Contour Ensembles and the Description of Gibbsian Probability Distributions at Low Temperature

Roberto Fernández

Slightly corrected version, August 1998

Foreword

Gibbsian probability distributions where introduced in statistical mechanics to describe the equilibrium properties of "macroscopic" physical samples, i.e. infinite-volume systems. In the last decades they have become widely used in a variety other settings as well (Gibbs sampling, simulated annealing, hidden Markov processes), and their theory has become a fertile chapter in probability theory (see for instance the comprehensive book by Georgii [18]). The goal of statistical mechanics —namely to catalogue and describe the different Gibbsian distributions admitted by each given system— has, therefore, become central for all these applications:

The methods used for the description of Gibbsian distributions can be roughly divided into two (interrelated) groups: (i) methods to estimate and bound decay properties of the correlation functions, and (ii) methods to determine the properties of the "typical" or "most probable" events (families of configurations). The course is intended as an introduction of a powerful method of this last, and less known, type.

The techniques to be discussed, globally known as *Pirogov-Sinai theory* or *contour arguments*, though elementary from the mathematical point of view, remain, unfortunately, well known only to a reduced clique of practitioners. As a byproduct of this state of affairs, the theory remains severely under-exploited. It is probably safe to state that the various known applications of the theory —determination of phase diagrams, study of interfaces among coexisting pure phases, determination of finite-volume corrections— are only a small fraction of the potential uses of it. One possible explanation for this situation is the lack of pedagogical and simplified expositions, free of the heavy notation and abstract bias that plague most papers on the subject.

The present course is intended as a modest contribution to revert this fact.

In consistency with this goal, the presentation will systematically be made *at the simplest possible level* so to convey the conceptual basis of the technique without unnecessary jargon or generality. The theory will be stated in its most classical and elementary version, and will be illustrated via the simplest non-trivial examples available. The course is aimed to a general audience of (potential or actual) users of Gibbsian theory, not restricted to specialists in rigorous statistical mechanics. Of course, a number of statistical mechanical notions will be used, but they will be introduced at the beginning.

In fact, a large part of the course (roughly half of it) is devoted to *prepare* the audience for the fundamental result of the theory. An effort is made to present the big picture in a brief but complete way: type of phenomena under study, different scenarios, scope of the theory, main tools, tricks of the trade. I attempt a clear exposition, using words besides formulas, even at the risk of exposing myself to criticism. Words are, by force, imprecise, when compared with a rigid definition-lemma-theorem-corollary approach. But this has been, precisely, one of the problems of the papers on the subject: They are mathematically impeccable, but hard to understand. These notes are an attempt more or less in the opposite direction.

My personal knowledge of the subject was acquired through many discussions with enlightened colleagues and through hard work with industrious coauthors. To my advisor, Joseph Slawny, I owe the crucial basic knowledge of almost everything in these notes: Gibbs states, Peierls argument, cluster expansions, Pirogov-Sinai theory. Later, my understanding of the latter benefited immensely from discussions with Christian Borgs, Roman Kotecký, Charles Pfister and Milos Zahradník (in alphabetic order), and my confidence in handling it developed during collaborations with Aernout van Enter, Nilanjana Datta, Jürg Fröhlich, Roman Kotecký, Luc Rey-Bellet and Alan Sokal. Of course they are not to blame for errors in these notes!

I take this opportunity to thank the organizers of the 21 Coloquio Matemático Brasileiro for inviting me to give this course and to the Institute for Theoretical Physics of the Swiss Federal Institute for Technology in Zürich (ETHZ) and the Instituto de Matemática e Es-

tatística of the University of São Paulo for hospitality while these notes were being written. I especially thank the members of the probability group at IME–USP for encouragement and support. I also acknowledge the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina, for granting me the leave of absence to visit the Universidade de São Paulo.

Table of Contents

Chapter 1

Introduction: Plan of the lectures

Contour arguments rely on the observation that, in many systems of interest, low-temperature Gibbs distributions are concentrated on configurations which are basically a single configuration plus a small fraction of small "fluctuations", also called "defects". The boundaries of these "fluctuations", define the *contours*. The contours are simple objects subjected to geometrical constraints, and the description of their distribution should be a tractable endeavor if the system is at "low temperature", i.e. when the "defects" are few and far-between. The method to be discussed consists, precisely, in exploiting this fact through the replacement of the original distributions by distributions of contours.

This technique was pioneered by Peierls in his study of the Ising model, later formalized more precisely by Griffiths and Dobrushin [21, 11]. The original argument benefited from the particular symmetries of the Ising model. The adaptation of the method to the treatment of non-symmetric models is not trivial, and was developed by Pirogov and Sinai [35, 36, 38]. Later, a particularly enlightening alternative version of the argument was put forward by Zahradník [44]. The most concise and simple exposition of this version is due to Borgs and Imbrie [3]. The course will be centered around a bare-bone presentation of this last work. Actually, there exists by now a new formulation of the theory due to Zahradník [46]. But this formulation is too new to me.

I prefer to stay with the by now classical approach which has been so prolific in the last decade.

The lectures have two well differentiated parts: Chapters 2 and 3 are preparatory chapters. They are oriented towards the unprepared reader, with little or no knowledge of statistical mechanics and Gibbs distributions, and who perhaps never heard of the Peierls argument and cluster expansions. They are written in some sense as overviews, hence they are less complete from the mathematical point of view. The actual Pirogov-Sinai theory is discussed in the last two chapters, to which the more informed reader can jump right ahead.

In Chapter 2 I introduce and discuss a family of concepts that are motivated in physics —energy, entropy, free-energy— whose knowledge is *not* assumed. The presentation is rather informal. The objective is to help the ignoramus to acquire a minimal intuitive grasp of the main concepts, so as to be able to read through the mathematics that follows. Some care is taken to present the general type of problems to which the theory applies, and to illustrate the limitations and scope of the theory.

The course continues, in Chapter 3, with the presentation of the two keystones on which the theory is built: the Peierls argument and the technology of cluster expansions. The first one is discussed in some detail, in its simplest version. The theory of cluster expansions is also presented in a rather complete manner *except* that I omit the proof of the main theorem. This proof has been extensively discussed in a number of references. I hope that, nevertheless, the audience get a reasonable idea of what can be accomplished with cluster expansions, and when and how they can be applied.

In Chapter 4, the full-fledged Pirogov-Sinai theory is introduced. A number of examples, that exhibits the different complications arising in the general case, is presented to motivate the more abstract setting of the theory. The main result of the theory is a criterion to determine when a configuration minimizing the energy function (ground-state configuration) gives rise to a low-temperature Gibbs distribution. A simplified proof will be discussed in Chapter 5, following the Zahradník-Borgs-Imbrie approach. The goal is to convey the logic of the argument and a basic understanding of its key elements. Hopefully, this will allow the student to access the various generalizations

available, and even try some of his/her own.

Unfortunately, I have not included any discussion of the many successful applications of the theory in different directions: Stability of phase diagrams, completeness of phase diagrams, finite-volume effects, applications to surface phenomena, extension to quantum mechanical systems. I leave this for a future opportunity with less scarce (writing) time.

Throughout the notes I have tried to be fair with the bibliography. As mentioned in the foreword, pedagogical reviews are not common in this topic. The review by Slawny [40] is lucid and clear, even when not totally self-contained, but it relies on the "old" approach of the original papers of Pirogov and Sinai. The latter, in Chapter 2 of his book [38], discusses this approach in a terse but mathematically careful manner. While the original paper of Zahradník [44], where he introduced the approach advocated here, is not particularly readable, his notes for the Cours de Troisième Cycle de la Physique en Suisse Romande [45] are full of useful comments and details. Unfortunately, they are not easily available. Recently, Kotecký has written two excellent reviews [27, 28] unfortunately also with limited distribution. Given that, the compact paper by Borgs and Imbrie [3] continues to be the clearest reference the average reader can access.

