Journal of Statistical Physics, Vol. 112, Nos. 5/6, September 2003 (© 2003)

Analogues of Non-Gibbsianness in Joint Measures of Disordered Mean Field Models

Christof Külske[1]

Received December 19, 2002; accepted March 11, 2003

It is known that the joint measures on the product of spin-space and disorder space are very often non-Gibbsian measures, for lattice systems with quenched disorder, at low temperature. Are there reflections of this non-Gibbsianness in the corresponding mean-field models? We study the continuity properties of the conditional probabilities in finite volume of the following mean field models:

behavior of these functions to discover non-trivial limiting objects. For (a) we find (1) discontinuous dependence for almost any realization and (2) dependence of the conditional probabilities on the phase. In contrast to that we see continuous behavior for (b) and (c), for almost any realization. This is in complete analogy to the behavior of the corresponding lattice models in high dimensions. It shows that non-Gibbsian behavior which seems a genuine lattice phenomenon can be partially understood already on the level of mean-field models.

KEY WORDS: Disordered systems; non-Gibbsian measures; mean field models; Morita-approach; random field model; decimation transformation; diluted ferromagnet.

1. INTRODUCTION

The relationship between mean field models and lattice models is an interesting meta-theme in statistical mechanics. The general wisdom is of course that (a) there should be mean-field like behavior in sufficiently high dimensions, as far as the phase-structure is concerned and (b) mean field models are often amenable to simple computations and explicit solutions.

[1] Weierstraß-Institut für Angewandte Analysis und Stochastik, Mohrenstraße 39, D-10117 Berlin, Germany; e-mail: kuelske@wias-berlin.de

0022-4715/03/0900-1079/0 © 2003 Plenum Publishing Corporation

In the limit of high dimensions the free energy of a spin system converges to its mean field value, and critical behaviour is supposed to be mean-field like. Moreover, mean field models sometimes possess independent applications outside of solid state physics, and deserve to be studied in their own right. The reader might think in this context also, e.g., on bond-percolation on the lattice vs. the random graph.[1] Simple mean-field models of disordered systems were also used earlier by the author to illustrate the (supposed) asymptotic large volume behavior of the Gibbs measures of the corresponding lattice models. This behavior was described by the corresponding metastates,[2, 3] a notion introduced in refs. 4 and 5. For an excellent overview about further related results also in more complicated situations we refer the readers to ref. 6 and the references therein.

On the other hand, there are cases (think of the famous Edward–Anderson spin-glass), where there are good reasons to question the equivalence between lattice and mean-field behavior in high dimensions. Also the corresponding mean field-solution[7] itself is mathematically not justified and not at all simple. (For the current status of the ongoing discussion on this fascinating topic see refs. 8 and 9.) Moreover, when conceptually subtle properties are investigated, it might not be straightforward or even prove impossible to translate well-defined questions on lattice models into questions on mean field models.

Our present paper is motivated by the study of non-Gibbsianness in lattice spin-systems. We pick three well-known models showing non-Gibbsian behavior of different character. Then we compare their lattice versions to their mean-field versions. As we will see there are close analogies in the behavior in these models, and in fact the mean field models do allow for very simple explicit computations. We believe that our examples contribute to a more intuitive understanding of some aspects of non-Gibbsianness and are also interesting in itself. It might seem surprising that we are looking for non-Gibbsianness in mean-field models. After all, there is no proper Gibbsian structure anyway. Our important point is the following: When "Non-Gibbsianness" for mean-field models is understood as "discontinuity of conditional probabilities as a function of the conditioning," it becomes a meaningful and natural notion. Of course the notion of continuity has to be taken in the appropriate sense. Indeed, the Gibbs measures of simple mean-field models (like the standard Curie–Weiss model) usually converge weakly to linear combinations of product measures. A (non-trivial) linear combination of product measures is non-Gibbsian and has each spin configuration as a discontinuity point.[10] So one could feel discouraged to look for non-trivial continuity properties in conditional probabilities of mean-field models. In contrast to that, the proceeding that is appropriate for our mean-field model is as follows: (1) Take the conditioning while staying in *finite volume*. (2) Observe that the conditional probabilities outside a finite set are automatically volume-dependent functions of the empirical average over all the (joint) spins in the conditioning. (3) Derive the large volume-asymptotics for these functions. (4) Consider their continuity properties. Look at the size of the set of their discontinuity points in the large volume-limit.

We stress that the functional dependence of the limiting form of the conditional probabilities can not be deduced from the sole information of the limiting product measures. This is particularly clear for the decimation transformation of the Curie–Weiss model: It has the same limiting measures as the Curie–Weiss model itself. However it has non-trivial continuity properties of its conditional probabilities in the above sense. This will become explicit later.

Our most interesting example however is the joint measure of the random field Ising lattice model. We will treat it more detail than the other two examples. The study of such joint measures arising in spin-systems with frozen disorder on the product space of disorder space was advocated a long time ago in the so-called Morita-approach to disordered systems[11] (see also refs. 12–14). Much later, starting from the first example of the dilute Ising ferromagnet,[15] rigorous investigations of these measures were performed. It was discovered in refs. 16 and 17 that such joint measures provide a whole class of examples of non-Gibbsian measures, in low temperature situations. (See also refs. 18 and 19.) This situation has some analogy with the much discussed non-Gibbsian behavior of images of low-temperature lattice spin measures under renormalization group transformations. (See refs. 20–22 and references therein.) The simplest example of such a transformation is just the projection to a sublattice, or decimation. The analogy is close in the region of interactions when the joint measure happens to be a Gibbs measure again (which is true for small enough interactions). Then one has regularity statements (uniqueness and Lipschitz-continuity, see ref. 23) that are parallel to the known statements for renormalized measures.[20] On the other hand, even if there is no interaction that is summable everywhere, by general arguments there always exists a potential that is at least summable *almost everywhere*.[17] However there is no a priori information on the decay. This abstract result does not have a counterpart for renormalized Gibbs measures.

Now, the joint measures of the random field Ising lattice model in dimensions greater or equal than 3, small randomness, provide a particularly nice example for various unusual "pathologies."

1. They are not Gibbs measures for any uniformly summable potential. Moreover, the set of configurations where the discontinuity of their

conditional probabilities happens is not negligible; it has even full measure (refs. 16 and 17).

2. The functional form of the conditional probabilities depends on whether the system is in the plus- or minus-phase (see refs. 17 and 24).

As explained in ref. 24 Property 2 implies the failure of the Gibbs variational principle. There is a potential nevertheless that converges even like a stretched exponential [which is a non-trivial result] on a full measure set. This indicates that having the existence of a potential is not of much use in general. We refer to this paper for a general discussion and for a restauration of the variational principle for a reasonable smaller class of generalized Gibbs measures. This class is defined in terms of the continuity properties of their conditional probabilities.

At first sight Properties 1 and 2 might seem not very intuitive, or at least unusual. In fact the expression of the conditional probabilities involves quantities that are not "explicitly" given. The aim of this paper is to show that these two properties in fact have analogous manifestations in the corresponding mean field model. A lot about this simple model is known,[2, 25, 26] and so we can partially draw from standard estimates. Our perspective however is new and the conditional probabilities of the joint measures we are looking at have not been considered.

Our point will become even clearer when we compare this model to mean-field versions of two other well-known examples of non-Gibbsian lattice measures. The first example is the mean-field analogue of the decimation transformation. For this measure the set of *continuity* points on the lattice is of full measure.[27] The other example is the mean field analogue of the "GriSing-field."[15] It is very simple to see that for both mean field models we get full measure *continuity* points. This complements our picture.

The paper is organized as follows. In Chapter 2 we introduce the models and state our results. In Chapter 3 we prove the statements about the random field model. We also provide some further discussion about the analogy to the lattice model. In Chapter 4 we give the remaining proofs of the statements on the decimated ferromagnet and the diluted ferromagnet.

2. MAIN RESULTS

In this section we look at the continuity properties of the conditional probabilities of our three models. We give precise estimates including error bounds only for the random field Ising model.

2.1. The Curie–Weiss Random Field Ising Model

The model is given by the Gibbs measures

$$\mu_{\beta,\varepsilon,N}[\eta_{[1,N]}](\sigma_{[1,N]}) := \frac{2^{-N} e^{\frac{\beta}{2N}(\sum_{i=1}^{N}\sigma_i)^2 + \beta\varepsilon\sum_{i=1}^{N}\eta_i\sigma_i}}{Z_{\beta,\varepsilon,N}[\eta_{[1,N]}]} \tag{2.1}$$

Let us look at the case of symmetric Bernoulli $\eta_i = \pm 1$ with equal probability $\mathbb{P}$, so that $\mathbb{P}(\eta_{[1,N]}) = 2^{-N}$. We define the corresponding *joint measure in finite volume N* by

$$K^N_{\beta,\varepsilon}[\eta_{[1,N]}, \sigma_{[1,N]}] := \mathbb{P}(\eta_{[1,N]}) \cdot \mu_{\beta,\varepsilon,N}[\eta_{[1,N]}](\sigma_{[1,N]}) \tag{2.2}$$

Of course this measure is permutation-invariant under *joint* permutation of the sites i of the *joint spin* (σ_i, η_i). This variable takes values in the set $\{(-1,-1), (-1,1), (1,-1), (1,1)\}$. We are interested in the behavior of the one-site conditional probability

$$K^N_{\beta,\varepsilon}[\sigma_1, \eta_1 \mid \sigma_{[2,N]}\eta_{[2,N]}]$$

$$\equiv K^N_{\beta,\varepsilon}[\sigma_1, \eta_1, \sigma_{[2,N]}, \eta_{[2,N]}] \Big/ \sum_{\substack{\tilde\sigma_1 = \pm 1\\ \tilde\eta_1 = \pm 1}} K^N_{\beta,\varepsilon}[\tilde\sigma_1, \tilde\eta_1, \sigma_{[2,N]}, \eta_{[2,N]}] \tag{2.3}$$

Without loss of generality we look here at the site $i = 1$. Define the function

$$m \mapsto \Phi^0_{\beta,\varepsilon,\alpha}(m) := \frac{m^2}{2} - \tilde{\mathbb{E}}_\alpha(\log\cosh(\beta(m + \varepsilon\tilde\eta_1))) \tag{2.4}$$

where $\tilde\eta_1 = \pm 1$ is a dummy random field variable with expectation $\tilde{\mathbb{E}}_\alpha(\tilde\eta_1) = \alpha$. Denote by $m^{\mathrm{RF}}(\beta,\varepsilon)$ the largest global minimizer for the symmetric case $\alpha = 0$. This is the "mean-field magnetization." In what follows we restrict ourselves to the two-phase region of the model. This is the region of the phase diagram where $\pm m^{\mathrm{RF}}(\beta,\varepsilon)$ are the only two different global minima.

Let us first recall how the measures look in the weak infinite-volume limit: From ref. 2 follows in particular that the finite-dimensional marginals of $K^N_{\beta,\varepsilon}$ converge to the symmetric linear combination of product measures $\frac{1}{2}(K^{\mathrm{prod},+}_{\beta,\varepsilon} + K^{\mathrm{prod},-}_{\beta,\varepsilon})$. Here $K^{\mathrm{prod},\pm}_{\beta,\varepsilon}$ are the product measures over the joint configurations whose single-site distributions are

$$K^{\mathrm{prod},\pm}_{\beta,\varepsilon}[\sigma_i, \eta_i] := \frac{\exp(\beta(\pm m^{\mathrm{RF}}(\beta,\varepsilon) + \varepsilon\eta_i)\,\sigma_i)}{4\cosh\beta(\pm m^{\mathrm{RF}}(\beta,\varepsilon) + \varepsilon\eta_i)} \tag{2.5}$$

Now we state and discuss the asymptotics of the conditional probabilities in finite volume. We start with the (more uninteresting) regime of atypically large modulus of the field sum.

