Delft University of Technology

Gaussian Concentration and Uniqueness of Equilibrium States in Lattice Systems

Chazottes, J.R.; Moles, J.; Redig, F.; Ugalde, E.

DOI
10.1007/s10955-020-02658-1

Publication date
2020

Document Version
Accepted author manuscript

Published in
Journal of Statistical Physics

Citation (APA)
Chazottes, J. R., Moles, J., Redig, F., & Ugalde, E. (2020). Gaussian Concentration and Uniqueness of Equilibrium States in Lattice Systems. *Journal of Statistical Physics*, *181*(6), 2131-2149. https://doi.org/10.1007/s10955-020-02658-1

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Gaussian concentration and uniqueness of equilibrium states in lattice systems

J.-R. Chazottes *[1], J. Moles [1,3], F. Redig [2], and E. Ugalde [3]

[1]Centre de Physique Théorique, CNRS, Institut Polytechnique de Paris, France
[2]Institute of Applied Mathematics, Delft University of Technology, The Netherlands
[3]Instituto de Física, Universidad Autónoma de San Luis Potosí, México

Dated: October 16, 2020

Abstract

We consider equilibrium states (that is, shift-invariant Gibbs measures) on the configuration space $S^{\mathbb{Z}^d}$ where $d \geq 1$ and S is a finite set. We prove that if an equilibrium state for a shift-invariant uniformly summable potential satisfies a Gaussian concentration bound, then it is unique. Equivalently, if there exist several equilibrium states for a potential, none of them can satisfy such a bound.

Keywords and phrases: concentration inequalities, relative entropy, blowing-up property, equilibrium states, large deviations, Hamming distance.

Contents

1 Introduction and main result **1**

2 Setting **5**

3 Gibbs measures and equilibrium states **8**

3.1 Potentials, specifications and relative entropy 8

3.2 Examples 10

3.2.1 Dobrushin's uniqueness 10

3.2.2 Ising model 10

3.2.3 Dyson model 12

3.2.4 Other examples 12

*Email: jeanrene@cpht.polytechnique.fr

4 Proof of the main result **12**
4.1 An abstract result 12
4.2 GCB implies blowing-up 13
4.3 Blowing-up implies exponential rate for frequencies 15
4.4 Exponential rate for frequencies implies positive relative entropy 17

5 A proof via large deviations **18**

A Appendix **19**
A.1 An estimate 19
A.2 Proof Lemma 4.1 20
A.3 A bound on relative entropy 21

1 Introduction and main result

The phenomenon we are interested in, which goes under the name of "concentration inequalities", is that if a function of many "weakly dependent" random variables does not depend too much on any of them, then it is concentrated around its expected value. A key feature of this phenomenon is that it is *non-asymptotic*, in contrast with the usual *limit* theorems where the number of random variables has to tend to infinity. Recall that the three main types of classical limit theorems are the law of large numbers, the central limit theorem, and large deviations. Another key feature of concentration inequalities is that they allow to deal with functions of random variables defined in an arbitrary way, provided they are "smooth enough", in contrast with classical limit theorems which deal with sums of random variables. Concentration inequalities made a paradigm shift in probability and statistics, but also in discrete mathematics, in geometry and in functional analysis, see e.g. [2, 10, 17, 25].

In this paper, we consider Gibbs measures on the configuration space $\Omega = S^{\mathbb{Z}^d}$ where S is a finite set and $d \geq 1$. Postponing precise definitions till next section, a probability measure μ satisfies a Gaussian concentration bound if there exists a constant $D > 0$ such that, for any local function $F : \Omega \to \mathbb{R}$,

$$\int \mathrm{e}^{F - \int F \,\mathrm{d}\mu} \,\mathrm{d}\mu \leq \mathrm{e}^{D \sum_{x \in \Lambda_F} \delta_x(F)^2} \tag{1}$$

where Λ_F is the (finite) set of sites $x \in \mathbb{Z}^d$ such that $\delta_x(F) \neq 0$, and $\delta_x(F)$ is the largest value of $|F(\omega) - F(\omega')|$ taken over the configurations ω and ω' differing only at site x. Note that D does not depend on F, and in particular not on Λ_F. By a standard argument (recalled later on), (1) implies a control on the fluctuations of F around $\int F \,\mathrm{d}\mu$: for all $u > 0$, we have

$$\mu\Big\{\omega \in \Omega : F(\omega) \geq \int F \,\mathrm{d}\mu + u\Big\} \leq \exp\left(-\frac{u^2}{4D \sum_{x \in \Lambda_F} \delta_x(F)^2}\right). \tag{2}$$

For instance, if $S = \{-1, +1\}$, F can be $\sum_{x \in \Lambda_n} \omega_x / |\Lambda_n|$, which is the magnetization inside the "cube" $\Lambda_n = [-n, n]^d \cap \mathbb{Z}^d$. One can check immediately that the previous bound is exponentially small in the volume $|\Lambda_n| = (2n+1)^d$. But we can consider much more general F's, and in particular nonlinear or implicitly defined functions. The bound (1) was first proved in [16] for potentials satisfying Dobrushin's uniqueness condition, with a constant D explicitly related to Dobrushin's interdependence matrix. This covers, for instance, finite-range potentials at sufficiently high temperature. Not surprisingly, one cannot expect that a Gaussian concentration holds for the (ferromagnetic) Ising model at temperatures below the critical one, because of the surface-order large deviations of the magnetization (see [5] for more details). In [5], the authors proved that a "stretched-exponential" concentration bound holds for the "+" phase and the "−" phase of this model at sufficiently low temperature. Concerning the various applications of these concentration bounds, we will not describe them here and refer the reader to [16, 6].

What happens for the Ising model raises the following general question:

> Suppose that a potential admits several Gibbs measures. Is it true that none of these measures satisfy a Gaussian concentration bound?

Equivalently, we ask:

> If a Gibbs measure for a given potential satisfies a Gaussian concentration bound, is it unique?

In this paper, we prove the following result which answers this question in the case of *shift-invariant* Gibbs measures, that is, *equilibrium states.*[1]

Main theorem. *If an ergodic equilibrium state for a shift-invariant absolutely summable potential satisfies a Gaussian concentration bound, then it must be the unique equilibrium state for this potential.*

Our theorem can be paraphrased by saying that nonuniqueness of the equilibrium states for a potential prevents Gaussian concentration. Observe that the statement is about *ergodic* equilibrium states because, as we prove below, if a shift-invariant probability satisfies a Gaussian concentration bound, then it must be ergodic (Proposition 2.2).

We will prove the above theorem in two different ways.
The first way is based on ideas which were put forward in ergodic theory to study the existence of finitary codings from a finite-valued i.i.d. process to certain ergodic processes. Two central notions turn out to be the "blowing-up property" and the "positive (lower) relative entropy property". Without

[1] In the class of potentials we consider, shift-invariant Gibbs measures coincide with equilibrium states [22, Theorem 4.2].

going into detail, it was proved in [21] that if a process is finitely determined then it has the blowing-up property, which in turn implies the positive relative entropy property. Here, we use the fact that the Gaussian concentration bound implies the blowing-up property, and then use part of the variational principle which says that the relative entropy of two distinct equilibrium states for the same potential is equal to zero. Hence the blowing-up property cannot hold, therefore it is not possible to have a Gaussian concentration bound. In fact, we establish an abstract result (Theorem 4.1) which states that if a probability measure satisfies a Gaussian concentration bound, then it has the positive lower relative entropy property. Technically speaking, we follow some methods that can be found in [21]. The passage from $d = 1$ to $d \geq 2$ (that is, going from processes to random measures) poses no difficulty. Although these methods are known to specialists of ergodic theory, they are probably not as well-known in the mathematical physics literature. For this reason and to make this article self-contained, we give a detailed and complete presentation of the methods. Let us also mention that the line of reasoning we follow here was used for the first time in [7] in the context of chains of unbounded memory, a.k.a., chains with complete connections. Using the methods of [21], it was proved in [1] that, for Markov random fields (including the nearest neighbour Ising model), a phase transition is an obstruction for the finitely determined property. A natural question is thus: How does the Gaussian concentration bound relate to the finitely determined property? A reasonable conjecture is that the former implies the latter, but we have not investigated this question.

