# Weierstraß-Institut
# für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633


# Analogues of non-Gibbsianness in joint measures of disordered mean-field models


Christof Külske*[1]



submitted: 17. Dezember 2002



[1] Weierstrass Institute
for Applied Analysis
and Stochastics
Mohrenstrasse 39
10117 Berlin, Germany
E-Mail: kuelske@wias-berlin.de



No. 800
Berlin 2002



2000 *Mathematics Subject Classification.* 82B44, 82B26, 82B20
1999 Physics and Astronomy Classification Scheme (PACS) 05.50.+q, 61.43.-j, 02.50.Cw.

*Key words and phrases.* Disordered systems, non-Gibbsian measures, mean field models, Morita-approach, random field model, decimation transformation, diluted ferromagnet.

*Work supported by the DFG priority program 'Interacting stochastic systems of high complexity'.

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

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

**Abstract:** 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 expectations in finite volume of the following mean field models: a) joint measures of random field Ising, b) joint measures of dilute Ising, c) decimation of ferromagnetic Ising. Observing that the conditional expectations are functions of the empirical mean of the conditionings we look at the large volume 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 expectation 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.

## 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. 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 [BCKS01]. Simple mean-field models of disordered systems were also used earlier by the author to illustrate the (supposed) asymptotic behavior of the Gibbs measures of the corresponding lattice models as described by the corresponding metastates [K97],[K98], a notion introduced in [NS96a,b]. For an excellent overview about further related results also in more complicated situations we refer the readers to [Bo01] 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 [MePaVi87] itself is mathematically not justified and not at all simple. (For the current status of the ongoing discussion on this fascinating topic see [NS02], [MPRRZ00].) 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, and 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 allow for very simple explicit computations. We believe that because of their simplicity 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, since there is no proper Gibbsian structure anyway. Our important point is that when "Non-Gibbsianness" for mean-field models is understood as "discontinuity of conditional expectations as a function of the conditioning", it becomes a meaningful and natural notion when it is taken in the appropriate sense. Note that 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 [EL96], so one could feel discouraged to look for non-trivial continuity properties in conditional expectations 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 expectations 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 and look at the size of the set of their discontinuity points in the large volume-limit.

The functional dependence of the limiting form of the conditional expectations 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, but non-trivial continuity properties of its conditional expectations in the above sense, as will become explicit later.

Our most interesting example however is the joint measure of the random field Ising lattice model, and 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 [Mo64] (see also [SP93],[Ku97],[KM94]). Much later, starting from the first example of the dilute Ising ferromagnet [EMSS00], it was discovered in [K99],[K01] that such joint measures provided a whole class of examples of non-Gibbsian measures, in low temperature situations, see also [EKM00],[KM00]. This situation has some analogy with the much discussed non-Gibbsian behavior of images of low-temperature lattice spin measures under renormalization group transformations (for the latter see [EFS93],[MRSV00],[BKL98] 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 [K02]) that are parallel to the known statements for renormalized measures [EFS93]. 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* [K01], however without any 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 expectations happens is not negligible but has even full measure ([K99], [K01]).

**2.** The functional form of the conditional expectations depends on whether the system is in the plus- or minus-phase (see [K01], [KLR02]).

As explained in [KLR02] Property 2 implies the failure of the Gibbs variational principle. Since 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. 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 that is defined in terms of the continuity properties of their conditional expectations.

At first sight Properties 1 and 2 might seem not very intuitive, or at least unusual.

In fact the expression of the conditional expectations 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 [SW85],[APZ92],[Ku96], and so we can partially draw from standard estimates, but our perspective is new and the conditional expectations 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 is the mean-field analogue of the decimation transformation. For this measure the set of *continuity* points on the lattice is of full measure [FLNR02]. Moreover we discuss the mean field analogue of the 'GriSing-field' [EMSS00]. It is very simple to see that for both mean field models we get full measure *continuity* points, and 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 and 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 expectations 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

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

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* $(\sigma_i, \eta_i)$ taking values in the set $\{(-1,-1),(-1,1),(1,-1),(1,1)\}$. We are interested in the behavior of the one-site conditional expectation

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

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

$$m \mapsto \Phi^0_{\beta,\varepsilon,\alpha}(m) := \frac{m^2}{2} - \tilde{\mathbb{E}}_\alpha\Big(\log\cosh(\beta(m+\varepsilon\tilde{\eta}_1))\Big) \tag{2.4}$$

where $\tilde{\eta}_1 = \pm 1$ is a dummy random field variable with expectation $\tilde{\mathbb{E}}_\alpha(\tilde{\eta}_1) = \alpha$. Denote by $m^{RF}(\beta,\varepsilon)$ 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 $\pm m^{RF}(\beta,\varepsilon)$ are the only two different global minima.

Let us first recall how the measures look in the weak infinite-volume limit: From [K97] follows in particular that the finite-dimensional marginals of $K^N_{\beta,\varepsilon}$ converge to the symmetric linear combination of product measures $\frac{1}{2}\Big(K^{prod,+}_{\beta,\varepsilon} + K^{prod,-}_{\beta,\varepsilon}\Big)$. Here $K^{prod,\pm}_{\beta,\varepsilon}$ are the product measures over the joint configurations with

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

Now we state and discuss the asymptotics of the conditional expectations 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).**

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

*where $m^{RF}(\beta,\varepsilon,\alpha)$ is the global minimizer of the function (2.4). The normalization is obtained by summing over $\sigma_1,\eta_1$. 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.* $m^{RF}(\beta,\varepsilon,\alpha)$ is the magnetization of a random field Curie-Weiss model with biased i.i.d. random fields such that $\mathbb{E}_\alpha(\eta_i) = \alpha \neq 0$.