Theorem 1 (Continuity for Large Modulus of the Field Sum)

$$\lim_{N\uparrow\infty} K^N_{\beta,\varepsilon}[\sigma_1,\eta_1 \mid \sigma_{[2,N]}\eta_{[2,N]}] = \frac{1}{\text{Norm}} \frac{\exp(\beta(\hat{m}+\varepsilon\eta_1)\,\sigma_1)}{\cosh\beta(m^{\text{RF}}(\beta,\varepsilon,\alpha)+\varepsilon\eta_1)}$$
$$\text{if} \quad \lim_{N\uparrow\infty}\frac{1}{N}\sum_{i=2}^{N}\sigma_i = \hat{m}, \qquad \lim_{N\uparrow\infty}\frac{1}{N}\sum_{i=2}^{N}\eta_i = \alpha \neq 0 \tag{2.6}$$

where $m^{\text{RF}}(\beta,\varepsilon,\alpha)$ is the global minimizer of the function (2.4). The normalization is obtained by summing over σ_1,η_1. In particular, the limiting expression (2.6) varies *continuously* as a function of the pair $(\hat{m},\alpha)$ of the empirical means of the conditioning joint spins, in the two connected components $\alpha>0$ and $\alpha<0$.

Remark. The quantity $m^{\text{RF}}(\beta,\varepsilon,\alpha)$ is the magnetization of a random field Curie–Weiss model with biased i.i.d. random fields, where the bias is given by $\mathbb{E}_\alpha(\eta_i)=\alpha\neq 0$.

Remark. We note that the conditional probability is continuous everywhere in the single variable $\hat{m}$. However, it is *not* continuous everywhere when we consider it as a function of α. This is identical to the lattice case. The continuity for $\alpha\neq 0$ follows from the elementary fact that, for β,ε in the two phase region (for the model with $\alpha=0$) the minimizer $m^{\text{RF}}(\beta,\varepsilon,\alpha)$ varies continuously as a function of α. It however jumps at $\alpha=0$ and changes sign. Indeed, we have $m^{\text{RF}}(\beta,\varepsilon,\alpha=0+)=m^{\text{RF}}(\beta,\varepsilon)=-m^{\text{RF}}(\beta,\varepsilon,\alpha=0-)$. Of course, the probability that the normalized η-sum takes values α away from zero is exponentially small in N, and so the regime we are looking at is atypical.

Heuristically speaking, in the large N-limit there will be all mass on the two conditionings $\alpha=0+$ and $\alpha=0-$. We will see and make precise below the following picture: Outside a set of zero measure the conditional probabilities (2.6) will acquire the two limiting forms for $\alpha=0+$ and $\alpha=0-$ that are manifestly different. So, an infinitesimal variation of α around zero leads to discontinuous behavior in the conditional probability. Since $\alpha=0$ is typical, we have an almost sure discontinuity as a function of the joint conditioning.

To understand better the nature of this discontinuity at $\alpha=0$ let us blow up the scale of the empirical mean of the random fields. We like to give a more precise statement here than in the simpler Theorem 1 given before. We even provide a uniform bound on the deviation from the limiting expression. Introduce the regular set of random field configurations

$$\bar{\mathscr{H}}(N) := \left\{ \eta : \left| \sum_{i=1}^{N} \eta_i \right| \leqslant N^{\frac{1+\delta}{2}} \right\} \tag{2.7}$$

for some fixed $0 < \delta < \frac{1}{6}$. We note that it is a very large set for large N. Indeed, we have $\mathbb{P}(\bar{\mathscr{H}}(N)) \geqslant 1 - 2e^{-\frac{N^\delta}{2}}$. Then the following holds.

Theorem 2 (Close-Up of Discontinuity Region—Almost Sure Discontinuity). We have the uniform approximation

$$\sup_{\eta \in \bar{\mathscr{H}}(N)} \sup_{\sigma_{[2,N]}} \Bigg| K^N_{\beta,\varepsilon}[\sigma_1, \eta_1 \mid \sigma_{[2,N]} \eta_{[2,N]}] \\ - K^\infty_{\beta,\varepsilon}\left[\sigma_1, \eta_1 \,\middle|\, \frac{1}{N} \sum_{i=2}^{N} \sigma_i, \sum_{i=2}^{N} \eta_i \right] \Bigg| \leqslant C(\beta, \varepsilon)\, N^{-\frac{1-6\delta}{4}} \tag{2.8}$$

Here the limiting expression is given by

$$K^\infty_{\beta,\varepsilon}[\sigma_1, \eta_1 \mid \hat{m}, w] := \frac{1}{\text{Norm}} \exp(\beta(\hat{m} + \varepsilon\eta_1)\,\sigma_1) \cdot q_{\beta,\varepsilon,\infty}(w)^{-\frac{\eta_1}{2}} \tag{2.9}$$

for any $\hat{m} \in [-1, 1]$, $w \in \mathbb{Z}$ where we have put

$$q_{\beta,\varepsilon,\infty}(w) = \frac{(r^*_{\beta,\varepsilon})^{\frac{w-1}{2}} + (r^*_{\beta,\varepsilon})^{-\frac{w-1}{2}}}{(r^*_{\beta,\varepsilon})^{\frac{w+1}{2}} + (r^*_{\beta,\varepsilon})^{-\frac{w+1}{2}}}, \qquad r^*_{\beta,\varepsilon} = \frac{\cosh(\beta(m^{\text{RF}}(\beta,\varepsilon) + \varepsilon))}{\cosh(\beta(m^{\text{RF}}(\beta,\varepsilon) - \varepsilon))} \tag{2.10}$$

Remark. Note that the limiting expression (2.9) is a function of the *non-normalized* sum of random fields in the conditioning. So, it is *not* continuous in the empirical mean of the random field conditioning $\frac{1}{N} \sum_{i=2}^{N} \eta_i$, due to the occurrence of the normalization $\frac{1}{N}$. Of course, it is continuous in the empirical mean of the spin conditioning $\hat{m}$.

We note that, for large N, the measure $\mathbb{P}$ gives mass to typical random field configurations on the scale of the central limit theorem, i.e., $\sum_{i=2}^{N} \eta_i \approx +C\sqrt{N}$ resp. $-C\sqrt{N}$. Since we have that

$$\lim_{w \to \pm\infty} q_{\beta,\varepsilon,\infty}(w) = (r^*_{\beta,\varepsilon})^{\mp 1} \tag{2.11}$$

the η-dependence acquires two limiting forms, outside of a set with vanishing $\mathbb{P}$-mass in the limit $N \uparrow \infty$. Using (2.11) they can be written in the form

$$K^{\infty}_{\beta,\varepsilon}[\sigma_1, \eta_1 \mid \hat{m}, \pm\infty] := \frac{1}{\text{Norm}} \frac{\exp(\beta(\hat{m}+\varepsilon\eta_1)\,\sigma_1)}{\cosh \beta(\pm m^{\text{RF}}(\beta,\varepsilon)+\varepsilon\eta_1)} \tag{2.12}$$

Note that the two forms coincide with the $\alpha \downarrow 0$ resp. $\alpha \uparrow 0$ limits of the r.h.s. of (2.6). For the sake of clarity let us make explicit the following trivial consequence of Theorem 2. It results from the fact that the convergence in (2.11) is exponentially fast in w.

Corollary to Theorem 2. We have the approximation

$$\sup_{\eta \in \mathscr{H}^{\pm}(N)} \sup_{\sigma_{[2,N]}} \left| K^{N}_{\beta,\varepsilon}[\sigma_1, \eta_1 \mid \sigma_{[2,N]}\eta_{[2,N]}] - K^{\infty}_{\beta,\varepsilon}\left[\sigma_1, \eta_1 \,\middle|\, \frac{1}{N}\sum_{i=2}^{N} \sigma_i, \pm\infty\right]\right| \leqslant C(\beta,\varepsilon)\, N^{-\frac{1-6\delta}{4}} \tag{2.13}$$

Here we have introduced two components of "regular realizations" given by

$$\mathscr{H}^{+}(N) := \left\{\eta : C_{-}(\beta,\varepsilon) \log N \leqslant \sum_{i=1}^{N} \eta_i \leqslant N^{\frac{1+\delta}{2}}\right\},$$
$$\mathscr{H}^{-}(N) := -\mathscr{H}^{+}(N) \tag{2.14}$$

with $C_{-}(\beta,\varepsilon) = (1+6\delta)/(4 \log \frac{1}{r^{*}_{\beta,\varepsilon}})$.

Remark. Note that $\mathbb{P}(\mathscr{H}^{+}(N) \cup \mathscr{H}^{-}(N)) \geqslant 1 - \text{const}\,\frac{\log N}{\sqrt{N}}$ goes to one for large N. However this convergence is much slower than the convergence for the set $\bar{\mathscr{H}}(N)$.

Remark. The corollary says in brief: "For all spin-conditionings σ and all random field conditionings outside of a set with vanishing mass in the large volume limit, the conditional probability is discontinuous in the empirical random field sum." This discontinuity in the random field-sum reflects the behavior of the lattice model. Also for the lattice model the conditional probabilities depend on the conditioning in a discontinuous (that is non-local) way. In the analogous expression for the lattice model the empirical spin-average is replaced by the sum over nearest neighbor spins, while the empirical random field average is replaced by a more complicated non-local function. For more on that, see the discussion at the end of Section 3.1.

2.2. Decimation of Standard Curie–Weiss

The decimation transformation of the usual ferromagnetic nearest neighbor Ising model at low temperature provides one of the most basic examples of a non-Gibbsian measure. It is however known in this example that its conditional probabilities are only discontinuous outside of a set of measure zero. Let us see the reflections of this in the corresponding mean-field model. Start with the ordinary Curie–Weiss model given by the Gibbs measures

$$\mu_{\beta,N}(\sigma_{[1,N]}) := \frac{2^{-N} e^{\frac{\beta}{2N}(\sum_{i=1}^{N}\sigma_i)^2}}{Z_{\beta,N}} \tag{2.15}$$

We look at the *decimated model* which is just the marginal distribution $\mu_{\beta,N}(\sigma_{[1,M+1]})$ on the first $M+1$ spins. We are interested in the asymptotic behavior of the conditional probability when $M = M_N$ grows like a multiple of N, as N tends to infinity. This provides a natural mean-field analogue of decimation to a sublattice. Denote by $m^{\text{CW}}(\beta)$ the ordinary Curie–Weiss magnetisation, i.e., the largest solution of the mean-field equation $m = \tanh(\beta m)$. It is well-known that $\lim_{N\uparrow\infty} \mu_{\beta,N} = \frac{1}{2}(\mu_\beta^+ + \mu_\beta^-)$. Here the measures appearing on the right hand side are the product measures given by their single-site distribution $\mu_\beta^\pm(\sigma_i) = \frac{e^{\pm\beta m^{\text{CW}}(\beta)\sigma_i}}{2\cosh(\beta m^{\text{CW}}(\beta))}$. This limit is in the sense of finite-dimensional marginals. So, the large M-limit of the decimated measure on finite dimensional marginals is given in terms of the same product measures, by definition. Moreover one has of course also the convergence of the empirical mean $\lim_{N,M\uparrow\infty} \mu_{\beta,N}(\frac{1}{M}\sum_{i=1}^{M+1}\sigma_i \in \cdot) = \frac{1}{2}(\delta_{m^{\text{CW}}(\beta)} + \delta_{-m^{\text{CW}}(\beta)})$. This makes particularly clear that it doesn't suffice to look at the limiting measures to see "non-Gibbsianness." Let us now give our results. We start with the behavior outside of the discontinuity region.