Chapter 2

Gibbs measures and other notions from statistical mechanics

2.1 The basic setup

2.1.1 Configurations

The starting structure is a space of the form

$$\Omega \;:=\; \Omega_0^{\mathbb{Z}^d} \tag{2.1}$$

where Ω_0 is a *finite* set (eg. $\Omega_0 = \{-1, 1\}$) and $\mathbb{Z}^d$ is just the set of integer-valued d-tuples. The set Ω_0 represents possible configurations of a certain "object" which, according to the applications, is called spin, molecule, pixel, etc. Let us, for concreteness, use the name "spin", in which case Ω_0 corresponds to the *single-spin space*. The space Ω —the *configuration space*— is thought as an infinite collection of copies of the space Ω_0, one at each point of $\mathbb{Z}^d$. This set $\mathbb{Z}^d$ is termed the *square lattice in d dimensions*, and its elements, for which we shall use letters towards the end of the alphabet (like x, y, u), are called *sites*. Other lattices are also used (triangular, honeycomb, etc), and most of what we shall discuss can be correspondingly adapted. For future purposes, we shall need a notion of distance in $\mathbb{Z}^d$. In these notes we shall take "dist(x, y)" to mean the Euclidean distance

restricted to $\mathbb{Z}^d$, and we adopt the associated notions of distance between sets and diameters of sets.

We shall use lowercase Greek letters like ω, σ or η for the elements of Ω, called *configurations*. A configuration $\omega \in \Omega$ is therefore a collection of values $\{\omega_x\}_{x\in\mathbb{Z}^d}$, each $\omega_x \in \Omega_0$, which specify the value of the spin at each site x of the lattice $\mathbb{Z}^d$. The fast way to say that our underlying space is of the form (2.1) is to say that we consider a *spin lattice system in d dimensions*.

Finite-volume configuration spaces

$$\Omega_\Lambda \;:=\; \Omega_0^\Lambda \;, \tag{2.2}$$

will also be used. Here $\Lambda \subset \mathbb{Z}^d$ is a finite set, often an hypercube. By the way, we shall adopt the traditional notion "$|\cdot|$" for the cardinality of sets. Hence, $|\Lambda| < \infty$. Finite-volume configurations will have the volume in question as a subscript: $\omega_\Lambda \in \Omega_\Lambda$. Moreover, for a configuration $\omega \in \Omega$ we shall also denote ω_Λ its "projection" (or "restriction") to Λ: $\omega_\Lambda = \{\omega_x\}_{x\in\Lambda}$. This is a widely adopted abuse of notation.

In our examples below, often $\Omega_0 = \{-1, 1\}$. The corresponding configuration space is referred to as the space of *Ising spins*. The value "+1" is interpreted as "spin up" and "−1" as "spin down". Alternatively, one can write $\Omega_0 = \{0, 1\}$, in which case Ω is interpreted as the space of *lattice-gas* configurations ("1"=presence and "0"=absence of a gas particle at the site). This is, in principle, just a trivial change of symbols. The change may be not so trivial, however, if used in conjunction with long-range interactions [26, Sections I.4 and III.4], [41]. Other examples to have in mind are *spin-1 systems* —$\Omega_0 = \{-1, 0, 1\}$— and *q-Potts spins* —$\Omega_0 = \{1, 2, \ldots, q\}$.

There is a natural notion of translation in configuration space. Namely, the *translate* of a configuration ω by a d-tuple $a \in \mathbb{Z}^d$ is the configuration

$$\tau_a\omega \;:=\; \{\omega_{x-a}\}_{x\in\mathbb{Z}^d} \tag{2.3}$$

obtained by shifting in block all spins to sites displaced by a. Here $x - a$ indicates the coordinate-wise difference. A configuration that is invariant under translations is called *translation-invariant*. These are very boring configurations: they are constant, i.e. the spins take

the same value at all sites. A little less simple are the *periodic* configurations. These are configurations invariant only under multiples of certain basic translations. More precisely, ω is periodic if there exists a d-tuple $b \in \mathbb{Z}^d$ such that ω is invariant under translations with a of the form

$$a \;=\; (m_1 b_1, m_2 b_2, \ldots, m_d b_d)\;, \tag{2.4}$$

for $m_1, \ldots, m_d \in \mathbb{Z}$. Such ω is obtained by periodic repetition of a configuration in the "fundamental cell"

$$F_b \;=\; \{[0, b_1 - 1] \times \cdots \times [0, b_d - 1]\} \cap \mathbb{Z}^d\;. \tag{2.5}$$

2.1.2 Local events and observables

A *local event* A is a family of configurations determined only through conditions involving the spins at a *finite* number of sites. The archetypical examples are the cylindrical events, formed by all configurations having fixed spin values at a certain finite set Λ. More precisely, given a configuration $\sigma_\Lambda \in \Omega_\Lambda$, the *cylinder with base* σ_Λ is the set $C(\sigma_\Lambda) \subset \Omega$,

$$C(\sigma_\Lambda) \;:=\; \left\{\omega \in \Omega : \omega_x = \sigma_x \;\forall x \in \Lambda\right\}\;, \tag{2.6}$$

formed by all configurations agreeing with σ on Λ. All local events are obtained by (finitely-many) unions and intersections of cylinders.

The notion of local event can be (apparently) generalized to that of *local observables*. A function $f\colon \Omega \to \mathbb{R}$ (or $\mathbb{C}$) is a local observable if it only depends on the spin values at a finite number of sites. More precisely, there must exist a finite $Y \subset \mathbb{Z}^d$ such that $f(\omega) = f(\sigma)$ whenever $\omega_Y = \sigma_Y$. I will often emphasize this property with the notational abuse $f(\omega) = f(\omega_Y)$. In particular each local event E defines a local observable via its characteristic (or indicator) function:

$$\mathbb{1}[E](\omega) \;:=\; \begin{cases} 1 & \text{if } \omega \in E \\ 0 & \text{otherwise}\;. \end{cases} \tag{2.7}$$

However, each local observable can be written as a linear combination of characteristic functions of local events. Hence the descriptions furnished by events and observables are equivalent.

In physical terms, the definition of events and observables corresponds to the specification of experiments or measurements (feasible or conceivable) for which the theory should account. The local events and observables are interpreted as *microscopic* experiments. One may wonder what types of observables correspond to *global* experiments, that is to measurements sensitive only to properties involving an infinite ("macroscopic") number of sites. Translation invariance may play a role here. It is natural to define the *translate of a function* $f: \Omega \to \mathbb{R}$ (or $\mathbb{C}$) by $a \in \mathbb{Z}^d$ as the function $\tau_a f$:

$$\tau_a f(\omega) \;:=\; f(\tau_{-a}\omega) \;, \tag{2.8}$$

defined by shifting. A translation-invariant function is a function that coincides with all its translates. The function is periodic if it is invariant only upon translations of the form (2.4). Functions that are periodic (and satisfy an added technical requirement of measurability) can be considered as global observables. Indeed, a periodic function takes the same value for configurations differing only at finitely-many spins (modulo sets of measure zero, see [18, Proposition 14.9]). This fact can be interpreted as insensitivity to microscopic changes.

There is an elegant measure-theoretical construction given mathematical consistency to all these definitions: The local events give rise to a σ-algebra via countably-many unions and intersections. Events (not necessarily local) are the elements of this algebra, and (not necessarily local) observables are functions measurable with respect to it. These considerations are largely beyond what is needed for these lectures, hence will not be pursued. Interested people can consult the treatise by Georgii [18], or Section 2.1 of [42] for a more concise exposition.

2.2 Finite-volume statistical mechanics

2.2.1 Interactions and Boltzmann-Gibbs weights

One considers a finite "window", that is a finite region $\Lambda \subset \mathbb{Z}^d$ outside which the spins are supposed to be frozen in some configuration $\eta_{\mathbb{Z}^d \setminus \Lambda}$. The statistical mechanical description of such a finite-volume system

amounts to prescribing probability weights for each configuration ω_Λ of the spins inside Λ. The correct form of these weights, from the point of view of physical applications, was determined by the Austrian Ludwig Boltzmann and the American Josiah Willard Gibbs towards the end of last century. The formalization of their prescription requires the introduction of an *interaction*. This is a family of functions

$$\Phi \;=\; \{\Phi_B\}_{B \subset \mathbb{Z}^d \text{ finite}} \tag{2.9}$$

where each

$$\Phi_B \colon \Omega \to \mathbb{R} \;, \tag{2.10}$$

depends only on spins in B:

$$\Phi_B(\omega) \;=\; \Phi_B(\sigma) \quad \text{if} \quad \omega_B = \sigma_B \;. \tag{2.11}$$

Once the interaction is given, one constructs the *Hamiltonian on Λ with external condition η*. This is the *function* $H_\Lambda^\eta \colon \Omega_\Lambda \to \mathbb{R}$ (or $\mathbb{C}$) defined by:

$$H_\Lambda^\eta(\omega_\Lambda) \;:=\; \sum_{B : B \cap \Lambda \neq \emptyset} \Phi_B(\omega_\Lambda \eta) \;, \tag{2.12}$$

where we have denoted by $\omega_\Lambda \eta$ the configuration coinciding with ω at the sites in Λ and with η at the sites outside Λ. The key feature of the Hamiltonians (2.12) is that they are a sum of *local* contributions Φ_B, each of which becomes independent of the region Λ and the external condition η once the former is large enough. To make this fact more explicit, let us write H_Λ^η in the form

$$H_\Lambda^\eta(\omega_\Lambda) \;:=\; \sum_{B : B \subset \Lambda} \Phi_B(\omega_\Lambda) \;+\; \sum_{\substack{B : B \cap \Lambda \neq \emptyset \\ B : B \cap (\mathbb{Z}^d \setminus \Lambda) \neq \emptyset}} \Phi_B(\omega_\Lambda \eta) \;, \tag{2.13}$$

The boundary of the region Λ is felt only by the last term on the right-hand side. The *value* of the Hamiltonian H_Λ^η on a particular configuration ω_Λ is usually called the *energy* of ω_Λ (for the external condition η outside Λ).