The other way of proving the main theorem is via large deviations. Let us sketch the proof. Let us first recall that, for equilibrium states, one can prove a "large deviation principle" for the empirical measure with a "rate function" which is the relative entropy. Roughly speaking, the content of such a result is that, if we take an equilibrium state μ, and an ergodic probability measure ν different from μ, then

$$\mathrm{Prob}_\mu\big(\omega_{\Lambda_n} \text{ is typical for } \nu\big) \asymp \mathrm{e}^{-|\Lambda_n| h(\nu|\mu)} \tag{3}$$

where $h(\nu|\mu)$ is the relative entropy of ν with respect to μ, which will be precisely defined below. The symbol "$\asymp$" means asymptotic equivalence on the logarithmic scale in the following sense: $a_{\Lambda_n} \asymp b_{\Lambda_n}$ means $\lim_{n\to+\infty} |\Lambda_n|^{-1} \log a_{\Lambda_n} = \lim_{n\to+\infty} |\Lambda_n|^{-1} \log b_{\Lambda_n}$. What "typical" means is in relation to the ergodic theorem: If ω_{Λ_n} is typical for ν, this means that the average $|\Lambda_n|^{-1} \sum_{x\in\Lambda_n} f(\theta_x\omega)$ of every continuous function $f : \Omega \to \mathbb{R}$ under the shift (translation) over Λ_n converges to $\int f \,\mathrm{d}\nu$, as $n \to +\infty$. The set of typical configurations for ν is of measure 0 for μ, and thus what (3) roughly says is that the probability that a configuration ω, taken from the probability distribution μ, looks in Λ_n like a typical configuration from ν decays exponentially in the volume $|\Lambda_n| = (2n + 1)^d$. Coming back to our

main result, the idea behind its proof is then simple: If μ and ν are two distinct (ergodic) equilibrium states for the same potential, then $h(\nu|\mu) = 0$, whence the probability in (3) decays sub-exponentially in Λ_n. Since $\mu \neq \nu$, there exists a local function f such that $\int f \, \mathrm{d}\nu = \int f \, \mathrm{d}\mu + \varepsilon$ for some $\varepsilon > 0$.

Now, suppose that μ satisfies a Gaussian concentration bound. Then $|\Lambda_n|^{-1} \sum_{x \in \Lambda_n} f \circ \theta_x$ must sharply concentrate around $\int f \, \mathrm{d}\mu$, in the sense that the μ-probability that this average is larger than $\int f \, \mathrm{d}\mu + \varepsilon$ (call this event E_n), decays exponentially fast in $|\Lambda_n|$. This is a consequence of (2). But, at the same time, by the large deviation principle, the μ-probability that this average is concentrated around $\int f \, \mathrm{d}\nu$ (call this event E'_n) decays only sub-exponentially fast. Since $E'_n \subset E_n$, we get a contradiction since $\mu(E'_n) \leq \mu(E_n)$, which is incompatible with the fact that $\mu(E'_n)$ is sub-exponentially small in $|\Lambda_n|$, whereas $\mu(E_n)$ is exponentially small in $|\Lambda_n|$.

Let us comment on the two proofs of the main theorem. The one based on large deviations is short and simple, given that we have a large deviation principle at our disposal. Moreover, it works for more general spin spaces S, in particular non-discrete spaces. The other proof is longer, but it puts forward the blowing-up property which is a remarkable property. It also connects Gaussian concentration with the positivity of the (lower) relative entropy (Theorem 4.1), which is of independent interest (and not tied to equilibrium states).

Finally, let us mention that, strictly speaking, we do not consider hard-core lattice gas models. But our result extends to that situation by ad-hoc modifications. Notice that the first proof works as well, whereas the proof based on large deviations also works, provided we invoke a large deviation principle proved in [12].

2 Setting

We set some basic notation. The configuration space is $\Omega = S^{\mathbb{Z}^d}$, where S is a finite set, and d an integer greater than or equal to 1. We endow Ω with the product topology that is generated by cylinder sets, which makes it a compact metrizable space. We denote by $\mathfrak{B}$ the Borel σ-algebra which coincides with the σ-algebra generated by cylinder sets.

An element x of $\mathbb{Z}^d$ (hereby called a site) can be written as a vector $(x_1, \ldots, x_d)$ in the canonical base of the lattice $\mathbb{Z}^d$. If Λ is a finite subset of $\mathbb{Z}^d$, denote by $|\Lambda|$ its cardinality. If Λ is a finite subset of $\mathbb{Z}^d$, we will write $\Lambda \Subset \mathbb{Z}^d$.

For $\Lambda \subset \mathbb{Z}^d$, $\sigma, \eta \in \Omega$, we denote by $\sigma_\Lambda \eta_{\Lambda^c}$ the configuration which agrees with σ on Λ and with η on Λ^c. We denote by $\mathfrak{B}_\Lambda$ the σ-algebra generated by the coordinate maps $\omega \mapsto \omega_x$ when x is restricted to Λ. We need to define centered "cubes": for every $n \in \mathbb{Z}_+$, let

$$\Lambda_n = \big\{ x \in \mathbb{Z}^d : -n \leq x_i \leq n, \, i = 1, 2, \ldots, d \big\}.$$

Given $\Lambda \Subset \mathbb{Z}^d$, an element p_Λ of S^Λ is called a pattern with shape Λ, or simply a pattern. We will write p_n for a pattern in S^{Λ_n}. We will also consider elements of S^Λ as configurations restricted to Λ. We will simply write ω instead of ω_Λ since we will always make clear to which set ω belongs. A pattern $p_n \in S^{\Lambda_n}$ determines a cylinder set $[p_n] = \{\omega \in \Omega : \omega_{\Lambda_n} = p_n\}$. More generally, given $\Lambda \subset \mathbb{Z}^d$ and $C \subseteq S^\Lambda$, let $[C] = \{\omega \in \Omega : \pi_\Lambda(\omega) \in C\}$ where π_Λ is the projection from Ω onto S^Λ.

Finally, the shift action $\Theta = (\theta_x, x \in \mathbb{Z}^d)$ is defined as usual: for each $x \in \mathbb{Z}^d$, $\theta_x : \Omega \to \Omega$ and $(\theta_x\,\omega)_y = \omega_{y-x}$, for all $y \in \mathbb{Z}^d$. A probability measure ν on $(\Omega, \mathfrak{B})$ is shift invariant if $\nu \circ \theta_x = \nu$ for all $x \in \mathbb{Z}^d$. We denote by $\mathcal{M}_\Theta(\Omega)$ the set of shift-invariant probability measures, which is a simplex. A shift-invariant probability measure is ergodic if it is trivial on the σ-algebra of all shift-invariant events $\{A \in \mathfrak{B} : \theta_x A = A \text{ for all } x \in \mathbb{Z}^d\}$. Ergodic measures are the extreme points of $\mathcal{M}_\Theta(\Omega)$; we denote them by $\operatorname{ex}\mathcal{M}_\Theta(\Omega)$.

We now define what we mean by a Gaussian concentration bound. Let $F : \Omega \to \mathbb{R}$ be a function and $x \in \mathbb{Z}^d$. The oscillation of F at x is defined by

$$\delta_x(F) = \sup\big\{|F(\omega) - F(\omega')| : \omega, \omega' \in \Omega \text{ differ only at site } x\big\}.$$

Given $\Lambda \Subset \mathbb{Z}^d$ and two configurations $\omega, \eta \in \Omega$ such that $\omega_{\Lambda^c} = \eta_{\Lambda^c}$, one has

$$|F(\omega) - F(\eta)| \le \sum_{x \in \Lambda} \delta_x(F)\, \mathbb{1}_{\{\omega_x \neq \eta_x\}}.$$

We introduce the space of local functions

$$\mathscr{L} = \bigcup_{\Lambda \Subset \mathbb{Z}^d} \mathscr{L}_\Lambda$$

where $F : \Omega \to \mathbb{R}$ belongs to $\mathscr{L}_\Lambda$ if there exists $\Lambda \Subset \mathbb{Z}^d$ (the dependence set of F) such that for all $\omega, \widetilde{\omega}, \widehat{\omega}$, $F(\omega_\Lambda \widetilde{\omega}_{\Lambda^c}) = F(\omega_\Lambda \widehat{\omega}_{\Lambda^c})$. (This means that F is $\mathfrak{B}_\Lambda$-measurable.) Equivalently, this means that $\delta_x(F) = 0$ for all $x \notin \Lambda$. It is understood that Λ is the smallest such set. Local functions are continuous, hence bounded since Ω is compact. (In fact, continuous functions are obtained as uniform limits of local functions.)

We write $\underline{\delta}(F)$ for the infinite array $(\delta_x(F), x \in \mathbb{Z}^d)$, and let

$$\|\underline{\delta}(F)\|_2^2 := \|\underline{\delta}(F)\|^2_{\ell^2(\mathbb{Z}^d)} = \sum_{x \in \mathbb{Z}^d} \delta_x(F)^2.$$

We use the notation $\mathbb{E}_\mu$ for the integration with respect to μ.

Definition 2.1 (Gaussian concentration bound)**.**
Let μ be a probability measure on $(\Omega, \mathfrak{B})$. We say that it satisfies a Gaussian

concentration bound if there exists $D = D(\mu) > 0$ such that, for all functions $F \in \mathscr{L}$, we have

$$\mathbb{E}_\mu\big[\exp\left(F - \mathbb{E}_\mu(F)\right)\big] \le \exp\left(D\|\underline{\delta}(F)\|_2^2\right). \tag{4}$$

For the sake of brevity, we will say that μ satisfies GCB(D)*, or simply GCB.*

A key point in this definition is that D is independent of F, in particular it is independent of the size of the dependence set of F.

Remark 2.1. *Let us notice that if* (4) *holds for all local functions, then it holds for a larger class of functions, namely continuous functions such that $\|\underline{\delta}(F)\|_2 < +\infty$ (see [6]).*

Inequality (4) easily implies the following tail inequality that we will use several times.