*Remark.* We note that the conditional expectation is continuous everywhere in the single variable $\hat{m}$, but it is not continuous everywhere when we consider it as a function of $\alpha$. This is identical to the lattice case. The continuity for $\alpha \neq 0$ follows from the elementary fact that, for $\beta,\varepsilon$ in the two phase region (for the model with $\alpha = 0$) the minimizer $m^{RF}(\beta,\varepsilon,\alpha)$ varies continuously as a function of $\alpha$. It however jumps at $\alpha = 0$ and changes sign, and we have $m^{RF}(\beta,\varepsilon,\alpha = 0+) = m^{RF}(\beta,\varepsilon) = -m^{RF}(\beta,\varepsilon,\alpha = 0-)$.

Of course, the probability that the normalized $\eta$-sum takes values $\alpha$ 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 expectations (2.6) will acquire the two limiting forms for $\alpha = 0+$ and $\alpha = 0-$ that are manifestly different. So, an infinitesimal variation of $\alpha$ around zero leads to discontinuous behavior in the conditional expectation. Since $\alpha = 0$ is typical, we have an almost sure discontinuity is 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 be more precise here than in the simpler Theorem 1 given before and provide a uniform bound on the deviation from the limiting expression.

Introduce the regular set of random field configurations

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

for some fixed $0 < \delta < \frac{1}{6}$. We note that it is a very large set for large $N$ since we have $\mathbb{P}(\bar{\mathcal{H}}(N)) \geq 1 - 2e^{-\frac{N^\delta}{2}}$.

**Theorem 2 (Close-up of discontinuity region - Almost Sure Discontinuity).** *We have the uniform approximation*

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

*Here the limiting expression is given by*

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

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

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

*Remark.* Note that (2.9) is a function of the *non-normalized* sum of random fields in the conditioning. So, while the limiting expression (2.9) is continuous in the empirical mean of the spin conditioning $\hat{m}$, 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 $\frac{1}{N}$.

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 $\eta$-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

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

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*

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

*with the two components of 'regular realizations' given by*

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

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

*Remark.* Note that $\mathbb{P}\left(\mathcal{H}^{+}(N)\cup\mathcal{H}^{-}(N)\right) \ge 1 - const\,\frac{\log N}{\sqrt{N}}$ goes to one for large $N$, but it does so much slower than the set $\bar{\mathcal{H}}(N)$.

*Remark.* The corollary says in brief: "For all spin-conditionings $\sigma$ and all random field conditioning outside of a set with vanishing mass in the large volume limit, the conditional expectation is discontinuous in the empirical random field sum." This discontinuity in the random field-sum reflects the behavior of the lattice model where the conditional expectations 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 in section 3.2.

### 2.2 Decimation of Standard Curie Weiss

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

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

We look at the *decimated model* which is just the marginal $\mu_{\beta,N}(\sigma_{[1,M+1]})$ on the first $M+1$ spins. We are interested in the asymptotic behavior of the conditional expectation when $M = M_N$ grows like a multiple of $N$ in the limit $N \uparrow \infty$. This is a natural analogue of decimation to a sublattice. Denote by $m^{CW}(\beta)$ the ordinary Curie-Weiss magnetisation, i.e. the largest solution of the mean-field equation $m = \tanh(\beta m)$. It is well-known that $\lim_{N\uparrow\infty} \mu_{\beta,N} = \frac{1}{2}\left(\mu_\beta^+ + \mu_\beta^-\right)$ with the product measures given $\mu_\beta^\pm(\sigma_i) = \frac{e^{\pm\beta m^{CW}(\beta)\sigma_i}}{2\cosh(\beta m^{CW}(\beta))}$. 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 that $\lim_{N,M\uparrow\infty} \mu_{\beta,N}(\sum_{i=1}^{M+1} \sigma_i \in \cdot) = \frac{1}{2}\left(\delta_{m^{CW}(\beta)} + \delta_{-m^{CW}(\beta)}\right)$. This makes particularly clear that it doesn't suffice to look at the limiting measures to see "non-Gibbsianness". Let us start with the behavior outside of the discontinuity region.

**Theorem 3 (Almost Sure Continuity in the Conditioning for Decimation).** *Assume that* $\lim_{N\uparrow\infty} M_N/N = 1-p$. *Then*

$$\begin{aligned} &\lim_{N\uparrow\infty} \mu_{\beta,N}(\sigma_1|\sigma_{[2,M_N+1]}) = \frac{e^{\beta h_{\beta,p}(\hat{m})\sigma_1}}{2\cosh(\beta h_{\beta,p}(\hat{m}))}, \quad \textit{if } \lim_{M\uparrow\infty} \frac{1}{M}\sum_{i=2}^{M+1} \sigma_i = \hat{m} \neq 0 \\ &\textit{where } h_{\beta,p}(\hat{m}) = p m^{CW}\left(p\beta, \frac{1-p}{p}\hat{m}\right) + (1-p)\hat{m} \end{aligned} \tag{2.16}$$

*where $m^{CW}(\beta', h')$ is the solution of the mean-field equation of the Curie-Weiss model in an external magnetic field $h'$, i.e. $m = \tanh(\beta'(m+h'))$, that has the sign of $h'$. In particular, the limiting form of the conditional expectation 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.* For a check-up we see that by conditioning on typical configurations we get back the product measures with mean-field magnetization, i.e.

