+1), \mathcal{H}_\sigma > x\right] \leq Ce^{-\frac{1}{2}\left(x^2+\frac{\delta_\ell^2}{1-(f(1-\rho))^2}\right)} \tag{3.164}$$

Thus for any ℓ_0 s.t. $\ell_0 < xf(1-\rho)$, we can obtain the bound

$$\sum_{K\in K_N(P)}\sum_{\ell=-\infty}^{\infty}\mathbb{P}\left[\mathcal{H}_k\in[\ell,\ell+1),\ \sup_{\sigma\in\mathcal{S}_N,k(\sigma)=k}\mathcal{H}_\sigma > x\right]$$

$$\leq \sum_{\ell\leq\ell_0} 2^N C e^{-\frac{1}{2}\left(x^2+\frac{(xf(1-\rho)-\ell+1)^2}{1-(f(R(1-\rho)))^2}\right)} + \sum_{\ell>\ell_0}|K_N(\rho)|2e^{-\ell^2/2} \tag{3.165}$$

$$= \sum_{\ell\leq\ell_0} 2^N C e^{-\frac{1}{2}\left(x^2+\frac{(xf(1-\rho)-\ell_0+\ell_0-\ell)^2}{1-(f(R(1-\rho)))^2}\right)} + \sum_{\ell>\ell_0}|K_N(\rho)|2e^{-\frac{(\ell_0+\ell-\ell_0)^2}{2}}$$

Clearly, the last expression is bounded by

$$NC\left(2^N e^{-\frac{1}{2}\left(x^2+\frac{(xf(1-\rho)-\ell_0)^2}{1-(f(R(1-\rho)))^2}\right)} + |K_N(\rho)|2e^{-\frac{\ell_0^2}{2}}\right) \tag{3.166}$$

Now set $\ell_0 = xf(1-\rho)\delta$, $0\leq\delta\leq 1$. Then (3.166) reads

$$NC\left(e^{N\ln 2-\frac{x^2}{2}\left(1+\frac{(1-\delta)^2 f^2(1-\rho)}{1-(f(R(1-\rho)))^2}\right)} + e^{N\phi(1-\rho)-\frac{x^2}{2}\delta^2 f^2(1-\rho)}\right) \tag{3.167}$$

The optimal choice for δ is obtained by equating the exponents in both summands. This yields (set $f^2\equiv f^2(1-\rho)$)

$$0 = N(\ln 2-\phi(1-\rho)) - \frac{x^2}{2}\left(1+\frac{(1-\delta)^2 f^2}{1-f^2}-\delta^2 f^2\right) \tag{3.168}$$

or

$$\frac{2N(\ln 2-\phi(1-\rho))(1-f^2)}{x^2} = (1-f^2\delta)^2 \tag{3.169}$$

leading to

$$\delta = \frac{1}{f^2}\left(1-\sqrt{\frac{2N(\ln 2-\phi(1-\rho))(1-f^2)}{x^2}}\right) \tag{3.170}$$

Inserting this into (3.167) we see that our probability is bounded by

$$\exp\left(N\phi(1-\rho)-\frac{1}{2f^2}\left(x-\sqrt{2N(\ln 2-\phi(1-\rho))(1-f^2)}\right)^2\right) \tag{3.171}$$

For this to be exponentially small, we must have that $x > x_c$ where

$$\begin{aligned} x_c =& f\sqrt{2N\phi(1-\rho)} + \sqrt{2N(\ln 2-\phi(1-\rho))(1-f^2)} \\ =& \sqrt{2N\ln 2}\left(f\sqrt{\phi(1-\rho)/\ln 2} + \sqrt{(1-\phi(1-\rho)/\ln 2)(1-f^2)}\right) \end{aligned} \tag{3.172}$$

The final result is now obtained by minimizing this expression with respect to ρ. While the exact computation of the minimizer is rather cumbersome, in case $f(y) = y^p$, using the leading asymptotics in ρ, i.e.

$$1 - f^2 = 2p\rho + O(\rho^2)$$

and

$$\phi(1-\rho) = \ln 2 + \frac{\rho}{2}\ln\rho - \rho\frac{\ln 2 + 1}{2} + O(\rho^2)$$

we get up to terms of order $\rho^2|\ln\rho|^2$,

$$\begin{aligned} x_c =&\sqrt{2N\ln 2}\left(\sqrt{2p\rho^2\frac{|\ln\rho| + \ln 2 + 1}{2\ln 2}} + \sqrt{1 - \frac{\rho(|\ln\rho| + \ln 2 + 1)}{2\ln 2}}(1-p\rho)\right) \\ =& \sqrt{2N\ln 2}\left(1 - \rho\left(\frac{|\ln\rho| + \ln 2 + 1}{4\ln 2} + p - \sqrt{p(\frac{|\ln\rho| + \ln 2 + 1}{\ln 2})}\right)\right) \end{aligned} \tag{3.173}$$

Setting $\rho = 2^{-\alpha^2 p}$, we get

$$x_c = \sqrt{2N\ln 2}\left(1 - 2^{-\alpha^2 p}\left(\frac{\alpha^2 p}{4} + \frac{\ln 2 + 1}{4\ln 2} + p - \sqrt{p\left(p\alpha^2 + \frac{\ln 2 + 1}{\ln 2}\right)}\right)\right) \tag{3.174}$$

For large p we can neglect the term of order ρ in the expansion of $\phi(1-\rho)$, which leads to the approximate formula

$$\begin{aligned} x_c &\approx \sqrt{2N\ln 2}\left(1 - 2^{-\alpha^2 p}p\left(\frac{\alpha^2}{4} + 1 - \alpha\right)\right) \\ &x\sqrt{2N\ln 2}\left(1 - 2^{-\alpha^2 p}p\,(\alpha-2)^2\right) \end{aligned} \tag{3.175}$$

Now the coefficient of ρ in this expression should be positive if we stand to gain over the trivial bound $\sqrt{2N\ln 2}$, which is the case provided $\alpha^2 \neq 4$. However, we must also make sure that $\delta < 1$. Going back to (3.170) and making the same approximations as before, this yields the condition

$$1 > \frac{1 - \sqrt{p\rho^2|\ln\rho|}}{1 - 2p\rho} = \frac{1 - \alpha p\rho}{1 - 2p\rho} \tag{3.176}$$

which is satisfied for $\alpha^2 > 4$. In fact, we can optimize over α in the expression (3.174) with the (unique admissible) solution

$$\alpha_p = 1 + \frac{1}{p\ln 2} + \sqrt{1 + \frac{1}{(p\ln 2)^2}} \tag{3.177}$$

This concludes the proof.$\diamondsuit$

Remark. In the case $p = 2$, i.e. the SK model proper, our bound (evaluated without approximations using MATHEMATICA), yields $\sup \frac{1}{N} H(\sigma) \leq \sqrt{2\ln 2} \times 0.999832$ which is considerably worse than the spherical bound of Comets [C1]. More generally, we see that the improvement in the upper bound over the trivial REM bound is rather marginal. While it may be possible with more work or some more clever tricks to sharpen this bound a bit, it is an unfortunate fact that no known methods will allow to completely close the gap between upper and lower bounds.

3.3.2. Truncated partition functions and overlaps.

We will now present Talagrand's approach to the p-spin model. As explained in the REM, this is based on the analysis of second moments of truncated partition functions.

The first result is the analogue to Proposition 3.1 in the REM.

Theorem 3.23: *In the p-spin SK model, (3.17) holds for all $\beta \leq \beta_p$ where*

$$\beta_p^2 \geq \hat{\beta}_p^2 \equiv \inf_{t\in[0,1]} \left(\phi(t) \frac{1+t^p}{t^p} \right) \tag{3.178}$$

Moreover,

$$\sqrt{2\ln 2}(1 - \frac{2^{-p-1}}{\ln 2}) \leq \beta_p \leq \sqrt{2\ln 2}(1 - c_p) \tag{3.179}$$

where c_p is the constant from Proposition 3.21.

Proof. We use the same definitions as in Section 3.2.1. Note that the estimate (3.26) carries over unaltered. A difference occurs only when computing the mean square of the partition function. Namely, while in the REM only the two cases $\sigma = \sigma'$ and $\sigma \neq \sigma'$ had to be distinguished, here we have to distinguish all different values for the overlap $R_N(\sigma,\sigma')$. We can use the convenient fact that $\mathcal{H}_\sigma + \mathcal{H}_{\sigma'}$ has the same distribution as $\sqrt{2+2R_N(\sigma,\sigma')}\mathcal{H}_\sigma$. This gives

$$\begin{aligned} \mathbb{E}\widetilde{Z}_{\beta,p,N}(c)^2 =& \mathbb{E}_\sigma \mathbb{E}_{\sigma'} \sum_{t=-1}^{1} \mathbb{1}_{R(\sigma,\sigma')=t} \mathbb{E}e^{\beta\sqrt{N}(\mathcal{H}_\sigma+\mathcal{H}_{\sigma'})} \mathbb{1}_{\mathcal{H}_\sigma<c\sqrt{N},\mathcal{H}'_\sigma<c\sqrt{N}} \\ \leq& \sum_{t=-1}^{1} \mathbb{E}e^{\beta\sqrt{N}\sqrt{2+2t^p}\mathcal{H}_\sigma} \mathbb{1}_{\mathcal{H}_\sigma<2c\sqrt{N}/\sqrt{2+2t^p}} \mathbb{E}_\sigma \mathbb{E}_{\sigma'} \mathbb{1}_{R(\sigma,\sigma')=t} \end{aligned} \tag{3.180}$$

where the sums over t are understood to be over the possible values of R_N. Up to sub-leading corrections,

$$\mathbb{E}e^{\beta\sqrt{N}\sqrt{2+2t^p}\mathcal{H}_\sigma} \mathbb{1}_{\mathcal{H}_\sigma<2c\sqrt{N/(2+2t^p)}} = \begin{cases} \frac{\sqrt{1+t^p}}{\sqrt{2\pi N}(\beta(1+t^p)-c)} e^{N\left(2\beta c-\frac{c^2}{1+t^p}\right)}, & \text{if } 1+t^p > c/\beta \\ e^{\beta^2 N(1+t^p)}, & \text{if } 1+t^p \leq c/\beta \end{cases} \tag{3.181}$$

Therefore, using Stirling's approximation for the binomial coefficients (3.115) we get:

(i) If $\beta < c$:

$$\begin{aligned} \mathbb{E}\widetilde{Z}^2_{\beta,p,N}(c) \leq& \frac{e^{\beta^2 N}}{\sqrt{2\pi N}} \left(\sum_{t:1+t^p<c/\beta} \frac{2e^{-N\phi(t)+N\beta^2 t^p}}{\sqrt{1-t^2}} \right. \\ &\left. + \sum_{t:1+t^p>c/\beta} \frac{2\sqrt{1+t^p}e^{-N\phi(t)+N\frac{c^2 t^p}{1+t^p}-N(c-\beta)^2}}{\sqrt{(1-t^2)}(\beta(1+t^p)-c)} \right) \end{aligned} \tag{3.182}$$

and

(ii) If $\beta > c$,

$$\mathbb{E}\widetilde{Z}^2_{\beta,p,N}(c) \leq \frac{e^{N(2\beta c-c^2)}}{2\pi N} \sum_{t=-1}^{1} \frac{2\sqrt{1+t^p}}{\sqrt{1-t^2}(\beta(1+t^p)-c)} e^{-N\phi(t)+N\frac{c^2 t^p}{1+t^p}} \tag{3.183}$$

We see that the expectation of the square of the truncated partition function is essentially equal to the square of the expectation, provided that c is chosen such that

$$\inf_t[-\phi(t) + \frac{c^2 t^p}{1+t^p}] \leq 0 \tag{3.184}$$

Since we must choose $c > \beta$ if we want that that $\mathbb{E}Z_{\beta,N} = \mathbb{E}\widetilde{Z}_{\beta,N}(c)$, we see that we can meet both conditions only if $\beta \leq \hat{\beta}_p$ as defined in (3.178). More precisely, analyzing the function in the exponent near $t = 0$, one finds:

Lemma 3.24: *For all $p \geq 3$, we have*

(i) If $\beta < c$ and c satisfies (3.184), then

$$\frac{\mathbb{E}\widetilde{Z}^2_{\beta,N}(c) - [\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2}{[\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2} \leq \begin{cases} C_p N^{-p/2+1}, & \text{for } p \text{ even,} \\ C_p N^{-p+2}, & \text{for } p \text{ odd,} \end{cases} \tag{3.185}$$

(ii) If $\beta \geq c$ and c satisfies (3.184), then

$$\frac{\mathbb{E}\widetilde{Z}^2_{\beta,N}(c)}{[\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2} \leq C_p N^{+1/2} \tag{3.186}$$

Proof. Part (ii) is obvious and we leave the details as an exercise. To prove the more subtle result (i), note that up to an exponentially small correction,

$$[\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2 = \sum_{t=-1}^{1} e^{\beta^2 N}\mathbb{E}_\sigma \mathbb{E}_{\sigma'} \mathbb{1}_{R(\sigma,\sigma')=t} \tag{3.187}$$

Therefore,

$$\begin{aligned} &\mathbb{E}\widetilde{Z}^2_{\beta,N}(c) - [\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2 \\ &\leq \sum_{t=-1}^{1} \left(\mathbb{E}e^{\beta\sqrt{N}\sqrt{2+2t^p}\mathcal{H}_\sigma}\mathbb{1}_{\mathcal{H}_\sigma < 2c\sqrt{N/(2+2t^p)}} - e^{\beta^2 N}\right)\mathbb{E}_\sigma \mathbb{E}_{\sigma'} \mathbb{1}_{R(\sigma,\sigma')=t} \\ &\leq \frac{e^{\beta^2 N}}{\sqrt{2\pi N}}\left(\sum_{t:1+t^p<c/\beta} \frac{2e^{-N\phi(t)}\left(e^{+N\beta^2 t^p} - 1\right)}{\sqrt{1-t^2}}\right. \\ &\qquad \left. + \sum_{t:1+t^p>c/\beta} \frac{2e^{-N\phi(t)+N\frac{c^2 t^p}{1+t^p} - N(c-\beta)^2}}{\sqrt{1-t^2}}\right) \end{aligned} \tag{3.188}$$

By (3.184) the second sum is exponentially small in N; for the first case, if p is even, use that on the domain of summation,

$$\left|\frac{2e^{-N\phi(t)}\left(e^{+N\beta^2 t^p} - 1\right)}{\sqrt{1-t^2}}\right| \leq \frac{2}{\sqrt{1-(c/\beta-1)^{2/p}}} N\beta^2 t^p e^{-N\left(\phi(t) - \beta^2 t^2 (c/\beta - 1)^{(p-2)/p}\right)} \tag{3.189}$$

The bound (3.185) is now obtained using that $\phi(t) \approx t^2/2$ for small t, comparing the sum in (3.188) to an integral and performing a simple change of variables. In the odd case, we expand the exponential to second order, and use that the first order term gives no contribution by symmetry. This yields the sharper result in that case.$\diamondsuit$

Remark. The bound (3.185) is the analogue of the exponential bound (3.42) valid in the REM. In [BKL] it has been shown that this variance estimate goes together with a central limit theorem for the free energy, analogous to parts (i) and (ii) of Theorem 3.2.

While (3.185) is weaker than its analog in the REM, it is still more than sufficient to prove (3.32). From here the first assertion of the theorem follows just like in the REM.

The lower bound on β_p follows from the analysis of the right-hand side of (3.178). The upper bound in (3.179) is a consequence of Proposition 3.21. Namely, if $\lim_{N\uparrow\infty} \mathbb{E}F_{\beta,N} = -\beta^2/2$ for all $\beta \leq \beta_p$. Then for such β,

$$\begin{aligned} \beta = & -\frac{d}{d\beta} \lim_{N\uparrow\infty} \mathbb{E}F_{\beta,N} = -\lim_{N\uparrow\infty} \frac{d}{d\beta}\mathbb{E}F_{\beta,N} = \lim_{N\uparrow\infty} N^{-1/2}\mu_{\beta,N}(\mathcal{H}_\sigma) \\ & \leq \limsup_{N\uparrow\infty} N^{-1/2} \sup_{\sigma\in\mathcal{S}_N} H_\sigma \leq \sqrt{2\ln 2}(1-c_p) \end{aligned} \tag{3.190}$$

This concludes the proof of the theorem in the case $p \geq 3$.

We cannot, however, resist the temptation to present an alternative proof that (a) does not need the *sharp* estimate of Lemma 3.24, and that also works in the case $p = 2$ (for $\beta < 1$). This clever trick was introduced by Talagrand already in [T1]. The argument uses an exponential estimate on the fluctuations of the free energy, which follows from a classical concentration of measure inequality for functions of i.i.d. Gaussian random variables, to be found e.g. in [LT], page 23:

Theorem 3.25: *Let $X_1, \dots, X_M$ be independent standard normal random variables, and let $f : R^M \to R$ be Lipshitz continuous with Lipshitz constant $\|f\|_{Lip}$. Set $g \equiv f(X_1, \dots, X_M)$. Then*

$$\mathbb{P}\left[|g - \mathbb{E}g| > x\right] \leq 2\exp\left(-\frac{x^2}{2\|f\|_{Lip}^2}\right) \tag{3.191}$$

Corollary 3.26: *For any value of β and for any p,*

$$\mathbb{P}\left[|F_{\beta,N}-\mathbb{E}F_{\beta,N}|>x\right]\leq 2\exp\left(-\frac{Nx^2}{2\beta^2}\right) \tag{3.192}$$

Proof. Just check that as a function of the N^p independent variables $J_{i_1,\dots,i_p}$ the free energy is Lipshitz with Lipshitz constant $\beta N^{-1/2}$. (Exercise!!). $\diamondsuit$

Remark. Exponential estimates of this type have played a major rôle in the development of the theory in the last years. The observation that free energies tend to satisfy such estimates appeared first, independently, in [BGP2] and [T1].

The crucial observation, first made in [T1], is that (3.192) and (3.185) are in contradiction, *unless* (3.17) holds. To see this, use the Paley-Zygmund inequality (3.153), which implies that, under our assumption,

$$\begin{aligned}
\frac{C}{4}&\leq\mathbb{P}\left[\widetilde{Z}_{\beta,N}(c)\geq q\mathbb{E}\widetilde{Z}_{\beta,N}(c)\right]\leq\mathbb{P}\left[Z_{\beta,N}\geq\frac{1}{2}\mathbb{E}\widetilde{Z}_{\beta,N}(c)\right]\\
&=\mathbb{P}\left[-F_{\beta,N}+\mathbb{E}F_{\beta,N}\geq\frac{1}{N}\ln\mathbb{E}\widetilde{Z}_{\beta,N}(c)+\mathbb{E}F_{\beta,N}-N^{-1}\ln 2\right]\\
&\leq 2\exp\left(-N\left(\frac{1}{N}\ln\mathbb{E}\widetilde{Z}_{\beta,N}(c)+\mathbb{E}F_{\beta,N}-N^{-1}\ln 2\right)^2/2\right)
\end{aligned} \tag{3.193}$$

provided $\frac{1}{N}\ln\mathbb{E}\widetilde{Z}_{\beta,N}(c)+\mathbb{E}F_{\beta,N}-N^{-1}\ln 2\geq 0$; if this condition fails, by Jensen's inequality $0\leq\mathbb{E}F_{\beta,N}+\beta^2/2\leq N^{-1}\ln 2$, otherwise, (3.193) implies that $|\mathbb{E}F_{\beta,N}-\beta^2/2|\leq CN^{-1/2}$. Thus the first assertion of the theorem is proven. $\diamondsuit$

Remark. The argument yielding the upper bound on β_p can be inverted to show that $\hat{\beta}_p$ is a upper bound for the ground state energy density, $-\liminf_{N\uparrow\infty}N^{-1}\sup_{\sigma\in\mathcal{S}_N}(-H_N^{p-SK}(\sigma))$. Namely, since

$$-\mathbb{E}\frac{d}{d\beta}F_{\beta,N}\leq N^{-1/2}\mathbb{E}\sup_{\sigma\in\mathcal{S}_N}H_\sigma \tag{3.194}$$

for all β, and for all $\beta<\beta_c$, it is true that $\mathbb{E}\frac{d}{d\beta}F_{\beta,N}=-\beta$, it follows that

$$N^{-1}\mathbb{E}\sup_{\sigma\in\mathcal{S}_N}-H_N^{p-SK}(\sigma)\geq\beta_c \tag{3.195}$$

This gives a much simpler proof with an improved estimate of Proposition 3.17. Thus, we have already proven the existence of a phase transition, and we have estimates for the critical temperature which become sharp in the limit $p \uparrow \infty$, and, not surprisingly, the limiting value is the same as in the REM. In fact, a simple argument which we leave as an exercise shows that the free energy converges, as $p \uparrow \infty$, to that of the REM! However, it is far more interesting to understand what happens to the Gibbs measure at the phase transition for fixed p. The key tool remains the analysis of the square of the truncated partition function. We have seen in the computation of the mean of $\widetilde{Z}^2_{\beta,N}(c)$ that in the case $p > 2$, practically all the contribution came from the pairs of spin-configurations with overlap close to zero, and, possibly, ± 1. One clearly gets the feeling that that this computation must have some more profound meaning, and that this should tell us something about the replica overlap distribution. Unfortunately, this is not totally trivial, the main obstacle being that, unlike in the REM, we do not know the precise value of $\sup_\sigma H_\sigma$. We will now show how to overcome, at least partly, this difficulty.

Theorem 3.27: *For any $\epsilon > 0$ there exists $p_0 < \infty$ such that for all $p \geq p_0$, and for all $0 \leq \beta < \infty$*

$$\lim_{N\uparrow\infty} \mathbb{E}\mu^{\otimes 2}_{N,\beta}(|R_N(\sigma,\sigma')| \in [\epsilon, 1-\epsilon]) = 0 \tag{3.196}$$

If, moreover, $\beta < \hat{\beta}_p$, then for any $\epsilon > 0$ there exists $p_0 < \infty$ such that for all $p \geq p_0$, such that for some $\delta > 0$, for all large enough N,

$$\mathbb{E}\mu^{\otimes 2}_{N,\beta}(|R_N(\sigma,\sigma')| \in [\epsilon, 1]) \leq e^{-\delta N} \tag{3.197}$$

Remark. Note that we prove this result without any restriction on the temperature, while Talagrand requires some upper bound on β both in [T4] and (largely improved) in the announcement [T5].

Proof. Let us first consider the high-temperature region. In view of the fact that we know how to compute the mean of the square of the *truncated* partition function and that this would give the desired result, we only need to (i) justify the truncation and (ii) show that the partition functions in the denominator can be replaced by a constant without harm. But (ii) follows from Lemma 3.24: by Chebyshev's inequality, it implies that for $\beta < c < \hat{\beta}_p$,

$$\begin{aligned}\mathbb{P}[Z_{\beta,N} < (1-\delta)\mathbb{E}\widetilde{Z}_{\beta,N}(c)] &\leq \mathbb{P}[\widetilde{Z}_{\beta,N}(c) < (1-\delta)\mathbb{E}\widetilde{Z}_{\beta,N}(c)] \\ &\leq \mathbb{P}\left[|\widetilde{Z}_{\beta,N}(c) - \mathbb{E}\widetilde{Z}_{\beta,N}(c)| \geq \delta\mathbb{E}\widetilde{Z}_{\beta,N}(c)\right] \\ &\leq \delta^{-2}N^{1-p/2}\end{aligned} \tag{3.198}$$

To justify point (i), we use (3.198) to show that

$$\begin{aligned}&\mathbb{E}\mu_{\beta,N}^{\otimes 2}\left(\{\mathcal{H}_\sigma > c\sqrt{N}\} \cup \{\mathcal{H}_{\sigma'} > c\sqrt{N}\}\right) \leq 2\mathbb{E}\mu_{\beta,N}\left(\{\mathcal{H}_\sigma > c\sqrt{N}\}\right) \\ &= 2\mathbb{E}\frac{\mathbb{E}_\sigma e^{\beta\mathcal{H}_\sigma)}\mathbb{1}_{\mathcal{H}_\sigma > c\sqrt{N}}}{Z_{\beta,N}} \\ &\leq 2\frac{\mathbb{E}\mathbb{E}_\sigma e^{\beta\mathcal{H}_\sigma}\mathbb{1}_{\mathcal{H}_\sigma > c\sqrt{N}}}{(1-\delta)\mathbb{E}\widetilde{Z}_{\beta,N}(c)} + \mathbb{P}[Z_{\beta,N} < (1-\delta)\mathbb{E}\widetilde{Z}_{\beta,N}(c)] \\ &\leq 2\frac{e^{-N(c-\beta)^2/2}}{(1-\delta)\sqrt{N}(c-\beta)} + \delta^{-2}N^{1-p/2}\end{aligned} \tag{3.199}$$

Now we can conclude readily that

$$\begin{aligned}\mathbb{E}\mu_N^{\otimes 2}(|R_N(\sigma,\sigma')| \in [\epsilon, 1]) \leq{}& \frac{\mathbb{E}\mathbb{E}_{\sigma,\sigma'}\mathbb{1}_{\mathcal{H}_\sigma < c\sqrt{N}, \mathcal{H}_{\sigma'} < c\sqrt{N}}\mathbb{1}_{R_N(\sigma,\sigma')\in[\epsilon,1]}e^{\beta\sqrt{N}(\mathcal{H}_\sigma + \mathcal{H}_{\sigma'})}}{(1-\delta)[\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2} \\ &+ 2\frac{e^{-N(c-\beta)^2/2}}{(1-\delta)\sqrt{N}(c-\beta)} + 2\delta^{-2}N^{1-p/2}\end{aligned} \tag{3.200}$$

Using the representation (3.182), we see that the principal term in (3.200) satisfies

$$\frac{\mathbb{E}\mathbb{E}_{\sigma,\sigma'}\mathbb{1}_{\mathcal{H}_\sigma < c\sqrt{N}, \mathcal{H}_{\sigma'} < c\sqrt{N}}\mathbb{1}_{R_N(\sigma,\sigma')\in[\epsilon,1]}e^{\beta\sqrt{N}(\mathcal{H}_\sigma + \mathcal{H}_{\sigma'})}}{(1-\delta)[\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2} \leq C\exp(-Nc\epsilon^2) \tag{3.201}$$

with constants of order unity. This means that ϵ can even be chosen as $\epsilon \sim N^{-1/2}\ln N$, showing that the measure at high temperatures concentrates as sharply on zero overlap as the uniform measure.