Proposition 2.1.
If a probability measure μ on $(\Omega, \mathfrak{B})$ satisfies GCB(D) *then, for all $u > 0$,*

$$\mu\big\{\omega \in \Omega : F(\omega) - \mathbb{E}_\mu(F) \ge u\big\} \le \exp\left(-\frac{u^2}{4D\|\underline{\delta}(F)\|_2^2}\right). \tag{5}$$

Proof. If $F \in \mathscr{L}$ then, for any $\lambda > 0$, λF obviously belongs to $\mathscr{L}$. Applying Markov's inequality and (4) we get

$$\begin{aligned}\mu\left\{\omega \in \Omega : F(\omega) - \mathbb{E}_\mu(F) \ge u\right\} &\le \exp\left(-\lambda u\right)\ \mathbb{E}_\mu\big[\exp\left(\lambda(F - \mathbb{E}_\mu(F))\right)\big] \\ &\le \exp\left(-\lambda u + D\|\underline{\delta}(F)\|_2^2\,\lambda^2\right).\end{aligned}$$

Minimizing over $\lambda > 0$ yields (5). $\square$
Observe that (5) can be applied to $-F$, which gives the same bound, and thus by a union bound we get (5) with $|F(\omega) - \mathbb{E}_\mu(F)|$ by multiplying the bound by 2.

The next result shows that a shift-invariant probability measure which satisfies GCB must be mixing.

Proposition 2.2.
Let μ be a shift-invariant probability measure on $(\Omega, \mathfrak{B})$ which satisfies GCB. Then μ is mixing.

Proof. First we remark that by Lemma A.1, combined with Proposition 2.1, we conclude that for every sequence $V_n, n \in \mathbb{N}$ of finite subsets of $\mathbb{Z}^d$ such that $|V_n| \to \infty$ as $n \to \infty$, and for every local function f, we have that $\frac{1}{|V_n|}\sum_{x\in V_n}\theta_x f$ converges to $\mathbb{E}_\mu(f)$ in μ-probability as $n \to \infty$.

We then argue by contradiction. Assume that μ is not mixing. Then there exist local functions f, g (without loss of generality both of μ expectation zero) and a sequence x_n, with $|x_n| \to \infty$ such that

$$\int f\theta_{x_n} g \,\mathrm{d}\mu$$

does not converge to zero as $n \to \infty$. By locality, both functions f, g are uniformly bounded, and therefore the sequence $\int f\theta_{x_n} g \,\mathrm{d}\mu, n \in \mathbb{N}$ is a bounded sequence. Therefore, there exists a subsequence y_n such that along that subsequence

$$\lim_{n\to\infty} \int f\theta_{y_n} g \,\mathrm{d}\mu = a \neq 0. \tag{6}$$

As a consequence,

$$\lim_{n\to\infty} \int f \frac{1}{n} \sum_{k=0}^{n} \theta_{y_k} g \,\mathrm{d}\mu = a \neq 0.$$

However, as we saw before, $\frac{1}{n}\sum_{k=0}^{n} \theta_{y_k} g$ converges to zero in probability by GCB. Then via dominated convergence we obtain

$$\lim_{n\to\infty} \left| \int f \frac{1}{n} \sum_{k=0}^{n} \theta_{y_k} g \,\mathrm{d}\mu \right| = 0$$

which contradicts (6). $\square$

Remark 2.2. *A stronger property than mixing is tail-triviality of μ, which in the context of Gibbs measures is equivalent with the fact that μ is an extreme point of the set of Gibbs measures. Though we strongly believe that GCB implies tail triviality and even more that μ is the unique Gibbs measure, at present we are not able to prove this.*

Let us finish this section by a variance inequality implied by (4). If μ satisfies GCB(D) then

$$\mathrm{Var}_\mu(F) := \mathbb{E}_\mu\big(F^2\big) - \mathbb{E}_\mu\big(F\big)^2 \leq 2D\|\underline{\delta}(F)\|_2^2 \tag{7}$$

for all functions $F \in \mathscr{L}$. The proof goes as follows. Take $\lambda > 0$ and apply inequality (4) to λF, subtract 1 on both sides, and then divide out by λ^2 the resulting inequality. Then (7) follows easily by Taylor expansion and letting λ tend to 0. (See [6] for details.)

3 Gibbs measures and equilibrium states

3.1 Potentials, specifications and relative entropy

We refer to [14] or [13] for details. We consider shift-invariant uniformly summable potentials. More precisely, a potential is a family of functions $(\Phi(\Lambda, \cdot))_{\Lambda \Subset \mathbb{Z}^d}$ such that, for each (nonempty) $\Lambda \Subset \mathbb{Z}^d$, the function $\Phi_\Lambda : \Omega \to \mathbb{R}$ is $\mathfrak{B}_\Lambda$-measurable. Shift-invariance means that $\Phi(\Lambda + x, \theta_x\omega) = \Phi(\Lambda, \omega)$

for all $\Lambda \Subset \mathbb{Z}^d$, $\omega \in \Omega$ and $x \in \mathbb{Z}^d$ (where $\Lambda + x = \{y + x : y \in \Lambda\}$). Uniform summability is the property that

$$\sum_{\substack{\Lambda \Subset \mathbb{Z}^d \\ \Lambda \ni 0}} \|\Phi(\Lambda, \cdot)\|_\infty < \infty.$$

We shall denote by $\mathscr{B}_\Theta$ the space of uniformly summable shift-invariant potentials.

Given $\Phi \in \mathscr{B}_\Theta$ and $\Lambda \Subset \mathbb{Z}^d$, the associated Hamiltonian in Λ with boundary condition $\eta \in \Omega$ is given by

$$\mathcal{H}_\Lambda(\omega|\eta) = \sum_{\substack{\Lambda' \Subset \mathbb{Z}^d \\ \Lambda' \cap \Lambda \neq \emptyset}} \Phi(\Lambda', \omega_\Lambda \eta_{\Lambda^c}) \,.$$

The corresponding Gibbsian specification is then defined as

$$\boldsymbol{\gamma}_\Lambda^\Phi(\omega|\eta) = \frac{\exp\left(-\mathcal{H}_\Lambda(\omega|\eta)\right)}{Z_\Lambda(\eta)}$$

where $Z_\Lambda(\eta)$ is the partition function in Λ (normalizing factor). We say that μ is a Gibbs measure for the potential Φ if, for every $\Lambda \Subset \mathbb{Z}^d$, $\boldsymbol{\gamma}_\Lambda^\Phi(\omega|\cdot)$ is a version of the conditional probability $\mu(\omega_\Lambda|\mathfrak{B}_{\Lambda^c})$. Equivalently, this means that for all $A \in \mathfrak{B}$, $\Lambda \Subset \mathbb{Z}^d$, one has the so-called "DLR equations"

$$\mu(A) = \int \sum_{\omega' \in S^\Lambda} \boldsymbol{\gamma}_\Lambda^\Phi(\omega'|\eta)\, \mathbb{1}_A(\omega'_\Lambda \eta_{\Lambda^c})\, \mathrm{d}\mu(\eta).$$

The set of Gibbs measures for a given potential Φ, denoted by $\mathcal{G}(\Phi)$, is never empty, but it may be not reduced to a singleton. It is a simplex. The set $\mathcal{G}_\Theta(\Phi) := \mathcal{G}(\Phi) \cap \mathcal{M}_\Theta(\Omega)$, that is, the set of shift-invariant Gibbs measures for Φ, is never empty. It is a simplex whose set of extreme points, denoted by $\operatorname{ex} \mathcal{G}_\Theta(\Phi)$, coincides with the set of ergodic Gibbs measures for Φ, that is, $\operatorname{ex} \mathcal{G}_\Theta(\Phi) = \mathcal{G}_\Theta(\Phi) \cap \operatorname{ex} \mathcal{M}_\Theta(\Omega)$. Of course, when $\mathcal{G}(\Phi)$ is a singleton, then the unique Gibbs measure is shift-invariant and ergodic.

We now define relative entropy which plays a central role in this paper. Let $\mu, \nu \in \mathcal{M}_\Theta(\Omega)$. For each $n \in \mathbb{N}$, we denote by ν_n (resp. μ_n) the probability measure induced by ν (resp. μ) on S^{Λ_n} by projection. Then the relative entropy of ν_n with respect to μ_n is defined by

$$H_n(\nu|\mu) = \sum_{p_n \in S^{\Lambda_n}} \nu([\,p_n]) \log \frac{\nu([\,p_n])}{\mu([\,p_n])}.$$

We denote by log the natural logarithm.

Definition 3.1 (Relative entropy).
Let $\mu, \nu \in \mathcal{M}_\Theta(\Omega)$. The lower and upper relative entropies of ν with respect to μ are

$$h_*(\nu|\mu) = \liminf_{n\to+\infty} \frac{H_n(\nu|\mu)}{(2n+1)^d} \quad \text{and} \quad h^*(\nu|\mu) = \limsup_{n\to+\infty} \frac{H_n(\nu|\mu)}{(2n+1)^d}. \tag{8}$$

When the limit exists, we put $h(\nu|\mu) = h_*(\nu|\mu) = h^*(\nu|\mu)$.