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

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

*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 expectation as a function of $\hat{m}$ corresponds to the situation in the lattice model. The reader familiar with the latter will recall that 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. Also in the mean-field model 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 iff the inverse temperature for the part of the system that is integrated out, $\beta p$ is smaller than one.

*Remark.* In contrast to 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 statement here for conditionings such that $\sum_{i=2}^{M_N} \sigma_i$ is subextensively small that shows how the interpolation at $\hat{m} = \pm 0$ looks on a finer scale. For reasons of simplicity let us formulate the result for conditionings such that $\sum_{i=2}^{M_N} \sigma_i = z$ is even constant. 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 that* $\lim_{N\uparrow\infty} M_N/N = 1 - p$. *Then*

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

Of course, in order for this statement to make, $M_N$ need to be all even, or all odd, depending on $z$. Again, the r.h.s. is a function of the non-normalized sum of the spins $z$ in the conditioning. 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 see how the mean-field version behaves. The model is given by the Gibbs measures

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

Here the $n_i$ or independent Bernoulli occupation numbers $\mathbb{P}[n_i = 1] = 1 - \mathbb{P}[n_i = 0] = p$. Of course, (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

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

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

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

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

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

$$\begin{aligned}&\lim_{N\uparrow\infty} K^N_{\beta,p}[\sigma_1, n_1 | \sigma_{[2,N]} n_{[2,N]}] = \frac{1}{Norm.}\, p^{n_1}(1-p)^{1-n_1}\frac{\exp\left(\beta \hat{m}\, n_1 \sigma_1\right)}{\cosh\big(\beta q\, m^{CW}(\beta q) n_1\big)}\\ &\textit{if } \lim_{N\uparrow\infty}\frac{1}{N}\sum_{i=2}^N n_i\sigma_i = \hat{m}, \qquad \lim_{N\uparrow\infty}\frac{1}{N}\sum_{i=2}^N n_i = q > 0\end{aligned} \tag{2.23}$$

*In particular, the limiting expression (2.6) varies continuously as a function of the pair $(\hat{m}, q)$ of the empirical magnetization and density of occupied sites.*

*Remark.* Doing our check-up of the formula by conditioning on typical configurations like we did for the other two models we see that indeed

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

*General Comment.* Theorems 1,3,5 immediately carry over from single-site conditional expectations to conditional expectations outside any finite subset. Here one obtains convergence to product measures ("propagation of chaos"), with the same parameters depending on the conditioning. E.g. for the decimation case we have

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

for the random field model and the dilute ferromagnet it is analogous. (2.25) follows recursively from Theorem 3 with the use of the formula $\rho(\sigma_{[1,k]}|\sigma_{[k+1,n]}) = \rho(\sigma_k|\sigma_{[1,k-1]}\sigma_{[k+1,n]})\times \left(\sum_{\tilde{\sigma}_k} \rho(\tilde{\sigma}_k|\sigma_{[1,k-1]}\sigma_{[k+1,n]})/\rho(\sigma_{[1,k-1]}|\tilde{\sigma}_k\sigma_{[k+1,n]})\right)^{-1}$ for any measure $\rho$, since 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. Now, we have for such 'typical' $\eta$ the approximation for the random finite volume Gibbs measures of the simple form

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

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

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

They play the role of infinite-volume Gibbs measures. [In this heuristic discussion $C$ indicates a positive random quantity of order unity.] This approximation is made rigorous in [K97] and asymptotic statements about metastates are derived from it (including the arcsine-law for the empirical metastate). Now, formally this approximation implies for the joint measure for typical $\eta$

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

These joint product measures have the concentration properties that

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

This clearly follows from the mean field equation which can be written $\pm m^{RF}(\beta,\varepsilon) = \int P(d\eta)\int \mu^{prod,\pm}_{\beta,\varepsilon}[\eta](d\sigma_i)\sigma_i$. Note that the empirical mean of the fields have arbitrary signs. By (2.28) it follows that the original joint measure then concentrates on the union of the two sets where $\frac{1}{N}\sum_{i=2}^N(\eta_i,\sigma_i) \approx \pm(C/\sqrt{N}, m^{RF}(\beta,\varepsilon))$, 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) the simple heuristic formula

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

But now we see that for joint configurations in these two joint lumps the expression given in the Corollary to Theorem 2 reduces to the *same* formula.

### 2.5 Random field model: Dependence of conditional expectations on the phase

A different feature of the lattice random field Ising model is the dependence of the conditional expectation of a joint measure on the phase $\mu^+$ resp. $\mu^-$. 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) in order to pick either the plus or the minus state. This leads to a different form of the limiting expression for the conditional expectation, depending on the sign of $s_N$. The precise result of this is given in Theorem 6.

So, let us define

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

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 expectations on the phase).** *Fix any constants $\alpha,\delta>0$ such that $\alpha+\delta<\frac{1}{2}$. Then we have the approximation*

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

*Remark.* This has to be interpreted as analogue of the dependence of the conditional expectation 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 2,6,1 (starting with the slightly more subtle estimates first). The first simple step is a formula for the one-site conditional expectations in finite volume. It is the simpler analogue of Proposition 3.1 in [K01] for lattice-spin models in finite volume for a mean-field model. We take the chance and provide some discussion about this formula, and its monotocity properties and also give a heuristic explanation for them. We also compare it to its lattice analogue and give a heuristic explanation of monotonicity and the dependence on the phase-phenomenon on the lattice. After that we provide details about the saddle-point estimates used to prove the theorems.