The Boltzmann-Gibbs probability weight for each $\omega_\Lambda \in \Omega_\Lambda$ is then

$$\rho_\Lambda^{(\beta)\,\eta}(\omega_\Lambda) \;:=\; \frac{\exp[-\beta H_\Lambda^\eta(\omega_\Lambda)]}{Z_\Lambda^{(\beta)\,\eta}} \;. \tag{2.14}$$

The factor $Z_\Lambda^{(\beta)\,\eta}$ is the obvious normalization factor

$$Z_\Lambda^{(\beta)\,\eta} := \sum_{\omega_\Lambda \in \Omega_\Lambda} \exp[-\beta H_\Lambda^\eta(\omega_\Lambda)] \,, \tag{2.15}$$

and the coefficient β is interpreted as the *inverse temperature*. (In physical units $\beta = 1/(kT)$, where T is the temperature and k the Boltzmann constant.) It is introduced for physical reasons, and it is a fact that properties of the systems change a lot when β changes from 0 (high temperature) to ∞ (zero temperature). In these lectures we shall be working in the neighborhood of $\beta = \infty$. Rigorously speaking, then, we shall be studying *families* of probabilistic models parameterized by a (in our case, very large) parameter β, all of them having the same interaction.

The interaction is what defines the model. In principle, it is the only input needed from physics; the rest of the game can be played strictly as an application of probability theory. This is, of course, an exaggeration. Physical intuition keeps providing crucial help, and, indeed, most methods used in the field were developed on the basis of physical ideas, even when they can a posteriori be restated in abstract probabilistic terms. Nevertheless, from the conceptual point of view, statistical mechanics can be considered a chapter of probability theory. It is a very special chapter though, because it has been a continuous source of concepts, methods and problems that have enriched the whole field of probability.

Most interactions studied in practice are very simple: The functions Φ_B are zero except for sets B involving a not very large number of sites, often just one and two. The sets B for which Φ_B is not zero are called *bonds*. One of the most famous examples is the Ising interaction (due, in fact, to Ising's advisor, Lenz). One considers Ising spins ($\sigma = +1$ or -1), and takes:

$$\Phi_B(\sigma) \;=\; \begin{cases} -J\,\sigma_x\sigma_y & \text{if } B = \{x, y\} \text{ nearest-neighbors} \\ -h\,\sigma_x & \text{if } B = \{x\} \\ 0 & \text{otherwise} \end{cases} \tag{2.16}$$

This model has, therefore, only two-site bonds, involving nearest-neighbor sites (that is, sites x, y such that $\mathrm{dist}(x, y) = 1$) , and single-site bonds. The "$-$" in front of J and h is an established convention.

A convenient, less rigorous but also less pedantic, way to give the information contained in (2.16) is to write

$$H \;=\; -J \sum_{\langle x,y \rangle} \sigma_x \sigma_y - h \sum_x \sigma_x \;. \tag{2.17}$$

This is roughly a remainder of the first summand ("bulk contribution") in (2.13). The notation $\langle x, y \rangle$ indicates $\mathrm{dist}(x, y) = 1$. Expression (2.17) in itself has no meaning —it is often called a *formal Hamiltonian*— but it is clear how to transform it into meaningful finite-volume Hamiltonians H^η_Λ: Constrain the sums to bonds intersecting Λ and assign the value η_y to sites y outside Λ. Other examples will be presented below.

All the interactions considered in these lectures will be of finite range. This means that there exists some $r \geq 0$ such that $\Phi_B = 0$ whenever $\mathrm{diam}\, B > r$. Moreover, our interactions will be *translation-invariant*, i.e.

$$\tau_a \Phi_B \;=\; \Phi_{B+a} \tag{2.18}$$

—where $B + a = \{x + a : x \in B\}$— or, at least, *periodic*, that is satisfying (2.18) form a subgroup of translations of the form (2.4).

Of course, the weights (2.14) yield a prescription to compute averages of observables: If f is a local observable,

$$\langle f \rangle^{(\beta)\,\eta}_\Lambda \;:=\; \sum_{\omega_\Lambda} f(\omega_\Lambda \eta)\, \rho^{(\beta)\,\eta}_\Lambda(\omega_\Lambda) \;. \tag{2.19}$$

For the corresponding probability of an event A, we shall use the notation

$$\begin{aligned} \mathrm{Prob}^{(\beta)\,\eta}_\Lambda(A) \;&:=\; \langle \mathbb{1}[A] \rangle^{(\beta)\eta}_\Lambda \\ &=\; \sum_{\omega_\Lambda :\, \omega_\Lambda \eta \in A} \rho^{(\beta)\,\eta}_\Lambda(\omega_\Lambda) \;. \end{aligned} \tag{2.20}$$

[We define $\sum_\emptyset = 0$.]

2.2.2 Zero temperature. Energy

The weights (2.14) take a particularly simple form in the limit $\beta \to \infty$. Such limit is interpreted as the statistical mechanical description of a

finite window at *zero temperature*. In this limit, the Boltzmann-Gibbs weights (2.14) become uniformly distributed over the configurations attaining the minimum energy value:

$$E^{\eta}_{0\,\Lambda} \;:=\; \min_{\sigma_\Lambda} H^{\eta}_{\Lambda}(\sigma_\Lambda)\;. \tag{2.21}$$

Resorting to standard physics terminology, let us call this minimum the *ground-state energy for the region* Λ *with external condition* η. More precisely, if we denote

$$G^{\eta}_{\Lambda} \;:=\; \left\{\omega_\Lambda : H^{\eta}_{\Lambda}(\omega_\Lambda) = E^{\eta}_{0\,\Lambda}\right\}\,, \tag{2.22}$$

then for $\beta \to \infty$ the weights (2.14) become

$$\rho^{(\infty)\,\eta}_{\Lambda}(\omega_\Lambda) \;=\; \begin{cases} 1/|G^{\eta}_{\Lambda}| & \text{if } \omega_\Lambda \in G^{\eta}_{\Lambda} \\ 0 & \text{otherwise}\,. \end{cases} \tag{2.23}$$

The configurations in G^{η}_{Λ} are the *(local) ground-state configurations for the region* Λ *with external condition* η. The determination of these ground-state configurations for general external condition may not be an easy matter. Let us consider the Ising model (2.17) with $h = 0$ and $J > 0$. We see that to minimize (2.17) one should align, as much as possible, neighbor spins. This is why models with $J > 0$ are termed *ferromagnetic*. If η is chosen as the configuration constantly equal to $+1$, then there is a unique ground state for every Λ, namely the configuration equal to $+1$ inside Λ too. However, for a square Λ ($d = 2$) and η equal to "+" on the top and left sides and "−" on the right and bottom, G^{η}_{Λ} has quite a few different configurations, namely all possible staircases between the upper right and lower left corners. The turning on of a magnetic field h renders the system insensitive to the external condition, if the region has a volume sufficiently larger than the area of the boundary. In such a case, there is a unique ground state regardless of the external condition, namely the configuration with all the spins equal to sgn$h$ ("parallel" to h).