Theorem 3 (Almost Sure Continuity in the Conditioning for Decimation). Assume that M_N tends to infinity such that $\lim_{N\uparrow\infty} M_N/N = 1-p$, with $0 \leqslant p \leqslant 1$. Then

$$\lim_{N\uparrow\infty} \mu_{\beta,N}(\sigma_1 \mid \sigma_{[2,M_N+1]}) = \frac{e^{\beta h_{\beta,p}(\hat{m})\sigma_1}}{2\cosh(\beta h_{\beta,p}(\hat{m}))}, \qquad \text{if} \quad \lim_{M\uparrow\infty} \frac{1}{M}\sum_{i=2}^{M+1}\sigma_i = \hat{m} \neq 0$$

$$\text{where} \quad h_{\beta,p}(\hat{m}) = \begin{cases} \hat{m} & \text{if} \quad p=0 \\ p\, m^{\text{CW}}\left(p\beta, \dfrac{1-p}{p}\hat{m}\right) + (1-p)\,\hat{m} & \text{if} \quad 0<p<1 \\ m^{\text{CW}}(\beta)\ \text{sign}\ (\hat{m}) & \text{if} \quad p=1 \end{cases} \tag{2.16}$$

Here $m^{CW}(\beta', h')$ is the solution of the mean-field equation $m = \tanh(\beta'(m+h'))$ that has the sign of h'. In particular, the limiting form of the conditional probability is continuous in $\hat{m}$ for $\hat{m} \neq 0$. It is discontinuous in $\hat{m}$ at $\hat{m} = 0$ for $p\beta > 1$.

Remark. The quantity $m^{CW}(\beta', h')$ is our course the mean-field magnetization of the Curie–Weiss model in an external magnetic field h'. For a quick check-up of the expression given in (2.16) let us condition on typical configurations. Then we get back the product measures with mean-field magnetization, i.e.,

$$\lim_{N\uparrow\infty} \mu_{\beta,N}(\sigma_1 \mid \sigma_{[2, M_N+1]}) = \mu_\beta^{\pm}(\sigma_i), \qquad \text{if} \quad \lim_{M\uparrow\infty} \frac{1}{M} \sum_{i=2}^{M+1} \sigma_i = \pm m^{CW}(\beta) \tag{2.17}$$

This is clear from the theorem since it is immediate to check that $h_{\beta,p}(\pm m^{CW}(\beta)) = \pm m^{CW}(\beta)$, for any $0 \leqslant p \leqslant 1$.

Remark. The continuity statements of the theorem are clear by the known properties of the mean-field solution $m^{CW}(p\beta, h')$ as a function of h': It is continuous for $h' \neq 0$ and jumps at $h' = \pm 0$ for $p\beta > 0$. The behavior of the conditional probability as a function of $\hat{m}$ corresponds to the situation in the lattice model. The reader familiar with the latter will recall the following. In the lattice model putting a checker-board configuration in the conditioning produces a point of discontinuity. The mechanism of non-Gibbsianness is to produce a phase transition in the system that is integrated out by varying this conditioning of the decimated system arbitrarily far away. In the mean-field model we have an analogous picture: Varying the neutral conditioning of the decimated system around $\hat{m} = 0$ produces a phase transition in the part of the system that is integrated out. This phase transition takes place if and only if the inverse temperature for the part of the system that is integrated out, βp is bigger than one.

We also note that the field $h_{\beta,p}(\hat{m})$ is continuous in p in the whole interval $[0, 1]$, for fixed $\hat{m} \neq 0$. In particular the special case $M+1 = N$ is contained in the case $p = 0$ in the statement of the theorem. Then the decimation transformation is the identity, i.e., no decimation takes place.

Remark. In contrast the behavior of the joint measures of the random field model, we see that the point of discontinuity $\hat{m} = 0$ is *atypical*: it is taken only with exponentially small probability. This also corresponds to the lattice model, where the "checker-board-like" points of discontinuity have zero mass w.r.t. the decimated measure.

For reasons of analogy to the random field model, let us also mention the following simple result. In Theorem 2 we gave a uniform approximation statement for conditionings such that $\sum_{i=2}^{N} \eta_i$ is subextensively small, in order to interpolate between $\alpha = \pm 0$. We could give a similar uniform approximation statement here. This would mean to look at conditionings such that $\sum_{i=2}^{M_N} \sigma_i$ is subextensively small, in order to show how the interpolation at $\hat{m} = \pm 0$ looks on a finer scale. For reasons of simplicity however let us formulate the result for conditionings such that $\sum_{i=2}^{M_N} \sigma_i = z$ is even constant when N tends to infinity. Of course, we could also give error estimates and allow for increasing values of z like we did before in Theorem 2 but we omit these details here.

Theorem 4 (Close-Up of Discontinuity Region). Assume again that M_N tends to infinity such that $\lim_{N\uparrow\infty} M_N/N = 1-p$, with $0 \leqslant p \leqslant 1$. Then

$$\lim_{N\uparrow\infty} \mu_{\beta,N}(\sigma_1 \mid \sigma_{[2,M_N+1]})$$
$$= \frac{e^{\beta z m^{\mathrm{CW}}(p\beta)} \mu^{+}_{\beta p}(\sigma_1) + e^{-\beta z m^{\mathrm{CW}}(p\beta)} \mu^{-}_{\beta p}(\sigma_1)}{2\cosh(\beta z m^{\mathrm{CW}}(p\beta))}, \qquad \text{if } \sum_{i=2}^{M_N+1} \sigma_i = z \quad \text{stays finite} \tag{2.18}$$

Of course, in order for this statement to make and keep z constant, M_N need to be all even, or all odd. Again, the r.h.s. of (2.18) is a function of the *non-normalized* sum of the spins in the conditioning. Note that its values for $z = \pm\infty$ coincide with the values of (2.16) for $\hat{m} = \pm 0$.

2.3. Joint Measures of the Mean-Field Diluted Ising Model

The lattice version of this model was the first example of a non-Gibbsian joint measure. Let us now see how the mean-field version behaves. The model is given by the Gibbs measures

$$\mu_{\beta,N}[n_{[1,N]}](\sigma_{[1,N]}) := \frac{2^{-N} e^{\frac{\beta}{2N}(\sum_{i=1}^{N} n_i\sigma_i)^2}}{Z_N[n_{[1,N]}]} \tag{2.19}$$

Here the n_i are independent Bernoulli occupation numbers with distribution $\mathbb{P}[n_i = 1] = 1 - \mathbb{P}[n_i = 0] = p$. Of course, the finite volume Gibbs measure (2.19) is nothing but the ordinary Curie–Weiss with the smaller inverse temperature $\beta' = \frac{\sum_{i=1}^{N} n_i}{N}\beta$ on the set of occupied sites, tensored with symmetric Bernoulli-spins on the sites of vacant sites. We define the corresponding joint measure in finite volume N by

$$K^N_{\beta,p}[n_{[1,N]},\sigma_{[1,N]}] := \mathbb{P}(n_{[1,N]})\cdot\mu_{\beta,N}[n_{[1,N]}](\sigma_{[1,N]}) \tag{2.20}$$

Then it is straightforward to see that we have

$$\lim_{N\uparrow\infty} K^N_{\beta,p} = \tfrac{1}{2}\,(K^{\text{prod},+}_{\beta,p} + K^{\text{prod},-}_{\beta,p}) \tag{2.21}$$

in the sense of finite-dimensional marginals of the joint variables, where

$$K^{\text{prod},\pm}_{\beta,p}[\sigma_i,n_i] := p^{n_i}(1-p)^{1-n_i}\,\frac{\exp(\pm\beta p\cdot m^{\text{CW}}(\beta p)\,n_i\sigma_i)}{2\cosh(\beta p\cdot m^{\text{CW}}(\beta p)\,n_i)} \tag{2.22}$$

The limit (2.21) is clear because the distribution of the occupation numbers concentrates exponentially fast on those configurations that have a fixed density p. Indeed, for those configurations we are back in the ordinary Curie–Weiss model with inverse temperature βp. Then we have

Theorem 5 (Almost Sure Continuity in the Conditioning for Dilute Ising)

$$\lim_{N\uparrow\infty} K^N_{\beta,p}[\sigma_1,n_1\mid\sigma_{[2,N]}n_{[2,N]}] = \frac{1}{\text{Norm}}\,p^{n_1}(1-p)^{1-n_1}\,\frac{\exp(\beta\hat m\,n_1\sigma_1)}{\cosh(\beta q\,m^{\text{CW}}(\beta q)\,n_1)}$$
$$\text{if}\quad \lim_{N\uparrow\infty}\frac{1}{N}\sum_{i=2}^N n_i\sigma_i = \hat m,\qquad \lim_{N\uparrow\infty}\frac{1}{N}\sum_{i=2}^N n_i = q>0 \tag{2.23}$$

In particular, the limiting expression (2.23) varies *continuously* as a function of the pair $(\hat m, q)$ of the empirical magnetization and density of occupied sites.

Remark. Let us do our check-up of the formula by conditioning on typical configurations, like we did if before for the other two models. Then we see that indeed

$$\lim_{N\uparrow\infty} K^N_{\beta,p}[\sigma_1,n_1\mid\sigma_{[2,N]}\eta_{[2,N]}]$$
$$=\begin{cases} K^{\text{prod},+}_{\beta,p}[\sigma_1,n_1] \\ K^{\text{prod},-}_{\beta,p}[\sigma_1,n_1]\end{cases}\quad \text{if}\quad \lim_{N\uparrow\infty}\frac{1}{N}\sum_{i=2}^N (n_i, n_i\sigma_i) = p\begin{cases}(1,+m^{\text{CW}}(\beta\,p))\\(1,-m^{\text{CW}}(\beta\,p))\end{cases} \tag{2.24}$$

General Comment. Theorems 1, 3, and 5 immediately carry over from *single-site* conditional probabilities to conditional probabilities *outside any finite set of sites*. Here one obtains convergence to product measures, whose parameters are given by the single-site expressions. E.g., for the decimation case we have

$$\lim_{N\uparrow\infty} \mu_{\beta,N}(\sigma_{[1,k]} \mid \sigma_{[k+1,M_N+1]})$$

$$= \prod_{i=1}^{k} \frac{e^{\beta h_{\beta,p}(\hat{m})\,\sigma_i}}{2\cosh(\beta h_{\beta,p}(\hat{m}))}, \qquad \text{if} \quad \lim_{M\uparrow\infty} \frac{1}{M} \sum_{i=k+1}^{M+1} \sigma_i = \hat{m} \neq 0 \tag{2.25}$$

In the case of the random field model and the dilute ferromagnet it is analogous. Formula (2.25) follows recursively from Theorem 3 with the use of the identity $\rho(\sigma_{[1,k]} \mid \sigma_{[k+1,n]}) = \rho(\sigma_k \mid \sigma_{[1,k-1]}\sigma_{[k+1,n]}) \times (\sum_{\tilde{\sigma}_k} \rho(\tilde{\sigma}_k \mid \sigma_{[1,k-1]}\sigma_{[k+1,n]})/\rho(\sigma_{[1,k-1]} \mid \tilde{\sigma}_k\sigma_{[k+1,n]}))^{-1}$ for any measure ρ. Then use that the change of any finite number of spins does not effect the limiting value $\hat{m}$.