Let us now turn to the more intricate and more delicate low temperature case. Again we have to justify truncation, and we need a uniform lower bound on the partition functions in the denominator. But first we must decide how to truncate. Note that so far we used truncation only for high temperatures and in such a way that the truncations did not alter the mean partition function. At low temperatures this will not do, and we must dare truncation with $c < \beta$. In the REM we have actually seen a much finer truncation at work that isolated the main contributions to the Gibbs measures coming from the extremal order statistics. This suggest that we should try to truncate at the ground state energy. Unfortunately, we do not know its limiting value, and we do not even know that it has a limit. Fortunately, however, we are able to show that its random fluctuations are quite small. This will give us an opportunity to actually use Theorem 3.25 in a situation where all else would fail.

Lemma 3.28: *For any $\epsilon > 0$, and for all N large enough,*

$$\mathbb{P}\left[\left|\sup_\sigma \mathcal{H}_\sigma - \mathbb{E}\sup_\sigma \mathcal{H}_\sigma\right| > x\sqrt{N}\right] \leq \exp\left(-N\frac{x^2}{2}\right) \tag{3.202}$$

Proof. This is in fact a simple corollary of Theorem 3.25. All we have to show is to compute the Lipshitz norm of $\sup_\sigma \mathcal{H}_\sigma$. But

$$\begin{aligned} &\left|\sup_\sigma \mathcal{H}_\sigma[\omega] - \sup_\sigma \mathcal{H}_\sigma[\omega']\right| \leq \sup_\sigma |\mathcal{H}\sigma[\omega] - \mathcal{H}_\sigma[\omega']| \\ &= N^{-p/2}\sup_\sigma\left|\sum_{i_1,\dots,i_p}(J_{i_1,\dots,i_p}[\omega] - J_{i_1,\dots,i_p}[\omega'])\sigma_{i_1}\dots\sigma_{i_p}\right| \\ &\leq N^{-p/2}\|J[\omega] - J[\omega']\|_2 N^{p/2} = \|J[\omega] - J[\omega']\|_2 \end{aligned} \tag{3.203}$$

which means that the Lipshitz norm of $\sup_\sigma(\mathcal{H}_\sigma))$ is equal to one. $\diamondsuit$

As a consequence, we can introduce without harm the indicator function of the event that $\sup_\sigma \mathcal{H}_\sigma \leq \mathbb{E}\sup_\sigma \mathcal{H}_\sigma + c_N$, where $c_N \sim C\ln N$. On the other hand, we can bound the partition functions in the denominator by $2^{-N}\exp\left(\beta\sqrt{N}(\mathbb{E}\sup_\sigma(\mathcal{H}_\sigma) - c_N)\right)$, since clearly

$$
\begin{aligned}
&\mathbb{P}\left[Z_{\beta,N} < 2^{-N}\exp\left(\beta\sqrt{N}(\mathbb{E}\sup_\sigma \mathcal{H}_\sigma) - c_N)\right)\right] \\
&\leq \mathbb{P}\left[2^{-N}\sup_\sigma e^{\beta\sqrt{\mathcal{H}_\sigma}} < 2^{-N}\exp\left(\beta\sqrt{N}(\mathbb{E}\sup_\sigma \mathcal{H}_\sigma) - c_N\right)\right] \\
&= \mathbb{P}\left[\sup_\sigma \mathcal{H}_N(\sigma) < \mathbb{E}\sup_\sigma(\mathcal{H}_\sigma) - c_N\right] \leq \exp\left(-\frac{c_N^2}{2}\right)
\end{aligned}
\tag{3.204}
$$

Now let $I \subset (-1,1)$. Set $E_N \equiv N^{-1/2}\mathbb{E}\sup_\sigma \mathcal{H}_\sigma$. Then

$$
\begin{aligned}
\mathbb{E}\mu_N^{\otimes 2}(R_N(\sigma,\sigma') \in I) = &\ \mathbb{E}\mathbb{1}_{\sup_\sigma \mathcal{H}_\sigma \leq \sqrt{N}E_N + c_N} \frac{\mathbb{E}_{\sigma,\sigma'} e^{\beta\sqrt{N}(\mathcal{H}_\sigma + \mathcal{H}_{\sigma'})}\mathbb{1}_{R_N(\sigma,\sigma')\in I}}{2^{-2N}e^{2\beta N(E_N - c_N/\sqrt{N})}} \\
&+ 2\exp\left(-\frac{c_N^2}{2}\right) \\
\leq &\ \frac{\mathbb{E}\mathbb{E}_{\sigma,\sigma'} e^{\beta\sqrt{N}(\mathcal{H}_\sigma + \mathcal{H}_{\sigma'})}\mathbb{1}_{\mathcal{H}_\sigma + \mathcal{H}_{\sigma'} \leq 2\sqrt{N}E_N + 2c_N)}\mathbb{1}_{R_N(\sigma,\sigma')\in I}}{2^{-2N}e^{2\beta N(E_N - c_N/\sqrt{N})}} \\
&+ 2\exp\left(-\frac{c_N^2}{2}\right)
\end{aligned}
\tag{3.205}
$$

Now assume that $\beta > E_N + c_N/\sqrt{N}$. Then, as in (3.183),

$$
\begin{aligned}
&\frac{\mathbb{E}\mathbb{E}_{\sigma,\sigma'} e^{\beta\sqrt{N}(\mathcal{H}_\sigma + \mathcal{H}_{\sigma'})}\mathbb{1}_{\mathcal{H}_\sigma + \mathcal{H}_{\sigma'} \leq 2(\sqrt{N}E_N + c_N)}\mathbb{1}_{R_N(\sigma,\sigma')\in I}}{2^{-2N}e^{2\beta N(E_N - c_N/\sqrt{N})}} \\
&\leq \frac{e^{N(2\ln 2 - (E_N + c_N/\sqrt{N})^2 + 4\beta N^{-1/2}c_N)}}{2\pi N} \sum_{t\in I} \frac{2\sqrt{1+t^p}\, e^{-N\phi(t) + N\frac{(E_N + c_N N^{-1/2})^2 t^p}{1+t^p}}}{(1-t^2)(\beta(1+t^p) - E_N - N^{-1/2}c_N)}
\end{aligned}
\tag{3.206}
$$

The pleasant aspect of this expression is that it is essentially independent of β, since c_N goes to zero with N. Moreover, we know that $\sqrt{2\ln 2} > E_N \geq \sqrt{2\ln 2}(1-c_p)$, with $c_p = 2^{-p}/\ln 2$. Thus, any interval I on which

$$
\phi(t) - N\frac{2\ln 2 t^p}{1+t^p} > c_p \sim 2^{-p}
\tag{3.207}
$$

gives only an exponentially small contribution. This applies to all intervals except $(-2^{-p/2}, 2^{-p/2})$ and $[1-2^{-p}, 1]$. This proves the theorem. $\diamondsuit$

Of course we would like to know more precisely how the the overlap is distributed on the remaining two little intervals. In particular it seems more than reasonable that the phase transition is accompanied, as in the REM, by a charging of mass to the neighbourhood of the value one, which would imply a "lumping phenomenon" of the Gibbs measure. But it looks at the moment rather hopeless to get such information from a computation like the preceding one; in particular, as long as we do not know the precise value of E_N (remember how sharply we had to estimate in the REM to get such information out!). Thus to get more information, one has to use some more subtle tricks. We have already seen in the REM, that integration by parts provides rather powerful insights. In fact, most of Section 3.2.5. can be carried over to the p-spin model with some changes.

3.3.3 Ghirlanda–Guerra identities and lump masses.

All we want to do now is to use the Gaussian integration by parts formula (3.122) for the p-spin model. Of course the i.i.d. Gaussians to use for that are no longer the Hamiltonian, but the independent couplings $J_{i_1,\dots,i_p}$. The analogue of Lemma 3.12 is

Lemma 3.29: *For any value of β, in the p-spin model,*

$$\mathbb{E}\frac{d}{d\beta}F_{\beta,N} = -\beta\left(1 - \mathbb{E}\mu_{\beta,N}^{\otimes 2}\left(R_N^p(\sigma,\sigma')\right)\right) \tag{3.208}$$

Proof. We just repeat the steps of the proof of Proposition 3.12 with the obvious modifications. Namely

$$\begin{aligned}
-\mathbb{E}\frac{d}{d\beta}F_{\beta,N} &= N^{-1/2}\mathbb{E}\frac{\mathbb{E}_\sigma \mathcal{H}_\sigma e^{\beta\sqrt{N}\mathcal{H}_\sigma}}{\mathbb{E}_\sigma e^{\beta\sqrt{N}\mathcal{H}_\sigma}} \\
&= N^{-(p+1)/2}\sum_{1\leq i_1,\dots,i_p\leq N}\mathbb{E}J_{i_1,\dots,i_p}\frac{\mathbb{E}_\sigma \sigma_{i_1},\dots\sigma_{i_p}e^{\beta\sqrt{N}\mathcal{H}_\sigma}}{\mathbb{E}_\sigma e^{\beta\sqrt{N}\mathcal{H}_\sigma}} \\
&= N^{-(p+1)/2}\sum_{1\leq i_1,\dots,i_p\leq N}\mathbb{E}\Bigg(\beta N^{-(p-1)/2}\frac{\mathbb{E}_\sigma \sigma_{i_1}^2,\dots\sigma_{i_p}^2 e^{\beta\sqrt{N}\mathcal{H}_\sigma}}{\mathbb{E}_\sigma e^{\beta\sqrt{H}_\sigma}}
\end{aligned}$$

$$
\begin{aligned}
&\left. - \beta N^{-(p-1)/2} \frac{\mathbb{E}_\sigma \mathbb{E}_{\sigma'} \sigma_{i_1}, \dots \sigma_{i_p} \sigma'_{i_1}, \dots \sigma'_{i_p} e^{\beta\sqrt{N}(\mathcal{H}_\sigma + \mathcal{H}_{\sigma'})}}{\mathbb{E}_\sigma \mathbb{E}_{\sigma'} e^{\beta\sqrt{N}(\mathcal{H}_\sigma + \mathcal{H}_{\sigma'})}} \right) \\
&= \beta - \beta \mathbb{E} \frac{\mathbb{E}_\sigma \mathbb{E}_{\sigma'} \left(N^{-1} \sum_{i=1}^N \sigma_i \sigma'_i \right)^p e^{\beta\sqrt{N}(\mathcal{H}_\sigma + \mathcal{H}_{\sigma'})}}{\mathbb{E}_\sigma \mathbb{E}_{\sigma'} e^{\beta\sqrt{N}(\mathcal{H}_\sigma + \mathcal{H}_{\sigma'})}}
\end{aligned}
\tag{3.209}
$$

This is (3.208).$\diamondsuit$

As first observed by Talagrand, the equality (3.208) implies that the replica overlap cannot remain concentrated on the value zero for all values of β. Namely, assume that for all values of $\beta \in [0, b]$, $\lim_{N\uparrow\infty} \mathbb{E}\mu_{\beta,N}^{\otimes 2} \left(R_N^p(\sigma, \sigma') \right) = 0$. Then it is plain from (3.208) that, in this interval, $\lim_{N\uparrow\infty} \mathbb{E}F_{N,\beta} = -\beta^2/2$. Thus the replica overlap cannot remain zero beyond the critical value. This actually suggests that for all values $\beta > \beta_p$, $\liminf_{N\uparrow\infty} \mathbb{E}\mu_{\beta,N}^{\otimes 2} \left(R_N^p(\sigma, \sigma') \right) > 0$, but there is unfortunately no monotonicity argument available to prove this[29]. However, the estimate on the ground state energy from Proposition 3.21 implies strict positivity at least soon after the critical point. In fact,

Proposition 3.30: *For all $\beta > \sqrt{2\ln 2}(1 - c_p)$, with c_p as in Proposition 3.21,*

$$
\liminf_{N\uparrow\infty} \mathbb{E}\mu_{\beta,N}^{\otimes 2} \left(R_N^p(\sigma, \sigma') \right) > 0 \tag{3.210}
$$

Proof. Just note that by the first line of (3.209) the left-hand side of (3.208) is equal to N^{-1} times the average of the Hamiltonian, which in turn is bounded by N^{-1} times the mean of the supremum of the Hamiltonian. The estimates from the proof of Proposition 3.21 then suffice that this is less than $\sqrt{2\ln 2}(1 - c_p)$, contradicting (3.208) unless (3.210) holds. $\diamondsuit$

If we combine Proposition 3.30 with Theorem 3.27, we see that above the critical value of β, the overlap has a non-trivial distribution supported on the neighbourhoods of zero, and ± 1, while intermediate values are excluded. As Talagrand observed in [T3], this fact alone allows to draw rather stringent conclusions about the Gibbs

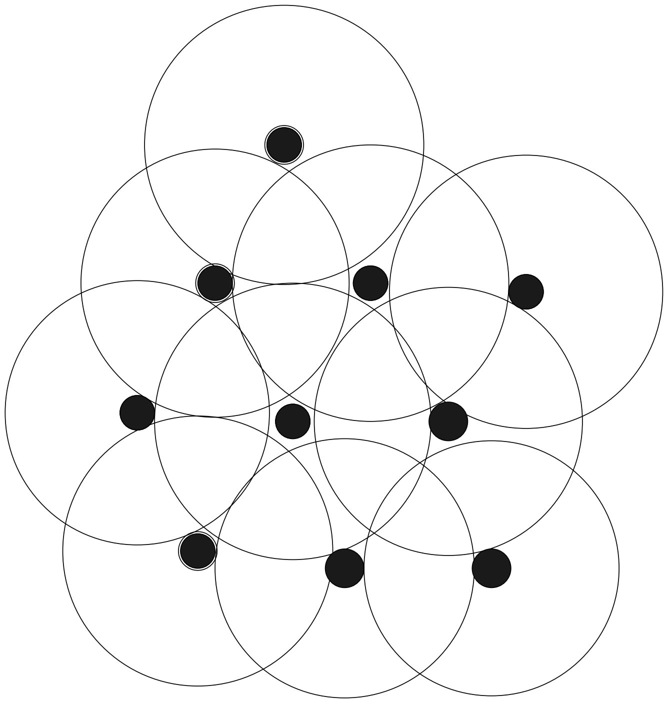
[29] That is to say, it is in principle possible that the overlap would exhibit a very erratic oscillating behaviour just above the critical value of β!

measures. Roughly speaking, and maybe not too surprisingly, they amount to saying that the Gibbs measure look like some softened up version of that of the REM. More precisely, one can conclude that asymptotically, as $N \uparrow \infty$, the Gibbs measure will be concentrated on a small subset of the hypercube that consists of of even smaller disjoint components that Talagrand called "lumps". Each lump is of size no bigger than $O(2^{-p})$, while the distance between any two lumps is at least $1 - O(2^{-p/2})$. More precisely, is possible to decompose the state space $\mathcal{S}_N$ into a collection of disjoint subsets $\mathcal{C}_k$ such that

(i)
$$\lim_{N\uparrow\infty} \mathbb{E}\mu_{\beta,N}^{\otimes 2}\left(\{(\sigma,\sigma') \,|\, |R_N(\sigma,\sigma')| > \epsilon\} \setminus \cup_k \mathcal{C}_k \times \mathcal{C}_k\right) = 0 \tag{3.211}$$

(where the $\mathcal{C}_k$ depend both on N and on the random parameter!), and

(ii) If $\sigma, \sigma' \in \mathcal{C}_k$, then $R_N(\sigma,\sigma') \geq 1 - \epsilon$.



Lumps surrounded by empty regions

These facts follow from quite simple geometric considerations that we invite the reader to work out for herself or to look up in [T3].

The most important contribution of Proposition 3.30 to this picture is that it implies that at least one of these lumps has positive mass. Of course one would expect that the distribution of the masses of the individual lumps should be somewhat similar to the situation in the REM as well, i.e. the total mass should be distributed according to some law on a countable set of lumps. In the REM this was the consequence of a fundamental theorem on the Poisson convergence of the extreme order statistics of an i.i.d. family of random variables. There seems no immediate way to obtain a similar prove in the p-spin case. On the other hand, we have seen that substantial information on the lump distribution could also be obtained via the Ghirlanda–Guerra identities. Talagrand has announced that this path is feasible also in the p-spin case, and we will follow him at least some steps along this road.

Theorem 3.31: *Assume that $\beta > \beta_p$. Let $\mathcal{C}_k$ be ordered such that for all k, $\mu_{N,\beta}(\mathcal{C}_k) \geq \mu_{N,\beta}(\mathcal{C}_{k+1})$. Then for all $k \in \mathbb{N}$, there exists $p_k < \infty$ such that for all $p \geq p_k$,*

$$\lim_{N\uparrow\infty} \mathbb{E}\mu_{N,\beta}\left(\cup_{l=1}^{k}\mathcal{C}_l\right) < 1 \tag{3.212}$$

except possibly for an exceptional set of β's of zero Lebesgue measure. Moreover, for k large, $p_k \sim \frac{2}{3}\frac{\ln k}{\ln 2}$.

Proof. In fact, just as we could prove an analogue of Lemma 3.12 for the p-spin model, we can also generalize, virtually without changing the proof, Proposition 3.15. This reads

Proposition 3.32: *In the p-spin model, for all but a countable set of values β, for any bounded function $h : \mathcal{S}_N^n \to \mathbb{R}$,*

$$\lim_{N\uparrow\infty}\Bigg| \mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\dots,\sigma^n)R_N^p(\sigma^k,\sigma^{n+1})\right) - \frac{1}{n}\mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\dots,\sigma^n)\left(\sum_{l\neq k}^{n} R_N^p(\sigma^l,\sigma^k) + \mathbb{E}\mu_{\beta,N}^{\otimes 2}(R_N^p(\sigma_1,\sigma_2))\right)\right)\Bigg| = 0 \tag{3.213}$$

Proof. We leave the proof, which follows that of Proposition 3.15 in detail, as an exercise. See also [L] for a good and general exposition.

Choosing h to be the indicator function

$$h(\sigma^1,\dots,\sigma^n) = \mathbb{I}_{\forall_{k\neq l} R_N(\sigma^k,\sigma^l)=q_{kl}} \tag{3.214}$$

we get that (for almost all β)

$$\begin{aligned} &\lim_{N\uparrow\infty} \mathbb{E}\mu_{N,\beta}^{\otimes n+1}\left[R_N^p(\sigma^k,\sigma^{n+1})\middle|\forall_{k\neq l} R_N(\sigma^k,\sigma^l) = q_{kl}\right] \\ &= \frac{1}{n}\sum_{l\neq k}^n q_{kl}^p + \frac{1}{n}\lim_{N\uparrow\infty} \mathbb{E}\mu_{N,\beta}^{\otimes 2}\left(R_N^p(\sigma^1,\sigma^2)\right) \end{aligned} \tag{3.215}$$

which is the relation (17) of [GG].

Assume that the assertion of Theorem 3.31 fails. Then there exists a first instance k^* such that

$$\lim_{N\uparrow\infty} \mathbb{E}\mu_N\left(\cup_{l=1}^{k^*}\mathcal{C}_l\right) = 1 \tag{3.216}$$

Now define events $\mathcal{Q}_{\epsilon_0}^{(n)} \in \mathcal{B}_n$ by

$$\mathcal{Q}_{\epsilon_0}^{(n)} \equiv \left\{\underline{R} \in [-1,1]^{n(n-1)/2} | \forall_{1\leq l<k\leq n} |R_{lk}| \leq \epsilon_0\right\} \tag{3.217}$$

The important observation is that, if $\{R_N(\sigma_l,\sigma_k)\}_{1\leq l<k\leq k^*} \in \mathcal{Q}_{\epsilon_0}^{(k^*)}$, then there exists some permutation $\pi \in S_{k^*}$ such that, with probability one $\sigma^k \in C_{\pi(k)}$ for all $k \leq k^*$. In particular

$$\begin{aligned} &\lim_{N\uparrow\infty} \mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{k^*+1})\mathbb{I}_{\{R_N(\sigma^l,\sigma_m)\}_{1\leq l<\leq k^*}\in\mathcal{Q}_{\epsilon_0}^{(k^*)}}\right] \\ &= \lim_{N\uparrow\infty} \mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{k^*+1})\mathbb{I}_{\exists_\pi \forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right] \end{aligned} \tag{3.218}$$

But

$$
\begin{aligned}
&\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{k^*+1})\mathbb{1}_{\exists_\pi \forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]\\
&\quad=\sum_{\pi\in S_{k^*}}\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{k^*+1})\mathbb{1}_{\forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]\\
&\quad=\sum_{\pi\in S_{k^*}}\sum_{j=1}^{k^*}\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{k^*+1})\mathbb{1}_{\sigma_{k^*+1}\in\mathcal{C}_{\pi(j)}}\mathbb{1}_{\forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]\\
&\quad=\sum_{\pi\in S_{k^*}}\sum_{j\neq k}^{k^*}\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{j})\mathbb{1}_{\sigma_{k^*+1}\in\mathcal{C}_{\pi(j)}}\mathbb{1}_{\forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]\\
&\quad+\sum_{\pi\in S_{k^*}}\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{k^*+1})\mathbb{1}_{\sigma_{k^*+1}\in\mathcal{C}_{\pi(k)}}\mathbb{1}_{\forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]
\end{aligned}
\tag{3.219}
$$

where we used the symmetry between replicas in the terms $j \neq k$ to exchange σ^{k^*+1} with σ^j. Note that for the first term we have the obvious (though not very good) bound

$$
\begin{aligned}
0&\leq\sum_{\pi\in S_{k^*}}\sum_{j\neq k}^{k^*}\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{j})\mathbb{1}_{\sigma_{k^*+1}\in\mathcal{C}_{\pi(j)}}\mathbb{1}_{\forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]\\
&\leq\epsilon_0^p\mathbb{E}\mu_{N,\beta}^{\otimes k^*}\left[\mathbb{1}_{\forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]\\
&=\epsilon_0^p\mathbb{E}\mu_{N,\beta}^{\otimes k^*}\left[\mathcal{Q}_{\epsilon}^{k^*}\right]
\end{aligned}
\tag{3.220}
$$

while the second term satisfies

$$
\begin{aligned}
&\sum_{\pi\in S_{k^*}}\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{k^*+1})\mathbb{1}_{\sigma_{k^*+1}\in\mathcal{C}_{\pi(k)}}\mathbb{1}_{\forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]\\
&\qquad\geq(1-\epsilon)^p\sum_{\pi\in S_{k^*}}\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[\mathbb{1}_{\sigma_{k^*+1}\in\mathcal{C}_{\pi(k)}}\mathbb{1}_{\forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]\\
&\qquad=\frac{1}{k^*}(1-\epsilon_1)^p\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[\mathbb{1}_{\forall_{l=1}^{k^*}\sigma^l\in\mathcal{C}_{\pi(l)}}\right]\\
&\qquad=\frac{1}{k^*}(1-\epsilon_1)^p\mathbb{E}\mu_{N,\beta}^{\otimes k^*}\left[\mathcal{Q}_{\epsilon_0}^{k^*}\right]
\end{aligned}
\tag{3.221}
$$

where we used the obvious permutation symmetry among the first k^* replicas. Let us now use (3.213) with f chosen as the indicator function of the event $\mathcal{Q}_{\epsilon_0}^{(k^*)}$. Clearly we get

$$\lim_{N\uparrow\infty} \mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[R_N^p(\sigma^k,\sigma^{k^*+1})\mathbb{1}_{\{R_N(\sigma^l,\sigma_m)\}_{1\leq l<\leq k^*}\in\mathcal{Q}_{\epsilon_0}^{(k^*)}}\right]$$
$$\leq \frac{1}{k^*}\lim_{N\uparrow\infty}\mathbb{E}\mu_{N,\beta}^{\otimes k^*+1}\left[\mathbb{1}_{\{R_N(\sigma^l,\sigma_m)\}_{1\leq l<\leq k^*}\in\mathcal{Q}_{\epsilon_0}^{(k^*)}}\right]\left((k^*-1)\epsilon_0^p+\mathbb{E}\mu_{N,\beta}^{\otimes 2}R^p(\sigma,\sigma')\right) \tag{3.222}$$

Comparing (3.220), (3.221) to (3.222) we see that

$$(1-\epsilon_1)^p \leq (k^*-1)\epsilon_0^p + \lim_{N\uparrow\infty}\mathbb{E}\mu_{N,\beta}^{\otimes 2}R^p(\sigma,\sigma') \leq (k^*-1+p_0)\epsilon^p + p_1 \tag{3.223}$$

This implies the lower bound

$$k^* \geq \frac{(1-\epsilon_1)^p - p_1}{\epsilon_0^p} \tag{3.224}$$

Quantitatively, this estimate can be refined to

$$k^* \geq C^{-1}2^{3p/2}((1-C2^{-p})^p - p_1) = 2^p p_0(1-O(2^{-2p})) \tag{3.225}$$

This proves the theorem. $\diamondsuit$

Remark. It is actually possible to prove, under the additional assumption that β is not too large, that for large enough p, (3.212) holds *for all* k, i.e. the number of lumps with positive mass is unbounded. This follows from an improved estimate on the lower bound of the interval of excluded replicas under such conditions that is given by Theorem 1.5 from [T3],resp. in improved form Theorem 8.8 of [T4].