2.2.3 Non-zero temperature. Entropy and free energy

As we have just seen, the zero-temperature distributions are concentrated on a particular set of configurations, in the sense of giving

zero probability to the event formed by all the other configurations. This set of configurations —the ground-state configurations— is determined purely on the basis of energy. When temperature is turned on, energy is no longer enough to determine the set of configurations where the Boltzmann-Gibbs weights are concentrated. One must consider another factor known as "entropy". To see this, let us consider the weight given by (2.14) to the event formed by all configurations with fixed energy:

$$\begin{aligned} \rho_\Lambda^{(\beta)\,\eta}(E) &:= \mathrm{Prob}_\Lambda^{(\beta)\,\eta}(\{\omega_\Lambda : H_\Lambda^\eta(\omega_\Lambda) = E\}) \\ &= \sum_{\omega_\Lambda : H_\Lambda^\eta(\omega_\Lambda) = E} \rho_\Lambda^{(\beta)\,\eta}(\omega_\Lambda) \; . \end{aligned} \tag{2.24}$$

But all the configurations in question have the same weight (2.14); hence

$$\rho_\Lambda^{(\beta)\,\eta}(E) \;=\; \frac{N_\Lambda^\eta(E)\exp[-\beta E]}{Z_\Lambda^{(\beta)\,\eta}} \; , \tag{2.25}$$

where we have denoted

$$N_\Lambda^\eta(E) \;:=\; \Big|\{\omega_\Lambda : H_\Lambda^\eta(\omega_\Lambda) = E\}\Big| \; . \tag{2.26}$$

It is natural now to denote

$$N_\Lambda^\eta(E) \;=:\; \exp[S_\Lambda^\eta(E)] \; , \tag{2.27}$$

so (2.25) becomes

$$\begin{aligned} \rho_\Lambda^{(\beta)\,\eta}(E) &= \frac{\exp\Big\{-[\beta E - S_\Lambda^\eta(E)]\Big\}}{Z_\Lambda^{(\beta)\,\eta}} \\ &=: \frac{\exp\Big\{-\beta\, F_\Lambda^{(\beta)\,\eta}(E)\Big\}}{Z_\Lambda^{(\beta)\,\eta}} \; . \end{aligned} \tag{2.28}$$

The quantity S_Λ^η is known as *entropy*, while

$$F_\Lambda^{(\beta)\,\eta}(E) \;=\; E - \frac{1}{\beta} S_\Lambda^\eta \tag{2.29}$$

is called *free energy*. Let us point out that there are different notions of entropy and free energy, tailored to different statistical mechanical

approaches. The entropy just introduced corresponds to the *Gibbs entropy for the microcanonical ensemble*. Likewise, the free energy $F_\Lambda^{(\beta)\,\eta}(E)$ defined in (2.29) corresponds to the Gibbs microcanonical free energy (for the given Λ, η and β). The lesson of (2.28)–(2.29) is that, for nonzero temperatures, all energies —not only the lowest ones— have nonzero probability. This probability, however, depends on a balance between the value of the energy and its associated entropy.

An increase in energy over ground-state values produces two competing changes in the probability weights: On the one hand, the damping due to the factor $e^{-\beta E}$ —the "energy cost"— and, on the other hand, the growth in the number $N(E)$ —the "entropy gain". Probability distributions at nonzero temperatures depend on both factors, that is, not just on energy but on *free* energy.

For *fixed* finite volumes, expression (2.29) shows that the influence of entropy becomes negligible at low temperatures ($\beta \to \infty$). But, as we shall see below, in the infinite-volume limit the situation is more subtle. Nevertheless, the techniques to be discussed in these notes will work only when the low-temperature dominance of the energy survives the infinite-volume limit.

More generally, one often deals with sums of the probability weights of configurations of a certain family. If, for conceptual or technical convenience, one writes such a sum as $\exp[-\beta\,\text{something}]$, this "something" is usually called free energy (with the right qualifiers, if one wants to be precise). This nomenclature can be interpreted as a reminder that both energy and entropy effects are being taken into account. In particular, the finite-volume partition function gives rise to a finite-volume free energy

$$\exp[-\beta\, F_\Lambda^{(\beta)\,\eta}] \;:=\; Z_\Lambda^{(\beta)\,\eta}\,. \tag{2.30}$$

The similarity in name and notation between the free energies (2.29) and (2.30) (note that only the "E" in the argument gives the clue) can be a source of confusion.

2.3 Infinite-volume statistical mechanics

2.3.1 Gibbs distributions

Infinite-volume statistical mechanics is obtained by letting $\Lambda \to \mathbb{Z}^d$ (in principle, in the sense of nets) in the theory described in the previous section. Through imprecise, this statement conveys an essential ingredient of the standard mathematical approach to infinite-volume phenomena: Rather than plunging into some abstract infinite-volume formalism, one humbly studies how things change as volume grows.

Some features, like the ones studied in these lectures, stabilize —rather quickly— as the volume increases. They are rigorously established by proving bounds that hold uniformly in the volume. Other features —like lack of analyticity in the presence of a phase transition— truly show up only at infinite volume. But, nevertheless, many rigorous studies of the critical exponents associated to this non-analyticity are based on correlation inequalities determined already for finite volumes [15]. On the other hand, there exists a genuinely infinite-volume method to study critical behavior, based on the so-called *renormalization transformations* (see, eg. [16] for a pedagogical introduction). This method, while conceptually very useful, is hard to formalize rigorously (for a list of rigorous works in this direction, see, for example, Section 6.1.1 of [42]).

In making precise the opening statement, we notice that we are left with very little choice if we want to make a sensible infinite-volume extension of the theory presented in Section 2.2. There is no hope of extending the notion of Hamiltonian to infinite volumes, as this would lead to infinite sums that will be almost surely divergent. As a consequence, the infinite-volume limits of the Boltzmann-Gibbs weights (2.14) are also useless: in most cases they are zero for all configurations ω. The only objects with meaningful infinite-volume limits are the expectations (2.19) or the measures (2.20).

Definition 2.1 *A probability distribution $\langle \cdot \rangle$ on Ω is called a Gibbs distribution for an interaction Φ and an inverse temperature β if there exists a sequence of volumes $\Lambda_n \underset{n}{\longrightarrow} \mathbb{Z}^d$ and a sequence of external*

conditions η_n such that

$$\lim_n \langle f \rangle_{\Lambda_n}^{(\beta)\,\eta_n} \;=\; \langle f \rangle \tag{2.31}$$

for each local observable f. When we need to be precise, we shall denote the limit distribution $\langle\,\cdot\,\rangle$ in the form $\langle\,\cdot\,\rangle^{(\beta)\,\{\eta_n\}}$. More generally a convex combination of distributions of the form $\langle\,\cdot\,\rangle^{(\beta)\,\{\eta_n\}}$ is also called Gibbsian.

In a more formal language, Gibbs measures are those measures obtained as weak limits (weak* limits, for functional analysts) of Boltzmann-Gibbs distributions (considered as measures on the whole of Ω), or convex combinations thereoff.

The objective of statistical mechanics is to determine all the Gibbs distributions for a given interaction at each given temperature. This is usually too big a job, so one settles for more limited objectives, like determining all translation-invariant (or periodic) Gibbs distributions, or at least checking whether there is a unique Gibbs distribution or more than one.

In these lectures we discuss a method to determine periodic Gibbs distributions at low temperature, using, as starting point, the zero-temperature Gibbs distributions briefly introduced in the following.

2.3.2 Zero temperature: Ground-state configurations and energy density

Rigid and non-rigid ground-state configurations

Already at zero temperature, the distributions defined by the infinite-volume limit of (2.19)–(2.20) can give rise to a rich structure of probability measures. Let us start with the simplest examples. Consider, for instance, the ferromagnetic (i.e. $J > 0$) Ising model (2.16) with $h = 0$, and take the "+" external condition, that is,

$$\eta_x \;=\; +1 \qquad x \in \mathbb{Z}^d\,. \tag{2.32}$$

Then, it is easy to see that

$$G_\Lambda^\eta \;=\; \{\eta\} \tag{2.33}$$

for all finite $\Lambda \subset \mathbb{Z}^d$. In words, there is a *unique* way to complete η inside a finite region Λ so to minimize the energy, namely taking η in the interior too. Given this lack of flexibility, an η satisfying (2.33) for Λ sufficiently large is called a *rigid ground-state configuration*. For such a configuration one easily concludes that

$$\langle f \rangle_\Lambda^{(\infty)\,\eta} \underset{\Lambda \to \mathbb{Z}^d}{\longrightarrow} f(\eta) \; ; \tag{2.34}$$

that is,

$$\langle \quad \rangle_\Lambda^{(\infty)\,\eta} \underset{\Lambda \to \mathbb{Z}^d}{\longrightarrow} \delta_\eta \tag{2.35}$$

as measures. Here δ_η is the probability distribution concentrated on η —$\delta_\eta(\eta) = 1$, $\delta_\eta(\omega) = 0$ $\forall \omega \neq \eta$. This limit distribution is, in fact, deterministic; it does not allow fluctuations. We conclude that each rigid ground-state configuration defines a *deterministic* zero-temperature Gibbs distribution.