2.4. A Simple Heuristic for the Random Field Model

Let us explain a simple heuristic that shows that Theorem 2 can be checked without computations in a limiting form for "most" joint configurations. It will make clear how the almost sure discontinuity in the random fields comes about in a qualitative way. For "typical" η the random finite volume Gibbs measures can be approximated in the simple form

$$\mu_{\beta,\varepsilon,N}[\eta] \approx \begin{cases} \mu^{\text{prod},+}_{\beta,\varepsilon}[\eta] \\ \mu^{\text{prod},-}_{\beta,\varepsilon}[\eta] \end{cases} \qquad \text{for} \quad \sum_{i=2}^{N} \eta_i \approx \begin{cases} +C\sqrt{N} \\ -C\sqrt{N} \end{cases} \tag{2.26}$$

Here $\mu^{\text{prod},\pm}_{\beta,\varepsilon}[\eta]$ are the η-dependent product measures on the σ-configurations given by

$$\mu^{\text{prod},\pm}_{\beta,\varepsilon}[\eta](\sigma_i) = \frac{\exp(\beta(\pm m^{\text{RF}}(\beta,\varepsilon) + \varepsilon\eta_i)\,\sigma_i)}{2\cosh\beta(\pm m^{\text{RF}}(\beta,\varepsilon) + \varepsilon\eta_i)} \tag{2.27}$$

They play the role of infinite-volume Gibbs measures. [In this heuristic discussion C indicates a positive random quantity of order unity.] The approximation (2.26) is made rigorous in ref. 2. In that paper also asymptotic statements about metastates are derived (including the arcsine-law for the empirical metastate). Now, for typical η, this approximation formally implies for the joint measure

$$K^N_{\beta,\varepsilon}[\sigma_{[1,N]}\eta_{[1,N]}] \approx \begin{cases} K^{\text{prod},+}_{\beta,\varepsilon}[\sigma_{[1,N]}\eta_{[1,N]}] \\ K^{\text{prod},-}_{\beta,\varepsilon}[\sigma_{[1,N]}\eta_{[1,N]}] \end{cases} \quad \text{for} \quad \sum_{i=2}^{N} \eta_i \approx \begin{cases} +C\sqrt{N} \\ -C\sqrt{N} \end{cases} \tag{2.28}$$

The joint product measures appearing on the r.h.s. have the concentration properties that

$$\begin{aligned} K^{\text{prod},+}_{\beta,\varepsilon}\left[\frac{1}{N}\sum_{i=2}^{N}(\eta_i,\sigma_i) \approx (\pm C/\sqrt{N}, +m^{\text{RF}}(\beta,\varepsilon))\right] \approx 1 \\ K^{\text{prod},-}_{\beta,\varepsilon}\left[\frac{1}{N}\sum_{i=2}^{N}(\eta_i,\sigma_i) \approx (\pm C/\sqrt{N}, -m^{\text{RF}}(\beta,\varepsilon))\right] \approx 1 \end{aligned} \tag{2.29}$$

This clearly follows from the mean field equation of the model which can be written $\pm m^{\text{RF}}(\beta,\varepsilon) = \int \mathbb{P}(d\eta) \int \mu^{\text{prod},\pm}_{\beta,\varepsilon}[\eta](d\sigma_i)\,\sigma_i$. Note that the empirical mean of the magnetic fields has arbitrary signs. By (2.28) it follows that the original joint measure (2.2) then concentrates on the union of the two sets where $\frac{1}{N}\sum_{i=2}^{N}(\eta_i,\sigma_i) \approx \pm(C/\sqrt{N}, m^{\text{RF}}(\beta,\varepsilon))$, forming a disconnected set of two "joint lumps" where the empirical means of the field and the magnetisation have the *same* sign. Conditioning on these typical configurations and using the product nature of (2.28) we obtain the simple heuristic formula

$$\begin{aligned} &K^N_{\beta,\varepsilon}[\sigma_1,\eta_1 \mid \sigma_{[2,N]}\eta_{[2,N]}] \\ &\approx \begin{cases} K^{\text{prod},+}_{\beta,\varepsilon}[\sigma_1,\eta_1] \\ K^{\text{prod},-}_{\beta,\varepsilon}[\sigma_1,\eta_1] \end{cases} \quad \text{if} \quad \frac{1}{N}\sum_{i=2}^{N}(\eta_i,\sigma_i) \approx \begin{cases} (+C/\sqrt{N}, +m^{\text{RF}}(\beta,\varepsilon)) \\ (-C/\sqrt{N}, -m^{\text{RF}}(\beta,\varepsilon)) \end{cases} \end{aligned} \tag{2.30}$$

But now we see the following: For joint configurations in these two joint lumps the expression given in the Corollary to Theorem 2 indeed reduces to the *same* formula.

2.5. Random Field Model: Dependence of Conditional Probabilities on the Phase

A different feature of the lattice random field Ising model is the dependence of the conditional probability of a joint measure on the phase

μ^+ resp. μ^-. It is this property that is responsible for the failure of the variational principle. We cannot put boundary conditions in a mean field model, but we can put an additional "infinitesimal" symmetry-breaking external field s_N that goes to zero like a suitable power of N (but dominates typical random field fluctuations). The purpose of this external field is to pick either the plus or the minus state. Putting this external field s_N then leads to a different form of the limiting expression for the conditional probability, depending on its sign. The precise result of this is given in Theorem 6.

So, let us define the measure

$$\mu_{\beta, \varepsilon, s, N}[\eta_{[1,N]}](\sigma_{[1,N]}) := \frac{2^{-N} e^{\frac{\beta}{2N}(\sum_{i=1}^{N}\sigma_i)^2 + \beta\sum_{i=1}^{N}(\varepsilon\eta_i + s)\sigma_i}}{Z_{\beta, \varepsilon, s, N}[\eta_{[1,N]}]} \tag{2.31}$$

and the corresponding joint measure $K^N_{\beta,\varepsilon,s}[\eta_{[1,N]}, \sigma_{[1,N]}] = \mathbb{P}(\eta_{[1,N]}) \cdot \mu_{\beta,\varepsilon,s,N}[\eta_{[1,N]}](\sigma_{[1,N]})$. Then we have

Theorem 6 (Dependence of Conditional Probabilities on the Phase). Fix any constants $\alpha, \delta > 0$ such that $\alpha + \delta < \frac{1}{2}$. Then the following approximation holds

$$\begin{aligned}\sup_{\substack{s:\\ N^{\frac{1}{2}+\delta} \leqslant s \leqslant N^{1-\alpha}}} \sup_{\eta \in \mathscr{H}(N)} \sup_{\sigma_{[2,N]}} \Bigg| & K^N_{\beta,\varepsilon,\pm s}[\sigma_1, \eta_1 \mid \sigma_{[2,N]}\eta_{[2,N]}] \\ & - K^\infty_{\beta,\varepsilon}\left[\sigma_1, \eta_1 \,\middle|\, \frac{1}{N}\sum_{i=2}^{N}\sigma_i, \pm\infty\right]\Bigg| \\ & \leqslant C(\beta,\varepsilon)\, N^{-\frac{1}{2}+2\delta} + C(\beta,\varepsilon)'\, N^{-\alpha}\end{aligned} \tag{2.32}$$

Remark. This has to be interpreted as analogue of the dependence of the conditional probabilities of the joint measures on the phase + resp. − in a lattice model. Note that the limiting expression is independent of the random field conditioning outside of a set of vanishing mass in the large N-limit. We could also proceed in the same way and add an infinitesimal external field for the decimated and the dilute Ising model. In both cases nothing interesting happens, and the limiting expressions stay the same as for zero field. We don't give an explicit analysis.

3. RANDOM FIELD: FURTHER COMMENTS AND PROOFS

In this section we prove the statements on the random field model in the order Theorems 2, 6, and 1. The first step is simple: We derive a formula for the one-site conditional probabilities in finite volume. The formula is the simpler mean-field analogue of Proposition 3.1 in ref. 17 for lattice-spin models in finite volume. We take the chance and provide some discussion about this formula, and its monotonicity properties and also give a heuristic explanation for them. Next we compare it to its lattice analogue. We also give a heuristic explanation of monotonicity and the dependence on the phase-phenomenon on the lattice. After that we provide technical details about the saddle-point estimates used to prove the theorems.

3.1. Representation of Conditional Probabilities and Monotonicity

Proposition 1. The one-site conditional probabilities can be written in the form

$$K^N_{\beta,\varepsilon}[\sigma_1,\eta_1 \mid \sigma_{[2,N]}\eta_{[2,N]}] = \frac{1}{\text{Norm}} \exp\left(\beta\left(\frac{1}{N}\sum_{i=2}^{N}\sigma_i + \varepsilon\eta_1\right)\sigma_1\right) \cdot q_{\beta,\varepsilon,N}\left(\sum_{i=2}^{N}\eta_i\right)^{\frac{\eta_1}{2}} \tag{3.1}$$

where

$$q_{\beta,\varepsilon,N}\left(\sum_{i=2}^{N}\eta_i\right) := \int \mu_{\beta,\varepsilon,N}[\eta_1=+1,\eta_{[2,N]}](d\hat\sigma_1)\, e^{-2\beta\varepsilon\hat\sigma_1} \tag{3.2}$$

with the obvious normalization obtained by summing over σ_1 and η_1.

Remark. The last definition is meaningful, because the Gibbs expectation on the r.h.s. depends only on the number of plus random fields, by permutation invariance of the model. Indeed, this number can equivalently be expressed by the sum over the random fields.

Remark. By monotonicity, $q_{\beta,\varepsilon,N}(w)$ is a decreasing function in w. This follows, e.g., from Theorem 4.8 in ref. 28 (Holley's theorem) because of the monotonicity of the single-site conditional probabilities, i.e., $\mu_{\beta,\varepsilon,N}[\eta_1=+1,\eta_{[2,N]}](\sigma_i=+\mid\sigma_{[1,N]\setminus i}) \leqslant \mu_{\beta,\varepsilon,N}[\eta_1=+1,\eta'_{[2,N]}](\sigma_i=+\mid\sigma'_{[2,N]\setminus i})$ for all $\eta_{[2,N]}\leqslant\eta'_{[2,N]}$, for all $\sigma_{[1,N]\setminus i}\leqslant\sigma'_{[1,N]\setminus i}$.

Remark. The reader may think of $h=\frac{1}{2}\log q$ as a field acting on η_1 that depends on the random fields appearing in the conditioning.