It is well-known that $h_*(\nu|\mu)$ and $h^*(\nu|\mu)$ are nonnegative numbers. Let $\Phi \in \mathscr{B}_\Theta$ and $\mu \in \mathcal{G}_\Theta(\Phi)$. It is proved in [14, Chapter 15] that for any $\nu \in \mathcal{M}_\Theta(\Omega)$ we have

$$h_*(\nu|\mu) = h^*(\nu|\mu) = h(\nu|\mu) = P(\Phi) + \sum_{\Lambda \ni 0} |\Lambda|^{-1} \mathbb{E}_\nu[\Phi(\Lambda, \cdot)] - h(\nu) \tag{9}$$

where $P(\Phi)$ is the pressure of Φ and $h(\nu)$ is the entropy of ν. (We do not need to give the precise definitions of these two quantities since they are not used explicitly in the sequel.) Notice that, given $\nu \in \mathcal{M}_\Theta(\Omega)$, $h(\nu|\mu)$ is the same number for all $\mu \in \mathcal{G}_\Theta(\Phi)$, so it is natural to define, for each $\nu \in \mathcal{M}_\Theta(\Omega)$,

$$h(\nu|\Phi) := h(\nu|\mu) \quad \text{where } \mu \text{ is any element in } \mathcal{G}_\Theta(\Phi).$$

We recall the definition of equilibrium states.

Definition 3.2 (Equilibrium states).
Let $\Phi \in \mathscr{B}_\Theta$. A shift-invariant probability measure ν such that $h(\nu|\Phi) = 0$ is called an equilibrium state for Φ.

We have the following fundamental result which is usually referred to as the variational principle for equilibrium states.

Theorem 3.1 ([14, Chapter 15], [22, Theorem 4.2]).
Let $\Phi \in \mathscr{B}_\Theta$. We have $h(\nu|\Phi) = 0$ if and only if $\nu \in \mathcal{G}_\Theta(\Phi)$. In particular, $\mathcal{G}_\Theta(\Phi)$ coincides with the set of equilibrium states for Φ.

3.2 Examples

3.2.1 Dobrushin's uniqueness

We first recall that a sufficient condition for GCB to hold is Dobrushin's uniqueness condition which guarantees that $\Phi \in \mathscr{B}_\Theta$ admits a unique Gibbs measure. Under this condition, this Gibbs measure satisfies a Gaussian concentration bound with a constant equal to $(2(1 - \mathfrak{c}(\boldsymbol{\gamma}^\Phi))^2)^{-1}$, where $\mathfrak{c}(\boldsymbol{\gamma}^\Phi) < 1$. This holds for instance for any finite-range potential $\beta\Phi$ provided that $\beta > 0$ is small enough. A basic example is the nearest-neighbor ferromagnetic Ising model. Note that if one adds a external uniform magnetic field h in this model, then Dobrushin's uniqueness condition can hold for every $\beta > 0$ if $|h|$ is large enough, hence we can have GCB at low temperature. We refer to [14, Chapter 8] for more details and examples.

3.2.2 Ising model

The nearest-neighbor ferromagnetic Ising model on $\mathbb{Z}^d$ with zero external magnetic field is a basic example illustrating the above definitions and results. Take $S = \{-1, +1\}$, $d \geq 2$, and

$$\Phi(\Lambda, \omega) = \begin{cases} -\,\omega_x\,\omega_y & \text{if} \quad \Lambda = \{x, y\} \text{ and } \|x - y\|_1 = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Denote by $\mu^+_{\beta\Phi}$ (resp. $\mu^-_{\beta\Phi}$) is the Gibbs measure obtained with the "+" boundary condition (resp. the "−" boundary condition). It is well-known that there exists $\beta_c = \beta_c(d)$ such that, for all $\beta > \beta_c$, $\mu^+_{\beta\Phi}$ and $\mu^-_{\beta\Phi}$ are distinct ergodic equilibrium states. According to our main theorem, they cannot satisfy GCB.

For $d = 2$, there is an inverse temperature β_c such that, for all $\beta \leq \beta_c$, $\mathcal{G}_\Theta(\beta\Phi)$ is a singleton, and, for all $\beta > \beta_c$, $\mathcal{G}_\Theta(\beta\Phi) = \{\lambda\mu^+_{\beta\Phi} + (1-\lambda)\mu^-_{\beta\Phi} : \lambda \in [0, 1]\}$ (hence $\operatorname{ex}\mathcal{G}_\Theta(\beta\Phi) = \{\mu^+_{\beta\Phi}, \mu^-_{\beta\Phi}\}$). It is known (see [11, p. 172]) that for $\beta = \beta_c = \log(1 + \sqrt{2})/2$, there is a unique Gibbs measure μ_{β_c} such that

$$\lim_{n\to\infty} \frac{1}{|\Lambda_n|} \mathrm{Var}_{\mu_{\beta_c}}\left(\sum_{x\in\Lambda_n} \omega_x\right) = +\infty. \tag{10}$$

Therefore, μ_{β_c} cannot satisfy a Gaussian concentration bound because, by (7), one would have

$$\limsup_{n\to\infty} \frac{1}{|\Lambda_n|} \mathrm{Var}_{\mu_{\beta_c}}\left(\sum_{x\in\Lambda_n} \omega_x\right) \leq 8D \tag{11}$$

which contradicts (10). To obtain (11), apply (7) to $F(\omega) = \sum_{x\in\Lambda_n} \pi_{\{0\}}(\theta_x\,\omega)$, where $\pi_{\{0\}}(\omega) = \omega_0$. Then use Lemma A.1 to get

$$\mathrm{Var}_{\mu_{\beta_c}}\left(\sum_{x\in\Lambda_n} \omega_x\right) \leq 8D|\Lambda_n|$$

for all n. Therefore, the Gibbs measure for the 2D Ising model at critical temperature does not satisfy a Gaussian concentration bound. It does not even satisfy the variance inequality (7).

Remark 3.1. *Let us mention that GCB holds for all $\beta < \beta_c$ in the 2D Ising model. This follows by combining several results taken from [20, 23, 24]. They involve log-Sobolev inequality, 'complete analyticity' in the sense of Dobrushin and Shlosman, and a particular phenomenon arising only in dimension two for finite-range potentials. We refer to [18] for details. This example suggests that GCB is equivalent to complete analyticity in any dimension and for any finite-range potential, but we are not able to prove this at present.*

For $d = 3$, the situation is more complicated at low temperatures. Indeed, in addition to $\mu^+_{\beta\Phi}$ and $\mu^-_{\beta\Phi}$, $\mathcal{G}(\beta\Phi)$ also contains, for β large enough, a family of Gibbs measures which are not shift-invariant, the so-called Dobrushin states. In other words, $\mathcal{G}(\beta\Phi)\backslash\mathcal{G}_\Theta(\beta\Phi) \neq \emptyset$. In the present paper, we do not deal with these non-shift invariant Gibbs measures for the Ising model (and other models). Indeed, whereas for translation invariant Gibbs states we obtain here a general uniqueness result, the situation becomes much more intricate for non-translation Gibbs states, and even more for non-translation invariant potentials. In fact, Dobrushin interface states can be shown to be incompatible with GCB, using the volume large deviation bound (19). However, other more subtle scenarios of non-uniqueness combined with a unique translation invariant Gibbs measure can occur, such as in [3], and with the techniques developed in this paper, we cannot show that GCB excludes such scenarios of non-uniqueness.

3.2.3 Dyson model

Consider the Dyson model: $S = \{-1, +1\}$, $d = 1$, and

$$\Phi(\Lambda, \omega) = \begin{cases} -\frac{\omega_x\,\omega_y}{|x-y|^\alpha} & \text{if} \quad \Lambda = \{x, y\} \text{ such that } x \neq y \\ 0 & \text{otherwise} \end{cases}$$

with $\alpha > 1$. As in the Ising model, consider the Gibbs measures $\mu^+_{\beta\Phi}$ and $\mu^-_{\beta\Phi}$ obtained as the infinite-volume limits of the corresponding specification with the "+" and the "−" boundary conditions, respectively. Let $1 < \alpha \leq 2$. There exists $\beta_c > 0$ such that, for all $\beta < \beta_c$, $\mu^+_{\beta\Phi} \neq \mu^-_{\beta\Phi}$, and $\operatorname{ex}\mathcal{G}_\Theta(\beta\Phi) = \{\mu^+_{\beta\Phi}, \mu^-_{\beta\Phi}\}$. We refer to [19] for the relevant references. By the above theorem, these two equilibrium states cannot satisfy a Gaussian concentration bound.

3.2.4 Other examples

For instance, consider the Potts model with $S = \{1, \ldots, N\}$ for sufficiently large N, for which there exists $0 < \beta_N < +\infty$ such that $\left|\operatorname{ex}\mathcal{G}_\Theta(\beta_N\Phi)\right| = N + 1$, one of these measures being symmetric under spin flip, and $\left|\operatorname{ex}\mathcal{G}_\Theta(\beta\Phi)\right| = N$ when $\beta > \beta_N$. Hence, there are $N + 1$ ergodic equilibrium states in this model which, at inverse temperature β_N, cannot satisfy a Gaussian concentration bound, and when $\beta > \beta_N$, there are N such ergodic equilibrium states. See [14, Chapter 19] for details and other examples.