Theorem 3.33: [T3] *There exists $p_0<\infty$ such that if $p\geq p_0$ and $\beta<2^{p/2-6}$, then for any $\epsilon>0$,*

$$\lim_{N\uparrow\infty}\mathbb{E}\mu^{\otimes 2}\beta, N\left(|R_N(\sigma,\sigma')|\in[\epsilon,1-2^{-p}]\right)=0 \tag{3.226}$$

I will not give the proof of this theorem, which is technically rather involved. Also, the condition that β should not be too large appears unnatural. Let me just mention that it makes use of the so-called *cavity method*, or induction over the volume. We will introduce this important method in the final part of these notes in the context of the Hopfield model only.

4. Mean-field models 2: Hopfield models

In this section we turn our attention to the second line of generalizations of the Curie-Weiss model we mentioned at the beginning of Section 3. That is we will study models that share with the Curie-Weiss model the feature that their Hamiltonian is expressed as a function of so-called "overlap parameters" or "macroscopic functions". The model we will focus on here is the so-called "Hopfield model" [Ho]. In this case, the macroscopic functions are just the "overlaps" with a random set of a-priori chosen spin configurations, typically denote by $\xi^1, \dots, \xi^M \in \mathcal{S}_N$ and called "patterns". We set

$$m_\mu(\sigma) \equiv R_N(\sigma, \xi^\mu) \tag{4.1}$$

If the ξ^μ are chosen at random, these quantities become random functions on the space of spin configurations, and a random Hamiltonian is defined e.g. via

$$H_N(\sigma) \equiv -N \sum_{\mu=1}^{M} m_\mu^2(\sigma) \tag{4.2}$$

This choice yields the famous Hopfield Hamiltonian.

4.1. Origins of the model.

The story of the Hopfield model is quite interesting and worth to be detailed. The name goes back to John Hopfield, a physicist at Caltech interested in modelling the behaviour of networks of neurons (for a more general survey, see [A], and from a more mathematical perspective [P]), such as the human brain. Now the brain is a really messy system, composed of a giant mesh of roughly 10^{10} cells, called neurons. These neurons are all linked up via so-called dendrites, essentially long organic wires that are capable of transmitting electric impulses from one neuron to another. What these neurons do basically is to send out and receive sequences of such electric pulses at various frequencies. The important thing is that the frequency (or firing rate in the jargon) at which a neuron sends out its pulses depends (among other things) in a rather complicated manner on the signals coming in from all the other neurons it is connected with. That is to say, each neuron is a small device that processes incoming information and transmits the result to other neurons. Apparently, the way these things are hooked up, this produces a device that can perform rather amazing computational tasks (like reading these pages and possibly making some sense out of them...). How can one possibly understand how such a system works? Clearly, already a single neuron is a rather complicated system whose dynamics is far from easy to analyse; trying to analyse the joint behaviour of billions of them looks thus hopeless. In such a situation physicists like to simplify the models, and to abstract from details while keeping what are believed to be the essential features. The first step in this simplification goes the same way as our old friend Ising: simplify the "state space" of a single neuron to the simplest possible one, $\{-1,1\}$, suggesting that the neuron fires "rapidly" or "slowly". This idea goes back at least to McCulloch and Pitts [MP] in 1943. Now it was known that a neuron changes its state at a rate depending on the compound, but weighted, effect of all its input signals, that are functions of the states of those neurons that are connected to it[30]. We can think of this effective input signal as some field

$$h_i = f_i\left(\{\sigma_j\}_{j\in N(i)}\right) \tag{4.3}$$

where $N(i)$ denotes the set of all neurons "firing" into the neuron i. The way the network processes information then depends on the structure of these neighbourhoods (the "graph" of connections), and the properties of the functions f_i. Setting these up corresponds in a way to programming the system. Clearly the simplest choice for a function f_i is a linear one, that is

$$f_i\left(\{\sigma_j\}_{j\in N(i)}\right) = \sum_{j\in N(i)} J_{ij}\sigma_j \tag{4.4}$$

It is known that the effects of a signal on a neuron can go both ways, inducing it to fire or to stop firing. Thus the coefficients J_{ij} should have the possibility of taking both signs, and different strengths. Thus we see that this field looks like the local field acting on site i in a *spin glass model.* Of course, contrary to a spin glass, the coupling should not be arbitrary, but be programmed to ensure a particular task. But how can such programming happen? As early as 1949, D. Hebb [He] suggested

[30]Connections in real neural networks are directed and not reciprocal, a fact that we shall ignore....

a sort of progressive self-programming of such a network that should give rise to a network functioning as a *memory.* The idea is that the connection between two neurons should be altered in a direction to "favour" the current states present, i.e. one should add a term proportional to $\sigma_i\sigma_j$. In this way, if the network over time has passed through a number M of different "states", denoted by $\xi^1, \dots, \xi^M$, the couplings would take the form

$$J_{ij} = \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu \tag{4.5}$$

This form of the coupling is called *Hebb's rule.* While it is a bit difficult to believe that things should be that simple, this rule has been a rather widely accepted one, and in any case is interesting enough for us to start investigating the ensuing model. Of course, if this idea was to be taken seriously, the couplings should evolve in time; we will, however, assume that the couplings have reached a state of saturation where they do not change anymore, and what we are interested in should be the evolution of the state of the network from some initial state $\sigma(0)$ with fixed couplings of the form (4.5). We still have to define the dynamics: Here Hopfield proposed in [Ho] a *Markov chain* where at independent exponentially distributed random times a clock rings at site i and the neuron changes its state $\sigma_i(t)$ to a new value ± 1 with rates proportional to $\exp(\pm\beta h_i)$. Hopfield's key observation was that (in the case of symmetric connections), such a Markov chain would have as its invariant measure a *Gibbs measure* corresponding to the Hamiltonian

$$H_N(\sigma) = -\sum_{i,j:j\in N(i)} J_{ij}\sigma_i\sigma_j \tag{4.6}$$

with J given by (4.5). Finally, simplifying the model further by assuming that any neuron is connected to any other, and normalizing properly, one arrives at the Hamiltonian

$$H_N(\sigma) = -\frac{1}{N}\sum_{\mu=1}^{M}\sum_{i,j=1}^{N} \xi_i^\mu \xi_j^\mu \sigma_i\sigma_j \tag{4.7}$$

which one sees easily to be equal to the expression given in (4.2). This is, in brief, the reasoning that led Hopfield to derive what could be called the "Ising model" of neural networks. The formal resemblance to spin glass models then of course sparked immediate interest among physicists who saw their chance to bring their expertise to bear in an entirely new context.

Interestingly enough, though, the Hamiltonian (4.7) was introduced already 6 years before Hopfield, in three papers by Figotin and Pastur [FP1-3] which had apparently received very little attention[31]. Not surprisingly, they were advocated as simple, exactly solvable models of *spin glasses!* Figotin and Pastur more or less gave a complete solution which, however, failed to exhibit the key features expected of spin glasses but revealed that the model behaves very much like the ordinary Curie–Weiss ferromagnet, except that the number of stable magnetized states was equal to $2M$ instead of 2. In a way, from this point of view the model was about as disappointing as the model introduced earlier on by Mattis [Ma] that corresponded to the case $M = 1$ and was seen to be totally equivalent to the Curie–Weiss model.

Fortunately, Hopfield did not repeat the analysis of Figotin and Pastur (this was done a few years later by various people), but performed numerical experiments[32]. Moreover, in these experiments he had an objective that was motivated by the interpretation of the model as a memory. This meant that, starting from an initial configuration somewhat close to one of the "patterns" ξ^μ, the system should approach ξ^μ and stay close to it for a long time, if not forever. Now Hopfield observed that this was indeed true (in a sense), but only if M was not too big: in fact the allowed value of M depended on the size N of the network and was roughly $M^* = 0.14N$. So something interesting happened, but only if M was taken as a function of N! Naturally, this fact had eluded Figotin and Pastur who studied the thermodynamic limit $N \uparrow \infty$ with M fixed!

The seemingly small modification brought to the model by considering M as a function $M(N)$ of the system size N thus actually turns what otherwise would be a rather simple mean-field model into something much more interesting and also

[31] This fact had been brought to my attention only in 1993 by L. Pastur; it seems that these papers were almost never cited before in the entire literature on Hopfield's model.

[32] This is not to discourage people from doing analytic work. However, the example shows that interesting facts are often found by experimenting with things one does not understand.

much more complicated to analyse. In [BG2] we have called such models *generalized random mean-field models*. I will not discuss here the more general setting introduced there but stay with the single example of the standard Hopfield model.

While Hopfield's results suggested that this model is much more interesting than was first thought, the choice of the function $M(N)$ provides a parameter that could make a rigorous analysis possible at least under certain conditions on the growth rate of this function. Indeed, at least the rigorous study of the model can be seen as a constant struggle to push our understanding to larger and larger growth rates, ranging from Figotin and Pastur's constant M through logarithmic [KP,vEvH,Ga] via sub-linear growth [BGP1] to what we can control today, linear growth $M(N) = \alpha N$ with sufficiently small α [BGP2,BG1,BG2, T2,T5]. I will in the sequel explain the most significant steps in this development.

4.2. Basic ideas: Finite M.

To get a feeling for the model, it will be worthwhile to discuss the case when M is finite first. Doing this, we face a small dilemma: there are two basic methods that can be used and have been used to study the model. One is based on the *large deviation theory*. This approach looks a priori more rational and has the advantage that it can be applied to a very general class of (generalized) mean-field models. Its strategy is to study first the distribution of the macroscopic order parameters $m_N(\sigma)$ under the Gibbs measure, i.e. the measures $\mathbb{Q}_N$ s.t.

$$\mathbb{Q}_{\beta,N}\left(m \in \mathcal{A}\right) \equiv \mu_{\beta,N}\left(m_N(\sigma) \in \mathcal{A}\right) \tag{4.8}$$

for any Borel set in $\mathbb{R}^M$. Now it is fairly straightforward to see that the family of measures $\mathbb{Q}_{\beta,N}$ satisfies a *large deviation principle*, i.e.

$$\begin{aligned} \sup_{m\in\mathcal{A}^O} \Psi_\beta(m) &= \liminf_{N\uparrow\infty} -\frac{1}{N}\ln \mathbb{Q}_{\beta,N}(\mathcal{A}) \\ \leq \limsup -\frac{1}{N}\ln \mathbb{Q}_{\beta,N}(\mathcal{A}) &= \sup_{m\in\bar{\mathcal{A}}} \Psi_\beta(m) \end{aligned} \tag{4.9}$$

with probability one, and for a *rate function* Ψ_β that is independent of the realization of the random variables ξ. This observation goes back to van Hemmen and co-workers [vH1,vH2,vEvH,vHGHK,vHvEC,] and, in greater generality, Comets [Co].

The computation of the rate function is greatly simplified by the fact that H_N is just a function of the m_N. In fact, finding the rate function is reduced to the *combinatorial problem* of counting the number of spin configurations σ that gives rise to the same value m_N. This problem is in principle elementary, even though the final expression is rather involved.

The second approach, based on what is frequently called the *Hubbard–Stratonovich (H-S) transformation* [Hu,St] looks a bit artificial, and works well only in cases where the Hamiltonian is a quadratic function of the order parameters. When it works, however, it is much simpler, and some of the latest results have only been obtained with this method (even though in the past it has always been possible to reproduce all results with both tools). Since the large deviation approach has been explained extensively in [BG3], I decided that in these notes I will stick with the Hubbard–Stratonovich approach. This approach was incidentally also the one used in [FP1,FP2]. One way to see the HS transformation is to say that it consists in constructing the convolution of the induced measure $\mathbb{Q}_{\beta,N}$ with a Gaussian measure of mean zero and variance $1/\beta N$,

$$\mathcal{Q}_{\beta,N} \equiv \mathbb{Q}_{\beta,N} \star \mathcal{N}(0, 1/\beta N) \tag{4.10}$$

The amazing fact is that this measure can be written down in a very explicit form (much more explicit than the measure $\mathbb{Q}_{\beta,N}$, due to the simple identity

$$e^{\frac{1}{2}x^2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dz e^{-\frac{1}{2}z^2 + xz} \tag{4.11}$$

applied to the Boltzmann factor:

$$e^{-\frac{1}{2}N\sum_{\mu=1}^{M}(m_N(\sigma)^\mu)^2} = \frac{1}{(2\pi)^{M/2}} \int_{-\infty}^{\infty} dz_1 \dots dz_M e^{-\frac{1}{2}\sum_\mu z_\mu^2 + N^{-1/2}\sum_{i=1}^{N}\sigma_i\left(\sum_\mu \xi_i^\mu z_\mu\right)} \tag{4.12}$$

The key point is that the exponent in this expression is now linear in the variables σ_i. It follows immediately, after a convenient change of variables, that

$$Z_{\beta,N} = \frac{1}{(2\pi\beta N)^{M/2}} \int_{-\infty}^{\infty} dz_1 \dots dz_M e^{-N\beta\left[\frac{1}{2}\sum_\mu z_\mu^2 - \frac{1}{\beta N}\sum_{i=1}^{N}\ln\cosh\left(\beta\sum_\mu \xi_i^\mu z_\mu\right)\right]} \tag{4.13}$$

and

$$\mathcal{Q}_{\beta,N}(dz) = \frac{\exp\left(-N\beta\Phi_{\beta,N}(z)\right)}{Z_{\beta,N}}dz \tag{4.14}$$

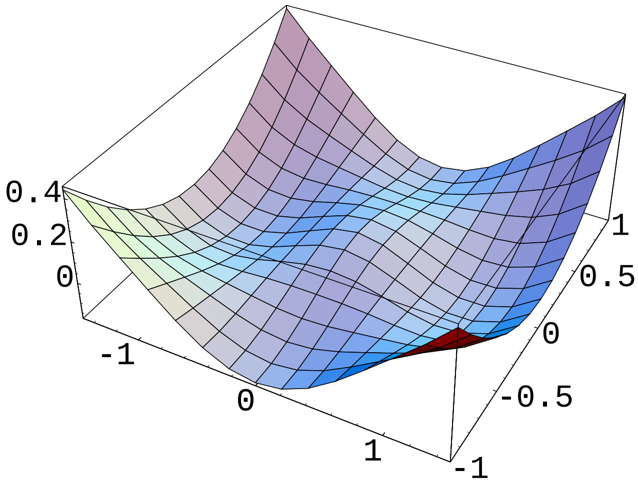
with

$$\Phi_{\beta,N}(z) \equiv \frac{1}{2}\sum_{\mu} z_\mu^2 - \frac{1}{\beta N}\sum_{i=1}^{N} \ln\cosh\left(\beta\sum_{\mu}\xi_i^\mu z_\mu\right) \tag{4.15}$$

This function will play a central rôle in the remainder of this section. In the context of the HS approach, it replaces the rate function in the large deviation approach; the big advantage of the HS method is that it can be written down explicitly for any values of N and M and looks thus much more suitable in the cases when M depends on N. Moreover, the rate function Ψ and the function Φ are quite intimately connected, as is explained at length in [BG2].

Looking at the function $\Phi_{\beta,N}$, we see that the second part of it is an empirical mean over the sample of N random vectors ξ_i. Thus we may expect that in the limit $N \uparrow \infty$, this function will converge to a deterministic one, namely

$$\Phi_\beta(z) \equiv \frac{1}{2}\|z\|_2^2 - \beta^{-1}\mathbb{E}\ln\cosh\left(\beta(\xi_1, z)\right) \tag{4.16}$$



The function $\Phi_2(z)$ for $M = 2$

This follows e.g. in the topology of uniform convergence on compact sets from the *law of large numbers* in Banach spaces (see e.g. [LT]). Since I will later show a related result in the case when M depends on N, I will not discuss this any further at this point. In any case, it is clear that the first thing to do is to understand how this function looks like if we want to understand the properties of the measure $\mathcal{Q}_{\beta,N}$ for large N.

The first observation, due to Figotin and Pastur, is that

Lemma 4.1: *For any $M \in \mathbb{N}$, if ξ_1^μ are i.i.d. Rademacher random variables taking the values ± 1 with equal probability, then the function Φ_β takes its minimal values $\phi_\beta(m^*) = \frac{1}{2}(m^*)^2 - \beta^{-1} \ln\cosh(\beta m^*)$, where m^* is the largest solution of the equation*

$$x = \tanh(\beta x) \tag{4.17}$$

on the set $\mathcal{M}_{\beta,M}$ given by

$$\mathcal{M}_{\beta,M} \equiv \bigcup_{\pm,\mu \in \{1,\dots,M\}} \{\pm m^* e^\mu\} \qquad (4.18),$$

where e^μ denotes the μ-th unit vector in $\mathbb{R}^M$. Note that $m^* = 0$, whence $|\mathcal{M}_{\beta,M}| = 1$ if and only if $\beta \leq 1$.

We even have more:

Lemma 4.2: *Under the assumptions of the preceding lemma, for all $\beta \neq 1$ there exists $c(\beta) > 0$ such that for all $M \in \mathbb{N}$,*

$$\Phi_\beta(z) - \phi_\beta(m^*) \geq c(\beta) \min_{y \in \mathcal{M}_{\beta,M}} \|z - y\|_2^2 \tag{4.19}$$

Proof. I will only give a short proof of Lemma 4.1. Note that

$$\|z\|_2^2 = \mathbb{E}\left((\xi_1, z)\right)^2 \tag{4.20}$$

so that

$$\Phi_\beta = \mathbb{E}\phi_\beta\left((\xi_1, z)\right) \tag{4.21}$$

where ϕ_β is the Curie–Weiss function defined above. This function attains its minima at the points $\pm m^*$. Thus it is clear that if z is such that the random variable (ξ_1, z) is supported on the set $\{-m^*, m^*\}$, then Φ_β attains its absolute minimum at this point; moreover, if such a value exists, then the absolute minimum is attained precisely on the set of values z for which this is true. Now if ξ_i^μ are Rademacher, then any z of the form $z = \pm m^* e^\mu$ has this property. Moreover, it is very easy to see that these are the only possible candidates. Namely, our condition is

$$\sum_{\mu=1}^{M} \xi_1^\mu z_\mu = \pm m^*, \; \forall \xi_i \in \{-1, 1\}^M \tag{4.22}$$

Without loss of generality we can assume $\xi_1^1 = 1$. Then (4.22) implies that

$$\begin{aligned} z_1 + b &= \pm m^* \quad \text{and} \\ z_1 - b &= \pm m^* \end{aligned} \tag{4.23}$$

Quite obviously this can only be true if either $z_1 = \pm m^*$ and $b = 0$, or $z_1 = 0$, and $b = \pm m^*$. In the second case we are done if $M = 2$, and we can proceed inductively otherwise. In the first case we argue that $b = 0$ implies that all z_μ, $\mu > 1$ must be zero. This is trivial in the case $M = 2$, while for $M > 2$ we can split b again into two pieces that would need to satisfy $z_2 + b_2 = 0$ and $z_2 - b_2 = 0$, which is obviously only possible if both $z_2 = 0$ and $b_2 = 0$. This proves Lemma 4.1. $\diamondsuit$

Noting the elementary fact that $\phi_\beta(x) - \phi_\beta(m^*) \geq a(\beta)(|x| - m^*)^2$, it is quite obvious that Lemma 4.2 holds with some constant $c(\beta)$. The proof given in [BG3] yields a numerical estimate on the constant, but I am not terribly happy with either the proof (that is rather cumbersome) or the estimate (that is rather poor). Thus I encourage the reader to do better by herself!

In the case of finite M, it follows readily from these observations that any ϵ-neighbourhood of the set $\mathcal{M}_{\beta,M}$ carries all but an exponentially small fraction of the total mass of the measures $\mathcal{Q}_{\beta,N}$, with probability that tends to one very rapidly. In particular, it is very easy to see that the measure conditioned on, say, any ball[33] of radius $r < m^*$ centered at $m^* e^\mu$ converges, as $N \uparrow \infty$, to the Dirac measure on the point $m^* e^\mu$, almost surely. The same is then true, naturally, for the measures $\mathbb{Q}_{\beta,N}$.

[33] Or, for that matter, any closed set containing the single point $m^* e^\mu$ from $\mathcal{M}_{\beta,M}$.

We see that the vicinities of the points $\pm m^* e^\mu$ play here the same rôle as the "lumps" in the REM or the p-spin SK model, with the difference that they are not randomly placed but deterministically put in by the construction of the model. A natural question is then whether we can control in this model the respective "lump-masses", that is whether we can control the behaviour of the unconditioned Gibbs measures. This problem was considered only rather late as an illustration of the concept of "metastates" in two papers by Külske [Ku1,Ku2]. It is clear that this requires a much more precise analysis of the function $\Phi_{\beta,N}$ than what we have given so far. How can this be obtained?