There are plenty of examples of rigid configurations. For the ferromagnetic Ising model with $h = 0$, one has: the "$-$" configuration ($\eta_x = -1$ for all $x \in \mathbb{Z}^d$), the "flat-interface" configurations

$$\omega_x \; = \; \begin{cases} +1 & \text{for } x_1 \geq 0 \\ -1 & \text{for } x_1 < 0 \end{cases} \tag{2.36}$$

(x_1 is the first component of $x \in \mathbb{Z}^d$), which is rigid for dimensions $d \geq 2$, and many more examples appearing at higher dimensions (see [14] for a catalogue).

In fact, these rigid configurations are what most people have in mind when thinking of zero-temperature statistical mechanics. They are often called simply *ground states*. One must keep in mind, however, that, in general, not everything is deterministic at zero temperature. Consider, for instance, the antiferromagnetic Ising model —that is an interaction (2.16) [or (2.17)] with $J < 0$— with a positive magnetic field $h = 2d\ |J|$:

$$H \; = \; -J \sum_{\langle x,y \rangle} \sigma_x \sigma_y - 2d\ |J| \sum_x \sigma_x \; . \tag{2.37}$$

After some juggling you can convince yourself that this model has *no* rigid ground-state configuration. Indeed, Dobrushin, Kolafa and

Shlosman [13] showed that this model has a unique zero-temperature Gibbs distribution which is far from deterministic, being, in fact, distributed uniformly on all configurations with no two nearest-neighbor spins equal to -1.

Another example is provided by the model having $\Omega_0 = \{-1, 0, 1\}$ and formal Hamiltonian

$$H \;=\; \sum_{\langle x,y\rangle} (\omega_x^2 - \omega_y^2)^2 \;. \tag{2.38}$$

This model has a unique rigid ground-state configuration, namely $\omega_x = 0$ for all $x \in \mathbb{Z}^d$. On the other hand (see Corollary B.14 in [42]), there is a non-deterministic translation-invariant zero-temperature Gibbs distribution supported (equally) on all configurations with no zero at any site.

For a more detailed account of the zero-temperature zoo, the reader is referred, for instance, to Section B.2 of [42]. The presence of several rigid ground-state configurations, and/or of some zero-temperature distribution supported on non-rigid configurations, is known as "degeneracy".

m-potentials

In these lectures we shall only be able to deal with deterministic zero-temperature distributions. Usually, they are determined by playing with the different terms of the interaction and seeing how to minimize (most of) them. The easiest possible case is when there are configurations that actually minimize *all* the terms of the interaction. Following Holsztynski and Slawny [25], we shall call *m-potentials* interactions with this property:

Definition 2.2 *An interaction $\Phi = \{\Phi_B\}$ is an m-potential if there exists a configuration ω simultaneously minimizing all functions Φ_B:*

$$\Phi_B(\omega) \;=\; \min_{\widetilde{\omega}\in\Omega} \Phi_B(\widetilde{\omega}) \quad \forall B \subset \mathbb{Z}^d \text{ finite} \;. \tag{2.39}$$

For m-potentials, it is enough for our purposes to study the set —to be denoted $\mathcal{M}(\Phi)$— of configurations satisfying (2.39). Indeed, in their article [25] Holsztynski and Slawny establish the following

Proposition 2.3 *If Φ is periodic and $\mathcal{M}(\Phi)$ is a non-empty finite set, then the configurations of $\mathcal{M}(\Phi)$ are all the periodic rigid ground-state configurations of the model.*

All models considered up to now are defined by m-potentials. Nevertheless, the interactions (2.37) and (2.38) do not satisfy Proposition 2.3 because in both cases $\mathcal{M}(\Phi)$ is not finite.

Energy density

In the finite-volume case, the Boltzmann-Gibbs distributions are concentrated on configurations minimizing the *total* energy. An analogous result holds in the infinite-volume case, but in relation with energy *density* (total energies are not well defined!).

Given an interaction Φ, the *energy density*, or *specific energy*, of a configuration ω is defined as the limit

$$e^{\Phi}(\omega) \;:=\; \lim_{\Lambda\to\mathbb{Z}^d} \frac{1}{|\Lambda|} \sum_{B\subset\Lambda} \Phi_B(\omega) \tag{2.40}$$

whenever the limit exists. In particular, if ω is periodic with fundamental cell F_b [see (2.5)], then this limit exists and equals

$$e^{\Phi}(\omega) \;=\; \frac{1}{|F_b|} \sum_{x\in F_b} \sum_{B\ni x} \frac{\Phi_B(\omega)}{|B|}\;. \tag{2.41}$$

If the interaction is just periodic, rather than translation-invariant, then b must be chosen as a common period for the interaction and the configuration in question, and one must replace $|B|$ by the number of periodic translates of B containing x.

It is known [14, 38], that periodic configurations must minimize the energy density (2.40) to have a chance of being in the support of zero-temperature Gibbs distributions. In particular, the rigid ground-state configurations that will concern us must have minimal energy density.

2.3.3 Nonzero temperature: Free energy density

Methods of study

At nonzero temperatures, Gibbsian distributions can no longer be deterministic. In fact, every local event has nonzero probability. These distributions are, therefore, nontrivial probability measures on an uncountable configuration space. Its rigorous study requires, in principle, the tools and concepts of full-fledged measure theory. In particular, there are some sets of configurations ("non-measurable sets") and some functions ("non-measurable functions") that can not be included at all in the formalism. This will not worry us here because we shall work only with local events and observables, all of which are measurable by construction. A more relevant observation, to understand the purposes of these lectures, is that infinite-volume Gibbs distributions are *not*, in general, the product of a density function times a product measure. [In measure-theoretical terms, Gibbs measures —for nonzero interactions— are *singular* with respect to any product measure.]

The study of Gibbs distributions, therefore, can not rely only on elementary considerations —like those used to study finite-volume distributions in Section 2.2.3. Most of the techniques developed to understand (infinite-volume) Gibbs distributions can be classified into two types: (1) study of correlations, and (2) determination of "typical configurations".

The first type of techniques refers to the study of averages of the form

$$\langle f\, g \rangle \tag{2.42}$$

—two-point correlation functions— or, more generally,

$$\langle g_1 \cdots g_n \rangle \tag{2.43}$$

—n-point correlation functions. Here the functions $f, g, g_1, \ldots, g_n$ are local observables —often very local, for instance depending only on spins at a single site.

The set of all correlations (2.43) uniquely determines the probability distribution $\langle \cdot \rangle$. Hence, in principle, the more correlations one studies, the more one knows about the measure $\langle \cdot \rangle$. Moreover, for most practical purposes the necessary information is already contained in the lowest orders of n. Indeed, on the one hand, except

for very particular values of the parameters β, h, etc —defining the so-called *critical points*— Gibbs distributions correspond to random variables with *weakly coupled fluctuations*. In such a situation, the 2-point functions (2.42) —more precisely the "truncated" correlations $\langle f\, g\rangle - \langle f\rangle\, \langle g\rangle$— provide crucial data, like whether clustering properties are present (hence the measure is extremal), or absent (which implies the existence of more than one Gibbs measure). On the other hand, at the critical points the behavior of the distributions is characterized by critical exponents whose determination usually requires only low-order —two- three- and four-point— correlation functions.

The main limitation of the approach based on correlation functions is not lack of information, but rather its reduced range of applicability. Correlations are studied via correlation inequalities that, once established, yield a plethora of information (see, for instance, [22], [19, Chapter 4] and [15, Chapters 12 and 13]). But these inequalities are valid only for very particular families of models —ferromagnetic, reflection-positive, models with the FKG property, etc.

On the other hand, the techniques of type (2), i.e. aimed to the determination of "typical" configurations, apply to more general interactions and yield a basically complete understanding of the distribution in question. In this lectures we study the mathematically "softer" version of these methods, based on the so-called cluster expansions. The use of expansions has a price: only regions where the expansions converge are accessible. In particular, usually only either very low or, as in these lectures, very high values of β can be treated. For intermediate values of β —where critical points appear— completely different arguments, based on multi-scaling or renormalization ideas, are needed. These intermediate-temperature techniques are mathematically much harder and quite beyond the scope of these notes.