4 Proof of the main result

4.1 An abstract result

The first way to prove the main theorem is to establish an abstract theorem which is of independent interest, and can be applied to equilibrium states. To state it, we need to define the "positive relative entropy property".

Definition 4.1 (Positive relative entropy property).
An ergodic probability measure μ on Ω is said to have the positive relative entropy property if $h_*(\nu|\mu) > 0$ for every ergodic probability measure $\nu \neq \mu$.

The lower relative entropy $h_*(\nu|\mu)$ is defined in (8). We can now state the following result. (Recall that by Proposition 2.2 we only need to consider *ergodic* measures.)

Theorem 4.1 (GCB implies positive relative entropy).
Let μ be an ergodic probability measure on Ω which satisfies GCB. Then μ has the positive relative entropy property.

Sketch of the proof. We outline the proof which consists in the following three steps. We first prove that GCB implies the so-called "blowing-up" property (Section 4.2). Then we prove that the blowing-up property implies the "exponential rate of convergence for frequencies" (Section 4.3). Finally we prove that the latter implies the positive relative entropy property (Section 4.4). □

Now the main theorem stated in the introduction is a corollary of the previous theorem.

Corollary 4.1.
Let $\Phi \in \mathscr{B}_\Theta$. If $|\operatorname{ex} \mathcal{G}_\Theta(\Phi)| \geq 2$ and $\mu \in \operatorname{ex} \mathcal{G}_\Theta(\Phi)$, then μ cannot satisfy a Gaussian concentration bound.

Proof. Fix an arbitrary $\mu \in \operatorname{ex} \mathcal{G}_\Theta(\Phi)$. By assumption there exists $\mu' \in \operatorname{ex} \mathcal{G}_\Theta(\Phi)$ such that $\mu \neq \mu'$. But, by Theorem 3.1, we have $h(\mu'|\Phi) = h_*(\mu'|\mu) = 0$, whence μ does not have the positive relative entropy property. Therefore, according to Theorem 4.1, μ cannot satisfy GCB. This proves that none of the elements of $\operatorname{ex} \mathcal{G}_\Theta(\Phi)$ can satisfy GCB. □

4.2 GCB implies blowing-up

Let $\Lambda \Subset \mathbb{Z}^d$ (finite subset of $\mathbb{Z}^d$). We define the (non-normalized) Hamming distance between two configurations ω and η in S^Λ by

$$\bar{\mathrm{d}}_\Lambda(\omega, \eta) = \sum_{x \in \Lambda} \mathbb{1}_{\{\omega_x \neq \eta_x\}} \in \{0, 1, \ldots, |\Lambda|\}.$$

So we count at how many sites the configurations ω and η in S^Λ differ. Clearly, we can see $\bar{\mathrm{d}}_\Lambda$ as a local function on $\Omega \times \Omega$. Given a subset $C \subseteq S^\Lambda$ define

$$\bar{\mathrm{d}}_\Lambda(\omega, C) = \inf_{\omega' \in C} \bar{\mathrm{d}}_\Lambda(\omega, \omega').$$

Given $\varepsilon \in [0,1]$, define the ε-neighborhood (or ε-blow-up) of C as

$$\langle C \rangle_\varepsilon = \left\{ \omega \in S^\Lambda : \bar{\mathrm{d}}_\Lambda(\omega, C) < \varepsilon |\Lambda| \right\} \subseteq S^\Lambda.$$

Recall that if $C \subseteq S^\Lambda$, $[C] = \{\omega \in \Omega : \pi_\Lambda(\omega) \in C\}$ where π_Λ is the projection from Ω onto S^Λ. We now define the blowing-up property.

Definition 4.2 (Blowing-up property)**.**
An ergodic probability measure μ on $(\Omega, \mathfrak{B})$ has the blowing-up property if given $\varepsilon > 0$ there is a $\delta > 0$ and an N such that if $n \geq N$ and $C \subseteq S^{\Lambda_n}$ then

$$\mu([C]) \geq \mathrm{e}^{-(2n+1)^d \delta} \quad \text{implies} \quad \mu(\langle C \rangle_\varepsilon) \geq 1 - \varepsilon.$$

Obviously, the blowing-up property can be formulated in terms of finite subsets of $\mathbb{Z}^d$ of arbitrary shape, instead of cubes, but we will not need this generalization. The blowing-up property roughly says that any collection of configurations on a large finite box which has a total measure which is not too exponentially small is such that most configurations are close to this collection in the Hamming distance. We have the following result.

Proposition 4.1.
Let $\Lambda \Subset \mathbb{Z}^d$ and $C \subseteq S^\Lambda$. Suppose that μ is a probability measure which satisfies GCB(D) *and such that $\mu([C]) > 0$. Then, we have*

$$\mu\big(\langle C \rangle_\varepsilon\big) \geq 1 - \exp\left[-\frac{|\Lambda|}{4D} \left(\varepsilon - \frac{2\sqrt{D \log(\mu([C])^{-1})}}{\sqrt{|\Lambda|}} \right)^2 \right] \tag{12}$$

whenever $\varepsilon > 2\sqrt{D \log(\mu([C])^{-1})/|\Lambda|}$. In particular, μ satisfies the blowing-up property.

Note that we do not require μ to be shift-invariant. As already mentioned, if μ is taken shift-invariant then it must be ergodic, this is enforced by the Gaussian concentration bound.

Proof. Consider the local function $F(\omega) = \bar{\mathrm{d}}_\Lambda(\omega, C)$. One easily checks that $\delta_x(F) \leq 1$ for all $x \in \Lambda$. Applying (5) gives

$$\mu\left\{ \omega \in \Omega : F(\omega) \geq u + \mathbb{E}_\mu(F) \right\} \leq \exp\left(-\frac{u^2}{4D|\Lambda|} \right) \tag{13}$$

for all $u > 0$. We now bound $\mathbb{E}_\mu(F)$ from above. Applying (4) to $-\lambda F$, for some $\lambda > 0$ to be chosen later on, we get

$$\exp\left(\lambda \mathbb{E}_\mu(F)\right) \mathbb{E}_\mu\big[\exp\left(-\lambda F\right) \big] \leq \exp\left(D\lambda^2 |\Lambda| \right).$$

Observe that by definition of F we have

$$\mathbb{E}_\mu\big[\exp(-\lambda F)\big] \geq \mathbb{E}_\mu\big[\mathbb{1}_C \exp(-\lambda F)\big] = \mu([C]).$$

Combining the two previous inequalities and taking the logarithm gives

$$\mathbb{E}_\mu(F) \leq \inf_{\lambda>0}\Big\{D\lambda|\Lambda| + \frac{1}{\lambda}\ln\big(\mu([C])^{-1}\big)\Big\}$$

which gives the following estimate by taking the value of λ minimizing the right-hand side

$$\mathbb{E}_\mu(F) \leq 2\sqrt{D|\Lambda|\ln\big(\mu([C])^{-1}\big)} =: v_0.$$

Therefore inequality (13) implies that

$$\mu\left\{\omega\in\Omega : F(\omega)\geq v\right\} \leq \exp\left(-\frac{(v-v_0)^2}{4D|\Lambda|}\right)$$

for all $v > v_0$. To finish the proof of (12), take $v = \varepsilon|\Lambda|$ and observe that, by definition of F, $\mu\left\{\omega\in\Omega : F(\omega)\geq v\right\} = \mu\big(\Omega\backslash\langle C\rangle_\varepsilon\big)$.
We now prove that μ satisfies the blowing-up property. Fix $\varepsilon > 0$ arbitrarily, and take C such that $\mu([C]) \geq \mathrm{e}^{-(2n+1)^d\delta}$ for some $\delta > 0$ to be chosen later on, subject to the condition $\varepsilon > 2\sqrt{D\delta}$. We now apply (12) with $\Lambda = \Lambda_n$ to get

$$\mu(\langle C\rangle_\varepsilon) \geq 1 - \exp\left(-\frac{(2n+1)^d}{4D}\Big(\varepsilon - 2\sqrt{D\delta}\Big)^2\right).$$

Taking $\delta = \varepsilon^2/(9D)$ yields

$$\mu(\langle C\rangle_\varepsilon) \geq 1-\varepsilon$$

for all $n \geq N := \lfloor(36D\varepsilon^{-2}\log\varepsilon^{-1})^{1/d}\rfloor/2$. $\square$

4.3 Blowing-up implies exponential rate for frequencies

Given $\omega\in\Omega$, $n > k \geq 0$, and a pattern $p_k \in S^{\Lambda_k}$, let

$$\mathfrak{f}_{n,k}(\omega; p_k) = \frac{|\{x\in\Lambda_{n-k} : (\theta_x\,\omega)_{\Lambda_k} = p_k\}|}{(2(n-k)+1)^d}. \tag{14}$$

In words, this is simply the frequency of occurrence of the pattern p_k if we look at the configuration ω restricted to the cube Λ_n. Let μ be an ergodic probability measure. By the multidimensional ergodic theorem, for μ-almost every ω, we have

$$\lim_{n\to\infty}\mathfrak{f}_{n,k}(\omega; p_k) = \mu([p_k]).$$

Given two probability measures μ and ν on Ω, recall that ν_k (resp. μ_k) is the probability measure induced by ν (resp. μ) on S^{Λ_k} by projection. The total variation distance between μ_k and ν_k is defined by

$$\|\mu_k - \nu_k\|_{\mathrm{TV}} = \frac{1}{2} \sum_{p_k \in S^{\Lambda_k}} |\mu([p_k]) - \nu([p_k])|.$$

We can now define the property of exponential rate of convergence for frequencies.