The first idea should be to use the functional central limit theorem (see e.g. [LT]) to extract the sub-leading corrections. Indeed the following holds:

$$\sqrt{N}\left(\Phi_{\beta,N}(z) - \Phi_\beta(z)\right) \xrightarrow{\mathcal{D}} g_\beta(z) \tag{4.24}$$

where g_β is a Gaussian process on $\mathbb{R}^M$ with covariance

$$C_\beta(z,z') = \mathbb{E}\ln\cosh(\beta(\xi_1,z))\ln\cosh(\beta(\xi_1,z')) - \mathbb{E}\ln\cosh(\beta(\xi_1,z))\mathbb{E}\ln\cosh(\beta(\xi_1,z')) \tag{4.25}$$

At first glance this would suggest that the fluctuations of $\Phi_{\beta,N}$ are of order $1/\sqrt{N}$, and thus the relative weights of the different "lumps" should differ by factors of order $\exp(\beta\sqrt{N})$. However, a closer inspection shows that this is not true. Namely, note that we are interested in the process g_β essentially only very near the points $z = \pm m^* e^\mu$. But at these points, the variance turns out to be zero, as better be the case, since at these points $\Phi_{\beta,N}$ is non-random! Note that this relies crucially on the fact that the random variables ξ_i^μ take only the values ± 1, and as soon other distributions are considered, this will change dramatically[34]. In any case, we see that the precision of the CLT is not enough to solve our problem, and we have to look for the next order corrections. In fact, given that the fluctuations are strictly zero at the points $\pm m^* e^\mu$, one might first suspect that maybe the weights could be all equal. However, the random effects will induce small shifts of the position of the

[34] A particularly interesting situation arises if the distribution of the ξ is taken to be Gaussian, see [BvEN].

minima of $\Phi_{\beta,N}$ away from these points, and we will have to control these shifts and the values of the function at these real minima to solve our problem. Since we expect these shifts to be very small (tending to zero with N, a natural approach is to use Taylor expansions. Let us consider the minimum near the point $m^* e^\mu$. Its location $z^{(\mu)}$ must satisfy the equations

$$z_\nu^{(\mu)} = \frac{1}{N}\sum_i \xi_i^\nu \tanh(\beta(\xi_i, z^{(\mu)}) \tag{4.26}$$

Now write $z^{(\mu)} = m^* e^\mu + \delta$. Then δ satisfies

$$\begin{aligned} \delta_\mu &= \frac{1}{N}\sum_i \xi_i^\mu \tanh(\beta(m^*\xi_i^\mu + (\delta, \xi_i))) - m^* \\ \delta_\nu &= \frac{1}{N}\sum_i \xi_i^\nu \tanh(\beta(m^*\xi_i^\mu + (\delta, \xi_i))), \quad \nu \neq \mu \end{aligned} \tag{4.27}$$

Taylor expanding, and using that $m^* = \tanh(\beta m^*)$, we get

$$\begin{aligned} \delta_\mu &= \beta\cosh^{-2}(m^*\beta)\frac{1}{N}\sum_i \xi_i^\mu(\delta,\xi_i) + O(\|\delta\|_2^2) \\ &= \beta\cosh^{-2}(m^*\beta)\delta_\mu + \beta\cosh^{-2}(m^*\beta)\frac{1}{N}\sum_{\nu\neq\mu}\sum_i \xi_i^\mu\xi_i^\nu\delta_\nu + O(\|\delta\|_2^2) \\ \delta_\nu &= \frac{1}{N}\sum_i \xi_i^\nu\xi_i^\mu \tanh(\beta m^*) + \frac{\beta}{N}\sum_i \xi_i^\nu \cosh^{-2}(\beta m^*)(\delta,\xi_i) + O(\|\delta\|_2^2), \quad \nu\neq\mu \\ &= \beta\cosh^{-2}(m^*\beta)\delta_\nu + \tanh(\beta m^*)\frac{1}{N}\sum_i \xi_i^\nu\xi_i^\mu \\ &+ \beta\cosh^{-2}(m^*\beta)\frac{1}{N}\sum_{\nu'\neq\nu}\sum_i \xi_i^\nu\xi_i^{\nu'}\delta_{\nu'} + O(\|\delta\|_2^2), \quad \nu\neq\mu \end{aligned} \tag{4.28}$$

Since $\frac{1}{N}\sum_i \xi_i^\nu\xi_i^\mu = O(N^{-1/2})$, one checks readily that to leading order the solution of these equations is

$$\begin{aligned} \delta_\mu &= 0 + O(1/N) \\ \delta_\nu &= \frac{\tanh(\beta m^*)}{1-\beta\cosh^{-2}(\beta m^*)}\frac{1}{N}\sum_i \xi_i^\nu\xi_i^\mu + O(1/N) \end{aligned} \tag{4.29}$$

It follows that

$$\Phi_{\beta,N}(z^{(\mu)}) - \phi_\beta(m^*) = \sum_\nu \delta_\nu^2 (1-\beta\cosh^{-2}(\beta m^*)) + o(1/N)$$
$$= \frac{1}{N}\frac{(m^*)^2}{1-\beta(1-(m^*)^2)}\sum_{\nu\neq\mu}\left(N^{-1/2}\sum_i \xi_i^\nu\xi_i^\mu\right)^2 + o(1/N) \tag{4.30}$$

where we used that $\cosh^{-2}(x) = 1-\tanh^2(x)$. As a consequence, the measure $\mathcal{Q}_{\beta,N}$ has a decomposition

$$\mathcal{Q}_{\beta,N} \approx \sum_{\mu=1}^M p_{\beta,N}(\mu)\left(\mathcal{Q}_{\beta,N}(\cdot|B_\epsilon(m^*e^\mu)) + \mathcal{Q}_{\beta,N}(\cdot|B_\epsilon(-m^*e^\mu)\right) \tag{4.31}$$

where the conditional measures converge almost surely to Dirac measures, and

$$p_{\beta,N}(\mu) = \frac{e^{\beta\frac{(m^*)^2}{1-\beta(1-(m^*)^2)}\sum_{\nu\neq\mu}\left(N^{-1/2}\sum_i\xi_i^\nu\xi_i^\mu\right)^2}}{\sum_{\mu=1}^M e^{\beta\frac{(m^*)^2}{1-\beta(1-(m^*)^2)}\sum_{\nu\neq\mu}\left(N^{-1/2}\sum_i\xi_i^\nu\xi_i^\mu\right)^2}} \tag{4.32}$$

We see that these weights are, as random variables, functions of the $M(M-1)$ sums of i.i.d. random variables

$$b_N^{\mu\nu} \equiv N^{-1/2}\sum_i \xi_i^\nu\xi_i^\mu \tag{4.33}$$

Since for $\mu<\nu$, these variables are uncorrelated (check!), it follows again by the central limit theorem, that this family of variables converges weakly to independent normal variables. This permits to formulate a convergence result in the spirit of the metastate formalism.

Theorem 4.3: [Ku1]*Assume that $\beta>1$ and $M<\infty$. Then*

$$\mathcal{Q}_{\beta,N} \stackrel{\mathcal{D}}{\to} \mathcal{Q}_\beta \tag{4.34}$$

where $\mathcal{Q}_\beta$ is the random measure on $\mathbb{R}^M$ given by

$$\mathcal{Q}_\beta \equiv \sum_{\mu=1}^M p_\beta^\mu(\delta_{m^*e^\mu} + \delta_{-m^*e^\mu}) \tag{4.35}$$

where

$$p_\beta^\mu \equiv \frac{e^{\beta \frac{(m^*)^2}{1-\beta(1-(m^*)^2)} \sum_{\nu\neq\mu} g_{\mu\nu}^2}}{\sum_{\mu=1}^M e^{\beta \frac{(m^*)^2}{1-\beta(1-(m^*)^2)} \sum_{\nu\neq\mu} g_{\mu\nu}^2}} \tag{4.36}$$

and, for $\mu < \nu$, the family $\{g_{\mu\nu}\}$ are independent standard Gaussian random variables.

Remark. The same result holds of course also with $\mathcal{Q}$ replaced by $\mathbb{Q}$.

It is clear that we could fix (condition on) a finite number of the components of ξ^μ without affecting at all the result. Thu the rôle of conditioning on the disorder that was emphasized in the construction of the metastates does not really come to bear in this setting. In that respect it will be more instructive to look at the Gibbs measures as measures on the original spin space. Since as usual we are interested in the convergence of finite-volume measures in the product topology, it will be enough to consider probabilities

$$\mu_{\beta,N}(\sigma_I = s_I) \tag{4.37}$$

for any finite $I \subset \mathbb{N}$ and $s_I \in \{-1,1\}^I$, and to prove joint convergence of arbitrary finite collections of such probabilities. A rather simple computation shows that these can be represented as follows:

$$\mu_{\beta,N}(\sigma_I = s_I) = \frac{\int dz e^{-\beta N \Phi_{\beta',N'}(z)} \prod_{i\in I} e^{\beta s_i(\xi_i,z)}}{\int dz e^{-\beta N \Phi_{\beta',N'}(z)} \prod_{i\in I} 2\cosh(\beta(\xi_i,z))} \tag{4.38}$$

where $N' = N - |I|$, $\beta' = \beta N/N'$. Note that of course the difference between N and N' and β and β' is negligible in the limit $N \uparrow \infty$. It is clear that Theorem 4.3 implies that $\mu_{\beta,N}$ converges in distribution to the random measure μ_β whose finite-dimensional marginals are given by

$$\mu_{\beta,N}(\sigma_I = s_I) = \int \mathcal{Q}_\beta(dz) \prod_{i\in I} \frac{e^{\beta s_i(\xi_i,z)}}{2\cosh(\beta(\xi_i,z))} \tag{4.39}$$

where ξ_i^μ are i.i.d. symmetric Rademacher r.v.'s and $\mathcal{Q}_\beta$ is the random measure from Theorem 4.3. We see that when looking at convergence in distribution, we loose the information on our patterns: the ξ_i^μ appearing in (4.39) have the same distribution as the patterns that we used to construct the Hamiltonian, but they are not the same random variables. But clearly we can do better. Namely, conditioning on the ξ_i^μ for all $i \in J$, for an arbitrarily large finite set $J \subset \mathbb{N}$ does not change anything concerning the convergence of the measures $\mathcal{Q}_{\beta,N}$ (since these finitely many variables give no contribution to the limits of the variables $b_N^{\mu\nu}$!), however, they do appear in (4.39). Therefore, the Aizenman-Wehr metastate will be the same random measure as μ_β, except that now the ξ_i^μ appearing on the right are precisely the original patterns from the definition of the Hamiltonian. Combining these observations, we can state the following theorem:

Theorem 4.4: [Ku1]*Assume that $\beta > 1$ and $M < \infty$. Then, given ξ,*

$$\mu_{\beta,N}[\xi] \stackrel{\mathcal{D}}{\to} \sum_{\nu=1}^{M} p_\beta^\nu \left(\mu_\beta^{+,\nu}[\xi] + \mu_\beta^{-,\nu}[\xi]\right) \tag{4.40}$$

where p_β^ν are defined in (4.36) and $\mu_\beta^{\pm,\nu}[\xi]$ are product measures on $\{-1,1\}^{\mathbb{N}}$ with marginals

$$\mu_\beta^{\pm,\nu}[\xi](\sigma_i = s_i) \equiv \frac{e^{\pm\beta m^* \xi_i^\nu s_i}}{2\cosh(\beta m^*)} \tag{4.41}$$

Remark. The product measures $\mu_\beta^{\pm,\nu}[\xi]$ can be naturally seen as the *pure states* for this model. Theorem 4.4 then says that the *metastate* is a random convex combination of these extremal states, in accordance with Theorem 2.5. One can easily construct sequences of measures converging almost surely to one of these extremal measures by adding an *external field term*

$$\pm h \sum_{i=1}^{N} \xi_i^\nu \sigma_i \tag{4.42}$$

to the Hamiltonian, and taking the limit $N \uparrow \infty$ first and $h \downarrow 0$ after this (Exercise!). This is in fact much simpler than to prove Theorems 4.3 and 4.4 and requires much less information about the fluctuations of the process $\Phi_{\beta,N}(z)$. It is also much more robust and can be proven even when M grows with N, as long as $\lim_{N\uparrow\infty} M/N = 0$ [BGP1].

Since all of these results follow just from the explicit representation of the lump-weights as functions of sums of i.i.d. r.v.'s, one can obtain further asymptotic results from the properties of this sum. In particular, from the *invariance principle* one can construct the so-called *superstate* (proposed in [BG3]), i.e. the measure-valued stochastic process constructed as the (conditioned) distributional limit of the process $\{\mu_{\beta,[tN]}\}_{t\in(0,1]}$ and which is obtained from the expression for μ_β by just replacing the Gaussian r.v.'s $g_{\mu\nu}$ by independent standard Brownian motions $b_{\mu\nu}(t)$. This and further results relating the different notions of metastates can be found in [Ku1,Ku2].

4.3 Growing M.

The finite M model is, as we have seen, readily solvable using just standard results from probability theory: the law of large numbers, central limit theorems, and the Laplace method. As soon as M turns into a function of N, these results are no longer immediately applicable and require substantial modifications. Before discussing the results obtained in this direction, let me identify the main steps in the analysis of the finite M case that need to be reconsidered.

(i) The first difficulty is the fact that the law of large numbers no longer provides the convergence of the function $\Phi_{\beta,N}$ to its deterministic limit. Indeed, since these functions are now defined on $\mathbb{R}^{M(N)}$, with $M(N)$ growing to infinity, we would not even know what we should mean by such a convergence.

(ii) Even if we control $\Phi_{\beta,N}$, the Laplace method will have to be adapted to a situation when the integral is over a space whose dimension grows together with the large parameter. This is, however, a rather minor difficulty.

(iii) A precise analysis of the local properties of the minima of $\Phi_{\beta,N}$ can no longer rely on simple Taylor expansions as used in the derivation of Theorem 4.3. The main point is that we have used that by the Schwartz inequality, $|(\xi_i,\delta)| \leq \sqrt{M}\|\delta\|_2$. If M is finite, this implies that if δ is small in norm, then all the shifts in the arguments of the functions appearing are also small; if M grows with N, this is no longer the case.

All these difficulties are getting the more serious, the faster M is allowed to grow. As a consequence, the history of the mathematical analysis of the Hopfield model is marked by a sequence of steps reaching larger and larger rates of growth: First results (using combinatorial large deviation methods) by Koch and Piasko [KP] and van Enter and van Hemmen [vEvH] reached logarithmic growth $M(N) \ll \ln N$. The next stage reached sub-linear growth $(M(N)/N \downarrow 0)$. After the first computation of the free energy (Koch [Ko] and Shcherbina and Tirozzi [ST]), the extremal Gibbs measures were constructed in a collaboration with Gayrard and Picco [BGP1], a large deviation principle was proven with Gayrard [BG5] (see also [CD] for an interesting variant), and finally a central limit theorem was proven in the same collaboration [BG5] (first results on the CLT, under stronger growth conditions, are due to Gentz [G1,G2]. An interesting result on a non-central limit theorem *at the critical temperature* was found mor recently by Gentz and Löwe [GL1,GL2]). The only result missing in this regime is the analogue of Theorem 4.3, which we have obtained only under more stringent growth conditions (namely $M(N) \ll \sqrt{N}$) in a collaboration with D. Mason [BM]. I will not go into the details of these results, but pass to the next step, the case when $M(N) = \alpha N$ with $\alpha > 0$, but small. Here I will distinguish two steps of progress: first, a priori estimates on the support of the Gibbs measures, and exponential estimates on the respective weights [BGP2,BG1], and second, the analysis of the conditional measures corresponding to one pattern, and the justification of the replica symmetric solution [T2,T5,BG2,BG3]. Let me mention that there is another line of research that relates solely to the analysis of local minima of the Hamiltonian of the model and that principally investigates the question of how large α can be chosen if one wants to guarantee the existence of local minima of H_N "near" the stored patterns (the so-called problem of the *storage capacity*) and that I will not discuss in these notes. Key references are [N2,KoPa,Lou,T2].

4.3.1. Fluctuations of Φ.

In some way one can say that the key to analyzing the Hopfield model with growing M is to understand how to use $\alpha = M/N$ as a small parameter instead of $1/N$. This means in particular that we would like to say that $\Phi_{\beta,N}$ is still close to its mean as long as α is small.

Our aim is to show that for small α, the minima of the function $\Phi_{\beta,N}$ are reasonably close to $\pm m^* e^\mu$, and that beyond a small neighbourhood of these points, the function grows somehow like $\mathbb{E}\Phi_{\beta,N}$. Clearly this requires an estimate on $\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z)$ that may get worse as z is farther away from the minima of $\mathbb{E}\Phi_{\beta,N}$. On the other hand, we need estimates that are uniform in z. That is to say, a desirable estimate will be of the form:

Proposition 4.5: *Let $M(N) = \alpha N$. Then there exists a constant $C < \infty$, such that for all $\beta > 1$,*

$$\mathbb{P}\left[\forall_{z:\,\mathrm{dist}(z,\mathcal{M}_{\beta,M(N)})>C\sqrt{\alpha}/m^*}\left\{\Phi_{\beta,N}(z) \geq \frac{1}{2}\mathbb{E}\Phi_{\beta,N}(z)\right\}\right] \geq 1 - e^{-M(N)/C} \tag{4.43}$$

Proof. Our problem is to control the fluctuations of a stochastic process defined on a space of dimension $M(N)$. In principle this is a classical problem in the theory of stochastic processes and there are well-developed tools available that we will not fail to employ: exponential estimates and *chaining* (see in particular [LT]). Let me briefly explain the ideas behind this. The probably most elementary estimate used in probability is that $\mathbb{P}[\max_{i\in I} X_i > x_i] \leq \sum_{i\in I}\mathbb{P}[X_i > x_i]$. This estimate tends to be good, if the variables X_i are independent. Clearly in our situation this is not directly applicable, since we are considering suprema over uncountable sets. The standard remedy is to consider a grid, and to group the points close to a grid-point together, hoping that they will not vary too much from the value on the grid-point, while bounding the maximum over the grid by the sum. Note that we have already used a similar procedure in the proof of Proposition 3.21.

Let us first look at an exponential estimate for the deviation at a fixed point.

$$\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z) = \frac{1}{N\beta}\sum_{i=1}^{N}\left(\ln\cosh(\beta(\xi_i, z)) - \mathbb{E}\ln\cosh(\beta(\xi_i, z))\right) \tag{4.44}$$

and so, as we have already seen above, this difference vanishes strictly whenever z has a single non-vanishing component. It will be crucial to exploit this and to get a bound that shows that the fluctuations decrease as we approach one of the minima of $\mathbb{E}\Phi_{\beta,N}$. Thus we exploit that with z^* being any of the values $\pm m^* e^\mu$, we have that

$$
\begin{aligned}
&\ln\cosh(\beta(\xi_i,z)) - \mathbb{E}\ln\cosh(\beta(\xi_i,z))\\
&= \ln\cosh(\beta(\xi_i,z)) - \ln\cosh(\beta(\xi_i,z^*)) - \mathbb{E}\left(\ln\cosh(\beta(\xi_i,z)) - \ln\cosh(\beta(\xi_i,z^*))\right)\\
&\equiv \beta[f_i(z,z^*) - \mathbb{E}f_i(z,z^*)]
\end{aligned}
\tag{4.45}
$$

Next we use Taylor's formula to bound

$$
|f_i(x,y)| \le |(\xi_i,(x-y))||\tanh(\beta(\xi_i,\bar{z}))| \le |(\xi_i,(x-y))| \tag{4.46}
$$

We will want to use use Chebyshev's inequality to estimate

$$
\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^N (f_i(x,y) - \mathbb{E}f_i(x,y)) \ge \delta\right] \le \inf_{t\ge 0} e^{-t\delta N}\prod_{i=1}^N \mathbb{E}e^{t(f_i(x,y)-\mathbb{E}f_i(x,y))} \tag{4.47}
$$

Thus we must estimate the Laplace transforms of f_i. Using the standard and trivial second order bound on the exponential function, $e^x \le 1 + x + \frac{1}{2}x^2 e^{|x|}$ together with (4.46), we get

$$
\begin{aligned}
\mathbb{E}e^{t(f_i(x,y)-\mathbb{E}f_i(x,y))} &\le 1 + \frac{t^2}{2}\mathbb{E}\left(f_i(x,y) - \mathbb{E}f_i(x,y)\right)^2 e^{t|f_i(x,y)-\mathbb{E}f_i(x,y)|}\\
&\le 1 + \frac{t^2}{2}\left[\mathbb{E}\left(f_i(x,y) - \mathbb{E}f_i(x,y)\right)^4 \mathbb{E}e^{2t|f_i(x,y)-\mathbb{E}f_i(x,y)|}\right]^{1/2}
\end{aligned}
\tag{4.48}
$$

where we have used the Cauchy–Schwarz inequality to separate the expectation of the polynomial and exponential terms for convenience. Now clearly,

$$
\mathbb{E}e^{2t|f_i(x,y)-\mathbb{E}f_i(x,y)|} \le \mathbb{E}e^{4t|f_i(x,y)|} \le \mathbb{E}e^{4t|(\xi_i,(x-y))|} \tag{4.49}
$$

and

$$
\mathbb{E}\left(f_i(x,y) - \mathbb{E}f_i(x,y)\right)^4 \le 7\mathbb{E}E\left(f_i(x,y)\right)^4 \le 7\mathbb{E}(\xi_i,(x-y))^4 \tag{4.50}
$$

Now we can use the Marcinkiewicz-Zygmund inequalities (see e.g. [CT], pp.366 ff.), in particular $\mathbb{E}(z,\xi_i)^k \le k^k\|z\|_2^{2k}$, and $\mathbb{E}e^{s|(z,\xi_i)|} \le 2e^{s^2\|z\|_2^2/2}$. This yields

$$
\mathbb{E}e^{t(f_i(x,y)-\mathbb{E}f_i(x,y))} \le 1 + t^2 32\|x-y\|_2^2 e^{4t^2\|x-y\|_2^2} \le \exp\left(32t^2\|x-y\|_2^2 e^{4t^2\|x-y\|_2^2}\right) \tag{4.51}
$$

We now get

Lemma 4.6: *Let z^* be any of the points $\pm m^* e^\mu$, and set $R \equiv R(z) \equiv \|z - z^*\|_2$. Then, for all $\delta \leq 1$*

$$\mathbb{P}\left[|\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z)| \geq \delta R\right] \leq 2e^{-\frac{1-e^{1/256}/2}{64}N\delta^2} \leq 2e^{-\frac{1}{150}N\delta^2} \tag{4.52}$$

Proof. Insert (4.51) into (4.46) and choose $t = \delta/(64R)$. This gives the bound for the upper deviation. For the lower deviation, the analogue procedure gives the same bound, and this implies (4.52). $\diamondsuit$

Lemma 4.6 shows that typical deviations at a given point z are of order $R/\sqrt{N}$, where R is the distance of z from the nearest coordinate axis. But this does not yet tell us anything about maximal fluctuations. The first idea would be to introduce a suitable lattice $\mathcal{W}$ in $\mathbb{R}^M$, to use Lemma 4.6 to bound the maximal fluctuations on the lattice (as a function of R), and to prove some uniform bound on $\Phi_{\beta,N}(x) - \Phi_{\beta,N}(y)$ that can then be used to control the deviations from the nearest lattice point. Using again (4.46), and the Cauchy–Schwarz inequality, we get easily

$$\begin{aligned} &|\Phi_{\beta,N}(x) - \Phi_{\beta,N}(y)| \\ &\leq \sqrt{\frac{1}{N}\sum_{i=1}^N (x-y,\xi_i)^2}\sqrt{\frac{1}{N}\sum_{i=1}^N \tanh^2(\beta(\bar{z},\xi_i))} \leq \sqrt{\frac{1}{N}\sum_{i=1}^N (x-y,\xi_i)^2} \end{aligned} \tag{4.53}$$

Even though this bound will not be sufficient to get optimal results (bounding tanh by 1 everywhere is exaggerated when β is close to one), it presents us with the occasion to consider an important object, namely the $M \times M$ random matrix A with elements

$$A_{\mu\nu} \equiv \frac{1}{N}\sum_{i=1}^N \xi_i^\mu \xi_i^\nu \tag{4.54}$$

In terms of this matrix we can of course write

$$\frac{1}{N}\sum_{i=1}^N (z,\xi_i)^2 = (z, Az) \tag{4.55}$$

and thus obtain from (4.53) the bound

$$|\Phi_{\beta,N}(x) - \Phi_{\beta,N}(y)| \leq \|x - y\|_2 \|A\|^{1/2} \tag{4.56}$$

where $\|A\|$ is the operator norm of the matrix A in $\ell_2(\mathbb{R}^M)$.

Random matrices of this form belong to one of the classical ensembles of random matrix theory, the so-called *Marchenko-Pastur matrices* [MP]. They are also well-known in statistics where they appear as *sample covariance matrices*. As a result, their spectral properties, and in particular their norm (coinciding with the maximal eigenvalue), have been widely investigated (a certainly incomplete selection of references is [Ge,Gi,YBK,Si]; some results have been rediscovered or even improved in the course of the investigation of the Hopfield model in [Ko,ST,BGP1,BG3,Nie]). In particular it is known that

Theorem 4.7: [YBK] *Let A be the $M \times M$ random matrix defined in (4.54) with ξ_i^μ i.i.d. random variables with mean zero, variance one, and $\mathbb{E}(\xi_i^\mu)^4 < \infty$. Assume that $\lim_{N\uparrow\infty} \frac{M}{N} = \alpha < \infty$. Then,*

$$\lim_{N\uparrow\infty} \|A\| = (1 + \sqrt{\alpha})^2 \quad a.s. \tag{4.57}$$

In fact, much more precise results are available, but they will not be relevant to us for the moment. In fact, in the remainder of these notes we will simply pretend that A has always norm bounded by $(1+\sqrt{\alpha})^2$, effectively placing ourselves on a subspace of full measure where this is true (for large enough N). It should be noted that the observation that this matrix and the bound on its norm are important goes back to two papers by Shcherbina and Tirozzi [ST] and Koch [Ko] and triggered much of the later progress.

Exercise: Use Lemma 4.6 and (4.56) with $\|A\| \leq (1+\sqrt{\alpha})^2$ to show that the assertion of Proposition 4.5 holds with the supremum taken over $z : \operatorname{dist}(z, \mathcal{M}_{\beta,M(N)}) > C\sqrt{\alpha}|\ln\alpha|$ for any fixed $\beta \geq \beta_0 > 1$.