### 3.1 Representation of conditional expectation and monotonicity

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

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

*where*

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

*with the obvious normalization obtained by summing over $\sigma_1$ and $\eta_1$.*

*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, and 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 [GHM00] (Holley's theorem) because of the monotonicity of the the single-site conditional expectations, i.e. $\mu_{\beta,\varepsilon,N}[\eta_1=+1,\eta_{[2,N]}](\sigma_i=+|\sigma_{[1,N]\setminus i})\leq\mu_{\beta,\varepsilon,N}[\eta_1=+1,\eta'_{[2,N]}](\sigma_i=+|\sigma'_{[2,N]\setminus i})$ for all $\eta_{[2,N]}\leq\eta'_{[2,N]}$, for all $\sigma_{[1,N]\setminus i}\leq\sigma'_{[1,N]\setminus i}$.

*Remark.* The reader may think of $h = \frac{1}{2}\log q$ field acting on $\eta_1$ 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 expectation. 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 $Z_{\beta,\varepsilon,N}[\eta_1 = -1, \eta_{[2,N]}]/Z_{\beta,\varepsilon,N}[\eta_1 = 1, \eta_{[2,N]}] = q_{\beta,\varepsilon,N}\Bigl(\sum_{i=2}^N \eta_i\Bigr)$.$\diamondsuit$

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

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

Note that we have the symmetry $B_{\beta,\varepsilon,N}[\sigma_{[2,N]}, \eta_{[2,N]}] = \Bigl(B_{\beta,\varepsilon,N}[-\sigma_{[2,N]}, -\eta_{[2,N]}]\Bigr)^{-1}$. To avoid trivial misunderstandings let us point out that, of course, by dropping the spin-conditioning, we get the unbiased expression $K_{\beta,\varepsilon,N}[\eta_1 = +|\eta_{[2,N]}]/K_{\beta,\varepsilon,N}[\eta_1 = -|\eta_{[2,N]}] = \frac{1}{2}$. The bias contains all the interesting non-trivial behavior of the conditional expectations (3.1) because of the trivial identity $K^N_{\beta,\varepsilon}[\sigma_1, \eta_1|\sigma_{[2,N]}\eta_{[2,N]}] = \mu_{\beta,\varepsilon,N}[\eta_1\eta_{[2,N]}](\sigma_1|\sigma_{[2,N]}) \cdot K^N_{\beta,\varepsilon}[\eta_1|\sigma_{[2,N]}\eta_{[2,N]}]$ 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

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

Here we have defined the function

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

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 this monotonicity. This might be a little more intuitive if we adopt a voter model interpretation of our model: Each index $i \in \{1, \dots, N\}$ stands for a voter. There are two political parties, called $1, -1$. Each voter $i$ has a frozen 'long-term preference' $\eta_i$ taking values in $1, -1$, modelling preferences to one of the political parties $1, -1$. Now, each voter $i$ acquires his current opinion $\sigma_i$ as a consequence of a random interacting decision making procedure according to the rules: Keep the long-term preference $\eta_i$ fixed but try to be similar to the current opinions $(\sigma_j)_{j\neq i}$ such that $\text{Prob}(\sigma_i = +|(\sigma_j)_{j\neq i})/\text{Prob}(\sigma_i = -|(\sigma_j)_{j\neq i}) = \exp\Bigl(\frac{2\beta}{N}\sum_{j\neq i}\sigma_j + 2\beta\varepsilon\eta_i\Bigr)$. The resulting measure describing the situation in equilibrium is then the model (2.1). The joint measure $K_{\beta,\varepsilon,N}$ then describes the two-step random procedure in which first the long-term preferences are produced according to the equidistribution and then the decision making process as described above happens.

Let us suppose that all persons participate in the interactive decision making process, but one person $i = 1$ is missing on election day. Suppose on election day not only the votes but moreover also the long-term preferences are asked (and faithfully answered). This gives a realization of the two quantities $\sum_{i=2}^N \sigma_i$ and $\sum_{i=2}^N \eta_i$, sampled from the joint measure. When we ask ourselves for the distribution of decision $\sigma_1$ and long-term preference $\eta_1$ of the missing person, given these two observed quantities, we see that it is described by (3.1). Let us interpret the monotonicity properties of the bias (3.3) for the long-term preference $\eta_1$ in this language. Note that the conditioning on $\sum_{i=2}^N \sigma_i$ destroys the independence of $\eta_1$ from $\sum_{i=2}^N \eta_i$. Indeed, when the latter sum *increases* we should expect that $\eta_1$ will acquire a *lesser* tendency to be plus when the same result of the election $\sum_{i=2}^N \sigma_i$ is observed. This is because increasing the $\eta$-sum will increase the tendency of $\sigma_i$'s to be plus and so the probability for $\eta_1 = -$ should also increase to make up for this. This is particularly clear in a situation where the decision happened to be a tie $\sum_{i=2}^N \sigma_i = 0$, while the preferences happened to be positive $\sum_{i=2}^N \eta_i > 0$. Then we expect quite naturally $\eta_1 = -$ with probability bigger than $\frac{1}{2}$, because we would ascribe to the value of $\eta_1$ some partial responsibility for the neutral outcome.