Proof of Proposition 1. This is just a computation. Write out the definition of the conditional probability. Apart from quotients of the exponential of the energy functions, quotients of quenched partition functions are appearing. All of these can be expressed in terms of $Z_{\beta,\varepsilon,N}[\eta_1=-1,\eta_{[2,N]}]/Z_{\beta,\varepsilon,N}[\eta_1=1,\eta_{[2,N]}]=q_{\beta,\varepsilon,N}(\sum_{i=2}^N \eta_i)$. ∎

In particular it is instructive to introduce the *bias* of the plus-random field at the site 1 defined by

$$B_{\beta,\varepsilon,N}[\sigma_{[2,N]}\eta_{[2,N]}]:=\frac{K_{\beta,\varepsilon,N}[\eta_1=+\mid\sigma_{[2,N]}\eta_{[2,N]}]}{K_{\beta,\varepsilon,N}[\eta_1=-\mid\sigma_{[2,N]}\eta_{[2,N]}]} \tag{3.3}$$

Note that we have the symmetry $B_{\beta,\varepsilon,N}[\sigma_{[2,N]},\eta_{[2,N]}]=(B_{\beta,\varepsilon,N}[-\sigma_{[2,N]},-\eta_{[2,N]}])^{-1}$. To avoid trivial misunderstandings let us point out the following: Of course, by dropping the spin-conditioning, we get the unbiased expression $K_{\beta,\varepsilon,N}[\eta_1=+\mid\eta_{[2,N]}]/K_{\beta,\varepsilon,N}[\eta_1=-\mid\eta_{[2,N]}]=\frac{1}{2}$. The bias contains all the interesting non-trivial behavior of the full conditional probabilities (3.1). This is because of the trivial identity $K_{\beta,\varepsilon}^N[\sigma_1,\eta_1\mid\sigma_{[2,N]}\eta_{[2,N]}]$ $=\mu_{\beta,\varepsilon,N}[\eta_1\eta_{[2,N]}](\sigma_1\mid\sigma_{[2,N]})\cdot K_{\beta,\varepsilon}^N[\eta_1\mid\sigma_{[2,N]}\eta_{[2,N]}]$ and the last probability is trivially related to the bias (3.3).

Carrying out the spin-sums in (3.1) we may write the bias in the form

$$B_{\beta,\varepsilon,N}[\sigma_{[2,N]}\eta_{[2,N]}]=r_{\beta,\varepsilon}\left(\frac{1}{N}\sum_{i=2}^N\sigma_i\right)\cdot q_{\beta,\varepsilon,N}\left(\sum_{i=2}^N\eta_i\right) \tag{3.4}$$

Here we have defined the function

$$\begin{aligned} r_{\beta,\varepsilon}(m)=r_{\beta,-\varepsilon}(-m)=r_{\beta,\varepsilon}(-m)^{-1}=r_{\beta,-\varepsilon}(m)^{-1}=\frac{\cosh(\beta(m+\varepsilon))}{\cosh(\beta(m-\varepsilon))}\\ =\cosh(2\beta\varepsilon)+\sinh(2\beta\varepsilon)\tanh(\beta(m+\varepsilon)) \end{aligned} \tag{3.5}$$

Note that $m\mapsto r_{\beta,\varepsilon}(m)$ is increasing and $r_{\beta,\varepsilon}(m=0)=1$, $\lim_{m\uparrow\infty} r_{\beta,\varepsilon}(m)=e^{2\beta\varepsilon}$.

So $B_{\beta,\varepsilon,N}$ is increasing in the spin-condition and *decreasing* in the random field-condition. Let us give a heuristic explanation for the latter. Conditioning the distribution $K_{\beta,\varepsilon,N}$ on $\sum_{i=2}^N\sigma_i$ destroys the independence of η_1 from $\sum_{i=2}^N\eta_i$: Indeed, when $\sum_{i=2}^N\eta_i$ *increases* we should expect that η_1 will acquire a *lesser* tendency to be plus when the *same* spin-sum $\sum_{i=2}^N\sigma_i$ is observed. This is because increasing the η-sum will increase the tendency of σ_i's for $i=2,\dots,N$ to be plus. So the probability for $\eta_1=-$ should also increase to make up for this. This phenomenon is particularly clear when we assume a neutral spin-sum $\sum_{i=2}^N \sigma_i = 0$, and a positive random field-sum $\sum_{i=2}^N \eta_i > 0$. Then we expect quite naturally $\eta_1 = -$ with probability bigger than $\frac{1}{2}$, because we would ascribe to the value of η_1 some partial responsibility for the neutral realization $\sum_{i=2}^N \sigma_i = 0$.

A somewhat similar "Bayesian destruction of independence" can also be seen in the related (dynamical) situation of ref. 29. (This catchy term goes back R. Schonmann in the context of ref. 15, as it was pointed out to the author by A. van Enter.)

Remark. To appreciate the analogy to the lattice nearest neighbor random field model, let us compare (3.1) to its lattice analogue. So, let us look at the random field Ising model with symmetric i.i.d. plus minus random field distribution $\mathbb{P}$. The Gibbs measures have formal Boltzmann weights $\propto \exp(\beta \sum_{\langle x, y\rangle} \sigma_x \sigma_y + \beta\varepsilon \sum_x \eta_x \sigma_x)$. Denote by $K^\mu(d\eta d\sigma) = \mathbb{P}(d\eta)\, \mu[\eta](d\sigma)$ the infinite-volume joint measure corresponding to a Gibbs measure $\mu[\eta](d\sigma)$. In particular one might think of the measure $\mu^+[\eta](d\sigma)$ and $\mu^-[\eta](d\sigma)$ obtained as weak limits for plus resp. minus boundary conditions. They are different for 3 or more dimensions, low temperature and small ε (see ref. 30). Now, with these notations the lattice analogue of (3.1) is

$$K^\mu[\sigma_x, \eta_x \mid \sigma_{x^c}\eta_{x^c}] = \frac{1}{\text{Norm}} \exp\left(\beta\left(\sum_{y\sim x} \sigma_y + \varepsilon\eta_x\right)\sigma_x\right) \cdot q_x^\mu(\eta_{x^c})^{\frac{\eta_x}{2}} \quad \text{where}$$

$$q_x^\mu(\eta_{x^c}) = \int \mu[\eta_1 = 1, \eta_{x^c}](d\hat{\sigma}_x)\, e^{-2\beta\varepsilon\hat{\sigma}_x} \tag{3.6}$$

This follows from ref. 17. In particular we have for the bias the expression

$$\frac{K^\mu[\eta_x = 1 \mid \sigma_{x^c}\eta_{x^c}]}{K^\mu[\eta_x = -1 \mid \sigma_{x^c}\eta_{x^c}]} = r_{\beta,\varepsilon}\left(\sum_{y\sim x} \sigma_y\right) \cdot q_x^\mu(\eta_{x^c}) \tag{3.7}$$

We note that all of the monotonicity arguments given for the mean field model stay correct; so the last expression is increasing in the nearest neighbor sum $\sum_{y\sim x} \sigma_y$ and decreasing in $\eta_{x^c} = (\eta_y)_{y\in\mathbb{Z}^d\setminus x}$. Of course the function $q_x^\mu(\eta_{x^c})$ is not explicitly given like the limiting forms of its mean-field analogue (when $N \uparrow \infty$). We also note that $q_x^{\mu^+}(\eta_{x^c}) \leqslant q_x^{\mu^-}(\eta_{x^c})$. This follows from the representation as finite volume limits with plus resp. minus boundary conditions, and the monotonicity in the boundary conditions.

The dependence of the conditional probability on the state μ is quite understandable, too. So, suppose we have observed a neutral realization of the spin-average of the neighbors of a site x, i.e., $\sum_{y\sim x}\sigma_y = 0$. Let us look at a typical realization of preferences η_{x^c} that are more or less neutral. Suppose we look at the state K^{μ^+}. It is overall plus like on the σ_y's; so we would expect $\eta_1 = +$ with a probability that is smaller than in the state K^{μ^-} that is overall minus like. This is natural because we would ascribe the value of η_x some partial responsibility for the neutral outcome of the neighbors of x.

3.2. Proofs

Proof of Theorem 2. We first remind the reader of some facts about the mean field random field Ising model. Let us put $\Phi^0_{\beta,\varepsilon}(m) \equiv \Phi^0_{\beta,\varepsilon,\alpha=0}(m)$ where the last function was introduced in (2.4). The large N behavior of the model is dominated by the minima of this function, for $\mathbb{P}$-typical η. This is seen by a Gaussian (Hubbard–Stratonovitch) transformation explained, e.g., in ref. 2. We won't repeat the details. This function $\Phi^0_{\beta,\varepsilon}$ has been well-analysed (see refs. 25 and 26). For "large magnetic fields" $\varepsilon > \frac{1}{2}$, it has only one global quadratic minimum at $m = 0$. For $0 \leqslant \varepsilon \leqslant \frac{1}{2}$ there exists a critical inverse temperature $\beta_c(\varepsilon)$ s.t. for $\beta > \beta_c(\varepsilon)$ the system has two symmetric global quadratic minima. We assume for this paper that β, ε are in this two phase regime. For $\beta < \beta_c(\varepsilon)$ the system has one global quadratic minimum at $m = 0$. At the phase transition line itself there are two regions: For small ε there is a unique global quartic minimum at $m = 0$, as for the usual CW ferromagnet; for large ε there are three global quadratic minima. These two line segments are separated by a tricritical point, where there is one global sixth order minimum.

Now, on the basis of Proposition 1, the proof of Theorem 2 follows immediately from the following proposition that provides control of the quantity $q_{\beta,\varepsilon,N}(w)$.

Proposition 2. Fix β, ε with $\beta > \beta_c(\varepsilon)$ and fix $0 < \delta < \frac{1}{6}$. Then there exists a constant $C(\beta, \varepsilon)$ and an integer $N_0 = N_0(\beta, \varepsilon)$ such that for all $N \geqslant N_0$ we have the uniform approximation

$$\sup_{w : |w| \leqslant N^{\frac{1+\delta}{2}}} |q_{\beta,\varepsilon,N}(w) - q_{\beta,\varepsilon,\infty}(w)| \leqslant C(\beta, \varepsilon)\, N^{-\frac{1}{4}+\frac{3\delta}{2}} \tag{3.8}$$

Remark. We have $r^*_{\beta,\varepsilon} \geqslant 1$ with strict inequality for $m^{\mathrm{RF}}(\beta, \varepsilon) > 0$.

Proof of Proposition 2. The first step is to use the following representation.

Lemma 1. For each $w \in \mathbb{Z}$ we have

$$q_{\beta,\varepsilon,N}(w) = \frac{\int \exp(-\beta N \Phi^0_{\beta,\varepsilon}(m))\, r_{\beta,\varepsilon}(m)^{\frac{w-1}{2}}\, dm}{\int \exp(-\beta N \Phi^0_{\beta,\varepsilon}(m))\, r_{\beta,\varepsilon}(m)^{\frac{w+1}{2}}\, dm} \tag{3.9}$$

But assuming the lemma, the form of the approximation of the Proposition is easily understood: Just approximate the integrals by their values at the unperturbed minimizer $\pm m^{\mathrm{RF}}(\beta, \varepsilon)$. This approximation is good as long as $|w|$ is not too big compared to N. Let us now proceed with the actual proof.