Since this one-step approach does not yield a result we deem sharp enough, we must use a refined approach known as chaining that consists of introducing a hierarchy of lattices. Let us denote by $\mathcal{W}_{M,r} \equiv (rM^{-1/2}\mathbb{Z})^M$ the hyper-cubic lattice of spacing $rM^{-1/2}$ in $\mathbb{R}^M$. Note that no point in $\mathbb{R}^M$ is farther away from $\mathcal{W}_{M,r}$ than $r/2$. It is not difficult to see that if $B_R(0)$ denotes the ball of radius R centered at the origin, then the number of lattice points in this ball satisfies the bound, for $R > r$,

$$|\mathcal{W}_{\beta,r} \cap B_R(0)| \leq e^{M[\ln(R/r)+2]} \tag{4.58}$$

Now choose a sequence of spacings $r_n = e^{-n}R$ and set $\mathcal{W}(n) \equiv \mathcal{W}_{M,r_n} \cap B_{r_{n-1}}$. For $x \in \mathbb{R}^M$, let $k_n(x) \in \mathcal{W}_{M,r_n}$ be the (in case of non-uniqueness, one of the) closest point(s) to x in $\mathcal{W}_{M,r_n}$. Note that by construction $\|k_n(x) - x\|_2 \leq r_n/2$ and $k_n(x) - k_{n-1}(x) \in \mathcal{W}(n)$. Clearly we have the telescopic expansion

$$\begin{aligned}
\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z) = {} & \Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z) - (\Phi_{\beta,N}(z^*) - \mathbb{E}\Phi_{\beta,N}(z^*)) \\
= {} & \Phi_{\beta,N}(k_0(z)) - \mathbb{E}\Phi_{\beta,N}(k_0(z)) - (\Phi_{\beta,N}(z^*) - \mathbb{E}\Phi_{\beta,N}(z^*)) \\
& + \Phi_{\beta,N}(k_1(z)) - \mathbb{E}\Phi_{\beta,N}(k_1(z)) - (\Phi_{\beta,N}(k_0(z)) - \mathbb{E}\Phi_{\beta,N}(k_0(z))) \\
& + \Phi_{\beta,N}(k_2(z)) - \mathbb{E}\Phi_{\beta,N}(k_2(z)) - (\Phi_{\beta,N}(k_1(z)) - \mathbb{E}\Phi_{\beta,N}(k_1(z))) \\
& + \dots\dots \\
& + \dots\dots \\
& + \Phi_{\beta,N}(k_n(z)) - \mathbb{E}\Phi_{\beta,N}(k_n(z)) - (\Phi_{\beta,N}(k_{n-1}(z)) - \mathbb{E}\Phi_{\beta,N}(k_{n-1}(z))) \\
& + \Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z) - (\Phi_{\beta,N}(k_n(z)) - \mathbb{E}\Phi_{\beta,N}(k_n(z)))
\end{aligned} \tag{4.59}$$

Now let $\delta_\ell > 0$, $\ell = 0, \dots, n$, be a sequence of numbers such that $\sum_{\ell=0}^n \delta_\ell = \delta$. Then the event that $\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z) \geq \delta$ occurs only if, at least for one $1 \leq \ell \leq n-1$,

$$\Phi_{\beta,N}(k_\ell(z)) - \mathbb{E}\Phi_{\beta,N}(k_\ell(z)) - (\Phi_{\beta,N}(k_{\ell-1}(z)) - \mathbb{E}\Phi_{\beta,N}(k_{\ell-1}(z))) \geq \delta_\ell, \tag{4.60}$$

$$\Phi_{\beta,N}(k_0(z)) - \mathbb{E}\Phi_{\beta,N}(k_0(z)) - (\Phi_{\beta,N}(z^*) - \mathbb{E}\Phi_{\beta,N}(z^*)) \geq \delta_0, \tag{4.61}$$

or

$$\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z) - (\Phi_{\beta,N}(k_n(z)) - \mathbb{E}\Phi_{\beta,N}(k_n(z))) \geq \delta_n \tag{4.62}$$

and consequently, the probability of the event in question is smaller than the sum of the probabilities of these $n+1$ events. The probability of the event (4.61) is bounded by Lemma 4.6. Using the uniform bound (4.56), if we choose $\delta_n > e^{-n}R(1+\sqrt{\alpha})$, the event (4.62) is excluded. By (4.44) the probabilities of the events (4.60) can be estimated using (4.47) and the arguments leading to Lemma 4.6. This gives

$$
\begin{aligned}
&\mathbb{P}\left[|\Phi_{\beta,N}(k_\ell(z)) - \mathbb{E}\Phi_{\beta,N}(k_\ell(z)) - (\Phi_{\beta,N}(k_{\ell-1}(z)) - \mathbb{E}\Phi_{\beta,N}(k_{\ell-1}(z)))| \geq \delta_\ell\right] \\
&\quad \leq 2\exp\left(-N\frac{1}{150}\frac{\delta_\ell^2}{r_{\ell-1}^2}\right) \\
&\quad = 2\exp\left(-N\frac{1}{y}150\frac{e^{2(\ell-1)}\delta_\ell^2}{R^2}\right)
\end{aligned}
\tag{4.63}
$$

On the other hand, when z varies over $B_R(z^*)$, the pairs $(k_\ell(z), k_{\ell-1}(z))$ take only a rather small number of values, since their difference lies in $\mathcal{W}(\ell)$. This yields that

$$
\text{Card}\,\{(k_\ell(z), k_{\ell-1}(z))|x \in B_R(z^*)\} \leq |\mathcal{W}_{M,r_{\ell-1}} \cap B_R(z^*)||\mathcal{W}(\ell)| \leq e^{M[\ln(R/r_{\ell-1})+5]}
\tag{4.64}
$$

Putting these observations together we get that

$$
\begin{aligned}
&\mathbb{P}\left[\sup_{z\in B_R(z^*)}|\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z)| \geq \sum_{\ell=0}^{n-1} e^{-n}R(1+\sqrt{\alpha}) + Re^{-n}\right] \\
&\quad \leq 2e^{M[\ln R/r_0+2]}\exp\left(-N\frac{1}{150}\frac{\delta_0^2}{R^2}\right) \\
&\quad + \sum_{\ell=1}^{n-1} e^{M[\ln(R/r_{\ell-1})+5]} 2\exp\left(-N\frac{1}{150}\frac{e^{2(\ell-1)}\delta_\ell^2}{R^2}\right)
\end{aligned}
\tag{4.65}
$$

If we chose $\delta_\ell = C\sqrt{\alpha}e^{-\ell}\sqrt{\ell}$ for some C large enough, then there is a constant $c > 0$, depending only on the choice of C, but not on α or R, such that

$$
\begin{aligned}
&\mathbb{P}\left[\sup_{z\in B_R(z^*)}|\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z)| \geq \sqrt{\alpha}(1+\sqrt{\alpha})RC\sum_{\ell=0}^{n-1} e^{-\ell}\ell^{1/2} + Re^{-n}\right] \\
&\leq 2ne^{-cM}
\end{aligned}
\tag{4.66}
$$

We see that it suffices to chose $n = -\frac{1}{2}\ln\alpha$ to achieve that

$$
\sqrt{\alpha}(1+\sqrt{\alpha})RC\sum_{\ell=0}^{n-1} e^{-\ell}\ell^{1/2} + Re^{-n} \leq C'R\sqrt{\alpha}
\tag{4.67}
$$

with $C' \sim C$ independent of α and R. One can easily see that (4.67) can be improved to show that for C large enough, there exists $c > 0$, s.t. for, say all $R \leq 1$

$$
\mathbb{P}\left[\sup_{z\in B_R(z^*)}|\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z)| \geq \sqrt{\alpha}CR(z)\right] \leq 2ne^{-cM}
\tag{4.68}
$$

and using very similar arguments,

$$\mathbb{P}\left[\sup_{z:\,\mathrm{dist}(z,\mathcal{M}_{\beta,M(N)})\geq 1} |\Phi_{\beta,N}(z) - \mathbb{E}\Phi_{\beta,N}(z)| \geq \sqrt{\alpha}CR(z)\right] \leq 2ne^{-cM} \tag{4.69}$$

We leave it to the reader to check that this implies the assertion of the Proposition for all β s.t. $m^*(\beta) \geq m > 0$; however, so far we have no uniform control on the constants when $m^*(\beta)$ tends to 0 (which happens as $\beta \downarrow 1$). Inspection of our proof shows that the only place where we have exaggerated is when in (4.53) we estimated $\tanh^2(\beta(\bar{z}, \xi_i)) \leq 1$. This estimate becomes poor when $\|\bar{z}\|_2$ tends to zero which precisely becomes relevant when $\beta \downarrow 1$. In this case we have to use $\tanh^2 x \leq x^2$ and replace (4.53) by

$$\begin{aligned} |\Phi_{\beta,N}(x) - \Phi_{\beta,N}(y)| &\leq \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x-y,\xi_i)^2}\sqrt{\beta^2\frac{1}{N}\sum_{i=1}^{N}(\bar{z},\xi_i)^2} \\ &\leq \beta\|A\|\|x-y\|_2 \max(\|x\|_2, \|y\|_2) \end{aligned} \tag{4.70}$$

We leave it to the reader to fill in the details showing that from this we obtain the assertion of the proposition in the case $1 < \beta < \beta_0$. $\diamondsuit$

Proposition 4.5 is a key result that allows immediately to conclude that the measure $\mathcal{Q}_{\beta,N}$ is concentrated on the union of $2M(N)$ disjoint balls of radius $c\sqrt{\alpha}/m^*$, provided $\alpha \leq \gamma(m^*)^2$, for finite positive constants γ and c. This is in perfect agreement with the prediction of the physicists [AGS], and even the scaling of the upper bound on α as $\beta \downarrow 0$ is in accordance with these predictions. Of course, our constants are pretty lousy, as one might expect. We state this result for easy reference as

Theorem 4.8: *There exist $0 < c_0, C, \gamma_a < \infty$ such that for all $\beta > 1$, $\sqrt{\alpha} < \gamma_a(m^*)^2$, and all ρ satisfying $c_0\left(\frac{\sqrt{\alpha}}{m^*} \wedge N^{-1/4}\right) < \rho < m^*/\sqrt{2}$, we have, with probability one, for all but a finite number of indices N,*

$$\mathcal{Q}_{N,\beta}\left(\cup_{\mu=1}^{M} \cup_{s=\pm 1} B_\rho(sm^* e^\mu)\right) \geq 1 - e^{-C(M\wedge N^{1/2})}. \tag{4.71}$$

The same result holds for the measures $\mathbb{Q}_{\beta,N}$.

Remark. A version of Theorem 4.8, with worse bounds on the radii of the balls and on the maximal value of α was first proven in [BGP2] (although the relevant estimates were already contained in [BGP1]). The correct asymptotics near $\beta = 1$ was proven in [BG1]. We have been following widely the version of the arguments given in [BG2]. An alternative proof was also given by Talagrand [T2].

As we have already explained in the analysis of the p-spin SK model, having established a result like Theorem 4.8, two questions remain open: How is the mass distributed over individual "lumps" (here balls), and what are the properties of the measure conditioned on one lump (ball)? In the case of finite M we could completely answer both questions. When M grows, both become much more subtle. As we will soon see, amazingly enough the second question can be answered in full under additional conditions on α and β. Concerning the first question, a full answer has been given only under very strong conditions on the growth rate of M, namely $M^2 \ll N$ in [BM]. The approach used there consisted essentially in pushing the analyses of the finite M case to its limits, employing in the process some very strong Gaussian approximation results. Since these appear to be rather special methods that work in a rather non-canonical regime, I will not include a discussion in these notes. On the other hand, there are some weaker, but very general results concerning these weights based on concentration of measure estimates that I will discuss in the next sub-section.

4.3.2 Logarithmic equivalence of the "lump"-weights.

Theorem 4.8 suggests quite naturally that as in the finite M case, there should be (at least) a pair of pure states corresponding to each pattern and its mirror image. However, on closer inspection one sees that this is somewhat premature. The point is that the theorem says nothing about the mass of any given pattern: it could well be that the mass of some of the balls is exponentially small (in N) and thus there would be no reason to give it preference over any other region in the state space. In particular, if one adopts the *external field* construction of extremal infinite-volume limits of Gibbs states, in such a situation we would not recover a limit state corresponding to such a pattern. This problem has been an obstruction for quite some time. Namely, a straightforward estimation (see [BGP1]) of the relative weights of these balls would only show that they differ by no more than a factor $\exp(O(M))$; this suffices in the case $M(N)/N \downarrow 0$. This allowed to construct the extremal measures under these hypothesis [BGP1], but it remained unclear what would happen if α is strictly positive. This problem was solved in [BGP2] where it was realized that the right tool to address this are *concentration inequalities*. By todays standards, the approach used in [BGP2] was rather clumsy, and due to some new concentration inequalities proven subsequently by Talagrand [T1] (cited as Theorem 2.11 in Section 2), this is now a very simple and standard routine.

Theorem 4.9: *Let ρ be as in Theorem 4.8 Set*

$$I_N^\mu \equiv \frac{1}{\beta N} \ln \int_{B_\rho(e^\mu m^*)} dz e^{-\beta N \Phi_{\beta,N}(z)} \tag{4.72}$$

Then for any $\mu, \nu \leq M(N)$,

$$\mathbb{P}\left[|I_N^\mu - I_N^\nu| \geq x\right] \leq 4 \exp\left(-N \frac{x^2}{128(m^*)^2}\right) \tag{4.73}$$

The same result holds if I_N^μ is replaced by $J_N^\mu \equiv (\beta N)^{-1} \ln \mathbb{Q}_{\beta,N}(B_\rho(e^\mu m^*))$.

Proof. Note that by symmetry $\mathbb{E}|I_N^\mu = \mathbb{E}I_N^\nu$, and so

$$|I_N^\mu - I_N^\mu| \leq= |I_N^\mu - \mathbb{E}I_N^\mu| + |I_N^\nu - \mathbb{E}I_N^\nu| \tag{4.74}$$

Thus

$$\mathbb{P}\left[|I_N^\mu - I_N^\nu| \geq x\right] \leq 2\mathbb{P}\left[|I_N^\mu - \mathbb{E}I_N^\mu| \geq x/2\right] \tag{4.75}$$

We want to use Theorem 2.11 to bound this probability. To do so we must prove a Lipshitz bound on I_N^μ. Note first that using Cauchy–Schwarz in a very similar way as in (4.53), we get that

$$|\Phi_{\beta,N}[\xi](z) - \Phi_{\beta,N}[\xi'](z)| \leq \|\xi - \xi'\|_2 \|z\|_2 \tag{4.76}$$

while

$$\begin{aligned} |I_N^\mu[\xi] - I_N^\mu[\xi']| &= (\beta N)^{-1} \left| \ln \left(\frac{\int_{B_\rho(m^* e^\mu)} dz e^{-\beta N \Phi_{\beta,N}[\xi](z)}}{\int_{B_\rho(m^* e^\mu)} dz e^{-\beta N \Phi_{\beta,N}[\xi'](z)}} \right) \right| \\ =&(\beta N)^{-1} \left| \ln \left(\frac{\int_{B_\rho(m^* e^\mu)} dz e^{-\beta N \Phi_{\beta,N}[\xi'](z) + \beta N [\Phi_{\beta,N}[\xi'](z) - \Phi_{\beta,N}[\xi](z)]}}{\int_{B_\rho(m^* e^\mu)} dz e^{-\beta N \Phi_{\beta,N}[\xi'](z)}} \right) \right| \\ \leq& \sup_{z \in B_\rho(m^* e^\mu)} |\Phi_{\beta,N}[\xi](z) - \Phi_{\beta,N}[\xi]'(z)| \leq \|\xi - \xi'\|_2 (m^* + \rho) \leq 2m^* \|\xi - \xi'\|_2 \end{aligned} \tag{4.77}$$

(4.73) is now straightforward.$\diamondsuit$

It remains to show that this result for the measure $\mathcal{Q}_{N,\beta}$ extends also to the measure $\mathbb{Q}_{\beta,N}$, and therefore to the Gibbs measure itself. But this is quite simple, using the fact that $\mathcal{Q}_{\beta,N}$ is a convolution of $\mathbb{Q}_{\beta,N}$ with a M-dimensional Gaussian measure with mean zero and covariance $\beta N \mathbb{1}$. This allows to bound

$$\mathbb{Q}_{\beta,N}(B_\rho(m^* e^\mu)) \leq \mathcal{Q}_{\beta,N}(B_{\rho+\delta}(m^* e^\mu)) + 2^M e^{-\beta N \delta^2/4} \tag{4.78}$$

that is up to an exponentially small correction, $\mathbb{Q}_{\beta,N}(B_\rho(m^* e^\mu))$ and $\mathcal{Q}_{\beta,N}(B_\rho(m^* e^\mu))$ differ at most by the $\mathcal{Q}_{\beta,N}$ mass of the shell between the radii $\rho-\delta$ and $\rho+\delta$. Choosing δ sufficiently small, and ρ not too small, this has exponentially small mass, and this implies the result for $\mathbb{Q}$. $\diamondsuit$

At this stage a reasonably satisfactory qualitative picture is reached that confirms the heuristic and numerical findings that for small α and not too small β Gibbs measures corresponding to each of the patterns (and their mirror images) exist, implying in some sense that the model, in this regime, does what it was conceived for, namely to "store a number of preselected random patterns".

4.4. The replica symmetric solution.

From the point of view of our general philosophy, having established certain localization properties of the Gibbs measures, we should now ask how the measure conditioned on one ball looked like. One natural approach (although, as it turned out, not the only one) appeared clearly to analyse more carefully the properties of the function $\Phi_{\beta,N}$ in the vicinity of its minima. The aim of such an analysis should should clearly be

(i) To localize more precisely the position of the minima (which so far are only localized in a ball of radius $\sim \sqrt{\alpha}$).

(ii) To determine the value of Φ at the minimum.

(iii) To determine whether or not there is a *unique minimum* in a sufficiently small ball around $m^* e^\mu$.

This analysis was started in the paper [BG1]. Our starting idea there was to extend the use of the Taylor expansion, which had been very useful for finite M beyond its natural realm of applicability. The idea behind this was rather simple: in the finite M case we used Taylor expansions in the arguments of functions like $\frac{1}{N}\sum_i \ln\cosh(\beta(z,\xi_i))$ and we took advantage of the uniform bound $|(z,\xi_i)| \leq \|z\|_1\lambda \leq \sqrt{M}\|z\|_2$. Now this bound is realized essentially when $\xi_i^\mu = \text{sign}\,(z_\mu)$, for all μ. But of course, for given z, unless the ξ_i are very untypical, it is impossible that this holds true for a large number of indices i. Rather, for most values of i, it should be true that $|(z,\xi_i)| \sim \|z\|_2$, leaving room for a Taylor expansion to work even when $M = \alpha N$.

While the main thrust of the paper [BG1] was directed towards answering the question (i), and to determine good bounds on the numerical constants allowing for the existence of local minima near $m^* e^\mu$, the most consequential result proved to be the answer to the third question which could accidentally also be given in a limited domain of α and β values. I will therefore concentrate on reviewing this issue.

4.4.1. Local convexity.

How does one prove that a function has a unique minimum in a region where the existence of a minimum is already established? In the absence of a better idea, prove that the function is convex in this region. This was done in [BG1], where the following result was proved.

Theorem 4.10: *There exist finite positive constants c_1, c_2, γ_a such that if*

(i) $\alpha \leq \gamma_a^2 m^*(\beta)^4$ and

(ii) $\beta > c\alpha^{-1}$, then for ρ as in Theorem 4.8, with probability one, for all but a finite number of values of N, the function $\Phi_{\beta,N}(z)$ is strictly convex on any of the balls $B_\rho(\pm m^* e^\mu)$, and there exists $\epsilon > 0$, s.t. the Hessian matrix $\nabla^2\Phi_{\beta,N}(z)$ has a smallest eigenvalue larger than ϵ for all $z \in B_\rho(\pm m^* e^\mu)$.

Remark. The lower bound (ii) on β may come as a surprise, but we will explain that it is qualitatively optimal, which we will see is a pity.

Proof. Let us consider without loss of generality the the neighbourhood of the point m^*e^1. It will be convenient to set $z = m^*e^1 + v$. We are interested in $\|v\|_2 \leq \rho$. Then we have

$$\nabla^2 \Phi_{\beta,N}(z) = \mathbb{1} - \frac{1}{N}\sum_{i=1}^{N} \frac{\beta}{\cosh^2(\beta(m^*\xi_i^1 + (\xi_i, v)))} \xi_i \xi_i \tag{4.79}$$

It is instructive to first consider the point $v = 0$. Here

$$\nabla^2 \Phi_{\beta,N}(m^*e^1) = \mathbb{1} - \frac{\beta}{\cosh^2(\beta(m^*))} A = \mathbb{1} - \beta(1-(m^*)^2)A \tag{4.80}$$

with A the matrix introduced earlier (4.54). Here we used that $\cosh^{-2} x = 1 - \tanh^2 x$ and $m^* = \tanh \beta m^*$. This matrix is positive if and only of

$$(1+\sqrt{\alpha})^2 \beta(1-(m^*(\beta))^2) < 1 \tag{4.81}$$

Note that with $\alpha = 0$, this is just the condition for the positivity of the second derivative at m^* in the Curie–Weiss model, and thus we know that for all $\beta > 1$, there exists $\alpha_0(\beta)$, such that (4.82) holds for $\alpha < \alpha_0(\beta)$. Moreover, as $\beta \uparrow \infty$, α_0 tends to infinity. Thus, so far we have not seen any sign of condition (ii).

To understand this point, it is best to think of β being large. Then positivity requires that the $\cosh^2$ in the denominators be large to compensate for the β in the numerator. This requires the argument to be roughly of order $\ln \beta$. Now assume that even for a single term in the sum in (4.79) $m^*\xi_i^1 + (\xi_i, v) \sim 0$. Then[35]

$$\nabla^2 \Phi_{\beta,N}(z) \leq \mathbb{1} - \frac{\beta}{N} \xi_i \xi_i \tag{4.82}$$

and since the matrix $\xi_i \xi_i$ has norm M, this cannot be positive definite if $\alpha\beta > 1$. On the other hand, if v has components $v_\mu = \xi_i^\mu \xi_i^1 m^*/M$, and $m^*\xi_i^1 + (\xi_i, v) = 0$, and $\|v\|_2 = m^*/\sqrt{M}$ so that the corresponding point is within the ball $B_\rho(e^1 m^*)$. Thus we see that Condition (ii) is surely necessary. To prove that it is also sufficient,

[35] We use the notation $A > B$ for matrices to mean $A - B$ is positive definite.

we must show that the condition $m^*\xi_i^1 + (\xi_i, v) \sim 0$ cannot be realized for too many indices i at the same time. To make this precise, fix $0 < \tau < 1$ and write

$$\begin{aligned} \nabla^2\Phi_{\beta,N}(z) =&\mathbb{1} - \frac{\beta}{N}\sum_{i=1}^N \xi_i\xi_i + \frac{\beta}{N}\sum_{i=1}^N \tanh^2(\beta(m^*\xi_i^1 + (\xi_i, v)))\xi_i\xi_i \mathbb{1}_{|(\xi_i,v)|\leq\tau} \\ &+ \frac{\beta}{N}\sum_{i=1}^N \tanh^2(\beta(m^*\xi_i^1 + (\xi_i, v)))\xi_i\xi_i \mathbb{1}_{|(\xi_i,v)|>\tau} \end{aligned} \tag{4.83}$$

Using positivity of $\xi_i\xi_i$ this can be bounded by

$$\begin{aligned} \nabla^2\Phi_{\beta,N}(z) \geq& \mathbb{1} - \frac{\beta}{N}\sum_{i=1}^N \xi_i\xi_i + \tanh^2(\beta m^*(1-\tau))\frac{\beta}{N}\sum_{i=1}^N \xi_i\xi_i \\ &+ \frac{\beta}{N}\sum_{i=1}^N \left(\tanh^2(\beta(m^*\xi_i^1 + (\xi_i, v))) - \tanh^2(\beta m^*(1-\tau))\right) \xi_i\xi_i \mathbb{1}_{|(\xi_i,v)|>\tau} \\ &\geq \mathbb{1} - \beta\left[1 - \tanh^2(\beta m^*(1-\tau))\right] A \\ &- \beta \tanh^2(\beta m^*(1-\tau))\frac{1}{N}\sum_{i=1}^N \xi_i\xi_i \mathbb{1}_{|(\xi_i,v)|>\tau} \end{aligned} \tag{4.84}$$

Clearly the only dangerous and difficult term is the last one. We see that, as β grows, it behaves like β times a certain matrix whose norm we therefore need to control. This is not an entirely trivial task, and I am not sure that we have found the most efficient way of dealing with it.

We start from the observation that for any symmetric matrix B, $\|B\| = \sup_{w:\|w\|_2=1}(w, Bw)$. Thus

$$\begin{aligned} &\sup_{v\in B_\rho}\left\| \tfrac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}\xi_i^T\xi_i \right\| \\ &= \sup_{v\in B_\rho}\sup_{w:\|w\|_2=\rho} \tfrac{1}{\rho^2}\tfrac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i, w)^2 \\ &\leq \tfrac{1}{\rho^2}\sup_{v\in B_\rho}\sup_{w\in B_\rho} \tfrac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i, w)^2 \end{aligned} \tag{4.85}$$

It will be convenient to use that

$$
\begin{aligned}
&\mathbb{I}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,w)^2 \\
&\qquad = \mathbb{I}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,w)^2 \left(\mathbb{I}_{\{|(\xi_i,w)|<|(\xi_i,v)|\}} + \mathbb{I}_{\{|(\xi_i,w)|\geq|(\xi_i,v)|\}}\right) \\
&\qquad \leq \mathbb{I}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,v)^2 + \mathbb{I}_{\{|(\xi_i,w)|>\tau m^*\}}(\xi_i,w)^2
\end{aligned} \tag{4.86)}
$$

which allows to bound (4.85). Thus

$$
2\rho^{-2} \sup_{v\in B_\rho} \tfrac{1}{N}\sum_{i=1}^N \mathbb{I}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,v)^2 \equiv 2 \sup_{v\in B_\rho} X_{\tau m^*}(v) \tag{4.87}
$$

Thus our task is to bound the supremum of the quantities X_a which actually can be seen as the partial second moments of the empirical measure of family of random variables (ξ_i, v). This was done in [BG1], and I give here only a sketch of the arguments. As in the analysis of the function $\Phi_{\beta,N}$ we are faced with the problem of controlling the supremum over a continuous family of random variables indexed by a high-dimensional set. Thus we may expect to have to use again the chaining technology. As we already know, this requires exponential estimates on the $X_a(v)$ as well as on differences $X_a(v) - X_a(v')$. Actually this is a rather tricky business, and things will be quite a lot more complicated, and we will be forced to study simultaneously a second object, the empirical distribution function of the same variables (ξ_i, v),

$$
Y_a(v) \equiv \tfrac{1}{N}\sum_{i=1}^N \mathbb{I}_{\{|(\xi_i,v)|>\tau m^*\}} \tag{4.88}
$$

Instead of considering simply the differences $X_a(v) - X_a(v')$, we will use the following lemma.