A somewhat similar "Bayesian destruction of independence" can also be seen in the related (dynamical) situation of [EFHR02]. (This catchy term goes back R.Schonmann in the context of [EMSS00], as it was pointed out to the author by A.van Enter.)

### 3.2 Comparison with lattice random field Ising model

To appreciate the analogy to the lattice nearest neighbor random field model, let us compare the above formula for the conditional expectation given in Proposition 1 with its lattice analogue. So, we 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 $\propto \exp\left(\beta \sum_{\langle x,y\rangle} \sigma_x\sigma_y + \beta\varepsilon \sum_x \eta_x\sigma_x\right)$. 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 $\varepsilon$ (see [BK88]).

Now, with these notations the lattice analogue of this formula is

$$K^\mu[\sigma_x, \eta_x | \sigma_{x^c}\eta_{x^c}] = \frac{1}{Norm.} \exp\Big(\beta\big(\sum_{y\sim x} \sigma_y + \varepsilon\eta_x\big)\sigma_x\Big) \cdot q_x^\mu(\eta_{x^c})^{\frac{\eta_x}{2}} \quad \text{where}$$
$$q_x^\mu(\eta_{x^c}) = \int \mu[\eta_1 = 1, \eta_{x^c}](d\hat\sigma_x) e^{-2\beta\varepsilon\hat\sigma_x} \tag{3.6}$$

This follows from [K01]. In particular we have for the bias the expression

$$\frac{K^{\mu}[\eta_x = 1|\sigma_{x^c}\eta_{x^c}]}{K^{\mu}[\eta_x = -1|\sigma_{x^c}\eta_{x^c}]} = r_{\beta,\varepsilon}\Big(\sum_{y\sim x}\sigma_y\Big)\cdot q^{\mu}_x(\eta_{x^c}) \tag{3.7}$$

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 $q^{\mu}_x(\eta_{x^c})$ is not explicitly given like the limiting forms of its mean-field analogue when $N\uparrow\infty$.

Moreover we have $q^{\mu^+}_x(\eta_{x^c}) \leq q^{\mu^-}_x(\eta_{x^c})$. This follows from the representation as finite volume limits with plus resp. minus boundary conditions, and the monotonicity in the boundary conditions.

Let us recall the voter interpretation given for the mean-field model and extend it to the lattice model. Then the dependence on the state $\mu$ is quite understandable, too. So, suppose we have observed a neutral outcome of the votes of the neighbors of $x$, i.e. $\sum_{y\sim x}\sigma_y = 0$. Let us look at a typical realization of preferences $\eta_{x^c}$ that are more or less neutral. Suppose we look at the state $K^{\mu^+}$. It is overall plus like on the $\sigma_y$'s; so we would expect $\eta_1 = +$ with a probability that is smaller than in the state $K^{\mu^-}$ that is overall minus like. This is natural because we would ascribe the value of $\eta_x$ some partial responsibility for the neutral outcome of the neighbors of $x$ and so it will get a different bias according to the choice of the infinite-volume state.

### 3.3 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 $\eta$. This is seen by a Gaussian (Hubbard-Stratonovitch) transformation explained e.g. in [K97]. We won't repeat the details. This function $\Phi^0_{\beta,\varepsilon}$ has been well-analysed (see [SW85],[APZ92]). For 'large magnetic fields' $\varepsilon > \frac{1}{2}$, it has only one global quadratic minimum at $m = 0$. For $0 \leq \varepsilon \leq \frac{1}{2}$ 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 papers that $\beta, \varepsilon$ 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 $\varepsilon$ there is a unique global quartic minimum at $m = 0$, as for the usual CW ferromagnet; for large $\varepsilon$ 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 $\beta,\varepsilon$ 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 $N\geq N_0$ we have the uniform approximation*

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

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

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

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

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

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 $\pm m^{RF}(\beta,\varepsilon)$. 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 [K97]. To make the connection to the formulae of [K97] the reader should note the following. In [K97] 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

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

which is immediate when writing the l.h.s. as a quotient of partition functions and performing the HS-transformation as explained in [K97]. $\diamondsuit$

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

$$I_\rho^\pm(w) := \int_{B_\rho(\pm m^{RF})} \exp\Big(-\beta N \left(\Phi_{\beta,\varepsilon,N}(m,w) - \Phi^0_{\beta,\varepsilon}(m^{RF})\right)\Big) dm$$
$$J_\rho(w) := \int_{R_\rho} \exp\Big(-\beta N \left(\Phi_{\beta,\varepsilon,N}(m,w) - \Phi^0_{\beta,\varepsilon}(m^{RF})\right)\Big) dm \tag{3.13}$$

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 [K97] that were shown by the use of the Taylor expansion around $\pm m^{RF}$ and estimates on Gaussian integrals. They say that

$$I_\rho^\pm(w) \geq \exp\left(\pm\beta L_-(m^{RF})w\right)\sqrt{\frac{2\pi}{\beta N b_+(\rho)}}\left(\exp\left(\frac{z(w)^2\beta N}{2b_+(\rho)}\right) - 2\exp\left(-\frac{\beta N b_+(\rho)\rho^2}{4}\right)\right) \tag{3.14}$$