Proof of Lemma 1. This is a simple identity following as the result of the well-known Gaussian transformation (Hubbard–Stratonovitch-transformation) for the partition function explained for our model, e.g., in ref. 2. To make the connection to the formulae of ref. 2 the reader should note the following. In ref. 2 we introduced the quantities

$$\begin{aligned} L_{\beta,\varepsilon,-}(m) &= \frac{1}{2\beta} \log r_{\beta,\varepsilon}(m) \\ \Phi_{\beta,\varepsilon,N}(m, w) &= \Phi^0_{\beta,\varepsilon}(m) - L_{\beta,\varepsilon,-}(m)\, \frac{w}{N} \end{aligned} \tag{3.10}$$

Then the claim of the lemma is equivalent to

$$q_{\beta,\varepsilon,N}(w) = \frac{\int \exp(-\beta N \Phi_{\beta,\varepsilon,N}(m, w-1))\, dm}{\int \exp(-\beta N \Phi_{\beta,\varepsilon,N}(m, w+1))\, dm} \tag{3.11}$$

which is immediate when writing the l.h.s. as a quotient of partition functions and performing the HS-transformation as explained in ref. 2. ▮

We continue with the proof of Proposition 2 and consider balls around the minima $\pm m^{\mathrm{RF}}$ with radii $\rho_N := N^{-\frac{1}{4}+\frac{\delta}{2}}$. We denote their complement by $R_\rho := (B_\rho(m^{\mathrm{RF}}) \cup B_\rho(-m^{\mathrm{RF}}))^c$. Then we may write

$$q_{\beta,\varepsilon,N}(w) = \frac{I^+_\rho(w-1) + I^-_\rho(w-1) + J_\rho(w-1)}{I^+_\rho(w+1) + I^-_\rho(w+1) + J_\rho(w+1)} \tag{3.12}$$

with the integrals

$$\begin{aligned} I^\pm_\rho(w) &:= \int_{B_\rho(\pm m^{\mathrm{RF}})} \exp(-\beta N(\Phi_{\beta,\varepsilon,N}(m, w) - \Phi^0_{\beta,\varepsilon}(m^{\mathrm{RF}})))\, dm \\ J_\rho(w) &:= \int_{R_\rho} \exp(-\beta N(\Phi_{\beta,\varepsilon,N}(m, w) - \Phi^0_{\beta,\varepsilon}(m^{\mathrm{RF}})))\, dm \end{aligned} \tag{3.13}$$

We will have to estimate $I_\rho^\pm(w)$ from above and below and $J_\rho(w)$ from above. We just recall the results of ref. 2 that were shown by the use of the Taylor expansion around $\pm m^{\mathrm{RF}}$ and estimates on Gaussian integrals. They say that

$$\begin{aligned} I_\rho^\pm(w) \geqslant{} & \exp(\pm\beta L_-(m^{\mathrm{RF}})\, w) \\ & \times \sqrt{\frac{2\pi}{\beta N b_+(\rho)}} \left(\exp\left(\frac{z(w)^2\, \beta N}{2b_+(\rho)}\right) - 2\exp\left(-\frac{\beta N b_+(\rho)\, \rho^2}{4}\right)\right) \end{aligned} \tag{3.14}$$

For the upper bound we simply write

$$I_\rho^\pm(w) \leqslant \exp(\pm\beta L_-(m^{\mathrm{RF}})\, w) \sqrt{\frac{2\pi}{\beta N b_-(\rho)}} \exp\left(\frac{z(w)^2\, \beta N}{2b_-(\rho)}\right) \tag{3.15}$$

Here we have put

$$z(w) := \frac{w}{N} L_-'(m^{\mathrm{RF}}) \tag{3.16}$$

and

$$\begin{aligned} b_+(\rho) &:= \sup_{v,\,|v|\leqslant\rho} \Phi^{0\prime\prime}(m^{\mathrm{RF}}+v) + \left|\frac{w}{N}\right| \sup_{v,\,|v|\leqslant\rho} |L_-''(m^{\mathrm{RF}}+v)| \\ b_-(\rho) &:= \inf_{v,\,|v|\leqslant\rho} \Phi^{0\prime\prime}(m^{\mathrm{RF}}+v) - \left|\frac{w}{N}\right| \sup_{v,\,|v|\leqslant\rho} |L_-''(m^{\mathrm{RF}}+v)| \end{aligned} \tag{3.17}$$

We note that $|b_+(\rho_N) - b_-(\rho_N)| \leqslant \mathrm{Const}(\beta, \varepsilon)\, N^{-\frac{1}{4}+\frac{\delta}{2}}$ because the difference is dominated by the variation of the second derivation in the ball with radius ρ_N. Note also $|z| \leqslant \mathrm{Const}\, N^{-\frac{1}{2}+\frac{\delta}{2}}$

For the integral over the complement of the balls we have further shown in ref. 2 that for $N \geqslant N_0(\beta, \varepsilon)$ we have that

$$J_{\rho_N}(w) \leqslant \exp(-\mathrm{const}(\beta, \varepsilon)\, N^{\frac{1}{2}+\delta}) \tag{3.18}$$

But from this it is not difficult to derive (3.8). We only show the lower bound. First note that the integral over the outer region can be ignored, using that $|w| \leqslant N^{\frac{1}{2}+\frac{\delta}{2}}$. Note that $\exp(\beta L_-(m^{\mathrm{RF}})\, w) = (r^*_{\beta,\varepsilon})^{\frac{1}{2}}$ and recall the definition of $q_{\beta,\varepsilon,\infty}(w)$. We have then by our upper and lower estimates on the integrals over the balls the bound of the form

$$\frac{I_\rho^+(w-1)+I_\rho^-(w-1)}{I_\rho^+(w+1)+I_\rho^-(w+1)} \geqslant q_{\beta,\varepsilon,\infty}(w)\cdot \frac{e^{\frac{z(w-1)^2\beta N}{2b_+(\rho)}}-e^{-\text{const}\,N\rho_N^2}}{e^{\frac{z(w+1)^2\beta N}{2b_-(\rho)}}}\sqrt{\frac{b_-(\rho_N)}{b_+(\rho_N)}}$$
$$\geqslant q_{\beta,\varepsilon,\infty}(w)(1-\text{Const}(\beta,\varepsilon)\,N^{-\frac{1}{4}+\frac{3}{2}\delta}) \qquad (3.19)$$

This bound is obtained with the use of the worst case bound $\frac{w^2}{N^2}N\,|b_+(\rho_N)-b_+(\rho_N)|\leqslant \text{Const}'(\beta,\varepsilon)\,N^{-\frac{1}{4}+\frac{3}{2}\delta}$. This finishes the proof of Proposition 2 and consequently also the proof of Theorem 2. ▮

Proof of the Corollary to Theorem 2. It is elementary to see that the *lower bound* $w\geqslant C_-(\beta,\varepsilon)\log N$ implies $0\leqslant (r^*_{\beta,\varepsilon})^{-1}-q_{\beta,\varepsilon,\infty}(w)\leqslant C'(\beta,\varepsilon)\,N^{-\frac{1}{4}+\frac{3\delta}{2}}$. Now use the upper bound $w\leqslant N^{\frac{1+\delta}{2}}$ to apply the approximation of Theorem 2 to finish the proof. ▮

Proof of Theorem 6. We need to generalize some of the steps given in the proof of Theorem 2 to the case of $s\neq 0$. We will be faster here than before. Suppose without loss of generality that $s>0$. Note first that in generalization of Proposition 1 we have the representation of the conditional probability in the form

$$K^N_{\beta,\varepsilon,s}[\sigma_1,\eta_1\mid\sigma_{[2,N]}\eta_{[2,N]}]$$
$$=\frac{1}{\text{Norm}}\exp\left(\beta\left(\frac{1}{N}\sum_{i=2}^N\sigma_i+s+\varepsilon\eta_1\right)\sigma_1\right)\cdot q_{\beta,\varepsilon,s,N}\left(\sum_{i=2}^N\eta_i\right)^{\frac{\eta_1}{2}} \qquad (3.20)$$

where

$$q_{\beta,\varepsilon,s,N}\left(\sum_{i=2}^N\eta_i\right)=\frac{Z_{\beta,\varepsilon,s,N}[\eta_1=-1,\eta_{[2,N]}]}{Z_{\beta,\varepsilon,s,N}[\eta_1=1,\eta_{[2,N]}]}$$
$$=\int\mu_{\beta,\varepsilon,s,N}[\eta_1=+1,\eta_{[2,N]}](d\hat\sigma_1)\,e^{-2\beta\varepsilon\hat\sigma_1} \qquad (3.21)$$

The Hubbard–Stratonovitch transformation gives in this situation the representation

$$q_{\beta,\varepsilon,s,N}(w)=\frac{\int\exp(-\beta N\,\bar\Phi_{\beta,\varepsilon,s}(m))\,r_{\beta,\varepsilon}(m)^{\frac{w-1}{2}}\,dm}{\int\exp(-\beta N\,\bar\Phi_{\beta,\varepsilon,s}(m))\,r_{\beta,\varepsilon}(m)^{\frac{w+1}{2}}\,dm} \qquad (3.22)$$

with the "tilted function"

$$m \mapsto \bar{\Phi}_{\beta, \varepsilon, s}(m) := \Phi^0_{\beta, \varepsilon}(m) - sm \tag{3.23}$$

This is not to be confused with (2.4). It remains to do a saddle-point approximation for the integrals in (3.22). This modification needs a little more care. We have

Proposition 3. Fix β, ε in the two-phase region. Fix two auxiliary constants $\alpha, \delta > 0$ such that $\alpha + \delta < \frac{1}{2}$. Then there exist constants $C(\beta, \varepsilon)$, $C'(\beta, \varepsilon)$ and an integer $N_0 = N_0(\beta, \varepsilon)$ such that for all $N \geqslant N_0$ we have the uniform approximation

$$\sup_{\substack{w:\, |w| \leqslant N^{\frac{1+\delta}{2}} \\ s:\, N^{\frac{1}{2}+\delta} \leqslant s \leqslant N^{1-\alpha}}} |q_{\beta, \varepsilon, s, N}(w) - (r^*_{\beta, \varepsilon})^{-1}| \leqslant C(\beta, \varepsilon)\, N^{-\frac{1}{2}+2\delta} + C'(\beta, \varepsilon)\, N^{-\alpha} \tag{3.24}$$

Proof. Denote by $m^* = m^*(\beta, \varepsilon, s) > 0$ the global minimizer of the function $m \mapsto \Phi^0_{\beta, \varepsilon, s}(m)$ for $s > 0$. In the two-phase region it is unique. m^* is the mean magnetization of the system in the presence of the field s. This time we consider only one ball around m^* with radius

$$\rho_N := N^{-\frac{1}{2}+\delta} \tag{3.25}$$

Then we may write

$$q_{\beta, \varepsilon, s, N}(w) = \frac{I^+_\rho(w-1, s) + I^c_\rho(w-1, s)}{I^+_\rho(w+1, s) + I^c_\rho(w+1, s)} = \frac{\dfrac{I^+_\rho(w-1, s)}{I^+_\rho(w+1, s)} + \dfrac{I^c_\rho(w-1, s)}{I^+_\rho(w+1, s)}}{1 + \dfrac{I^c_\rho(w+1, s)}{I^+_\rho(w+1, s)}} \tag{3.26}$$

with the integrals

$$I^+_\rho(w, s) := \int_{|m-m^*| \leqslant \rho} dm \exp(-\beta N(\bar{\Phi}_{\beta, \varepsilon, s, N}(m, w) - \Phi^0_{\beta, \varepsilon, s}(m^*)))$$

$$I^c_\rho(w, s) := \int_{|m-m^*| \geqslant \rho} dm \exp(-\beta N(\bar{\Phi}_{\beta, \varepsilon, s, N}(m, w) - \Phi^0_{\beta, \varepsilon, s}(m^*)))$$

$$\text{with} \quad \bar{\Phi}_{\beta, \varepsilon, s, N}(m, w) = \bar{\Phi}_{\beta, \varepsilon, s}(m) - L_{\beta, \varepsilon, -}(m)\, \frac{w}{N} \tag{3.27}$$