Lemma 4.11: *Let $a_1, b_1 > 0$, $v, \epsilon \in B_\rho$. Then*

$$
X_{a_1+b_1}(v+\epsilon) \leq X_{a_1}(v) + 2\sqrt{X_{a_1}(v)(\epsilon, A\epsilon)} + 2a_1^2 Y_{b_1}(\epsilon) + 3(\epsilon, A\epsilon) \tag{4.89}
$$

and

$$
Y_{a_1+b_1}(v+\epsilon) \leq Y_{a_1}(v) + Y_{b_1}(\epsilon) \tag{4.90}
$$

Proof. The basic idea behind this lemma is the simple observation that $|(\xi_i, v+\epsilon)| > a_1 + b_1$ can only be true if $|(\xi_i, v)| > a_1$, or if $|(\xi_i, v)| \leq a_1$ and $|(\xi_i, \epsilon)| > b_1$, that is

$$\mathbb{1}_{|(\xi_i,v+\epsilon)|>a_1+b_1} \leq \mathbb{1}_{|(\xi_i,v)|>a_1} + \mathbb{1}_{|(\xi_i,v)|\leq a_1}\mathbb{1}_{|(\xi_i,\epsilon)|>b_1} \tag{4.91}$$

(4.90) is already obvious from this. To get (4.89) we still have to work work with $(\xi_i, v+\epsilon)^2$. Squaring out the sum and using again (4.91) together with the Schwarz inequality then gives the result rather easily. $\diamondsuit$

The second basic ingredient are the exponential bounds on both $X_a(v)$ and $Y_a(v)$.

Lemma 4.12: *Set $p_a \equiv 2\exp(-a^2/2)$. Then for all v with $\|v\| = 1$,*

$$\mathbb{P}[X_a(v) \geq x] \leq \exp\left(N[2p_a^{1/2} - x/4]\right) \tag{4.92}$$

and for $x \geq p_a$,

$$\mathbb{P}[Y_a(v) \geq x] \leq \exp\left(N[(2p_a)^{1/2} - xa^2/4]\right) \tag{4.93}$$

The proof of this lemma can be found in [BG1] (Lemma 4.2). It is, as usual, a somewhat involved application of the exponential Chebyshev inequality. Note that p_a is roughly equal to the mean of $Y_a(v)$. Note also that the corresponding estimates for other values of $\|v\|_2$ follow by scaling since $Y_a(Cv) = Y_{a/C}(v)$ and $X_a(Cv) = C^2 X_{a/C}(v)$.

We now have all the ingredients for the analysis of the supremum together: Choosing a lattice $\mathcal{W}_{M,r_1}$, we can control $X_a(v)$ everywhere in terms of $X_{a_1}(v_1)$ with v_1 on the lattice and the supremum of $Y_{a-a_a}(\epsilon)$ with $\|\epsilon\|_2 \leq r_1$. This latter supremum is then controlled via the usual chaining, using (4.90).

This allows to prove the following estimate:

Proposition 4.13: *There exists finite positive constants C, c such that if*

$$\Gamma(\alpha, a) = Ce^{-ca^2} + C\alpha|\ln\alpha| \tag{4.94}$$

Then for all $\rho > 0$

$$\mathbb{P}\left[\sup_{v\in B_\rho} X_a(v) \geq \rho^2\Gamma(\alpha, a/\rho)\right] \leq Ce^{\alpha N} \tag{4.95}$$

Remark. Proposition 4.8 in [BG1] is a quantitatively more precise version of this statement.

We can now combine this proposition with (4.87) and (4.84) to get immediately a uniform lower bound on the Hessian of Φ:

Lemma 4.14: *Let* $\Gamma(\alpha, a)$ *be as in Proposition 4.13. Then, with probability large than* $1 - Ce^{-\alpha N}$*,*

$$\begin{aligned}&\inf_{z\in B_\rho(m^*e^1)} \nabla^2\Phi_{\beta,N}(z)\\ &\quad\geq 1-\beta[1-\tanh^2(\beta(m^*(1-\tau)))](1+\sqrt{\alpha})^2 - \beta\tanh^2(\beta(m^*(1-\tau)))\Gamma(\alpha, \tau m^*/\rho)\end{aligned} \tag{4.96}$$

If we choose $\rho = c\gamma m^*$ the lower bound is

$$\begin{aligned}&1-\beta[1-\tanh^2(\beta(m^*(1-\tau)))](1+\sqrt{\alpha})^2\\ &\quad -\gamma^2(m^*)^2\beta\tanh^2(\beta(m^*(1-\tau)))\Gamma(\alpha,\tau/\gamma)\\ &\sim 1-\beta[1-\tanh^2(\beta(m^*(1-\tau)))](1+\sqrt{\alpha})^2\\ &\quad - C\beta\tanh^2(\beta(m^*(1-\tau)))(e^{-c/\gamma^2}+\alpha|\ln\alpha|)\end{aligned} \tag{4.97}$$

Thus we see that this bound is strictly positive on a nonempty domain of parameters α and β, which has the shape claimed in the theorem. $\diamondsuit\diamondsuit$

4.4.2. A heuristic derivation of the replica symmetric solution.

Before investigating the consequences of this convexity result, it may be instructive to go through some very heuristic considerations that, however, will illustrate our goal.

To this end we go back to the equations (4.26) determining the location of a minimum of $\Phi_{\beta,N}$ near a point $e^\mu m^*$. To simplify the notation, we consider the case

$\mu = 1$, and we may take without loss of generality $\xi_i^1 \equiv 1$. Let us write this time $z^{(1)} = m_1 + x$ where $x_1 = 0$. Then we get instead of (4.27) the system of equations

$$\begin{aligned} m_1 &= \frac{1}{N}\sum_{i=1}^{N} \tanh(\beta(m_1 + (x, \xi_i))) \\ x_\nu &= \frac{1}{N}\sum_{i=1}^{N} \xi_i^\nu \tanh(\beta(m_1 + (x, \xi_i))) \end{aligned} \tag{4.98}$$

Let us denote by R_N the empirical measure

$$R_N \equiv \frac{1}{N}\sum_{i=1}^{N} \delta_{(x,\xi_i)} \tag{4.99}$$

Then m_1 is only a function of this empirical measure through the equation

$$m_1 = \int R_N(dg) \tanh(\beta(m_1 + g)) \tag{4.100}$$

Now it is not difficult to see that, provided all x_ν tend to zero sufficiently fast as $N \uparrow \infty$, R_N converges to a Gaussian distribution with mean zero and variance $\|x\|_2^2$. Thus if we knew that this convergence held for the (random!) solution of these equations, all we would need to determine was this variance. This should be the rôle of the remaining equations. Naively one might want to simply square both sides and sum over ν, but a bit more care must be taken to disentangle the dependence between the argument of the tanh and the coefficient ξ_i^ν. Thus it will be more useful to write

$$(x, \xi_i) = x_\nu \xi_i^\nu + (x^{(\nu)}, \xi_i) \tag{4.101}$$

and to Taylor expand

$$\begin{aligned} x_\nu =& \frac{1}{N}\sum_{i=1}^{N} \xi_i^\nu \tanh(\beta(m_1 + (x^{(\nu)}, \xi_i))) + \frac{1}{N}\sum_{i=1}^{N} \xi_i^\nu \beta x_\nu \xi_i^\nu \cosh^{-2}(\beta(m_1 + (x^{(\nu)}, \xi_i))) \\ =& x_\nu \beta \frac{1}{N}\sum_{i=1}^{N} \cosh^{-2}(\beta(m_1 + (x^{(\nu)}, \xi_i))) + \frac{1}{N}\sum_{i=1}^{N} \xi_i^\nu \tanh(\beta(m_1 + (x^{(\nu)}, \xi_i))) \end{aligned} \tag{4.102}$$

which can be written, using that $1-\cosh^{-2}(y)=\tanh^2(y)$, as

$$x_\nu\left(1-\beta+\beta\frac{1}{N}\sum_{i=1}^N\tanh^2(\beta(m_1+(x^{(\nu)},\xi_i)))\right)=\frac{1}{N}\sum_{i=1}^N\xi_i^\nu\tanh(\beta(m_1+(x^{(\nu)},\xi_i))) \tag{4.103}$$

Now if we ignore the small difference between (ξ_i,x) and $(\xi_i,x^{(\nu)})$, the coefficient of x_ν is just $1-\beta+\beta\int R_N(dg)\tanh^2(\beta(m_1+g))$. Then squaring and summing over ν gives

$$\begin{aligned}\|x\|_2^2\left(1-\beta+\beta\int R_N(dg)\tanh^2(\beta(m_1+g))\right)^2&=\alpha\int R_N(dg)\tanh^2(\beta(m_1+g))\\&+\frac{1}{N^2}\sum_{i\neq j}\sum_{\nu=2}^M\xi_i^\nu\xi_j^\nu\tanh(\beta(m_1+(x^{(\nu)},\xi_i)))\tanh(\beta(m_1+(x^{(\nu)},\xi_j)))\end{aligned} \tag{4.104}$$

The important point now is that the second term in the numerator on the right has mean zero, by construction. Thus if it was true that in the limit $N\uparrow\infty$, $\|x\|_2^2$ is almost surely constant and equal to its mean value, then this limit must satisfy

$$\|x\|_2^2=\mathbb{E}\|x\|_2^2=\frac{\alpha\int R_\infty(dg)\tanh^2(\beta(m_1+g))}{\left(1-\beta+\beta\int R_\infty(dg)\tanh^2(\beta(m_1+g))\right)^2} \tag{4.105}$$

assuming that R_∞ is a Gaussian distribution with mean zero and variance $\|x\|_2^2\equiv\alpha r$, we see that we arrive at a closed system of equations for the two parameters m_1 and r, namely

$$\begin{aligned}m_1&=\frac{1}{\sqrt{2\pi}}\int dge^{-g^2/2}\tanh(\beta(m_1+\sqrt{\alpha r}))\\r&=\frac{q}{(1-\beta+\beta q)^2},\quad\text{with}\\q&\equiv\frac{1}{\sqrt{2\pi}}\int dge^{-g^2/2}\tanh^2(\beta(m_1+\sqrt{\alpha r}))\end{aligned} \tag{4.106}$$

The solution of this system of equations is known as the "replica-symmetric" solution, first derived by Amit, Gutfreund, and Sompolinsky [AGS] using what is called the "replica trick". Although the derivation of these equations I gave above may look somewhat less striking than the replica method, it is hardly more rigorous, given the numerous ad hoc assumptions we had to make. It is surprising that they can indeed be derived rigorously, as we shall explain in the sequel.

4.4.3. The cavity method 1.

To see how our convexity results are related to the question of the replica symmetric solution, we have to take a step back and look at an approach to the analysis of disordered mean field models that was originally introduced by Parisi et al. (see [M-PV]) as an alternative to the replica method, called the *cavity method.* This method is in principle nothing else than induction, more precisely induction over the volume, N, of the system. The basic idea is simple. Let f_N be any thermodynamic quantity of interest. Suppose we could derive a relation of the form $f_{N+1} = F(f_N) + o(1)$. Then, if f_N converges to limit, this limit must be a fixed point of the map F. Moreover, under certain hypothesis, we may even be able to show that f_N will converge by virtue of this recursion relation. Of course the difficulty will be that one will not be able in general to find such relations, in particular it will not be true that f_{N+1} will be a function of f_N only. However, at least heuristically, i.e. ignoring the problem of proving that the error terms are really $o(1)$, it is indeed possible to obtain such recursions for certain, sufficiently large sets of thermodynamic quantities.

Let us look at this problem in our model. The way to proceed is in fact not quite obvious, but we will to some extend be guided by the preceding heuristic discussion. Note that we are now interested in local properties of the Gibbs measure, that is we want to consider the measure restricted to, say, the ball $B_\rho(m^*e^1)$. We denote by $\mu^{(1,1)}_{\beta,N}$, $\mathcal{Q}^{(1,1)}_{\beta,N}$, etc. the conditioned measures

$$\mu_{\beta,N}\left(\cdot|m_N(\sigma)\in B_\rho(m^*e^1)\right), \quad \text{resp.} \quad \mathcal{Q}_{\beta,N}\left(\cdot|z\in B_\rho(m^*e^1)\right) \tag{4.107}$$

. Consider the Hamiltonian in a volume $N+1$. We may write it as

$$\begin{aligned} H_{N+1}(\sigma) &= -\frac{1}{2(N+1)}\sum_{\mu=1}^{M}\sum_{i,j=1}^{N+1}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j \\ &= -\frac{1}{2(N+1)}\sum_{\mu=1}^{M}\sum_{i,j=1}^{N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j - \frac{1}{N+1}\sigma_{N+1}\sum_{\mu=1}^{M}\sum_{j=1}^{N}\xi_{N+1}^\mu\xi_j^\mu\sigma_j - \frac{M}{2(N+1)} \\ &= \frac{N+1}{N}H_N(\sigma) - \frac{N}{N+1}\sigma_{N+1}(\xi_{N+1}, m_N(\sigma)) - \frac{M}{2(N+1)} \end{aligned} \tag{4.108}$$

It will also be important to note that

$$\|m_{N+1}(\sigma) - m_N(\sigma)\|_2 \leq \frac{1}{N}\|m_N(\sigma)\|_2 + \frac{\sqrt{M}}{N} \tag{4.109}$$

Let us now consider

$$\begin{aligned}\frac{Z_{\beta,N+1}}{Z_{\beta,N}} &= \sum_{\sigma_{N+1}=\pm1} \mathbb{E}_\sigma e^{\beta'\sigma_{N+1}(\xi_{N+1},m_N(\sigma))} e^{-\beta' N N+1 H_N(\sigma)} Z_{\beta',N} \frac{Z_{\beta',N}}{Z_{\beta,N}} e^{\frac{M}{2(N+1)}} \\ &= \sum_{\sigma_{N+1}=\pm1} \mu_{\beta',N}\left(e^{\beta'\sigma_{N+1}(\xi_{N+1},m_N(\sigma))}\right) \frac{Z_{\beta',N}}{Z_{\beta,N}} e^{\frac{M}{2(N+1)}}\end{aligned} \tag{4.110}$$

where $\beta' \equiv \frac{N}{N+1}\beta$. Similarly we get

$$\mu^{(1,1)}_{\beta,N+1}(\sigma_{N+1}) = \frac{\sum_{s=\pm1} s\mu^{(1,1)}_{\beta',N}\left(e^{\beta' s(\xi_{N+1},m_N(\sigma))}\right)}{\sum_{s=\pm1} \mu^{(1,1)}_{\beta',N}\left(e^{\beta' s(\xi_{N+1},m_N(\sigma))}\right)} \tag{4.111}$$

up to a small error (of order $\exp(-\alpha N)$) coming from the fact that the conditioning on the left is on the vector $m_{N+1}(\sigma)$ while on the right it is on $m_N(\sigma)$. But due to (4.109), this difference gives only contributions of the size of the order of the mass of the shell $\rho - \frac{\sqrt{M}}{N} \leq \mu_N(\sigma)\|_2 \leq \rho + \frac{\sqrt{M}}{N}$ which is exponentially small. We will ignore all errors of that order in the sequel.

(4.111) can be easily extended to a formula representing all finite-dimensional marginal distributions. It shows that a central rôle is played by the Laplace transform

$$\begin{aligned}\widetilde{\mathcal{L}}_{\beta,N}(t) &\equiv \mu^{(1,1)}_{\beta',N}\left(e^{\beta' t(\xi_{N+1},m_N(\sigma))}\right) = \mathbb{Q}^{(1,1)}_{\beta',N}\left(e^{\beta' t(\xi_{N+1},m)}\right) \\ &= e^{\alpha/2}\mathcal{Q}^{(1,1)}_{\beta',N}\left(e^{\beta' t(\xi_{N+1},z)}\right) \equiv e^{\alpha/2}\mathcal{L}_{\beta,N}(t)\end{aligned} \tag{4.112}$$

Thus we can rewrite (4.111) as

$$\mu^{(1,1)}_{\beta,N+1}(\sigma_{N+1}) = \frac{\sum_{s=\pm1} s\, \mathcal{Q}^{(1,1)}_{\beta',N}\left(e^{\beta' s(\xi_{N+1},z)}\right)}{\sum_{s=\pm1} \mathcal{Q}^{(1,1)}_{\beta',N}\left(e^{\beta' s(\xi_{N+1},z)}\right)} \tag{4.113}$$

Being able to replace the measure $\mathbb{Q}$ by $\mathcal{Q}$ will actually be very useful (although it shouldn't). Formula (4.111) appeared in the work of Pastur et al. [PS,PST] where it was realized that being able to control the Laplace transform $\mathcal{L}_{\beta,N}$ was a key step to getting the replica symmetric solution. This was later pushed by Talagrand towards full rigour. Let us see why this is the case.

We compute first the mean of $\mathcal{L}_{\beta,N}(t)$ with respect to the variables ξ_{N+1}. This is possible since $\mathcal{Q}_{\beta,N}$ is independent of them. Of course we get simply

$$\mathbb{E}_{\xi_{N+1}} \mathcal{L}_{\beta,N}(t) = \mathcal{Q}_{\beta,N}\left(e^{\sum_\mu \ln\cosh(\beta t z_\mu)}\right) \tag{4.114}$$

Now this would make sense if $\mathcal{L}$ could then be shown to be self-averaging. But this is not the case. In a way one can see that taking this average throws out too much information. One thing one can try then is to extract first a random part, and try to show that what is left is self-averaging. A natural possibility is to center the variable z in the exponent. Let Z denote the r.v. whose distribution is $\mathcal{Q}^{(1,1)}_{\beta,N}$, and write $\bar{Z} \equiv Z - \mathcal{Q}^{(1,1)}_{\beta,N}(Z)$; actually, at least at this point we start to feel that our notations are getting too heavy. Let us therefore denote henceforth by $\mathbb{E}_{\mathcal{Q}}$ the expectation with respect to the random measure $\mathcal{Q}^{(1,1)}_{\beta,N}$. Then we can of course write

$$\mathcal{L}_{\beta,N}(t) = e^{t(\xi_{N+1},\mathbb{E}_{\mathcal{Q}} Z)}\mathbb{E}_{\mathcal{Q}}\, e^{t(\xi_{N+1},\bar{Z})} \tag{4.115}$$

As in (4.114) we get of course that

$$\mathbb{E}\mathbb{E}_{\mathcal{Q}}\, e^{t(\xi_{N+1},\bar{Z})} = \mathbb{E}_{\mathcal{Q}}\left(e^{\sum_\mu \ln\cosh(\beta t \bar{Z}_\mu)}\right) \tag{4.116}$$

Now, if all $\bar{Z}_\mu$ are small, $\ln\cosh(\beta t\bar{Z}_\mu) \sim \frac{1}{2}\beta^2 t^2 \|\bar{Z}\|_2^2$. Assume for a moment that the distribution of $\bar{Z}$ was an M-dimensional Gaussian distribution with variance of order $1/N$. Then the length of $\bar{Z}$ would be sharply concentrated about its mean, since its distribution would have a density proportional to $\exp(-N(r^2/2 - \ln r))$. Thus in such a case we would get that

$$\mathbb{E}\mathbb{E}_{\mathcal{Q}}\, e^{t(\xi_{N+1},\bar{Z})} = e^{\frac{1}{2}\beta^2 t^2 \mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2 + o(1)} \tag{4.117}$$

4.4.4. Brascamp-Lieb inequalities.

Of course, $\mathcal{Q}_{\beta,N}^{(1,1)}$ is not a Gaussian distribution, but maybe it is sufficiently similar to one so that we still get (4.117)? Indeed, we the local convexity proven in Theorem 4.10 does imply (4.117) as well as a number of similar results that we will need in the sequel. Interestingly, the proof of this fact uses a rather sophisticated result form functional analysis, the so-called Brascamp-Lieb inequalities [BL], and I doubt that an elementary proof (e.g. as in the Gaussian case) could be given. Unfortunately, while convexity gives a very elegant way of advancing here, it is known not to be necessary, neither for the Brascamp-Lieb inequalities, nor for the replica symmetric solution to be correct. The proof I will present here does therefore not give the best available conditions. In principal there are two ways to get to better the results: (i) to give a proof that does not use Brascamp-Lieb inequalities (and (4.117)). This was Talagrand's original approach [T2]; in its original version, this gave however conditions that were comparable to those under which convexity holds; improved conditions, closer to those expected by physicists, required substantially more work [T5] than the already very difficult [T2][36]. (ii) There has been considerable work done to establish Brascamp-Lieb inequalities without the assumption of convexity [BJS]. In fact the real conditions for the B-L inequalities to hold concern the minimal eigenvalue of a certain matrix-differential operator. [BJS] have establish criteria that allow to bound this operator from below even even when convexity fails, but so far nobody has been able to show that they are useful in our situation. This remains an interesting question to study.

Let us now state the Brascamp-Lieb inequalities in their original form.

Lemma 4.15: *Let $V : \mathbb{R}^M \to \mathbb{R}$ be nonnegative and strictly convex with $Hess(V(x)) \geq \epsilon$. Denote by $\mathbb{E}_V$ the expectation with respect to the probability measure $e^{-NV(x)}dx / \int e^{-NV(y)}dy$. Let $f : \mathbb{R}^M \to \mathbb{R}$ be any continuously differentiable function that is square integrable w.r.t. $\mathbb{E}_V$. Then*

$$\mathbb{E}_V (f - \mathbb{E}_f)^2 \leq \frac{1}{\epsilon N} \mathbb{E}_V \|\nabla f\|_2^2 \tag{4.118}$$

[36] Recently, M. Talagrand has, however, told me, that a new, simplified approach will soon appear.

Remark. It is not difficult to see that the result holds also, up to an exponentially small error term, if V is an extended convex function, i.e. it is a convex function on its domain D and equal to $+\infty$ outside of D. See e.g. [BG3] for details. In our situation this is what we will actually have to use.

Remark. The condition $\text{Hess}(V(x))$ can be replaced by the condition that the smallest eigenvalue of the matrix-differential operator

$$(-\Delta + \nabla V \cdot \nabla) \otimes \mathbb{1} + \text{Hess}(V(x)) \tag{4.119}$$

is larger than ϵ. It is easy to see that the first two terms are a positive operator, so that if the Hessian matrix is strictly positive, this gives an immediate bound. However, as pointed out above, this condition is not necessary.

The following simple applications of Lemma 4.15 show how we will actually use these inequalities (see [BG2]).

Corollary 4.16: *Let $\mathbb{E}_V$ be as in Lemma 4.15. Then*

(i) $\mathbb{E}_V \, \|x - \mathbb{E}_V x\|_2^2 \leq \frac{M}{\epsilon N}$

(ii) $\mathbb{E}_V \, \|x - \mathbb{E}_V x\|_4^4 \leq 4 \frac{M}{\epsilon^2 N^2}$

(iii) if any function f such that $V_t(x) \equiv V(x) - tf(x)/N$ for $t \in [0,1]$ is still strictly convex and Hess$(\nabla^2 V_t) \geq \epsilon' > 0$, then

$$0 \leq \ln \mathbb{E}_V e^f - \mathbb{E}_V f \leq \frac{1}{2\epsilon' N} \sup_{t \in [0,1]} \mathbb{E}_{V_t} \|\nabla f\|_2^2 \tag{4.120}$$

In particular

(iii) $\ln \mathbb{E}_V e^{(t,(x - \mathbb{E}_V x))} \leq \frac{\|t\|_2^2}{2\epsilon N}$

(iv) $\ln \mathbb{E}_V e^{\|x - \mathbb{E}_V x\|_2^2} - \mathbb{E}_V \, \|x - \mathbb{E}_V x\|_2^2 \leq \frac{M}{\epsilon^2 N^2}$

The point of these relations is that the measure $\mathbb{E}_V$ behaves with regard to its covariance structure essentially like a Gaussian measure. A first important consequence of this corollary is (4.117) that follows easily from these estimates and the bound $x^2/2 - x^4/4 \leq \ln\cosh(x) \leq x^2/2$ (Exercise!). Moreover, the same tool allows to estimate the variance of $\mathbb{E}_Q e^{t(\xi_{N+1},\bar{Z})}$, namely

$$\mathbb{E}\left(\mathbb{E}_Q e^{t(\xi_{N+1},\bar{Z})} - \mathbb{E}\mathbb{E}_Q e^{t(\xi_{N+1},\bar{Z})}\right)^2 \leq \frac{C}{N} \tag{4.121}$$

4.4.5. The local mean field.

These results combine to the following following important observation.