For the upper bound we simply write

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

Here we have put

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

and

$$b_+(\rho) := \sup_{v,|v|\leq\rho} \Phi^{0''}(m^{RF}+v) + \left|\frac{w}{N}\right| \sup_{v,|v|\leq\rho}\left|L''_-(m^{RF}+v)\right|$$
$$b_-(\rho) := \inf_{v,|v|\leq\rho} \Phi^{0''}(m^{RF}+v) - \left|\frac{w}{N}\right| \sup_{v,|v|\leq\rho}\left|L''_-(m^{RF}+v)\right| \tag{3.17}$$

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

For the integral over the complement of the balls we have further shown in [K97] that for $N \geq N_0(\beta,\varepsilon)$ we have that

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

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 $|w| \leq N^{\frac{1}{2}+\frac{\delta}{2}}$. Note that $\exp(\beta L_-(m^{RF})w) = (r^*_{\beta,\varepsilon})^{\frac{1}{2}}$ 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

$$\begin{aligned}\frac{I_\rho^+(w-1) + I_\rho^-(w-1)}{I_\rho^+(w+1) + I_\rho^-(w+1)} &\geq q_{\beta,\varepsilon,\infty}(w)\cdot\frac{e^{\frac{z(w-1)^2\beta N}{2b_+(\rho)}} - e^{-const\,N\rho_N^2}}{e^{\frac{z(w+1)^2\beta N}{2b_-(\rho)}}}\sqrt{\frac{b_-(\rho_N)}{b_+(\rho_N)}} \\ &\geq q_{\beta,\varepsilon,\infty}(w)\left(1 - Const\,(\beta,\varepsilon)N^{-\frac{1}{4}+\frac{3}{2}\delta}\right)\end{aligned} \tag{3.19}$$

This bound is obtained with the use of the worst case bound $\frac{w^2}{N^2}N|b_+(\rho_N) - b_+(\rho_N)| \leq Const'(\beta,\varepsilon)N^{-\frac{1}{4}+\frac{3}{2}\delta}$. 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* $w \geq C_-(\beta,\varepsilon)\log N$ implies $0 \leq (r^*_{\beta,\varepsilon})^{-1} - q_{\beta,\varepsilon,\infty}(w) \leq C'(\beta,\varepsilon)N^{-\frac{1}{4}+\frac{3\delta}{2}}$. Now use the upper bound $w \leq N^{\frac{1+\delta}{2}}$ 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 expectation in the form

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

where

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

The Hubbard-Stratonovitch transformation gives in this situation the representation

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

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 $\beta,\varepsilon$ 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 $N \geq N_0$ we have the uniform approximation*

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

*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

$$q_{\beta,\varepsilon,s,N}(w) = \frac{I_\rho^+(w-1,s) + I_\rho^c(w-1,s)}{I_\rho^+(w+1,s) + I_\rho^c(w+1,s)} = \frac{\frac{I_\rho^+(w-1,s)}{I_\rho^+(w+1,s)} + \frac{I_\rho^c(w-1,s)}{I_\rho^+(w+1,s)}}{1 + \frac{I_\rho^c(w+1,s)}{I_\rho^+(w+1,s)}} \tag{3.26}$$

with the integrals

$$\begin{aligned} I_\rho^+(w,s) &:= \int_{|m-m^*|\le\rho} dm \exp\Bigl(-\beta N \left(\bar\Phi_{\beta,\varepsilon,s,N}(m,w) - \Phi^0_{\beta,\varepsilon,s}(m^*)\right)\Bigr) \\ I_\rho^c(w,s) &:= \int_{|m-m^*|\ge\rho} dm \exp\Bigl(-\beta N \left(\bar\Phi_{\beta,\varepsilon,s,N}(m,w) - \Phi^0_{\beta,\varepsilon,s}(m^*)\right)\Bigr) \\ \text{with} \quad & \bar\Phi_{\beta,\varepsilon,s,N}(m,w) = \bar\Phi_{\beta,\varepsilon,s}(m) - L_{\beta,\varepsilon,-}(m)\frac{w}{N} \end{aligned} \tag{3.27}$$

The r.h.s. of (3.26) 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 $|v| \le \rho$,

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

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

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

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

$$\begin{aligned} I_\rho^+(w,s) &\ge \exp\left(\beta L_-(m^*)w\right)\sqrt{\frac{2\pi}{\beta N b_+}}\left(\exp\left(\frac{z^2\beta N}{2b_+}\right) - 2\exp\left(-\frac{\beta N b_+\rho^2}{4}\right)\right) \\ &\le \exp\left(\beta L_-(m^*)w\right)\sqrt{\frac{2\pi}{\beta N b_-}}\exp\left(\frac{z^2\beta N}{2b_-}\right) \end{aligned} \tag{3.30}$$

As in (3.19) we obtain the estimate

$$
\begin{aligned}
&\frac{I^{+}_{\rho_N}(w-1,s)}{I^{+}_{\rho_N}(w+1,s)} \geq \Big(r_{\beta,\varepsilon}\big(m^*(\beta,\varepsilon,s)\big)\Big)^{-1} \cdot \frac{e^{\frac{z(w-1)^2\beta N}{2b_+(\rho_N)}} - e^{-const\, N\rho_N^2}}{e^{\frac{z(w+1)^2\beta N}{2b_-(\rho_N)}}} \sqrt{\frac{b_-(\rho_N)}{b_+(\rho_N)}} \\
&= \Big(r_{\beta,\varepsilon}\big(m^*(\beta,\varepsilon,s)\big)\Big)^{-1} \big(1 - Const\,(\beta,\varepsilon)\,\mathcal{O}(N^{\delta}\rho_N)\big)
\end{aligned}
\tag{3.31}
$$