The r.h.s. of (3.26) is approximately equal to the first term in the numerator and this term is close to $(r^*_{\beta,\varepsilon})^{-1}$. Let us discuss this in more detail. By the Taylor expansion around the s-dependent minimizer we have for $|v|\leqslant\rho$,

$$\bar\Phi_{\beta,\varepsilon,s,N}(m^*+v,w)-\bar\Phi_{\beta,\varepsilon,s}(m^*)+\frac{w}{N}L_-(m^*)$$
$$=\frac{\bar\Phi''_{\beta,\varepsilon,s}(m^*+\theta v)}{2}v^2-\frac{w}{N}L'_-(m^*)\,v-\frac{w}{N}\frac{L''_-(m^*+\theta' v)}{2}v^2 \quad (3.28)$$

with some $0\leqslant\theta,\theta'\leqslant 1$. Thus, on $|v|\leqslant\rho$ we have the upper and lower bounds

$$\text{l.h.s. of (3.28)}\leqslant\frac{b_+}{2}v^2-zv,\quad \geqslant\frac{b_-}{2}v^2-zv\quad\text{with}$$
$$z:=\frac{w}{N}L'_-(m^*),$$
$$b_+:=b_+(\rho):=\sup_{v,\,|v|\leqslant\rho}\bar\Phi''_{\beta,\varepsilon,s}(m^*+v)+\left|\frac{w}{N}\right|\sup_{v,\,|v|\leqslant\rho}|L''_-(m^*+v)| \quad (3.29)$$
$$b_-:=b_-(\rho):=\inf_{v,\,|v|\leqslant\rho}\bar\Phi''_{\beta,\varepsilon,s}(m^*+v)-\left|\frac{w}{N}\right|\sup_{v,\,|v|\leqslant\rho}|L''_-(m^*+v)|$$

We obtain from this for $\rho\geqslant 4|z|/b_+$ (which holds by our choice on ρ for N sufficiently large) after standard estimates on the tails of Gaussian estimates the bounds

$$I^+_\rho(w,s)\geqslant\exp(\beta L_-(m^*)\,w)\sqrt{\frac{2\pi}{\beta N b_+}}\left(\exp\left(\frac{z^2\beta N}{2b_+}\right)-2\exp\left(-\frac{\beta N b_+\rho^2}{4}\right)\right)$$
$$\leqslant\exp(\beta L_-(m^*)\,w)\sqrt{\frac{2\pi}{\beta N b_-}}\exp\left(\frac{z^2\beta N}{2b_-}\right) \quad (3.30)$$

As in (3.19) we obtain the estimate

$$\frac{I^+_{\rho_N}(w-1,s)}{I^+_{\rho_N}(w+1,s)}\geqslant(r_{\beta,\varepsilon}(m^*(\beta,\varepsilon,s)))^{-1}$$
$$\cdot\frac{e^{\frac{z(w-1)^2\beta N}{2b_+(\rho_N)}}-e^{-\text{const}\,N\rho_N^2}}{e^{\frac{z(w+1)^2\beta N}{2b_-(\rho_N)}}}\sqrt{\frac{b_-(\rho_N)}{b_+(\rho_N)}}$$
$$=(r_{\beta,\varepsilon}(m^*(\beta,\varepsilon,s)))^{-1}\,(1-\text{Const}(\beta,\varepsilon)\,\mathcal{O}(N^\delta\rho_N)) \quad (3.31)$$

and a similar upper bound.

Next we show that all other corrections are of lower order. Look at the term in the denominator on the r.h.s. of (3.26). It is not difficult to see that

$$\frac{I_\rho^c(w, s)}{I_\rho^+(w, s)} \leqslant e^{-\text{const}\,(\beta, \varepsilon)\, Ns} + e^{-\text{const}\,(\beta, \varepsilon)\, \rho^2 N} \tag{3.32}$$

The first term comes from the fact that we have for the difference between the local minima

$$\bar{\Phi}_{\beta, \varepsilon, s}(m_-^*(\beta, \varepsilon, s)) - \bar{\Phi}_{\beta, \varepsilon, s}(m^*(\beta, \varepsilon, s)) \sim 2s\, m^*(\beta, \varepsilon, s=0) \tag{3.33}$$

for $s \downarrow 0$ where we have denoted by $m_-^*(\beta, \varepsilon, s)$ the local minimum that is close to $-m_-^*(\beta, \varepsilon, s=0)$ for $s \downarrow 0$. This term dominates the w-dependent contributions because $|w| \ll s$ for large N. The second term in (3.32) is an estimate on the tail of a Gaussian random variable and it comes from the distance from the local minimum to the range of integration. The second term in the numerator of (3.26) is just bounded by a constant times the same expression, by the fact that the function r is bounded.

Finally we use that we have for the s-dependent shift of the minimum

$$|m^*(\beta, \varepsilon, s) - m^*(\beta, \varepsilon, s=0)| \leqslant \text{Const}(\beta, \varepsilon)\, s \tag{3.34}$$

and thus

$$|(r_{\beta, \varepsilon}(m^*(\beta, \varepsilon, s_N)))^{-1} - (r_{\beta, \varepsilon}^*)^{-1}| \leqslant \text{Const}(\beta, \varepsilon)'\, s_N \leqslant \text{Const}(\beta, \varepsilon)'\, N^{-\alpha} \tag{3.35}$$

This finishes the proof of Proposition 3, and consequently also the proof of Theorem 6. ▮

Proof of Theorem 1. We need to look at $q_{\beta, \varepsilon, N}(w)$ for $w_N = \alpha N$, with $\alpha > 0$. Indeed, using saddle point approximation arguments on the expression (3.12) like we did in the more complicated situation in the proof of Proposition 2 we arrive at

$$\lim_{N \uparrow \infty} q_{\beta, \varepsilon, N}(w_N) = r_{\beta, \varepsilon}(m^{\text{RF}}(\beta, \varepsilon, \alpha))^{-1} \tag{3.36}$$

recalling that $m^{\text{RF}}(\beta, \varepsilon, \alpha)$ is the (unique) minimizer of the function $m \mapsto \Phi_{\beta, \varepsilon, \alpha}^0(m)$. For the saddle point approximation to work we need to make sure that the minimum is unique, the function in the exponent is quadratic around the minimum, and also uniformly bounded below by (say) a quadratic function. All of these elementary analytical properties of the function $\Phi^0_{\beta,\varepsilon,\alpha}(m)$ are true (however we don't give details of this here). ∎

4. DECIMATION AND DILUTED FERROMAGNET–PROOFS

4.1. Decimation

Proof of Theorem 3 and Theorem 4. We start with the representation of the finite volume conditional probability for the decimation transformation. It reads

$$\mu_{\beta,N}(\sigma_1 \,|\, \sigma_{[2,M+1]}) = \int \tilde{\mu}_{\beta,N,M}\left(dm \,\Big|\, \sum_{i=2}^{M+1} \sigma_i\right) \frac{e^{\beta m \sigma_1}}{2\cosh(\beta m)} \tag{4.1}$$

where

$$\tilde{\mu}_{\beta,N,M}(dm \,|\, z) = \frac{1}{\text{Norm}} \exp\left(-\frac{\beta N m^2}{2} + (N-M)\log\cosh(\beta m) + \beta z m\right) dm \tag{4.2}$$

This is seen by Hubbard–Stratonovitch transformation. Let us assume that $0<p<1$ at first. To check the formula let us derive (2.17), giving back the unbiased measure with mean-field magnetisation for typical conditioning. To see this we write the negative exponent of the exponential under the integral in the form $\beta(N-M)(\frac{m^2}{2}-\frac{1}{\beta}\log\cosh(\beta m))+\frac{\beta M}{2}(m-\frac{z}{M})^2$ plus an m-independent constant.

We see from this: Conditioning on the positive or negative mean-field solution gives that the integral is concentrated on this conditioning, in the limit of large N. So we have

$$\tilde{\mu}_{\beta,N,M}(dm \,|\, z) \to \begin{cases} \delta_{+m^{\mathrm{CW}}(\beta)} \\ \delta_{-m^{\mathrm{CW}}(\beta)} \end{cases} \quad \text{if} \quad \frac{z}{M} \to \begin{cases} +m^{\mathrm{CW}}(\beta) \\ -m^{\mathrm{CW}}(\beta) \end{cases} \tag{4.3}$$

To prove Theorem 3 for general $\hat{m} \neq 0$ we use the following alternative representation of the measure $\tilde{\mu}_{\beta,N,M}$. It is obtained after a change of variable in the form

$$\int \tilde{\mu}_{\beta,N,M}(dm \,|\, z)\,\varphi(m) = \int \tilde{\mu}'_{\beta',h',N-M}(dy)\,\varphi\left(\frac{N-M}{N}\,y + \frac{z}{N}\right) \quad \text{with}$$

$$\beta' = \frac{N-M}{N}\,\beta, \qquad h' = \frac{z}{N-M} \tag{4.4}$$

where

$$\tilde{\mu}'_{\beta', h', N'}(dy) = \frac{1}{\text{Norm}} \exp\left(-\beta' N' \left(\frac{y^2}{2} - \frac{1}{\beta'} \log \cosh(\beta'(y+h')) \right) \right) dy \tag{4.5}$$

The function appearing in the exponent of (4.5) is well known. Its minimizer $m^{\mathrm{CW}}(\beta', h')$ is the mean-field magnetization of the Curie–Weiss ferromagnet in an external magnetic field h'. For $h' > 0$ it is well-known that the minimizer is unique. Letting N, M tend to infinity such that $\frac{N-M}{N} \to p > 0$ we thus get that

$$\tilde{\mu}_{\beta, N, M=(1-p)N}(dm \mid z) \to \delta_{p m^{\mathrm{CW}}(p\beta, \frac{1-p}{p}\hat{m}) + (1-p)\hat{m}}(dm) \qquad \text{if } \frac{z}{M} \to \hat{m} \tag{4.6}$$

Let's now turn to the case $p = 0$. We write the negative exponent of the exponential of (4.2) in the form $\beta N(\frac{1}{2}(m - \frac{z}{N})^2 - \frac{1}{\beta}\frac{N-M}{N} \log \cosh(\beta m))$ plus an m-independent constant. We have $z/N \to \hat{m}$ and $\frac{N-M}{N} \to 0$. So the measure $\tilde{\mu}_{\beta, N, M}(dm \mid z)$ converges weakly to $\delta_{\hat{m}}$. This finishes the proof for $p = 0$.