Lemma 4.17: *Whenever the conclusions of Theorem 4.10 hold, there exists a constant $C < \infty$ such that*

$$\ln \mathcal{L}_{\beta,N}(t) = \beta t(\xi_{N+1}, \mathbb{E}_{\mathbb{Q}}(z)) + \frac{t^2\beta^2}{2}\mathbb{E}_Q\|\bar{Z}\|_2^2 + R_N \tag{4.122}$$

where

$$\mathbb{E}R_N^2 \leq \frac{C}{N} \tag{4.123}$$

Remark. Note that this lemma can be seen as a statement about the distribution of the random field (ξ_{N+1}, Z) under the Gibbs measure, stating that it is asymptotically Gaussian with mean $(\xi_{N+1}, \mathbb{E}_{\mathbb{Q}}(Z))$ and variance $\mathbb{E}_Q\|\bar{Z}\|_2^2$. Its mean is still a random variable that we will now have to investigate.

From our previous results we certainly expect that the vector $\mathbb{E}_Q z$ has one component (the first one) of order m^*, while all other components should be 'microscopic', i.e. tend to zero as $N \uparrow \infty$. Thus we write

$$(\xi_{N+1}, \mathbb{E}_{\mathbb{Q}}(Z)) = \xi_{N+1}^1 \mathbb{E}_{\mathbb{Q}}(Z_1) + (\hat{\xi}_{N+1}, \mathbb{E}_{\mathbb{Q}}(\hat{Z})) \tag{4.124}$$

where $\hat{Z} = 0$ and $\hat{Z}_\nu = Z_\nu$ for $\nu \neq 1$. It is now very natural to expect that the second term in (4.124), being a sum of αN independent random variables (under the conditional distribution given $\xi_1, \dots, \xi_N$), will converge in distribution to a *Gaussian random variable* with mean zero and variance $\|\mathbb{E}_Q\hat{Z}\|_2^2$. If, moreover, as one might also expect, the quantity $\|\mathbb{E}_Q\hat{Z}\|_2^2$ converges to a constant almost surely, as $N \uparrow \infty$, this second term would in fact converge in distribution to the same Gaussian also unconditionally. In that case the entire Laplace transform $\mathcal{L}_{\beta,N}(t)$ would be fully characterized in terms of the three constants $m_1(N) \equiv \mathbb{E}_{\mathbb{Q}}(Z_1), U_N \equiv \|\mathbb{E}_{\mathcal{Q}}\hat{Z}\|_2^2$, and $T_N \equiv \mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2$.

Thus we are left with three problems to solve: (1) Show that the central limit theorem alluded to holds[37]. (2) Show that the three quantities mentioned above are *self-averaging*. (3) Proof that these converge and characterize their limits. Technically, both (1) and (2) will rely essentially on concentration of measure estimates. Problem (3) will then be solved by the cavity method, i.e. we will derive a system of recursive equations that can be proven to have a unique stable fix-point in the domain where these quantities are a priori located[38].

We will immediately formulate a somewhat more general version of this central limit theorem (which we will actually need to construct the metastate).

Proposition 4.18: *Let $I \subset \mathbb{N}\backslash\{1,\dots,N\}$ be finite, independent of N. For $i \in I$, set $X_i(N) \equiv \frac{1}{\sqrt{T_N}}\sum_{\mu=2}\xi_i^\mu \mathbb{E}_{\mathcal{Q}} Z_\mu$. Whenever the conclusions of Theorem 4.10 hold, either this family converges to a family of i.i.d. standard normal random variables, or $\sqrt{T_N}X_i(N)$ converges to zero in probability.*

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

$$\sum_{\mu=2}^{M}[\mathbb{E}\mathbb{E}_{\mathcal{Q}} Z_\mu]^2 \leq \mathbb{E}\sum_{\mu=2}^{M}[\mathbb{E}_{\mathcal{Q}} Z_\mu]^2 \leq \rho^2 \tag{4.125}$$

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

To use information on the mean values, we will need a concentration estimate

[37] This fact is assumed in [PST] without proof. It is however rather delicate and requires concentration estimates that were not available at that time.

[38] This approach is in principle contained in [PST]; Talagrand gave the first fully rigorous version, without using the a priori estimates furnished by the Brascamp-Lieb inequalities, making the entire inductive scheme even more complicated.

for derivatives of self-averaging quantities (since all expectations w.r.t. $\mathbb{E}_{\mathcal{Q}}$ can be represented as derivatives of some log-Laplace transforms).

Lemma 4.19: *Assume that $f(x)$ is a random function defined on some open neighbourhood $U \subset \mathbb{R}$. Assume that f satisfies for all $x \in U$ that for all $0 \leq r \leq 1$,*

$$\mathbb{P}\left[|f(x) - \mathbb{E}f(x)| > r\right] \leq c \exp\left(-\frac{Nr^2}{c}\right) \tag{4.126}$$

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

$$\mathbb{P}\left[|f'(x) - \mathbb{E}f'(x)| > \delta N^{-\zeta/2}\right] \leq \frac{32C^2}{\delta^2} N^{\zeta} \exp\left(-\frac{\delta^4 N^{1-2\zeta}}{256c}\right) + p \tag{4.127}$$

The rather elementary proof of this lemma can be found in [BG2] or [BG3].

We will now use Lemma 4.19 to control $\mathbb{E}_{\mathcal{Q}} Z_\mu$. We define

$$f(x) = \frac{1}{\beta N} \ln \int_{B_\rho^{(1,1)}} d^M z e^{\beta N x z_\mu} e^{-\beta N \Phi_{\beta,N,M}(z)} \tag{4.128}$$

and denote by $\mathbb{E}_{\mathcal{Q},x}$ the corresponding modified expectation. By exactly the same arguments as in the proof of Theorem 4.9, $f(x)$ verifies (4.126). Moreover, $f'(x) = \mathbb{E}_{\mathcal{Q},x} Z_\mu$ and

$$f''(x) = \beta N \mathbb{E}_{\mathcal{Q},x} \left(Z_\mu - \mathbb{E}_{\mathcal{Q},x} Z_\mu\right)^2 \tag{4.129}$$

Amazingly enough, it is again the Brascamp-Lieb inequalities that allow us to bound this second derivative:

$$\mathbb{E}_{\mathcal{Q},x} \left(Z_\mu - \mathbb{E}_{\mathcal{Q},x} Z_\mu\right)^2 \leq \frac{1}{\epsilon N \beta} \tag{4.130}$$

and so $f''(x) \leq c = \frac{1}{\epsilon}$.

Thus we arrive at

Corollary 4.20: *There are finite positive constants c, C such that, for any $0 < \zeta \leq \frac{1}{2}$, for any μ,*

$$\mathbb{P}\left[|\mathbb{E}_{\mathcal{Q}} Z_\mu - \mathbb{E}\mathbb{E}_{\mathcal{Q}} Z_\mu| \geq N^{-\zeta/2}\right] \leq C N^{\zeta} \exp\left(-\frac{N^{1-2\zeta}}{c}\right) \tag{4.131}$$

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, $|\mathbb{E}_{\mathcal{Q}} Z_\mu - \mathbb{E}\mathbb{E}_{\mathcal{Q}} Z_\mu| \leq N^{-1/8}$. Then $\mathbb{P}[\Omega_N^c] \leq O\left(e^{-N^{1/2}}\right)$.

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

$$\begin{aligned}\mathbb{E}\prod_{j\in I} e^{it_j X_j(N)} &= \mathbb{E}_{\xi_{I^c}}\left[\mathbb{1}_{\Omega_N}\mathbb{E}_{\xi_I} e^{i\sum_{j\in I} t_j X_j(N)} + \mathbb{1}_{\Omega_N^c}\mathbb{E}_{\xi_I} e^{i\sum_{j\in I} t_j X_j(N)}\right] \\ &= \mathbb{E}_{\xi_{I^c}}\left[\mathbb{1}_{\Omega_N}\prod_{\mu\geq 2}\prod_{j\in I}\cos\left(\frac{t_j}{\sqrt{T_N}}\mathbb{E}_{\mathcal{Q}} Z_\mu\right)\right] + O\left(e^{-N^{1/2}}\right)\end{aligned} \tag{4.132}$$

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

$$\sum_{\mu=2}^{M}(\mathbb{E}_{\mathcal{Q}} Z_\mu)^4 \leq N^{-1/4}\sum_{\mu=2}^{M}(\mathbb{E}_{\mathcal{Q}} Z_\mu)^2 \leq N^{-1/4}T_N \tag{4.133}$$

Moreover, for any finite t_j, for N large enough, $\left|\frac{t_j}{\sqrt{T_N}}\mathbb{E}_{\mathcal{Q}} Z_\mu\right| \leq 1$. Thus, using that $|\ln\cos x + x^2/2| \leq cx^4$ for $|x| \leq 1$, and that

$$\begin{aligned}&\mathbb{E}_{\xi_{I^c}}\mathbb{1}_{\Omega_N}\mathbb{E}_\eta e^{i\sum_{j\in I} t_j X_j(N)} \\ &\leq e^{-\sum_{j\in I} t_j^2/2}\sup_{\Omega_N}\left[\prod_{j\in I}\exp\left(c\frac{t_j^4 N^{-1/4}}{T_N}\right)\right]\mathbb{P}_\xi(\Omega_N)\end{aligned} \tag{4.134}$$

Clearly, the right hand side converges to $e^{-\sum_{j\in I} t_j^2/2}$, provided only that $N^{1/4}T_N \uparrow \infty$. Otherwise, $\mathbb{E}T_N X_i(N)^2 =\downarrow 0$. Thus the lemma is proven. $\diamondsuit$

4.4.6. Gibbs measures and metastates.

We now control the convergence of our Laplace transform except for the two parameters $m_1(N) \equiv \mathbb{E}_{\mathcal{Q}} Z_1$ and $T_N \equiv \sum_{\mu=2}^{M}\left[\mathbb{E}_{\mathcal{Q}} Z_\mu\right]^2$. 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).

To this end we first note that all of the preceding discussion may be carried out without substantial changes for the Laplace transforms of the local mean fields acting on a finite family of singled out spins $\sigma_i, i \in I \subset \mathbb{N}$ (the details of the computations can be found in [BG3]). As a result one obtains the following expression for the Gibbs mass of cylinder events:

$$\mu^{(1,1)}_{\Lambda,\beta,\rho}[\omega]\left(\{\sigma_I = s_I\}\right) = \frac{e^{\beta'_N \sum_{i\in I} s_i\left[m_1(N)\xi_i^1 + X_i(N)\sqrt{T_N}\right] + R_N(s_I)}}{2^I \mathbb{E}_{\sigma_I} e^{\beta'_N \sum_{i\in I} \sigma_i\left[m_1(N)\xi_i^1 + X_i(N)\sqrt{T_N}\right] + R_N(\sigma_I)}} \tag{4.135}$$

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 $\{g_i\}_{i\in\mathbb{N}}$ is a family of i.i.d. standard Gaussian random variables. Note that the first three of these assertions are proven, while the last two are for the moment *assumed.* From this representation we obtain:

Proposition 4.21: *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 $I \subset \mathbb{N}$*

$$\mu^{(1,1)}_{\Lambda,\beta,\rho}\left(\{\sigma_I = s_I\}\right) \to \prod_{i\in I} \frac{e^{\beta s_i\left[m_1\bar{\xi}_i^1 + g_i\sqrt{\alpha r}\right]}}{2\cosh\left(\beta\sigma_i\left[m_1\bar{\xi}_i^1 + g_i\sqrt{\alpha r}\right]\right)} \tag{4.136}$$

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

Rademacher random variables, independent of the g_i and having the same distribution as the variables ξ_i^1.

To arrive at the convergence in law of the random Gibbs measures, it is enough to show that (4.136) holds jointly for any finite family of cylinder sets, $\{\sigma_i = s_i, \forall_{i\in I_k}\}, I_k \subset \mathbb{N}$, $k = 1, \dots, \ell$ (C.f. [Ka], Theorem 4.2). But this is easily seen to hold from the same arguments. Therefore, denoting by $\mu_{\infty,\beta}^{(1,1)}$ the random measure

$$\mu_{\infty,\beta}^{(1,1)}[\omega](\sigma) \equiv \prod_{i\in\mathbb{N}} \frac{e^{\beta\sigma_i[m_1\xi_i^1[\omega]+\sqrt{\alpha r}g_i[\omega]]}}{2\cosh\left(\beta[m_1\xi_i^1[\omega]+\sqrt{\alpha r}g_i[\omega]]\right)} \tag{4.137}$$

we have

Theorem 4.22: *Under the assumptions of Proposition 4.21, and with the same notation,*

$$\mu_{\Lambda,\beta,\rho}^{(1,1)} \to \mu_{\infty,\beta}^{(1,1)}, \quad \textit{in law, as } \Lambda \uparrow \infty \quad , \tag{4.138}$$

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 4.23: *Let $\kappa_\beta(\cdot)[\omega]$ denote the Aizenman-Wehr metastate. Under the hypothesis of Proposition 4.22, for almost all ω, for any continuous function $F : \mathbb{R}^k \to \mathbb{R}$, and cylinder functions f_i on $\{-1,1\}^{I_i}$, $i = 1, \dots, k$, one has*

$$\begin{aligned}
&\int_{\mathcal{M}_1(\mathcal{S}_\infty)} \kappa_\beta(d\mu)[\omega] F\left(\mu(f_1), \dots, \mu(f_k)\right) \\
&= \int \prod_{i\in I} d\mathcal{N}(g_i) F\Bigg(\mathbb{E}_{s_{I_1}} f_i(s_{I_1}) \prod_{i\in I_1} \frac{e^{\beta\left[\sqrt{\alpha r}g_i + m_1\xi_i^1[\omega]\right]}}{2\cosh\left(\sqrt{\alpha r}g_i + m_1\xi_i^1[\omega]\right)}, \dots \\
&\qquad \dots, \mathbb{E}_{s_{I_k}} f_k(s_{I_k}) \prod_{i\in I_k} \frac{e^{\beta\left[\sqrt{\alpha r}g_i + m_1\xi_i^1[\omega]\right]}}{2\cosh\left(\sqrt{\alpha r}g_i + m_1\xi_i^1[\omega]\right)} \Bigg)
\end{aligned} \tag{4.139)}$$

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

Proof. This theorem is proven just as Theorem 4.22, except that the "almost sure version" of the central limit theorem, Proposition 4.18, 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.

Some remarks concerning the implications of this proposition are in place. First, it shows (modulo a small argument that can be found in [BG3]) 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).

4.4.7. The cavity method 2.

We now conclude our analysis by showing that the quantities $U_N \equiv \mathbb{E}_{\mathcal{Q}} \|\bar{Z}\|_2^2$, $m_1(N) \equiv \mathbb{E}_{\mathcal{Q}} Z_1$ and $T_N \equiv \sum_{\mu=2}^M [\mathbb{E}_{\mathcal{Q}} Z_\mu]^2$ 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. The basic ideas of this section are otherwise due to Pastur, Shcherbina, and Tirozzi [PST], and Talagrand [T2].

We first need some more concentration of measure results.

Proposition 4.24: *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 $0 < \zeta \leq \frac{1}{2}$,*

$$\mathbb{P}\left[|A_N - \mathbb{E}A_N| \geq N^{-\zeta/2}\right] \leq CN^{\zeta} \exp\left(-\frac{N^{1-2\zeta}}{c}\right) \tag{4.140)}$$

Proof. The proofs of these three statements are all very similar to that of Corollary 4.20. Indeed, for $m_1(N)$, (4.140) 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 (4.128). They are

$$g(x) \equiv \frac{1}{\beta N} \ln \mathbb{E}_Q^{\otimes 2} e^{\beta N x(\bar{Z}, \bar{Z}')} \tag{4.141}$$

in the case of T_N and

$$\tilde{g}(x) \equiv \frac{1}{\beta N} \ln \mathbb{E}_Q^{\otimes 2} e^{\beta N x \|\bar{Z}\|_2^2} \tag{4.142}$$

The proof then proceeds as in that of Corollary 4.20. $\diamondsuit$

We now turn to the induction part of the proof and derive a recursion relation for the three quantities above. To simplify notation we will whenceforth set $\eta \equiv \xi_{N+1}$. Let us define

$$u_N(\tau) \equiv \ln \mathbb{E}_Q e^{\beta\tau(\eta, Z)} \tag{4.143}$$

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

Lemma 4.25: *Under our general assumptions*

(i) $\frac{1}{\beta\sqrt{T_N}} \frac{d}{d\tau}(v_N(\tau) - \beta\tau\eta^1 \mathbb{E}_Q Z_1)$ *converges weakly to a standard Gaussian random variable.*

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

Proof. (i) is obvious from Proposition 4.19 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* $\mathrm{var}\,(w_N(\tau)) \leq \frac{C}{\sqrt{N}}$, then 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 4.19 (see e.g. [T6], Proposition 5.4). On the other hand, $|\mathbb{E}w_N(\tau) - \frac{\tau^2\beta^2}{2}\mathbb{E}\mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2| \leq \frac{K}{\sqrt{N}}$, by Lemma 5.3, which, together with the boundedness of the second derivative of $w_N(\tau)$ implies that $|\frac{d}{d\tau}\mathbb{E}w_N(\tau) - \tau\beta^2\mathbb{E}\mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2| \downarrow 0$. This means that var $(w_N(\tau)) \leq \frac{C}{\sqrt{N}}$ implies the lemma. Since we already know from (4.123) that $\mathbb{E}R_N^2 \leq \frac{C}{N}$, it is enough to prove var $\left(\mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2\right) \leq \frac{C}{\sqrt{N}}$. 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, it is true that $|U_N - \mathbb{E}U_N| \leq N^{-1/4}$, $|T_N - \mathbb{E}T_N| \leq N^{-1/4}$, etc. This subspace has probability one by our estimates.

Let us note that $\mathbb{E}_{\mathcal{Q}}Z_\mu$ and $\int d\mathcal{Q}^{(1,1)}_{N,\beta,\rho}(m)m_\mu$ differ only by an exponentially small term. Thus

$$\mathbb{E}_{\mathcal{Q}}Z_\mu = \frac{1}{N}\sum_{i=1}\xi_i^\mu \int \mu^{(1,1)}_{N,\beta,\rho}(d\sigma)\sigma_i + O\left(e^{-cM}\right) \tag{4.144}$$

Since we want to perform induction over N, we will have to add an index referring to the volume to the mesures $\mathcal{Q}$. Note that by symmetry, from (4.144) we get

$$\mathbb{E}\mathbb{E}_{\mathcal{Q}_{N+1}}(Z_\mu) = \mathbb{E}\eta^\mu \int \mu^{(1,1)}_{N+1,\beta,\rho}(d\sigma)\sigma_{N+1} + O\left(e^{-cM}\right) \tag{4.145}$$

Using (4.113) and the definition of u_N, this gives

$$\mathbb{E}\mathbb{E}_{\mathcal{Q}_{N+1}}(Z_\mu) = \mathbb{E}\eta^\mu \frac{e^{u_N(1)} - e^{u_N(-1)}}{e^{u_N(1)} + e^{u_N(-1)}} + O\left(e^{-cM}\right) \tag{4.146}$$

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 4.17 this gives

$$\mathbb{E}\mathbb{E}_{\mathcal{Q}_{N+1}}(Z_\mu) = \mathbb{E}\eta^\mu \tanh\left(\beta(\eta^1\mathbb{E}m_1(N) + \sqrt{\mathbb{E}T_N}X_{N+1}(N))\right) + O(N^{-1/4}) \tag{4.147}$$

Using further Proposition 4.18 we get a first recursion for $m_1(N)$:

$$m_1(N+1) = \int d\mathcal{N}(g)\tanh\left(\beta(\mathbb{E}m_1(N) + \sqrt{\mathbb{E}T_N}g)\right) + o(1) \tag{4.148}$$

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. To simplify the notation, we ignore all the $o(1)$ error terms and put them back in the end only. Also, the remarks concerning β and α made above apply throughout.

Note that $T_N = \|\mathbb{E}_{\mathcal{Q}} Z\|_2^2 - (\mathbb{E}_{\mathcal{Q}} Z_1)^2$ and

$$\begin{aligned}\mathbb{E}\|\mathbb{E}_{\mathcal{Q}_{N+1}} Z\|_2^2 &= \sum_{\mu=1}^{M} \mathbb{E}\left(\frac{1}{N+1}\sum_{i=1}^{N+1}\xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i)\right)^2 \\ &= \frac{M}{N+1}\mathbb{E}\left(\mu_{\beta,N+1,M}^{(1,1)}(\sigma_{N+1})\right)^2 \\ &+ \sum_{\mu=1}^{M} \mathbb{E}\xi_{N+1}^\mu \mu_{\beta,N+1,M}^{(1,1)}(\sigma_{N+1})\left(\frac{1}{N+1}\sum_{i=1}^{N}\xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i)\right)\end{aligned} \tag{4.149}$$

Using (4.113) as in the step leading to (4.146), we get for the first term in (4.149)

$$\mathbb{E}\left(\mu_{\beta,N+1,M}^{(1,1)}(\sigma_{N+1})\right)^2 = \mathbb{E}\tanh^2\left(\beta(\eta_1\mathbb{E}_{\mathcal{Q}} Z_1 + \sqrt{\mathbb{E}T_N X_{N+1}(N)})\right) \equiv \mathbb{E}Q_N \tag{4.150}$$

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

$$\begin{aligned}\sum_{\mu=1}^{M}\xi_{N+1}^\mu\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i)\right) &= \frac{\sum_{\sigma_{N+1}} \mathbb{E}_{\mathcal{Q}}\,(\xi_{N+1}, Z) e^{\beta\sigma_{N+1}(\xi_{N+1},Z)}}{\sum_{\sigma_{N+1}} \mathbb{E}_{\mathcal{Q}}\, e^{\beta\sigma_{N+1}(\xi_{N+1},Z)}} \\ =\beta^{-1}\frac{\sum_{\tau=\pm1} u_N{}'(\tau) e^{u_N(\tau)}}{\sum_{\tau=\pm1} e^{u_N(\tau)}}\end{aligned} \tag{4.151}$$

Together with Lemma 4.25 one concludes that in law up to small errors

$$\begin{aligned}\sum_{\mu=1}^{M}\xi_{N+1}^\mu\left(\frac{1}{N+1}\sum_{i=1}^{N}\xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i)\right) &= \xi_{N+1}^1\mathbb{E}_{\mathcal{Q}} Z_1 + \sqrt{\mathbb{E}T_N}X_N \\ + \beta\mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2\tanh\beta\left(\xi_{N+1}^1\mathbb{E}_{\mathcal{Q}} Z_1 + \sqrt{\mathbb{E}T_N}X_N\right)\end{aligned} \tag{4.152}$$

and so

$$\begin{aligned}\mathbb{E}\|\mathbb{E}_{\mathcal{Q}_{N+1}} Z\|_2^2 = \alpha\mathbb{E}Q_N + \mathbb{E}\Bigg[&\tanh\beta\left(\xi_{N+1}^1\mathbb{E}_{\mathcal{Q}} Z_1 + \sqrt{\mathbb{E}T_N}X_N\right) \\ &\times\left[\xi_{N+1}^1\mathbb{E}_{\mathcal{Q}} Z_1 + \sqrt{\mathbb{E}T_N}X_N\right]\Bigg] \\ &+ \beta\mathbb{E}\mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2\tanh^2\beta\left(\xi_{N+1}^1\mathbb{E}_{\mathcal{Q}} Z_1 + \sqrt{\mathbb{E}T_N}X_N\right)\end{aligned} \tag{4.153}$$

Using the self-averaging properties of $\mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2$, the last term is of course essentially equal to

$$\beta\mathbb{E}\mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2\mathbb{E}Q_N \tag{4.154}$$

The reappearance of $\mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2$ (remember that this was the variance of the local mean field!) may seem 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

$$\begin{aligned}\mathbb{E}\mathbb{E}_{\mathcal{Q}_{N+1}}\|Z\|_2^2 =& \alpha + \mathbb{E}\Bigg[\tanh\beta\left(\xi^1_{N+1}\mathbb{E}_{\mathcal{Q}}Z_1 + \sqrt{\mathbb{E}T_N}X_N\right)\\ &\times\left[\xi^1_{N+1}\mathbb{E}_{\mathcal{Q}}Z_1 + \sqrt{\mathbb{E}T_N}X_N\right]\Bigg] + \beta\mathbb{E}\mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2\mathbb{E}Q_N\end{aligned} \tag{4.155}$$

so that setting $U_N \equiv \mathbb{E}_{\mathcal{Q}}\|\bar{Z}\|_2^2$, we get, subtracting (4.153) from (4.155), the simple recursion

$$\mathbb{E}U_{N+1} = \alpha(1-\mathbb{E}Q_N) + \beta(1-\mathbb{E}Q_N)\mathbb{E}U_N \tag{4.156}$$

From this we get (since all quantities considered are self-averaging, we drop the $\mathbb{E}$ to simplify the notation), setting $m_1(N) \equiv \mathbb{E}_{\mathcal{Q}}Z_1$,

$$\begin{aligned}T_{N+1} &= -(m_1(N+1))^2 + \alpha Q_N + \beta U_N Q_N\\ &+ \int d\mathcal{N}(g)[m_1(N) + \sqrt{T_N}g]\tanh\beta(m_1(N) + \sqrt{T_N}g)\\ &= m_1(N+1)(m_1(N) - m_1(N+1)) + \beta U_N Q_N + \beta T_N(1-Q_N) + \alpha Q_N\end{aligned} \tag{4.157}$$

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

$$\begin{aligned}m_1(N+1) &= \int d\mathcal{N}(g)\tanh\beta\left(m_1(N) + \sqrt{T_N}g\right) + O(N^{-1/4})\\ T_{N+1} &= m_1(N-1)(m_1(N) - m_1(N+1)) + \beta U_N Q_N + \beta T_N(1-Q_N) + \alpha Q_N\\ &\quad + O(N^{-1/4})\\ U_{N+1} &= \alpha(1-Q_N) + \beta(1-Q_N)U_N + O(N^{-1/4})\\ Q_{N+1} &= \int d\mathcal{N}(g)\tanh^2\beta\left(m_1(N) + \sqrt{T_N}g\right) + O(N^{-1/4})\end{aligned} \tag{4.158}$$

If the solutions to this system of equations converge, 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

$$m_1 = \int d\mathcal{N}(g) \tanh(\beta(m_1 + \sqrt{\alpha r} g)) \tag{4.159}$$

$$q = \int d\mathcal{N}(g) \tanh^2(\beta(m_1 + \sqrt{\alpha r} g)) \tag{4.160}$$

$$r = \frac{q}{(1 - \beta + \beta q)^2} \tag{4.161}$$

which are the equations for the replica symmetric solution of the Hopfield model found by Amit et al. [AGS], and also through our heuristic speculations in Section 4.4.2!