Energy versus entropy

By their very definition, the natural way to understand infinite-volume Gibbs distributions and their typical configurations is by watching the behavior of the finite-volume averages (2.19) [or the measures (2.20)] when the region Λ tends to $\mathbb{Z}^d$. As an enlightening guideline, let me present a simplified *heuristic* discussion of possible scenarios, based

on energy-entropy considerations introduced in section 2.2.3.

Let us first place ourselves at zero temperature and consider *deterministic* zero-temperature Gibbs distributions (i.e. rigid ground-state configurations). If we turn on (a little) the temperature we expect Gibbs distributions which, in some sense, do not differ much from the starting δ-like distributions. The first observation is that in the limit $\Lambda \to \mathbb{Z}^d$ the probability (2.20) of a single configuration —that is, of the event formed by only one configuration— converges to zero. Therefore, the Gibbs distributions at non-zero temperature must include fluctuations, that is, they must be supported on configurations where groups of spins are misaligned with respect to the ground-state configuration. Moreover, these spin fluctuations have to be distributed throughout $\mathbb{Z}^d$: the probability of an event formed by all the configurations which outside a given finite region M are frozen in some fixed manner, also tends to zero. [Such an event is a finite union of single-configuration events, each of which has zero probability.]

For high β (low temperature), one may expect the fluctuations to appear mostly in small groups of spins in widely separated regions. Therefore, we may expect to associate to each rigid (periodic) ground-state configuration η, a low-temperature Gibbs distribution $\langle \quad \rangle^\eta$ whose typical configurations look like a "sea" of spins aligned as in η with small and far-between "islands" of misaligned spins. If this is the case, the zero-temperature (deterministic) Gibbs distribution δ_η is said to *survive* at the given temperature, or that one has a *stable* η *phase* (phase = extremal periodic Gibbs measure).

As β decreases (i.e. the temperature increases), the set of stable phases is expected to change. To understand this —again *heuristically*— it is enough to consider the effect of fluctuations localized around the origin. They have an energy cost but also increase entropy. If the sea-with-islands picture holds, similar fluctuations will be present throughout the whole of $\mathbb{Z}^d$ with a certain density. Therefore, both the finite-volume energy and the finite-volume entropy of Section 2.2.3 should typically be of the order $|\Lambda|$ times the changes due to fluctuations around zero. In other words, the analysis of fluctuations involving the origin should provide us with an estimation of energy, entropy and (some sort of) free energy *densities*. These local, energy raising, fluctuations are called *excitations*, or *elementary*

excitations, of the ground-state configuration of reference. They are often classified according to their energy, those leading to the smallest change in energy being called *lowest excitations*.

Let me illustrate the argument with a concrete example. Consider the model defined by $\Omega_0 = \{-1, 0, 1\}$ and formal Hamiltonian

$$H \;=\; \frac{1}{2} \sum_{\langle x,y \rangle} (\omega_x - \omega_y)^2 \tag{2.44}$$

(spin-1 Blume-Capel model). It has three translation-invariant rigid ground-state configurations: the all-"+1", all-"0" and all-"−1". The lowest excitations involving the origin consist, precisely, in flipping the spin at the origin itself. There is a difference, however: The all-"0" configuration has *two* elementary excitations

$$\begin{array}{l} \omega_0 = +1, \\ \omega_x = 0 \ , \ |x| = 1 \end{array} \quad \text{and} \quad \begin{array}{l} \omega_0 = -1, \\ \omega_x = 0 \ , \ |x| = 1 \ , \end{array} \tag{2.45}$$

both with energy d, while the all-"+1" and the all-"−1" configurations have only *one* excitation of such low energy:

$$\begin{array}{l} \omega_0 = 0, \\ \omega_x = +1 \ , \ |x| = 1 \end{array} \tag{2.46}$$

and

$$\begin{array}{l} \omega_0 = 0, \\ \omega_x = -1 \ , \ |x| = 1 \ , \end{array} \tag{2.47}$$

respectively (the other possibility, namely putting a "−" in a sea of "+" or vice-versa, has a higher energy cost $4d$). Therefore, the lowest-energy excitation for the all-"0" ground-state configuration has a higher entropy ($\log 2$ versus 0), and hence a lower "free energy", than that of the other two periodic rigid configurations. The precise definition of this "free energy" will be a central issue in later chapters. The argument suggests, then, that as soon as the temperature is turned on, the all-"0" configuration is favored, from the point of view of the "free energy", and hence is the only one to survive.

If one tries to force either an all-'+" or an all-'−" sea by taking the respective boundary conditions, the corresponding Boltzmann-Gibbs distributions will tend, as $\Lambda \to \mathbb{Z}^d$, to favor configurations in which

close to the boundary there is a massive flip into an all-"0"-plus-fluctuations configuration. Such configurations have an energy cost of the order of the number of spins in the boundary, but a "*free* energy" gain of the order of the volume of the region Λ. This gain takes over as the volume grows, and the limit yields the "0"-phase distribution. We shall see, in this lectures, the not-so-trivial steps needed to make a rigorous proof out of the part of the preceding argument dealing with the stability of the "0"-phase. The proof of the unstability of the putative "+"- and "−"- phases is an additional development of the theory that we shall have no time to discuss.

Of course, it can also happens that the number of stable phases grow at a certain temperature. For example, let me add one more term to the previous model (2.44):

$$H \;=\; \frac{1}{2}\sum_{\langle x,y\rangle}(\omega_x-\omega_y)^2 - g\sum_x \omega_x^2\,. \tag{2.48}$$

For $g>0$ the model has only two rigid ground-state configurations: the all-"+" and the all-"−". It turns out that for each g positive but small, there is a value $\beta^{(3)}(g)$ such that for $\beta>\beta^{(3)}(g)$ both the all-"+" and all-"−" phases remain stable, at $\beta=\beta^{(3)}(g)$ the *three* phases —all-"+", all-"0" and all-"−"— are stable, and for $\beta<\beta^{(3)}(g)$ (but still large enough) only the all-"0" phase is stable. While for $g>0$ the all-"0" ground-state has higher specific energy, the entropic advantage of its low-lying excitations does decrease the "free energy" and, as temperature grows, makes it comparable to, and later less than, that of the other possible phases.

Other possible scenarios

A much more involved scenario takes place in case of infinite degeneracy —like the one for the antiferromagnetic Ising model (2.37). In the first place, one has to deal with non-deterministic zero-temperature Gibbs distributions. In such case, it is not clear, in general, whether the notion of excitation and the picture of a sea with islands are meaningful or useful. Moreover, in cases of extreme degeneracy, one may find a very complicated pattern of cascades of low-temperature transitions to different families of Gibbs distributions

Another scenario, useful to keep in mind as a reference, is when entropy so overwhelmingly wins over energy that no deterministic zero-temperature distribution survives at nonzero temperature. The archetypical example of this phenomenon is the $1-d$ ferromagnetic Ising model in zero field. The model has exactly two periodic (in fact, translation-invariant) rigid ground-state-configurations: the all-"+" and the all-"−". Let us apply the energy-entropy argument to, say, the all-"+" configuration. The lowest excitations involving the origin are obtained by flipping any consecutive string of spins at and around the origin

$$\omega_x = \begin{cases} -1 & \text{for } a \le x \le b \\ +1 & \text{otherwise} \end{cases} \tag{2.49}$$

($a \le b$). All these excitations have an energy cost 4J, independent of the (finite) volume $|\Lambda|$ being considered. But there are $O(|\Lambda|)$ of them, hence they embody an entropy gain of order $\log|\Lambda|$. As $\Lambda \to \mathbb{Z}$, entropy beats energy and long fluctuations become overwhelmingly probable. The sea-with-island picture breaks down completely; the all-"+" zero-temperature distribution does not survive the addition of temperature. By symmetry, the same happens with the all-"−" configuration. Therefore, this heuristic argument indicates that as soon as the temperature is turned on the ordered zero-temperature distributions disappear and the system settles in a disordered Gibbs distribution. This fact can actually be proven, through the proof —by now a simple exercise for advanced undergraduate students— is not based on energy-entropy considerations, but rather using transfer-matrix techniques. I propose, as an interesting exercise, the construction of a rigorous proof of the unstability of both rigid configurations, based on the previous heuristic argument.