Next we consider $p = 1$. We look again at the formula given in the text below (4.2) for the negative exponent in (4.2). From this we see that the measure $\tilde{\mu}_{\beta, N, M}(dm \mid z)$ converges weakly to $\delta_{m^{\mathrm{CW}}(\beta) \cdot \operatorname{sign}(\hat{m})}$, whenever $M \uparrow \infty$. The assumption $M \uparrow \infty$ is necessary, for finite M the limiting measure is a convex combination of $\delta_{+m^{\mathrm{CW}}(\beta)}$ and $\delta_{-m^{\mathrm{CW}}(\beta)}$. This proves Theorem 3. To prove Theorem 4 where z stays fixed, we write

$$\mu_{\beta, N}(\sigma_1 \mid \sigma_{[2, M+1]}) = \frac{\int \tilde{\mu}_{\beta, N, M}(dm \mid z = 0)\, e^{\beta z m} e^{\beta m \sigma_1} / (2 \cosh(\beta m))}{\int \tilde{\mu}_{\beta, N, M}(dm \mid z = 0)\, e^{\beta z m}} \tag{4.7}$$

Then we use $\tilde{\mu}_{\beta, N, M_N}(\cdot \mid z = 0) \to \frac{1}{2}(\delta_{-p m^{\mathrm{CW}}(p\beta)} + \delta_{p m^{\mathrm{CW}}(p\beta)})$, for any $0 \leqslant p \leqslant 1$. ∎

4.2. Diluted Ferromagnet

Proof of Theorem 5. Here the formula for the one-site conditional probabilities can be written in the form

$$K^N_{\beta, p}[\sigma_1, n_1 \mid \sigma_{[2,N]} n_{[2,N]}] \\ = \frac{1}{\text{Norm}} p^{n_1} (1-p)^{1-n_1} \exp\left(\frac{\beta}{N} \sum_{i=2}^{N} n_i \sigma_i \cdot n_1 \sigma_1 \right) \cdot w_{\beta, N} \left(\sum_{i=2}^{N} n_i \right)^{n_1} \tag{4.8}$$

where

$$w_{\beta,N}(M) := \left(\int \mu_{\beta',M}(d\hat{\sigma}) \exp\left(\frac{\beta'}{M} \sum_{i=1}^{M} \hat{\sigma}_i \right) \right)^{-1} \qquad \text{with} \quad \beta' = \beta M/N \tag{4.9}$$

The normalization is given by summing over $n_1 = 0, 1$ and $\sigma_1 = \pm 1$.

Remark. In particular we have for the "occupation-bias" the formula

$$\frac{K^N_{\beta,p}[n_1 = 1 \mid \sigma_{[2,N]} n_{[2,N]}]}{K^N_{\beta,p}[n_1 = 0 \mid \sigma_{[2,N]} n_{[2,N]}]} = \frac{p}{1-p} \cosh\left(\frac{\beta}{N} \sum_{i=2}^{N} n_i \sigma_i \right) \cdot w_{\beta,N}\left(\sum_{i=2}^{N} n_i \right) \tag{4.10}$$

To derive (4.8) with (4.9) proceed as in the random field model. To express the resulting fractions of partition functions use the following: For any $n_{[2,N]}$ such that $\sum_{i=2}^{N} n_i = M$ we have that $Z_N[n_1 = 0, n_{[2,N]}]/Z_N[n_1 = 1, n_{[2,N]}] = e^{-\frac{\beta}{2N}} w_{\beta,N}(M)$.

Now, it is easy to see that $\lim_{M \uparrow \infty} \int \mu_{\beta',M}(d\hat{\sigma}) \exp(\frac{\beta'}{M} \sum_{i=1}^{M} \hat{\sigma}_i) = \cosh(\beta' m^{\mathrm{CW}}(\beta'))$ since the distribution of the magnetization of the ordinary Curie–Weiss model concentrates on plus or minus the mean-field value. We also have $\lim_{N \uparrow \infty} w_{\beta,N}(qN)^{n_1} = \cosh(\beta q \, m^{\mathrm{CW}}(\beta q))^{n_1} = \cosh(\beta q \, m^{\mathrm{CW}}(\beta q) \, n_1)$. This proves Theorem 5. ∎

Remark. Let us finally compare the formula for the conditional probabilities with the lattice analogue with formal Boltzmann weights $\propto \exp(\beta \sum_{\langle x,y \rangle} n_x \sigma_x n_y \sigma_y)$. For a joint measure K^μ obtained from a Gibbs measure μ it acquires the following form

$$K^\mu[\sigma_x, n_x \mid \sigma_{x^c} n_{x^c}] = \frac{1}{\text{Norm}} \exp\left(\beta \sum_{y \sim x} n_y \sigma_y \cdot n_x \sigma_x \right) \cdot w^\mu_x(n_{x^c})^{n_x} \quad \text{where}$$

$$w^\mu_x(n_{x^c}) = \left(\int \mu[n_x = 0, n_{x^c}](d\hat{\sigma}) \exp\left(\beta \sum_{y \sim x} n_y \hat{\sigma}_y \cdot \hat{\sigma}_x \right) \right)^{-1} \tag{4.11}$$

This representation also follows from ref. 17. In particular we have from that for the "occupation-bias"

$$\frac{K^\mu[n_x = 1 \mid \sigma_{x^c} n_{x^c}]}{K^\mu[n_x = 0 \mid \sigma_{x^c} n_{x^c}]} = \frac{p}{1-p} \cosh\left(\beta \sum_{y \sim x} n_y \sigma_y \right) \cdot w^\mu_x(n_{x^c}) \tag{4.12}$$

As for the joint measures of the random field model mean-field and lattice expressions look similar. Now, writing $w_x^{\mu}(n_{x^c}) = \int \mu[n_x = 1, n_{x^c}](d\hat{\sigma}) \exp(-\beta \sum_{y\sim x} n_y \hat{\sigma}_y \cdot \hat{\sigma}_x))$ and choosing the specific configuration of ref. 15 for n_{x^c} made of two separate clusters it is possible to see that (4.12) is indeed not a quasilocal function. This however is an example of a typical lattice effect and can not be mimicked in the mean field model.

ACKNOWLEDGMENTS

The author thanks A.van Enter for interesting comments on an earlier draft.

Work supported by the DFG priority program "Wechselwirkende stochastische Systeme hoher Komplexität."

REFERENCES

1. C. Borgs, J. T. Chayes, H. Kesten, and J. Spencer, The birth of the infinite cluster: Finite-size scaling in percolation, *Comm. Math. Phys.* **224**:153–204 (2001).
2. C. Külske, Metastates in disordered mean-field models: Random field and Hopfield models, *J. Stat. Phys.* **88**:1257–1293 (1997).
3. C. Külske, Metastates in disordered mean-field models II: The superstates, *J. Stat. Phys.* **91**:155–176 (1998).
4. C. M. Newman and D. L. Stein, Non-mean-field behavior of realistic spin glasses, *Phys. Rev. Lett.* **76**:515–518 (1996).
5. C. M. Newman and D. L. Stein, Spatial inhomogeneity and thermodynamic chaos, *Phys. Rev. Lett.* **76**:4821–4824 (1996).
6. A. Bovier, *Statistical Mechanics of Disordered Systems*, MaPhySto lecture notes, Vol. 10, available at http://www.maphysto.dk
7. M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*, World Scientific Lecture Notes in Physics, Vol. 9 (1987).
8. C. M. Newman and D. L. Stein, The state(s) of replica symmetry breaking: Mean field theories vs. short-ranged spin glasses, *J. Stat. Phys.* **106**:213–244 (2002).
9. E. Marinari, G. Parisi, F. Ricci-Tersenghi, J. J. Ruiz-Lorenzo, and F. Zuliani, Replica symmetry breaking in short-range spin glasses: Theoretical foundations and numerical evidences, *J. Stat. Phys.* **98**:973–1047 (2000).
10. A. C. D. van Enter and J. Lörinczi, Robustness of the non-Gibbsian property: Some examples, *J. Phys. A* **29**:2465–2473 (1996).
11. T. Morita, Statistical mechanics of quenched solid solutions with application to magnetically dilute alloys, *J. Math. Phys.* **5**:1402–1405 (1964).
12. M. Serva and G. Paladin, Gibbs thermodynamic potentials for disordered systems, *Phys. Rev. Lett.* **70**:105–108 (1993).
13. R. Kühn, Equilibrium ensemble approach to disordered systems I: General theory, exact results, *Z. Phys.* **100**:231–242 (1996).
14. R. Kühn and G. Mazzeo, Critical behavior of the randomly spin diluted 2D Ising model: A grand ensemble approach, *Phys. Rev. Lett.* **73**:2268–2271 (1994).
15. A. C. D. van Enter, C. Maes, R. H. Schonmann, and S. Shlosman, The Griffiths singularity random field, *On Dobrushin's Way: From Probability Theory to Statistical Physics*,

Amer. Math. Soc. Transl. Serie 2, Vol. 198 (Amer. Math. Soc. Providence, P.I., 2000), pp. 51–58.
16. C. Külske, (Non-)gibbsianness and phase transitions in random lattice spin models, *Markov Process. Related Fields* **5**:357–383 (1999).
17. C. Külske, Weakly Gibbsian representations for joint measures of quenched lattice spin models, *Probab. Theory Related Fields* **119**:1–30 (2001).
18. A. C. D. van Enter, C. Maes, and C. Külske, Comment on ref. 14, *Phys. Rev. Lett.* **84**:6134 (2000).
19. R. Kühn and G. Mazzeo, Reply to ref. 18, *Phys. Rev. Lett.* **84**:6135 (2000).
20. A. C. D. van Enter, R. Fernández, and A. D. Sokal, Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitations of Gibbsian theory, *J. Stat. Phys.* **72**:879–1167 (1993).
21. C. Maes, F. Redig, S. Shlosman, and A. van Moffaert, Percolation, path large deviations and weak Gibbsianity, *Comm. Math. Phys.* **209**:517–45 (2000).
22. J. Bricmont, A. Kupiainen, and R. Lefevere, Renormalization group pathologies and the definition of Gibbs states, *Comm. Math. Phys.* **194**:359–388 (1998).
23. C. Külske, Regularity properties of potentials for joint measures of random spin systems, WIAS preprint0781, available at http://www.ma.utexas.edu/mp\–arc/preprint 02–519, to be published in *Markov Process. Related Field.*
24. C. Külske, A. Le Ny, and F. Redig, Variational principle for generalized Gibbs measures, Eurandom preprint 2002-035 (2002), available at http://www.ma.utexas.edu/mp\–arc/preprint 02–466.
25. S. R. Salinas and W. F. Wreszinski, On the mean-field Ising model in a random external field, *J. Stat. Phys.* **41**:299–313 (1985).
26. J. M. G. Amaro de Matos, A. E. Patrick, and V. A. Zagrebnov, Random infinite-volume Gibbs states for the Curie–Weiss random field Ising model, *J. Stat. Phys.* **66**:139–164 (1992).
27. R. Fernández, A. Le Ny, and F. Redig, Variational principle and almost quasilocality for renormalized measures, *J. Stat. Phys* **111**:465–477 (2003).
28. H. O. Georgii, O. Häggström, and C. Maes, The random geometry of equilibrium phases, in *Phase Transitions and Critical Phenomena,* C. Domb and J. L. Lebowitz (eds.), Vol. 18 (Academic Press, London, 2000), pp. 1–142.
29. A. C. D. van Enter, R. Fernández, F. den Hollander, and F. Redig, Possible loss and recovery of Gibbsianness during the stochastic evolution of Gibbs measures, *Comm. Math. Phys.* **226**:101–130 (2002).
30. J. Bricmont and A. Kupiainen, Phase transition in the 3d random field Ising model, *Commun. Math. Phys.* **142**:539–572 (1988).