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 [T2]. 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 functions F_p such that

$$X_{n+1}(p_{n+1}) = F_{p_n}(X_n(p_n)) + O(n^{-1/4}) \tag{4.162}$$

In this setting, we have the following lemma.

Lemma 4.26: *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*

$$\lim_{n\uparrow\infty} X_n(p) = X^*(p) \tag{4.163}$$

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

$$\lim_{n\uparrow\infty} X_{n+k(n)}(p) = X^*(p) \tag{4.164}$$

To see this, note first that $|p_{n+k(n)}-p_n|\leq\frac{k(n)}{n}$. By (4.162), we have that using the Lipshitz properties of F

$$X_{n+k(n)}(p) = F_p^{k(n)}(X_n(p_n)) + O(n^{-1/4}) \tag{4.165}$$

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 (4.165) implies (4.164). But (4.164) 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 (4.157). 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 to 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 fixed points of (4.157), 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, we need to verify the local convexity in the corresponding domains since this was used in the derivation of the equations (4.157).

General References.

[BP] A. Bovier and P. Picco (Eds.). Mathematical aspects of spin glasses and neural networks. Progress in Probability, Birkhäuser, Boston, 1997.

[Ge1] H.-O. Georgii. Gibbs measures and phase transitions. de Gruyter Studies in Mathematics, 9. Walter de Gruyter & Co., Berlin-New York, 1988.

[F] J. Fröhlich. Mathematical aspects of the physics of disordered systems. With the collaboration of A. Bovier and U. Glaus. Phénomènes critiques, systèmes alátoires, théories de jauge, Part I, II (Les Houches, 1984), 725–893, North-Holland, Amsterdam, 1986.

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

[Pr1] Ch.J. Preston. Gibbs states on countable sets. Cambridge Tracts in Mathematics, No. 68. Cambridge University Press, London-New York, 1974.

[Pr2] Ch.J. Preston. Random fields. Lecture Notes in Mathematics, Vol. 534. Springer-Verlag, Berlin-New York, 1976.

[Ru1] D. Ruelle. Statistical mechanics: Rigorous results. W. A. Benjamin, Inc., New York-Amsterdam 1969.

[Ru2] D. Ruelle. Thermodynamic formalism. The mathematical structures of classical equilibrium statistical mechanics. Encyclopedia of Mathematics and its Applications, 5. Addison-Wesley Publishing Co., Reading, Mass., 1978.

[Sim] B. Simon. The statistical mechanics of lattice gases. Vol. I. Princeton Series in Physics. Princeton University Press, Princeton, NJ, 1993.

[Sin] Ya. G. Sinaĭ. Theory of phase transitions: rigorous results. International Series in Natural Philosophy, 108. Pergamon Press, Oxford-Elmsford, N.Y., 1982.

[T2000] M. Talagrand. Mean field models for spin glasses: a first course. Lectures at the Saint Flour summer school on probability theory, 2000.

References for Section 1

[BD] L.A. Bassalygo, R.L. Dobrushin. Uniqueness of a Gibbs field with a random potential—an elementary approach. Theory Probab. Appl. 31 (1987), 572–589.

[vdBM] J. van den Berg, C. Maes. Disagreement percolation in the study of Markov fields. Ann. Probab. 22 (1994), 749-763.

[Cu] P. Curie, Ann. de Chimie et Physique 5 (1985), 289.

[Do1] R.L. Dobrushin. Description of a random field by means of conditional probabilities and conditions for its regularity. Teor. Verojastnost i Primenen 13 (1968), 201-229 (Theor. Probab. Appl. 13 (1968) 197-224.

Definition of a system of random variables by means of conditional distributions. Teor. Verojastnost i Primenen 15 (1970), 469-497.

[Do2] R.L. Dobrushin. Existence of a phase transition in the two dimensional Ising model. Soviet Phys. Dokl. 10 (1965), 111-113.

Existence of a phase transition in the three dimensional Ising model. Theor. Probab. Appl. 10 (1965), 193-213.

[DS] R.L. Dobrushin, S.B. Shlosman. "Non-Gibbsian" states and their Gibbs description. Comm. Math. Phys. 200 (1999), 125–179.

[vEFS] A.C.D. van Enter, R. Fernández, A. Sokal. Regularity properties and pathologies of position-space renormalization-group transformations: scope and limitations of Gibbsian theory. J. Statist. Phys. 72 (1993), 879-1167.

[Ge1] H.-O. Georgii. Gibbs measures and phase transitions. de Gruyter Studies in Mathematics, 9. Walter de Gruyter & Co., Berlin-New York, 1988.

[Gri] R. B. Griffiths. Peierls proof of spontaneous magnetization in a two-dimensional Ising ferromagnet. Phys. Rev. 136 (1964) A437-A439.

[Gu] Y.M. Guttmann. The concept of probability in statistical physics. Cambridge University Press, Cambridge, 1999.

[Is] Ernst Ising. Beitrag zur Theorie des Ferro- und Paramagnetismus. Doktorarbeit, Hamburg, 1924; Zeitschrift. f. Physik 31 (1925), 253.

[Isr] R.B. Israel. Convexity in the theory of lattice gases. Princeton Series in Physics. With an introduction by Arthur S. Wightman. Princeton University Press, Princeton, N.J., 1979.

[Ko] S. Kobe. Ernst Ising - physicist and teacher. J. Statist. Phys. 88 (1997) 991-995.

[LR] O. Lanford, D. Ruelle. Observables at infinity and states with short range correlations in statistical mechanics. Comm. Math. Phys. 13 1969 194-215.

[MMR] C. Maes, F. Redig, A. van Moffaert. Almost Gibbsian versus weakly Gibbsian measures. Stochastic Process. Appl. 79 (1999), 1–15. g, F.;

[Pei] R. Peierls. Proc. Cambridge Philos. Soc. 32 (1936), 477-481.

[Pf] Ch. Pfister. Thermodynamical aspects of classical lattice systems. Lectures at the 4-th Brazilian school on probability theory, Mambucaba, 2000.

[Pr1] Ch. J. Preston. Gibbs states on countable sets. Cambridge Tracts in Mathematics, No. 68. Cambridge University Press, London-New York, 1974.

[Pr2] Ch.J. Preston. Random fields. Lecture Notes in Mathematics, Vol. 534. Springer-Verlag, Berlin-New York, 1976.

[PS] S. A. Pirogov, Ya. G. Sinaĭ. Phase diagrams of classical lattice systems. Teoret. Mat. Fiz. 25 (1975), 358-369.

[Ru1] D. Ruelle. Statistical mechanics: Rigorous results. W. A. Benjamin, Inc., New York-Amsterdam 1969.

[Ru2] D. Ruelle. Thermodynamic formalism. The mathematical structures of classical equilibrium statistical mechanics. Encyclopedia of Mathematics and its Applications, 5. Addison-Wesley Publishing Co., Reading, Mass., 1978.

[Sim] B. Simon. The statistical mechanics of lattice gases. Vol. I. Princeton Series in Physics. Princeton University Press, Princeton, NJ, 1993.

[Ue1] D. Ueltschi. Discontinuous phase transitions in quantum lattice systems. Ph. D. thesis, EPFL-Lausanne 1998.

[We] P. Weiss. J. de Physique 6 (1907), 661.

[Za1] M. Zahradník. An alternate version of Pirogov-Sinaĭ theory. Comm. Math. Phys. 93 (1984), 559-581.

[Za2] M. Zahradník. A short course on the Pirogov-Sinaĭ theory. Rend. Mat. Appl. 18 (1998), 411-486.

References for Section 2

[ACCN] M. Aizenman, J.T. Chayes, L. Chayes, C.M. Newman. The phase boundary in dilute and random Ising and Potts ferromagnets. J. Phys. A 20 (1987), L313-L318.

[AW1] M. Aizenman, J. Wehr. Rounding effects of quenched randomness on first-order phase transitions. Comm. Math. Phys. 130 (1990), no. 3, 489-528.

[AW2] J. Wehr, M. Aizenman. Fluctuations of extensive functions of quenched random couplings. J. Statist. Phys. 60 (1990), 287-306.

[ARS] J.E. Avron, G. Roepstorff, L.S. Schulman. Ground state degeneracy and ferromagnetism in a spin glass. J. Statist. Phys. 26 (1981), 25-36.

[vdBM] J. van den Berg, C. Maes. Disagreement percolation in the study of Markov fields. Ann. Probab. 22 (1994), 749-763.

[Be] A. Berretti. Some properties of random Ising models. J. Statist. Phys. 38 (1985), 483-496.

[BK] J. Bricmont, A. Kupiainen. Phase transition in the 3d random field Ising model. Comm. Math. Phys. 116 (1988), 539-572.

[BoKu1] A. Bovier, Ch. Külske. A rigorous renormalization group method for interfaces in random media. Rev. Math. Phys. 6 (1994), 413-496.

[BoKu2] A. Bovier, Ch. Külske. Stability of hierarchical interfaces in a random field model. J. Statist. Phys. 69 (1992), 79-110.

[BoKu3] A. Bovier, Ch. Külske. Hierarchical interfaces in random media. II. The Gibbs measures. J. Statist. Phys. 73 (1993), 253-266.

[BoKu4] A. Bovier, Ch. Külske. There are no nice interfaces in $(2+1)$-dimensional SOS models in random media. J. Statist. Phys. 83 (1996), 751-759.

[BRZ] P.M. Bleher, J. Ruiz, V.A. Zagrebnov. One-dimensional random-field Ising model: Gibbs states and structure of ground states. J. Statist. Phys. 84 (1996), 1077-1093.

[Ch] J. Chalker. J. Phys. C 16 (1983), 6615

[vEG] A.C.D. van Enter, R.B. Griffiths. The order parameter in a spin glass. Comm. Math. Phys. 90 (1983), 319–327.

[FKG] C.M. Fortuin, P.W. Kasteleyn, J. Ginibre. Correlation inequalities on some partially ordered sets. Comm. Math. Phys. 22 (1971), 89-103.

[FFS] D.S.Fisher, J. Fröhlich, Th. Spencer. The Ising model in a random magnetic field. J. Statist. Phys. 34 (1984), 863-870.

[FZ1] J. Fröhlich, B. Zegarliński. Spin glasses and other lattice systems with long range interactions. Comm. Math. Phys. 120 (1989), 665-688.

[FZ2] J. Fröhlich, B. Zegarliński. The high-temperature phase of long-range spin glasses. Comm. Math. Phys. 110 (1987), 121-155.

[Ge1] H.-O. Georgii. Gibbs measures and phase transitions. de Gruyter Studies in Mathematics, 9. Walter de Gruyter & Co., Berlin-New York, 1988.

[Ge2] H.-O. Georgii. Spontaneous magnetization of randomly dilute ferromagnets. J. Statist. Phys. 25 (1981), 369-396

[Gi] G. Gielis. Spin systems with random interactions: the Griffiths' regime. Ph.D. thesis, Leuven 1998.

[dHK] W.Th.F. den Hollander, M. Keane. Inequalities of FKG type. Physica 138A (1986), 167-182.

[IM] Y. Imry. S. Ma. Random-field instability of the ordered state of continuous symmetry, Phys. Rev. Lett. 35 (1975), 1399-1401.

[I] J.Z. Imbrie. The ground state of the three-dimensional random-field Ising model. Comm. Math. Phys. 98 (1985), 145-176.

[KM] A. Klein, S. Masooman. Taming Griffiths' singularities in long range random Ising models. Comm. Math. Phys. 189 (1997), 497-512.

[LML] J.L. Lebowitz, A. Martin-Löf. On the uniqueness of the equilibrium state for Ising spin systems. Comm. Math. Phys. 25 (1972), 276-282.

[NS1] C.M. Newman, D.L. Stein. Metastate approach to thermodynamic chaos. Phys. Rev. E (3) 55 (1997), 5194-5211.

[NS2] C.M. Newman, D.L. Stein. Thermodynamic chaos and the structure of short-range spin glasses. Mathematical aspects of spin glasses and neural networks, 243-287, Progr. Probab., 41, Birkhäuser Boston, Boston, MA, 1998.

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

[Pr1] Ch.J. Preston. Gibbs states on countable sets. Cambridge Tracts in Mathematics, No. 68. Cambridge University Press, London-New York, 1974

[Ru3] D. Ruelle. On the use of "small external fields" in the problem of symmetry breakdown in statistical mechanics. Ann. Phys. 69 (1972), 364-374.

[T1] M. Talagrand. A new look at independence. Ann. Probab. 24 (1996), 1-34.

[Z] B. Zegarliński. Random spin systems with long-range interactions. Mathematical aspects of spin glasses and neural networks, 289-320, Progr. Probab., 41, Birkhäuser Boston, Boston, MA, 1998.

References for Section 3

[A] R. J. Adler. On excursion sets, tube formulas and maxima of random fields. Ann. Appl. Probab. 10 (2000), 1-74.

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

[ALR] M. Aizenman, J.L. Lebowitz, D. Ruelle Some rigorous results on Sherrington-Kirkpatrick spin glass model. Commun. Math. Phys. 112 (1987), 3-20.

[BKL] A. Bovier, I. Kurkova, M. Löwe. The fluctuations of the free energy in the REM and the p-spin SK models, WIAS-preprint (2000), to appear in Ann. Probab.

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

[C1] F. Comets. A spherical bound for the Sherrington-Kirkpatrick model. Astérisque 236 (1996), 103-108.

[CN] F. Comets, J. Neveu. The Sherrington-Kirkpatrick Model of Spin Glasses and Stochastic Calculus: The High Temperature Case. Commun Math. Phys. 166 (1995), 549-564.

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

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

[DW] T.C. Dorlas, J.R. Wedagedera. Large deviations and the random energy model. Int. J. Mod. Phys. B 15 (2001), 1-15.

[E] R.S. Ellis. Entropy, large deviations, and statistical mechanics. Grundlehren der Mathematischen Wissenschaften 271. Springer-Verlag, New York-Berlin, 1985.

[Ei] Th. Eisele. Theodor On a third-order phase transition. Comm. Math. Phys. 90 (1983), 125-159.

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

[GM] D.J. Gross, M. Mézard. The simplest spin glass. Nucl. Phys. B 240 (1984), 431-452.

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

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

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

[LT] M. Ledoux, M. Talagrand. Probability in Banach spaces. Springer, Berlin-Heidelberg-New York, 1991.

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

[OP] E.Olivieri, P. Picco. On the existence of thermodynamics for the random energy model. Comm. Math. Phys. 96 (1984), 125-144.

[Ru4] D. Ruelle. A mathematical reformulation of Derrida's REM and GREM. Math. Phys 108 (1987), 225-239 .

[Re] S.I. Resnick. Extreme values, regular variation, and point processes. Springer-Verlag, New York-Berlin, 1987.

[Shi] A.N. Shiryaev. Probability theory. Springer Verlag, New-York-Berlin, 1984.

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

[T0] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. I.H.E.S. Publ. Math. No. 81 (1995), 73-205.

[T2] M. Talagrand. Rigorous results for the Hopfield model with many patterns. Probab. Theory Related Fields 110 (1998), 177-276.

[T3] M. Talagrand. Rigorous low-temperature results for the mean field p-spins interaction model. Probab. Theory Related Fields 117 (2000), 303-360.

[T4] M. Talagrand. On the p-spin interaction model at low temperature. C. R. Acad. Sci. Paris Sér. I Math. 331 (2000), 939-942.

References for Section 4

[A] D.J. Amit. Modelling brain function, Cambridge University Press, Cambridge (1989).

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

[BJS] V. Bach, T. Jecko, J. Sjöstrand. Correlation asymptotics of classical lattice spin systems with nonconvex Hamilton function at low temperature. Ann. Henri Poincaré 1 (2000), 59-100.

[BvEN] A. Bovier, A.C.D. van Enter, B. Niederhauser. Stochastic symmetry-breaking in a Gaussian Hopfield model. J. Statist. Phys. 95 (1999), 181-213.

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

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

[BG3] A. Bovier, V. Gayrard. Metastates in the Hopfield model in the replica symmetric regime. Math. Phys. Anal. Geom. 1 (1998), 107-144.

[BG4] A. Bovier, V. Gayrard. An almost sure large deviation principle for the Hopfield model. Ann. Probab 24 (1996), 1444-1475.

[BG5] A. Bovier, V. Gayrard. An almost sure central limit theorem for the Hopfield model. Markov Proc. Rel. Fields 3(1997), 151-174.

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

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

[BM] A. Bovier, D. Mason. Extreme value behaviour in the Hopfield model. to appear in Ann. Appl. Probab. (2001).

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

[Co] F. Comets. Large deviation estimates for a conditional probability distribution. Applications to random Gibbs measures. Probab. Theor. Rel. Fields 80 (1989), 407-432.

[CD] F. Comets and A. Dembo. Ordered overlaps in disordered mean-field models. preprint 1999.

[vEvH] A.C.D.van Enter J.L.van Hemmen. Chopper model for pattern recognition. Phys. Rev. A 34 (1986), 2509-2512.

[FP1] L.A. Pastur, A.L. Figotin. Exactly soluble model of a spin glass. Sov. J. Low Temp. Phys. 3(6) (1977), 378-383.

[FP2] L.A. Pastur, A.L. Figotin. On the theory of disordered spin systems. Theor. Math. Phys. 35 (1978), 403-414.

[FP3] L.A. Pastur, A.L. Figotin. Infinite range limit for a class of disordered spin systems. Theor. Math. Phys. 51 (1982), 564-569.

[Ga] V. Gayrard. Thermodynamic limit of the q-state Potts-Hopfield model with infinitely many patterns. J. Statist. Phys. 68 (1992), 977-1011.

[Ge] S. Geman. A limit theorem for the norms of random matrices. Ann. Probab. 8 (1980), 252-261.

[Gen1] B. Gentz. A central limit theorem for the overlap in the Hopfield model. Ann. Probab.

24 (1996), 1809-1841.

[Gen2] B. Gentz. An almost sure central limit theorem for the overlap parameters in the Hopfield model. Stochastic Process. Appl. 62 (1996), 243–262.

[GL1] B. Gentz, M. Löwe. The fluctuations of the overlap in the Hopfield model with finitely many patterns at the critical temperature. Probab. Theory Related Fields 115 (1999), 357–381.

[GL2] B. Gentz, M. Löwe. Fluctuations in the Hopfield model at the critical temperature. Markov Process. Related Fields 5 (1999), 423-449.

[Gi] V.L. Girko. Limit theorems for maximal and minimal eigenvalues of random matrices. Theor. Probab. Appl. 35 (1989), 680-695.

[H] B. Helffer. Recent results and open problems on Schrödinger operators, Laplace integrals, and transfer operators in large dimension. In Schrödinger operators, Markov semigroups, wavelet analysis, operator algebras, 11-162, Math. Top., 11, Akademie Verlag, Berlin, 1996.

[He] D. Hebb. The organisation of behaviour: a neurophysiologiocal theory. Wiley, New York (1949).

[vH1] J.L. van Hemmen. Equilibrium theory of spin-glasses: mean-field theory and beyond. In Heidelberg Colloquium on Spin Glasses, eds. J.L. van Hemmen and I.Morgenstern, 203-233 (1983), LNP 192 Springer, Berlin-Heidelberg-New York, 1983.

[vH2] J.L. van Hemmen. Spin glass models of a neural network. Phys. Rev. A 34 (1986), 3435-3445.

[vHGHK] J.L. van Hemmen, D. Grensing, A. Huber, R. Kühn. Elementary solution of classical spin-glass models. Z. Phys. B 65 (1986), 53-63.

[vHvEC] J.L. van Hemmen, A.C.D. van Enter, J. Canisius. On a classical spin-glass model. Z. Phys. B 50 (1983), 311-336.

[HH] P. Hall, 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 (1982), 2554-2558.

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

[Hu] J. Hubbard. Calculation of partition functions. Phys. Rev. Lett. 3 (1959), 77-78.

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

[Ku1] Ch. Külske. Metastates in disordered mean field models: random field and Hopfield models. J. Statist. Phys. 88 (1997), 1257-1293.

[Ku2] Ch. Külske. Metastates in disordered mean-field models. II. The superstates. J. Statist. Phys. 91 (1998), 155-176.

[Ko] H. Koch. A free energy bound for the Hopfield model. J. Phys. A 26 (1993), L353-L355.

[KP] H. Koch, J. Piasko. Some rigorous results on the Hopfield neural network model. J. Statist. Phys. 55 (1989), 903-928.

[KoPa] J. Komlos, R. Paturi. Convergence results in a autoassociative memory model. Neural Networks 1(1988), 239-250.

[Lou] D.Loukianova. Lower bounds on the restitution error in the Hopfield model. Probab.

Theory Related Fields 107 (1997), 161-176.

[LT] M. Ledoux, M. Talagrand. Probability in Banach spaces. Springer, Berlin-Heidelberg-New York (1991).

[Ma] D.C. Mattis. Solvable spin system with random interactions. Phys. Lett. 56A (1976), 421-422 .

[MP] A.S. McCulloch, W. Pitts. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5 (1943), 115-133.

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

[N2] Ch.M. Newman. Memory capacity in neural network models: Rigorous lower bounds. Neural Networks 1(1988), 223-238.

[Nie] B. Niederhauser. On the p-spin Hopfield model. Ph.D. thesis, TU-Berlin, 2000.

[P] D. Petritis. Thermodynamic formalism of neural computing. Dynamics of complex interacting systems (Santiago, 1994), 81-146, Nonlinear Phenom. Complex Systems, 2, Kluwer Acad. Publ., Dordrecht, 1996.

[PS] L. Pastur, M. Shcherbina. Absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model. J. Stat. Phys. 62 (1991), 1-19.

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

[Si] J. Silverstein. Eigenvalues and eigenvectors of large sample covariance matrices. Contemporary Math. 50 (1986), 153-159.

[St] R.L. Stratonovich. On a method of calculating quantum distribution functions. Doklady Akad. Nauk S.S.S.R. 115(1957), 1097 [translation: Soviet Phys. Doklady 2 (1958), 416-419]

[ST] M. Shcherbina, B. Tirozzi. The free energy of a class of Hopfield models. J. Statist. Phys. 72 (1993), 113-125.

[T1] M. Talagrand. A new look at independence. Ann. Probab. 24 (1996), 1-34.

[T2] M. Talagrand. Rigorous results for the Hopfield model with many patterns. Probab. Theory Related Fields 110 (1998), 177-276.

[T5] M. Talagrand. Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model. Ann. Probab., to appear (2001).

[T6] M. Talagrand. The Sherrington-Kirkpatrick model: A challenge for mathematicians. Probab. Theory Related Fields 110 (1998), 109-176.

[YBK] Y.Q. Yin, Z.D. Bai, P.R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. Probab. Theor. Rel. Fields 781 (1988), 509-521.