Another aspect that will be totally left out in these lectures is the fate of non-periodic rigid ground-state configurations, like the flat-interface configurations (2.36). The lack of periodicity renders its study considerably more complicated and extremely case-dependent. Low-temperature Gibbs distributions associated to flat interfaces have been the object of a number of rigorous studies [12, 1, 24, 34]. Its study is related to rather subtle aspects of the theory of random surfaces.

Free energy

To finish this section, let us come back to the issue of the different notions of free energy. Above I have loosely spoken of a certain "free-energy" density, associated to families of configurations obtained by adding local fluctuations to a rigid ground-state configurations. We have in mind a density that depends of the reference "sea" configuration, and whose minimization is associated to stability. In contrast, what texts in statistical mechanics usually call free-energy density is the limit

$$f(\beta,\ldots) \;=\; -\frac{1}{\beta}\lim_{\Lambda\to\mathbf{Z}^d}\frac{1}{|\Lambda|}\ln Z_\Lambda^{(\beta)\,\eta} \tag{2.50}$$

based on the free energy (2.30). Here "..." stands for all other parameters present in the interaction. As the notation indicates this limit density does *not* depend on the external condition η. For spin models, this free-energy density corresponds to what physicists call *Helmholtz free energy density*, while in the case of lattice gas models —eg. $\Omega_0 = \{0,1\}$— it is interpreted as minus the *pressure*.

The limit (2.50) constitutes, therefore, an example of a *thermodynamic potential*, that is, a quantity that makes the connection between statistical mechanics and the preexisting, very successful but essentially phenomenological, thermodynamical approach. Its (one-sided) derivatives have important physical interpretations (specific heat, magnetization, magnetic susceptibility). Nevertheless, the insensitivity to external conditions shows that this object can not be directly used to determine, via minimization, the stability of phases. It turns out that a variational principle can be set out, in a rather abstract framework, so that periodic Gibbs distributions emerge as minimizers, in the space of probability measures, of a free-energy density like (2.50) considered as a functional on the space of interactions (see, eg. [26, Chapters I, II and V], [18, Chapters 15 and 16] for a detailed exposition, and [42, Section 2.6] for a quicker account). Such an approach is too abstract for our purposes here.

Rather, we shall see below how one can rigorously define (in fact in a highly non-unique manner) a notion of free-energy density related with the sea-with-island picture, which depends on the reference "sea" configuration, and which being minimal implies stability. In this sense, it could be called *metastable* free-energy density. Moreover, the mini-

mal metastable free-energy density coincides with the physical density (2.50).

Chapter 3

The basic technology: The Peierls argument and cluster expansions

3.1 Introduction

The two basic ingredients behind the description of low-temperature Gibbs distributions via contour ensembles —or, more generally, gases of defects [17]— are (1) the Peierls argument, and (2) the technique of cluster expansions.

Peierls introduced his argument in 1936 [33] to show that the 2-d Ising model does exhibit phase coexistence at low temperature. It was a crucial contribution, as it confirmed that the comparatively simple picture proposed by Gibbs and Boltzmann was indeed sufficient to describe the existence of phase transitions. Before this, Ising had shown, in his doctoral dissertation, that the one-dimensional version of the model does *not* have a phase transition at nonzero temperatures. Ising was under the impression that the same would be true for higher dimensions, and this fueled the suspicion that perhaps a more complicated statistical theory —possibly involving multiple Hamiltonians, or different Hamiltonians for different temperature regimes— was necessary to account for phase transitions. Peierls, with his argument, showed this not to be the case.

Despite its importance, the original argument needed some math-

ematical debugging which was performed, almost 30 years later, independently by Dobrushin [11] and Griffiths [21]. Our presentation below is along the lines of the former.

The cluster expansion was introduced by Ursell in 1927, and its use is also associated with the name Mayer (1937). The form used in rigorous statistical mechanics is due to Glimm, Jaffe and Spencer [20]. Its theory has undergone the following developments:

(i) The expansion was generalized and systematized by Malyshev [30] and Seiler [37].

(ii) Cammarota [7] clarified and simplified the key combinatorial bound showing that there is an underlying *identity* (tree-graph identity), explicitly exhibiting the cancellations among different terms of the expansion. Good references for this new proof are the monographies of Brydges [6] and of Pfister [34, Section 3].

(iii) A compact and elegant approach was put forward by Kotecký and Preiss [29]. This approach often yields the best bounds for the convergence region, but it requires that the clusters be formed from only a *finite* number of possible "polymers" . Some generalizations need to dispense with this limitation [8].

(iv) A further development of the approach of Kotecký and Preiss has been recently presented by Roland Dobrushin [9]. In particular I recommend his posthumous work, *Perturbation methods of the theory of Gibbsian fields* [10], to all the probabilists in the audience.

Our presentation below is based on the approach of the references cited in (ii).

3.2 The Peierls argument

3.2.1 Setting and result

The argument applies to the ferromagnetic Ising model at zero field; that is, to the model with interaction (2.16) with $J > 0$ and $h =$

0. This model has two periodic (in fact, translation-invariant) rigid ground-state configurations: the all-"+" and the all-"−" configurations. The argument shows that both deterministic measures survive at low temperatures giving rise to two *different* Gibbs distributions. More precisely, the argument shows that there is a range of inverse temperatures $(\infty, \beta_{\rm P})$ such that for each β in this range there are two *different* Gibbs distributions, $\langle\ \ \rangle^{(\beta)+}$ and $\langle\ \ \rangle^{(\beta)-}$, which are the "sea-with-islands" version of the corresponding ground-state configurations. The "+" distribution is characterized by a strictly positive magnetization:

$$\langle \omega_0 \rangle^{(\beta)+} \ > \ 0 \ , \tag{3.1}$$

while the "−" distribution has a strictly negative magnetization:

$$\langle \omega_0 \rangle^{(\beta)-} \ < \ 0 \ . \tag{3.2}$$

The argument is not useful to estimate the critical temperature (the temperature at which both Gibbs distributions cease to be different), but it does give an essentially complete characterization of the configurations typical for each of the distributions. Both, the fact that the two ground-states configurations survive, and the mathematical implementation of the argument, rely on symmetry considerations.

3.2.2 Dissection of the argument

Let me now present the argument, clearly distinguishing its different components.

Step 0. Existence of $\langle\cdot\rangle^{(\beta)+}$ and $\langle\cdot\rangle^{(\beta)-}$

The (infinite-volume) distribution $\langle\cdot\rangle^{(\beta)+}$ is constructed by taking the limit of finite-volume distributions with "+"-boundary conditions. The argument assumes that such a limit does exist. This existence can be proven, indeed rather easily, using correlation inequalities. Let us denote by ω^A, for a finite $A \subset \mathbb{Z}^d$, the function

$$\omega^A \ := \ \prod_{x \in A} \omega_x \ . \tag{3.3}$$

Then, the 2nd Griffiths inequality [22] implies that

$$\langle \omega^A \rangle_\Lambda^{(\beta)+} \;\geq\; \langle \omega^A \rangle_\Delta^{(\beta)+} \qquad \text{whenever } \Lambda \subset \Delta \,. \tag{3.4}$$

This implies that the limit

$$\lim_{\Lambda \to \mathbb{Z}^d} \langle \omega^A \rangle_\Lambda^{(\beta)+} \;=:\; \langle \omega^A \rangle^{(\beta)+} \tag{3.5}$$

exists, by monotonicity, for all finite $A \subset \mathbb{Z}^d$. As every local function is a (finite) linear combination of functions ω^A, this proves the existence of $\langle \cdot \rangle^{(\beta)+}$ as a probability distribution. The existence of $\langle \cdot \rangle^{(\beta)-}$ is proven either analogously or just saying "by symmetry". The construction shows that both measures $\langle \cdot \rangle^{(\beta)+}$ and $\langle \cdot \rangle^{(\beta)-}$ are translation invariant.

Step 1. Definition of contour ensembles

If the sea-with-islands picture is expected, it is mathematically convenient to change variables from spins to something associated with the islands. Such new objects should have a simpler distribution, as the islands occur so seldomly at low temperature. These new variables are the *contours* which we now introduce.