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

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

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 $|w| << s$ 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

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

and thus

$$
\left|\Big(r_{\beta,\varepsilon}\big(m^*(\beta,\varepsilon,s_N)\big)\Big)^{-1} - (r^*_{\beta,\varepsilon})^{-1}\right| \leq Const\,(\beta,\varepsilon)' s_N \leq Const\,(\beta,\varepsilon)' N^{-\alpha} \tag{3.35}
$$

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

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

recalling that $m^{RF}(\beta,\varepsilon,\alpha)$ is the (unique) minimizer of the function $m \mapsto \Phi^0_{\beta,\varepsilon,\alpha}(m)$. 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). $\diamondsuit$

## 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 expectation for the decimation. It reads

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

where

$$\tilde\mu_{\beta,N,M}\big(dm\big|z\big) = \frac{1}{Norm.}\exp\Big(-\frac{\beta N m^2}{2} + (N-M)\log\cosh(\beta m) + \beta z m\Big)\,dm \tag{4.2}$$

This is seen by Hubbard-Stratonovitch transformation. To check the formula let us at first 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 $\beta(N-M)\Big(\frac{m^2}{2}-\frac{1}{\beta}\log\cosh(\beta m)\Big)+\frac{\beta M}{2}\Big(m-\frac{z}{M}\Big)^2$ plus an $m$-independent constant.

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

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

To prove Theorem 3 in generality use the alternative representation of the measure $\tilde\mu_{\beta,N,M}$ that is obtained after a change of variable in the form

$$\begin{aligned}&\int \tilde\mu_{\beta,N,M}\big(dm\big|z\big)\varphi(m) = \int \tilde\mu'_{\beta',h',N-M}(dy)\,\varphi\Big(\frac{N-M}{N}y+\frac{z}{N}\Big) \quad \text{with}\\ &\beta' = \frac{N-M}{N}\beta, \quad h' = \frac{z}{N-M}\end{aligned} \tag{4.4}$$

where

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

The function appearing in the exponent of (4.5) is well known. Its minimizer $m^{CW}(\beta', h')$ 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

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

This proves Theorem 3. To prove Theorem 4 where $z$ stays fixed, we write

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

and use $\tilde{\mu}_{\beta,N,M=(1-p)N}\big(\cdot\,\big|z=0\big) \to \frac{1}{2}(\delta_{-pm^{CW}(p\beta)} + \delta_{pm^{CW}(p\beta)})$. $\diamondsuit$

## 4.2 Diluted ferromagnet

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

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

where

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

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

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

To derive (4.8) with (4.9) proceed as in the random field model. To express the resulting fractions of partition functions use that 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 $\lim_{M\uparrow\infty} \int \mu_{\beta',M}(d\hat{\sigma}) \exp\Big(\frac{\beta'}{M}\sum_{i=1}^M \hat{\sigma}_i\Big) = \cosh(\beta' m^{CW}(\beta'))$ since the distribution of the magnetization of the ordinary Curie-Weiss model concentrates on plus or minus the mean-field value. We also have $\lim_{N\uparrow\infty} w_{\beta,N}(qN)^{n_1} = \cosh(\beta q\, m^{CW}(\beta q))^{n_1} = \cosh(\beta q\, m^{CW}(\beta q) n_1)$. This proves Theorem 5.$\diamondsuit$

Let us finally compare the formula for the conditional expectation with the lattice analogue with formal Boltzmann weights $\propto \exp\bigl(\beta \sum_{\langle x,y\rangle} n_x\sigma_x n_y\sigma_y\bigr)$. It reads for the joint measure $K^\mu$ obtained from a Gibbs measure $\mu$

$$
\begin{aligned}
&K^\mu[\sigma_x, n_x|\sigma_{x^c} n_{x^c}] = \frac{1}{Norm.} \exp\Bigl(\beta \sum_{y\sim x} n_y\sigma_y \cdot n_x\sigma_x\Bigr) \cdot w_x^\mu(n_{x^c})^{n_x} \quad \text{where}\\
&w_x^\mu(n_{x^c}) = \Bigl(\int \mu[n_x=0, n_{x^c}](d\hat\sigma) \exp\Bigl(\beta \sum_{y\sim x} n_y\hat\sigma_y \cdot \hat\sigma_x\Bigr)\Bigr)^{-1}
\end{aligned}
\tag{4.11}
$$

This representation also follows from [K01]. In particular we have from that for the "occupation-bias"

$$
\frac{K^\mu[n_x=1|\sigma_{x^c} n_{x^c}]}{K^\mu[n_x=0|\sigma_{x^c} n_{x^c}]} = \frac{p}{1-p} \cosh\Bigl(\beta \sum_{y\sim x} n_y\sigma_y\Bigr) \cdot w_x^\mu(n_{x^c}) \tag{4.12}
$$

As for the joint measures of the random field model mean-field and lattice expressions look similar. Now, writing $w_x^\mu(n_{x^c}) = \int \mu[n_x=1, n_{x^c}](d\hat\sigma) \exp\Bigl(-\beta \sum_{y\sim x} n_y\hat\sigma_y \cdot \hat\sigma_x\Bigr)\Bigr)$ and choosing the specific configuration of [EMSS00] 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.