Definition 4.3 (Exponential rate of convergence for frequencies)**.** *An ergodic probability measure μ on $(\Omega, \mathfrak{B})$ has the exponential rate of convergence property for frequencies if, given k and $\varepsilon > 0$, there is a $\delta > 0$ and an N such that*

$$\mu\left\{\omega \in \Omega : \|\mathfrak{f}_{n,k}(\omega;\cdot) - \mu_k\|_{\mathrm{TV}} \geq \varepsilon\right\} \leq \mathrm{e}^{-(2n+1)^d \delta}, \ \forall n \geq N.$$

We now prove that this property is implied by the blowing-up property.

Proposition 4.2. *Let μ be an ergodic probability measure on $(\Omega, \mathfrak{B})$. If μ has the blowing-up property, then it has the exponential rate of convergence property for frequencies.*

Proof. We adapt a proof of [21] to our setting. Take $\varepsilon > 0$ and $k \geq 0$, and for any $n \geq 0$ let

$$\mathcal{B}(n,k,\varepsilon) = \left\{\omega \in \Omega : \|\mathfrak{f}_{n,k}(\omega;\cdot) - \mu_k\|_{\mathrm{TV}} \geq \varepsilon\right\}.$$

Note that we can naturally identify this subset of Ω with a subset of S^{Λ_n}. We have the following lemma whose proof is given in Appendix A.2 below.

Lemma 4.1. *Let $\varepsilon > 0$ and $k \geq 0$, and define*

$$\varrho = \frac{2\varepsilon}{5(2k+1)^d}.$$

There exists $\breve{N} > k$ such that, if $n \geq \breve{N}$ and $\bar{\mathrm{d}}_{\Lambda_n}(\omega,\eta) \leq \varrho\,(2n+1)^d$, then

$$\|\mathfrak{f}_{n,k}(\omega;\cdot) - \mathfrak{f}_{n,k}(\eta;\cdot)\|_{\mathrm{TV}} \leq \frac{\varepsilon}{2}.$$

The lemma implies that

$$\langle \mathcal{B}(n,k,\varepsilon)\rangle_\varrho \subseteq \mathcal{B}\Big(n,k,\frac{\varepsilon}{2}\Big). \tag{15}$$

Now, by the blowing-up property (Definition 4.2), there is a $\delta > 0$ and an N such that, if $n \geq N$ and $C \subseteq S^{\Lambda_n}$ so that $\mu([C]) \geq \mathrm{e}^{-(2n+1)^d\delta}$, then

$\mu(\langle C\rangle_\varrho) \geq 1-\varrho \geq 1-\varepsilon$. Therefore, if $n \geq N$ and if we suppose that $\mu\big(\mathcal{B}(n,k,\varepsilon)\big) \geq \mathrm{e}^{-(2n+1)^d\delta}$, then we get

$$\mu\big(\langle \mathcal{B}(n,k,\varepsilon)\rangle_\varrho\big) \geq 1-\varepsilon.$$

This bound together with the inclusion (15) implies that for all $n \geq \max(N, \breve{N})$

$$\mu\Big(\mathcal{B}\Big(n,k,\frac{\varepsilon}{2}\Big)\Big) \geq 1-\varepsilon.$$

But this contradicts the multidimensional ergodic theorem which ensures that, for each $\varepsilon > 0$ and each k,

$$\lim_{n\to\infty} \mu\Big(\mathcal{B}\Big(n,k,\frac{\varepsilon}{2}\Big)\Big) = 0.$$

Indeed, we have

$$\mu\Big(\mathcal{B}\Big(n,k,\frac{\varepsilon}{2}\Big)\Big) \leq \sum_{p_k\in S^{\Lambda_k}} \mu\left\{\omega\in\Omega : |\mathfrak{f}_{n,k}(\omega;p_k) - \mu([p_k])| \geq \frac{\varepsilon}{|S|^{(2k+1)^d}}\right\}$$

and each term in this finite sum goes to 0 as $n \to +\infty$. The proposition is proved. $\square$

4.4 Exponential rate for frequencies implies positive relative entropy

We now have the following proposition.

Proposition 4.3. *Let μ be an ergodic probability measure on $(\Omega, \mathfrak{B})$. If it has the exponential rate of convergence property for frequencies, then, given $\varepsilon > 0$ and k, there is a $\delta > 0$ such that, if ν is an ergodic probability measure such that $h_*(\nu|\mu) < \delta$ then $\|\nu_k - \mu_k\|_{\mathrm{TV}} < \varepsilon$.*

Proof. We adapt a proof of [21]. Let ρ be a positive number that we will specify later on, and suppose that

$$\frac{H_n(\nu|\mu)}{(2n+1)^d} < \frac{\rho^2}{2}.$$

If n is large enough, we have by Markov's inequality[2] and Lemma A.2 (see appendix below) that there exists a set $G_n \subseteq S^{\Lambda_n}$ such that

$$\nu([G_n]) > 1-\rho$$

[2]Let $f \geq 0$ be an integrable function on a probability space $(\mathcal{X}, \Sigma, \mu)$. If $\int f\,\mathrm{d}\mu \leq \rho^2$ then $f(x) \leq \rho$, except for a set of measure at most ρ.

and

$$\mathrm{e}^{-(2n+1)^d\rho}\,\nu([\omega_{\Lambda_n}]) \le \mu([\omega_{\Lambda_n}]) \le \mathrm{e}^{(2n+1)^d\rho}\,\nu([\omega_{\Lambda_n}]), \quad \omega \in [G_n]. \tag{16}$$

Now, for each fixed k, and for n large enough, the multidimensional ergodic theorem applied to ν tells us that there is a set $\widetilde{G}_n \subseteq G_n$ such that

$$\nu\big([\widetilde{G}_n]\big) > 1 - 2\rho \tag{17}$$

$$\|\mathfrak{f}_{n,k}(\omega;\cdot) - \nu_k\|_{\mathrm{TV}} < \rho, \quad \omega \in [\widetilde{G}_n]. \tag{18}$$

If μ has the exponential rate of convergence property for frequencies then

$$\mu\left\{\omega \in \Omega : \|\mathfrak{f}_{n,k}(\omega;\cdot) - \mu_k\|_{\mathrm{TV}} \ge \rho\right\} \le \mathrm{e}^{-(2n+1)^d\tau}$$

for some $\tau > 0$ and all n sufficiently large. Now, conditions (16) and (17) imply that

$$\mu\big([\widetilde{G}_n]\big) > (1-2\rho)\,\mathrm{e}^{-(2n+1)^d\rho}\,.$$

Therefore, if ρ is small enough and n large enough, there exists $\omega \in [\widetilde{G}_n]$ such that $\|\mathfrak{f}_{n,k}(\omega;\cdot) - \mu_k\|_{\mathrm{TV}} < \rho$. Indeed, it is enough to check that

$$\mu\big([\widetilde{G}_n]\big) > \mu\left\{\omega \in \Omega : \|\mathfrak{f}_{n,k}(\omega;\cdot) - \mu_k\|_{\mathrm{TV}} \ge \rho\right\}$$

which implies that

$$\big[\widetilde{G}_n\big] \cap \left\{\omega \in \Omega : \|\mathfrak{f}_{n,k}(\omega;\cdot) - \mu_k\|_{\mathrm{TV}} \ge \rho\right\}$$

has strictly positive μ-measure. Since (18) holds for this same ω, we thus arrive by the triangle inequality at the estimate $\|\nu_k - \mu_k\|_{\mathrm{TV}} < 2\rho$. Therefore, we proved that, given ε and k, if ν is ergodic such that $h_*(\nu|\mu) < \varepsilon^2/8$, then $\|\nu_k - \mu_k\|_{\mathrm{TV}} < \varepsilon$, which ends the proof. $\square$

Now we can state the main proposition of this section.

Proposition 4.4. *If μ has the exponential rate of convergence property for frequencies then it has the positive relative entropy property.*

Proof. Suppose that μ has not the positive relative entropy property, but satisfies the exponential rate of convergence property for frequencies. We will obtain a contradiction. By assumption there is an ergodic probability measure $\widehat{\nu} \neq \mu$ such that $h_*(\widehat{\nu}|\mu) = 0$, and there is a $\hat{k}$ and an $\widehat{\varepsilon} > 0$ such that $\|\widehat{\nu}_{\hat{k}} - \mu_{\hat{k}}\|_{\mathrm{TV}} \ge \widehat{\varepsilon}$. Now we apply Proposition 4.3. By the exponential rate of convergence property for frequencies, given $\hat{\varepsilon}$ and $\hat{k}$, there is a $\hat{\delta} > 0$ such that if ν satisfies $h_*(\nu|\mu) < \hat{\delta}$ then $\|\nu_{\hat{k}} - \mu_{\hat{k}}\|_{\mathrm{TV}} < \widehat{\varepsilon}$. We can take $\nu = \widehat{\nu}$, hence we arrive at $\|\widehat{\nu}_{\hat{k}} - \mu_{\hat{k}}\|_{\mathrm{TV}} < \widehat{\varepsilon}$, a contradiction. $\square$

5 A proof via large deviations

We present another proof of the main theorem, based on large deviations. Suppose that μ is an ergodic equilibrium state which satisfies a Gaussian concentration bound. Now assume that there exists $\mu' \in \operatorname{ex}\mathcal{G}_\Theta(\Phi)$ such that $\mu \neq \mu'$. We are going to arrive at a contradiction. Notice that we can suppose that μ' is ergodic without loss of generality, because of the ergodic decomposition [14, Theorem 14.17, p. 298], and the fact that the map $\nu \mapsto h(\nu|\Phi)$, $\nu \in \mathcal{M}_\Theta(\Omega)$, is affine (which is a consequence of [14, Theorem 15.20, p. 318] and (9)).

Since $\mu \neq \mu'$, there exists a local function $f : \Omega \to \mathbb{R}$ such that $\mathbb{E}_\mu(f) \neq \mathbb{E}_{\mu'}(f)$. Without loss of generality, assume that $\mathbb{E}_{\mu'}(f) > \mathbb{E}_\mu(f)$. So there exists $\varepsilon > 0$ such that

$$\mathbb{E}_\mu(f) + \varepsilon = \mathbb{E}_{\mu'}(f).$$

We want to apply Proposition 2.1 to $F = S_n f$, where, for each $n \geq 0$, $S_n f = \sum_{x\in\Lambda_n} f \circ \theta_x$. We claim that

$$\|\underline{\delta}(S_n f)\|_2^2 \leq (2n+1)^d \|\underline{\delta}(f)\|_1^2$$

where $\|\underline{\delta}(f)\|_1 := \sum_{x\in\mathbb{Z}^d} \delta_x(f)$ (which is finite since f is local). See Appendix A.1 for the proof. Letting

$$\mathcal{E}_{n,\varepsilon} := \left\{\omega \in \Omega : \frac{S_n f(\omega)}{(2n+1)^d} \geq \mathbb{E}_\mu(f) + \frac{\varepsilon}{3}\right\}.$$

we get

$$\mu(\mathcal{E}_{n,\varepsilon}) \leq \exp\left(-\frac{(2n+1)^d \varepsilon^2}{36D\|\underline{\delta}(f)\|_1^2}\right). \tag{19}$$

Now let

$$\mathcal{E}'_{n,\varepsilon} := \left\{\omega \in \Omega : \frac{S_n f(\omega)}{(2n+1)^d} \in \left]\mathbb{E}_{\mu'}(f) - \frac{\varepsilon}{3}, \mathbb{E}_{\mu'}(f) + \frac{\varepsilon}{3}\right[\right\}$$

Since

$$\mathcal{E}'_{n,\varepsilon} \subset \mathcal{E}_{n,\varepsilon}$$

we deduce that

$$\limsup_{n\to+\infty} \frac{1}{(2n+1)^d} \log \mu(\mathcal{E}'_{n,\varepsilon}) \leq -\frac{\varepsilon^2}{36D\|\underline{\delta}(f)\|_1^2} < 0. \tag{20}$$

Now we use the large deviation principle satisfied by μ (see [14, Section 15.5]) which implies that

$$\liminf_{n\to+\infty} \frac{1}{(2n+1)^d} \log \mu(\mathcal{E}'_{n,\varepsilon}) \geq -\inf_{u\in\left]\mathbb{E}_{\mu'}(f)-\frac{\varepsilon}{3}, \mathbb{E}_{\mu'}(f)+\frac{\varepsilon}{3}\right[} I_f(u) \tag{21}$$

where

$$I_f(u) = \inf \{h(\nu|\mu) : \nu \in \mathcal{M}_\Theta(\Omega), \mathbb{E}_\nu(f) = u\}.$$

The right-hand side of (21) is larger than $-I_f(v)$ for any value of v taken in the interval, so in particular it is larger than $-I_f(\mathbb{E}_{\mu'}(f))$, which is equal to 0 because $h(\mu'|\mu) = 0$ by Theorem 3.1. Hence we obtain

$$\liminf_{n\to+\infty} \frac{1}{(2n+1)^d} \log \mu(\mathcal{E}'_{n,\varepsilon}) = 0$$

which contradicts (20).

A Appendix

A.1 An estimate

We first recall Young's inequality for convolutions [4, p. 316]. Let $\underline{u} = (u_x)_{x\in\mathbb{Z}^d}$ and $\underline{v} = (v_x)_{x\in\mathbb{Z}^d}$. Formally define their convolution $\underline{u} * \underline{v}$ by

$$(\underline{u} * \underline{v})_x = \sum_{y\in\mathbb{Z}^d} u_{x-y} v_y, \ x \in \mathbb{Z}^d.$$

If $\underline{u} \in \ell^p(\mathbb{Z}^d)$ and $\underline{v} \in \ell^q(\mathbb{Z}^d)$, where $p, q \geq 1$, then $\underline{u} * \underline{v} \in \ell^r(\mathbb{Z}^d)$ where $r \geq 1$ is such that $1 + r^{-1} = p^{-1} + q^{-1}$, then we have

$$\|\underline{u} * \underline{v}\|_{\ell^r(\mathbb{Z}^d)} \leq \|\underline{u}\|_{\ell^p(\mathbb{Z}^d)} \|\underline{v}\|_{\ell^q(\mathbb{Z}^d)}.$$

Lemma A.1. *Let $f : \Omega \to \mathbb{R}$ such that $\|\underline{\delta}(f)\|_1 := \sum_{x\in\mathbb{Z}^d} \delta_x(f) < +\infty$. Then for any $\Lambda \Subset \mathbb{Z}^d$ we have*

$$\|\underline{\delta}(S_\Lambda f)\|_2^2 \leq |\Lambda| \, \|\underline{\delta}(f)\|_1^2.$$

Proof. Since $\delta_z(S_\Lambda f) \leq \sum_{x\in\Lambda} \delta_{z-x}(f)$, we apply Young's inequality with $r = 2, p = 2, q = 1$, $u_x = \mathbb{1}_\Lambda(x)$, and $v_x = \delta_x(f)$ to get the desired estimate. □

A.2 Proof Lemma 4.1

The version of this lemma in dimension $d = 1$ is stated without proof in [21]. Since it is not completely obvious, we give it here for any $d \geq 1$.

We fix $\varepsilon > 0$ and $k \geq 0$. The frequency of a pattern $p_k \in S^{\Lambda_k}$ in ω (see (14)) can rewritten as

$$\mathfrak{f}_{n,k}(\omega; p_k) = \frac{1}{(2(n-k)+1)^d} \sum_{x\in\Lambda_{n-k}} \mathbb{1}_{\{(\theta_x\omega)_{\Lambda_k} = p_k\}}.$$

By definition we have

$$\|\mathfrak{f}_{n,k}(\omega;\cdot)-\mathfrak{f}_{n,k}(\eta;\cdot)\|_{\mathrm{TV}} = \frac{1}{2(2(n-k)+1)^d}\sum_{p_k\in S^{\Lambda_k}}\left|\sum_{x\in\Lambda_{n-k}}\left(\mathbb{1}_{\{(\theta_x\omega)_{\Lambda_k}=p_k\}}-\mathbb{1}_{\{(\theta_x\eta)_{\Lambda_k}=p_k\}}\right)\right|. \tag{22}$$

Letting

$$\mathcal{I}_{\omega,\eta,n}=\left\{x\in\Lambda_{n-k}:(\theta_x\omega)_{\Lambda_k}=(\theta_x\eta)_{\Lambda_k}\right\}$$

we get

$$\begin{aligned}
&\sum_{p_k\in S^{\Lambda_k}}\left|\sum_{x\in\Lambda_{n-k}}\left(\mathbb{1}_{\{(\theta_x\omega)_{\Lambda_k}=p_k\}}-\mathbb{1}_{\{(\theta_x\eta)_{\Lambda_k}=p_k\}}\right)\right| \\
&\qquad=\sum_{p_k\in S^{\Lambda_k}}\left|\sum_{x\in\mathcal{I}^c_{\omega,\eta,n}}\left(\mathbb{1}_{\{(\theta_x\omega)_{\Lambda_k}=p_k\}}-\mathbb{1}_{\{(\theta_x\eta)_{\Lambda_k}=p_k\}}\right)\right| \\
&\qquad\le\sum_{p_k\in S^{\Lambda_k}}\sum_{x\in\mathcal{I}^c_{\omega,\eta,n}}\left|\mathbb{1}_{\{(\theta_x\omega)_{\Lambda_k}=p_k\}}-\mathbb{1}_{\{(\theta_x\eta)_{\Lambda_k}=p_k\}}\right| \\
&\qquad\le\sum_{p_k\in S^{\Lambda_k}}\sum_{x\in\mathcal{I}^c_{\omega,\eta,n}}\left(\mathbb{1}_{\{(\theta_x\omega)_{\Lambda_k}=p_k\}}+\mathbb{1}_{\{(\theta_x\eta)_{\Lambda_k}=p_k\}}\right) \\
&\qquad=\sum_{x\in\mathcal{I}^c_{\omega,\eta,n}}\sum_{p_k\in S^{\Lambda_k}}\left(\mathbb{1}_{\{(\theta_x\omega)_{\Lambda_k}=p_k\}}+\mathbb{1}_{\{(\theta_x\eta)_{\Lambda_k}=p_k\}}\right) \\
&\qquad=2\left|\mathcal{I}^c_{\omega,\eta,n}\right|.
\end{aligned}$$