## REFERENCES:

[APZ92] J.M.G. Amaro de Matos, A.E.Patrick, V.A.Zagrebnov, Random Infinite-Volume Gibbs States for the Curie-Weiss Random Field Ising Model., J.Stat.Phys **66**, 139-164 (1992)

[BCKS01] C.Borgs, J.T.Chayes, H.Kesten, J.Spencer, *"The birth of the infinite cluster: finite-size scaling in percolation"*, Comm. Math. Phys. **224** no. 1, 153-204 (2001)

[BK88] J.Bricmont, A.Kupiainen, *"Phase transition in the 3d random field Ising model"*, Commun. Math. Phys. **142** 539-572 (1988)

[BKL98] J.Bricmont, A.Kupiainen, R.Lefevere, *" Renormalization group pathologies and the definition of Gibbs states"*, Commun. Math. Phys. **194** 359-388 (1998)

[Bo01] A.Bovier, *"Statistical Mechanics of Disordered Systems"*, MaPhySto lecture notes no. 10, available at http://www.maphysto.dk/

[EFS93] 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)

[EFHR02] A.C.D. van Enter, R. Fernández, F. den Hollander, F. Redig, *"Possible loss and recovery of Gibbsianness during the stochastic evolution of Gibbs measures"*, Comm. Math. Phys. **226** 101-130 (2002)

[EL96] A.C.D.van Enter, J.Lörinczi, *"Robustness of the non-Gibbsian property: some examples"* J. Phys. A **29** no. 10, 2465-2473 (1996)

[EMSS00] 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, 51–58, Amer. Math. Soc. Transl. Serie 2, **198**, Amer. Math. Soc. Providence, P.I. (2000)

[EKM00] A.C.D.van Enter, C.Maes and C.Külske, *Comment on* [KM94], Phys.Rev.Lett. **84**, 6134 (2000)

[FLNR02] R. Fernández, A.Le Ny and F.Redig, *"Variational principle and almost quasilocality for renormalized measures"*, Eurandom preprint 2002-032 (2002)

[GHM00] H.O.Georgii, O.Häggström, C.Maes, *"The random geometry of equilibrium phases"*, in: C.Domb. J.LO.Lebowitz (eds.), Phase Transitions and Critical Phenomena Vol. **18**, 1-142, Academic Press, London (2000)

[Ku96] R.Kühn, *"Equilibrium Ensemble Approach to Disordered Systems I: General Theory, Exact Results"*, Z. Phys. **100**, 231-242 (1996)

[K97] C.Külske, *"Metastates in Disordered Mean-Field Models: Random Field and Hopfield Models"*, J.Stat.Phys. **88** 5/6, 1257-1293 (1997)

[K98] C.Külske, *"Metastates in Disordered Mean-Field Models II: The Superstates"*, J.Stat.Phys. **91** 1/2, 155-176 (1998)

[K99] C.Külske, *"(Non-) Gibbsianness and Phase Transitions in Random Lattice Spin Models"*, Markov Proc. Rel. Fields **5**, 357-383 (1999)

[K01] C.Külske, *"Weakly Gibbsian Representations for joint measures of quenched lattice spin models"*, Probab. Theor. Rel. Fields **119**, 1-30 (2001).

[K02] C.Külske, *"Regularity properties of potentials for joint measures of random spin systems"*, WIAS preprint 781, available at http://www.ma.utexas.edu/mp-arc/ preprint 02-519 .

[KLR02] C.Külske, A.Le Ny, 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.

[KM94] 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)

[KM00] R.Kühn and G.Mazzeo, *Reply to [EKM00]*, Phys.Rev.Lett. **84**, 6135 (2000)

[MePaVi87] M. Mézard, G. Parisi, M. A. Virasoro, *"Spin glass theory and beyond"*, World Scientific Lecture Notes in Physics 9. (1987)

[MPRRZ00] E.Marinari, G.Parisi, F.Ricci-Tersenghi, J.J.Ruiz-Lorenzo, F.Zuliani, *"Replica symmetry breaking in short-range spin glasses: theoretical foundations and numerical evidences"*, J. Stat. Phys. **98** no. 5-6, 973-1047 (2000).

[MRSV00] C. Maes, F. Redig, S. Shlosman and A. van Moffaert, *"Percolation, path large deviations and weak Gibbsianity"*, Comm. Math. Phys. **209** 517-45 (2000)

[Mo64] T. Morita, *"Statistical Mechanics of quenched solid solutions with application to magnetically dilute alloys"*, J. Math. Phys. **5**, 1402-1405 (1964)

[NS96a] C. M. Newman, D. L. Stein, *"Non-mean-field behavior of realistic spin glasses"*, Phys. Rev. Lett. **76**, 515-518 (1996).

[NS96b] C. M. Newman, D. L. Stein, *"Spatial inhomogeneity and thermodynamic chaos"*, Phys. Rev. Lett. **76**, 4821-4824 (1996).

[NS02] C.M.Newman, D.L. Stein, *"The state(s) of replica symmetry breaking: mean field theories vs. short-ranged spin glasses"*, J.Stat.Phys. **106**, no. 1-2, 213–244 (2002)

[SP93] M.Serva, G.Paladin, *"Gibbs thermodynamic potentials for disordered systems"*, Phys.Rev.Lett. **70**, 105-108 (1993)

[SW85] S.R.Salinas, W.F.Wreszinski, *"On the Mean-Field Ising Model in a Random External Field"*, J.Stat.Phys **41**, 299-313 (1985)