Hence we obtain from (22)

$$\|\mathfrak{f}_{n,k}(\omega;\cdot)-\mathfrak{f}_{n,k}(\eta;\cdot)\|_{\mathrm{TV}}\le\frac{\left|\mathcal{I}^c_{\omega,\eta,n}\right|}{(2(n-k)+1)^d}. \tag{23}$$

We now look for an upper bound for $\left|\mathcal{I}^c_{\omega,\eta,n}\right|$. If $(\theta_x\omega)_{\Lambda_k}=p_k$ and $(\theta_x\eta)_{\Lambda_k}\neq p_k$, then $\omega_y\neq\eta_y$ for at least one site $y\in\Lambda_k+x$. Such a y can produce as many as $(2k+1)^d$ sites such that $(\theta_x\omega)_{\Lambda_k}=p_k$ and $(\theta_x\eta)_{\Lambda_k}\neq p_k$. Hence

$$\begin{aligned}
\left|\mathcal{I}^c_{\omega,\eta,n}\right| &\le(2k+1)^d\left|\{x\in\Lambda_{n-k}:\omega_x\neq\eta_x\}\right| \\
&\le(2k+1)^d\left|\{x\in\Lambda_n:\omega_x\neq\eta_x\}\right| \\
&\le(2k+1)^d\,\bar{\mathrm{d}}_{\Lambda_n}(\omega,\eta).
\end{aligned}$$

Hence (23) yields

$$\|\mathfrak{f}_{n,k}(\omega;\cdot)-\mathfrak{f}_{n,k}(\eta;\cdot)\|_{\mathrm{TV}}\le\frac{(2k+1)^d}{(2(n-k)+1)^d}\,\bar{\mathrm{d}}_{\Lambda_n}(\omega,\eta).$$

Obviously there exists $\breve{N} > k$ such that for all $n \geq \breve{N}$ we have

$$\left(\frac{2n+1}{2(n-k)+1}\right)^d \leq \frac{5}{4}$$

therefore, if we take

$$\bar{\mathrm{d}}_{\Lambda_n}(\omega, \eta) \leq \frac{2\varepsilon}{5(2k+1)^d}\,(2n+1)^d$$

we finally obtain

$$\|\mathfrak{f}_{n,k}(\omega;\cdot) - \mathfrak{f}_{n,k}(\eta;\cdot)\|_{\mathrm{TV}} \leq \frac{\varepsilon}{2}$$

for all $n \geq \breve{N}$, which concludes the proof of the lemma.

A.3 A bound on relative entropy

Recall that 'log' stands for the natural logarithm. We were not able to find a reference for a proof of the following estimate, so we prove it for the reader's convenience.

Lemma A.2. *Let ν and μ be probability measures on a finite set A. Then*

$$\sum_{a\in A} \nu(\{a\}) \left|\log \frac{\nu(\{a\})}{\mu(\{a\})}\right| \leq H(\nu|\mu) + \frac{2}{\mathrm{e}} \tag{24}$$

where

$$H(\nu|\mu) = \sum_{a\in A} \nu(\{a\}) \log \frac{\nu(\{a\})}{\mu(\{a\})}.$$

Proof. Define

$$A^- = \left\{a \in A : \log \frac{\nu(\{a\})}{\mu(\{a\})} < 0\right\}.$$

Now

$$\begin{aligned}\sum_{a\in A} \nu(\{a\}) \left|\log \frac{\nu(\{a\})}{\mu(\{a\})}\right| &= \sum_{a\in A\setminus A^-} \nu(\{a\}) \log \frac{\nu(\{a\})}{\mu(\{a\})} + \sum_{a\in A^-} \nu(\{a\}) \log \frac{\mu(\{a\})}{\nu(\{a\})} \\ &= H(\nu|\mu) + 2\sum_{a\in A^-} \nu(\{a\}) \log \frac{\mu(\{a\})}{\nu(\{a\})}.\end{aligned}$$

By the concavity of the logarithm function and Jensen's inequality we get

$$\begin{aligned}\sum_{a\in A^-} \nu(\{a\}) \log \frac{\mu(\{a\})}{\nu(\{a\})} &= \nu(A^-) \sum_{a\in A^-} \frac{\nu(\{a\})}{\nu(A^-)} \log \frac{\mu(\{a\})}{\nu(\{a\})} \\ &\leq \nu(A^-) \log\left(\frac{\mu(A^-)}{\nu(A^-)}\right) \\ &\leq \nu(A^-) \log\left(\frac{1}{\nu(A^-)}\right) \leq \mathrm{e}^{-1}\end{aligned}$$

where we used the elementary inequality $-x \log x \leq \mathrm{e}^{-1}$, $x \geq 0$. Therefore we arrive at (24). $\square$

Acknowledgements. We thank Pierre Collet for stimulating discussions. The authors also thank the anonymous referee for very useful comments.

References

[1] J. van den Berg, J. Steif. On the Existence and Nonexistence of Finitary Codings for a Class of Random Fields. Ann. Probab. **27** (1999), no. 3, 1501–1522.

[2] S. Boucheron, G. Lugosi, P. Massart. *Concentration inequalities.* A nonasymptotic theory of independence. Oxford University Press, 2013.

[3] C. Borgs, Translation symmetry breaking in four dimensional lattice gauge theories. Comm. Math. Phys. **96** (1984), 251–284.

[4] P. Bullen. *Dictionary of inequalities.* Monographs and Research Notes in Mathematics. CRC Press, second edition, 2015.

[5] J.-R. Chazottes, P. Collet, C. Külske, F. Redig. Concentration inequalities for random fields via coupling. Probab. Theory Related Fields **137** (2007), no. 1-2, 201–225.

[6] J.-R. Chazottes, P. Collet, F. Redig. On concentration inequalities and their applications for Gibbs measures in lattice systems. J. Stat. Phys. **169** (2017), issue 3, pp 504–546.

[7] J.-R. Chazottes, S. Gallo, D. Takahashi. Optimal Gaussian concentration bounds for stochastic chains of unbounded memory. Preprint, 2019.

[8] R. L. Dobrushin, S. Shlosman. Completely analytical Gibbs fields. Statistical physics and dynamical systems (Köszeg, 1984), 371–403, Progr. Phys. **10**, Birkhäuser, 1985.

[9] R. L. Dobrushin, S. Shlosman. Completely analytical interactions: constructive description. J. Statist. Phys. **46** (1987), no. 5-6, 983–1014.

[10] D.P. Dubhashi, A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press, 2009.

[11] R. Ellis. *Entropy, large deviations, and statistical mechanics.* Classics in Mathematics. Springer-Verlag, Berlin, 2006.

[12] A. Eizenberg, Y. Kifer, and B. Weiss. Large deviations for $\mathbb{Z}^d$-actions. Comm. Math. Phys. **164** (1994), 433-454.

[13] S. Friedli, Y. Velenik. *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction.* Cambridge University Press, 2017.

[14] H.-O. Georgii. *Gibbs Measures and Phase Transitions.* Second edition. Walter de Gruyter, 2011.

[15] G. Keller. *Equilibrium States in Ergodic Theory.* London Mathematical Society Student Texts vol. 42, 1998.

[16] C. Külske. Concentration inequalities for functions of Gibbs fields with applications to diffraction and random Gibbs measures. Comm. Math. Phys. **239**, 29–51 (2003).

[17] M. Ledoux. *The concentration of measure phenomenon*, Mathematical Surveys and Monographs **89**. American Mathematical Society, Providence R.I., 2001.

[18] J. Moles. On concentration inequalities for lattice systems and thermodynamic formalism. PhD thesis, Institut Polytechnique de Paris, 2020.

[19] A. Le Ny. Gibbs Measures for Long-Range Ising Models (2019). arXiv:1911.01860 [math-ph].

[20] F. Martinelli, E. Olivieri, R. H. Schonmann. For 2-D lattice spin systems weak mixing implies strong mixing. Comm. Math. Phys. **165** (1994), no. 1, 33–47.

[21] K. Marton, P. Shields. The positive-divergence and blowing-up properties. Israel J. Math. (1994) **86**, Issue 1–3, pp 331–348.

[22] D. Ruelle, D. (2004). *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics.* Cambridge University Press, 2004.

[23] D. Stroock, B. Zegarliński. The equivalence of the logarithmic Sobolev inequality and the Dobrushin-Shlosman mixing condition. Comm. Math. Phys. **144** (1992), no. 2, 303–323.

[24] D. Stroock, B. Zegarliński. The logarithmic Sobolev inequality for discrete spin systems on a lattice. Comm. Math. Phys. **149** (1992), no. 1, 175–193.

[25] M. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint.* (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press, 2019.