Our objective is to analyze limits of distributions $\langle \omega_0 \rangle_\Lambda^{(\beta)+}$. Consider, therefore, the space Ω_Λ^+ of configurations in $\mathbb{Z}^d$ which are equal to "+1" outside Λ. For each $\omega \in \Omega_\Lambda^+$, we have that

$$H_\Lambda^+(\omega) \;=\; 2J N_{+-}(\omega) + E_\Lambda^+ \,, \tag{3.6}$$

where N_{+-} is the number of misaligned nearest neighbors:

$$N_{+-} \;:=\; |\{x, y \,:\, |x-y| = 1 \text{ and } \omega_x \neq \omega_y\}| \,, \tag{3.7}$$

and E_Λ^+ is the energy in Λ of the all-"+" configuration. This term is independent of ω, and in the present symmetric case will turn out to be unimportant. In fact, in many elementary presentations this term is turned off by redefining the Hamiltonian so that both the all-"+" and the all-"−" configurations get zero energy [that is, taking the equivalent interaction $H \;=\; -J \sum_{\langle x,y \rangle} (\sigma_x \sigma_y - 1)$]. This trick ceases to be useful if one wants to study full phase diagrams (for instance if one

considers also a magnetic field h). For future reference is opportune to remark that E^+_Λ is of the form

$$E^+_\Lambda = 2dJ|\Lambda| + \mathrm{O}(|\partial\Lambda|) \tag{3.8}$$

$$= e(+)\,|\Lambda| + \mathrm{O}(|\partial\Lambda|)\ . \tag{3.9}$$

Here $e(+)$ is the energy density (2.40) of the all-"+" configuration, $\partial\Lambda$ stands for the *internal* boundary of Λ (set of sites in Λ with a nearest-neighbor outside it), and "$\mathrm{O}(|\partial\Lambda|)$" is a correction that does not grow faster than the number of sites in $\partial\Lambda$ (for a parallelepiped it equals $|\partial\Lambda|/2$).

Peierls realized of the convenience of doing a *geometric* accounting of the number $N_{+-}(\omega)$ by drawing, in two dimensions, a unit segment perpendicular to the line joining misaligned nearest neighbors. It is intuitively obvious that these segments join up to form closed curves (the curves must be closed because outside Λ all spins are "+"). Its connected components are called *Peierls contours*. In three dimensions one considers plaquettes perpendicular to bonds with misaligned spins, and the contours are closed surfaces. The rigorous formalization of these facts is straightforward but tedious. One needs to introduce the *dual lattice* whose sites have half-integer coordinates. Contours have vertices in this lattice. The fact that they are closed —i.e. that each dual vertex can only be the vertex of an *even* number of segments (plaquettes)— also requires a little proof. Let me omit the details; readers can consult them in [18, Section 6.2] or [10, Section 5B]. I wish, however, to comment on the notion of connectedness.

Given a configuration ω and the family of segments (plaquettes in higher dimensions) separating misaligned spins in ω, it is natural to call two segments connected if they share a vertex. This defines a notion of connection for finite families of segments. Two families of segments are disconnected if no dual site can be found that is simultaneously a vertex of at least one segment of each family. Each family of segments (plaquettes) can be divided into maximally connected components (that is, on contained in any other connected set). If we call contours to these connected components, the resulting contours can include loops. An alternative, loop-free, definition is possible in two dimensions if we introduce a convention to "chop-off" the corners where four segments meet. One must be careful here, not every rule is

possible. See for instance [10, Section 5B] for a rule that works. The advantage of doing this is that loop-free contours are closed paths, and hence is much easier to count their number. Nevertheless, for the general case, and also for Ising in higher dimensions, there is no similar algorithm. As I am aiming for generality, I will not insist on the "chopping-off" trick.

Let me use the letter γ for individual contours, and $\Gamma(\omega)$ for the family of contours corresponding to a configuration $\omega \in \Omega_\Lambda^+$. Whatever the definition of connectedness, each contour γ is a closed curve (surface, $d-1$-dimensional manifold). Each γ is closed, hence (Jordan curve theorem) it divides $\mathbb{Z}^d$ in several components: one infinite, called the exterior and denoted $\operatorname{Ext}\gamma$, and several with finite cardinality (only one if the contour does not intersect itself), collectively called the interior of γ and denoted $\operatorname{Int}\gamma$. By construction, there is a map that to each $\omega \in \Omega_\Lambda^+$ it associates a family of non-intersecting contours $\Gamma(\omega)$. A less trivial fact is that this map is invertible: Any family of pairwise disjoint closed curves in (the dual of) Λ is the family of contours of some configuration. This configuration is obtained by starting with "+" on the outside of all contours and flipping each time a contour is traversed. A little proof is needed to show that this algorithm does not lead to contradictory requirements (see [10, pp. 25–26]). Hence, if we denote $\mathcal{C}_\Lambda^+$ the set formed by families of pairwise disjoint contours, we have a bijection

$$\begin{array}{ccc} \Omega_\Lambda^+ & \longleftrightarrow & \mathcal{C}_\Lambda^+ \\ \omega & \rightleftharpoons & \Gamma(\omega)\ , \end{array} \tag{3.10}$$

such that

$$N_{+-}(\omega) \;=\; \sum_{\gamma\in\Gamma(\omega)} |\gamma|\ , \tag{3.11}$$

where $|\gamma|$ is the *length* (area) of the curve (surface) γ, i.e. the number of unit segments (plaquettes) comprising it.

Thus, the energy of a configuration $\omega \in \Omega_\Lambda^+$ takes a very simple expression in terms of contours:

$$H_\Lambda^+(\omega) \;=\; 2J \sum_{\gamma\in\Gamma(\omega)} |\gamma| + E_\Lambda^+\ , \tag{3.12}$$

We see that the Boltzmann-Gibbs weight factorizes when expressed in terms of contours:

$$\exp\left[-\beta H_\Lambda^+(\omega)\right] \;=\; \exp\left[-\beta E_\Lambda^+\right] \prod_{\gamma\in\Gamma(\omega)} W^{(\beta)}(\gamma)\;, \tag{3.13}$$

where the contour weights are:

$$W^{(\beta)}(\gamma) \;=\; \exp\left[-\beta 2J\,|\gamma|\right]\;. \tag{3.14}$$

The set $\mathcal{C}_\Lambda^+$ of contour configurations endowed with the weights

$$W^{(\beta)}(\Gamma) \;=\; \frac{\prod_{\gamma\in\Gamma(\omega)} W^{(\beta)}(\gamma)}{\Theta_\Lambda^{(\beta)\,+}}\;, \tag{3.15}$$

with

$$\Theta_\Lambda^{(\beta)\,+} \;:=\; 1+\sum_{\Gamma\in\mathcal{C}_\Lambda^+}\prod_{\gamma\in\Gamma(\omega)} W^{(\beta)}(\gamma)\;, \tag{3.16}$$

constitutes a probability space —in physicists' nomenclature an ensemble— called the (+)-*contour ensemble in* Λ. It is a much simpler ensemble than the original set of spin configurations with Boltzmann-Gibbs weights, because contours have *no interaction* except for being forbidden to intersect (*volume exclusion*). They constitute a *gas of particles with hard-core interactions*.

It is now clear how to proceed. We should relate expectations for the original Ising system with expectations for the contour ensemble, which look much simpler to evaluate.

A similar construction can be done for configurations in Ω_Λ^-, that is, with "−" external conditions. One obtains in this fashion a set $\mathcal{C}_\Lambda^-$ of families of contours. Nevertheless, it is immediate that

$$\mathcal{C}_\Lambda^- = \mathcal{C}_\Lambda^+ \tag{3.17}$$

and that the (−)-contours get the same weights (3.14). Hence, the ensemble of (−)-contours coincides with that of the (+)-contours. This exceptional fact is a consequence of the flipping symmetry of the Ising model at zero field and of the nearest-neighbor character of the interaction.

Step 2. Bound on the magnetization

The bound is obtained through a number of sub-steps.

Sub-step 2.1. Expression in terms of contour ensembles

The bijection (3.10) implies that any observable g in Λ can be written purely as a function $\widetilde{g}$ of contour configurations. Then equation (3.13) implies that

$$\langle g \rangle_\Lambda^{(\beta)+} \;=\; \sum_{\Gamma \in \mathcal{C}_\Lambda^+} \widetilde{g}(\Gamma)\, \frac{\prod_{\gamma \in \Gamma} W^{(\beta)}(\gamma)}{\Theta_\Lambda^{(\beta)+}} \;. \tag{3.18}$$

[We adopt the convention $\prod_\emptyset := 1$.] Let us use this to prove the inequality (3.1).

The starting point is the simple identity

$$\begin{aligned} \langle \omega_0 \rangle_\Lambda^{(\beta)+} \;&=\; \mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = +1) \;-\; \mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = -1) \\ &=\; 1 \;-\; 2\,\mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = -1) \;. \end{aligned} \tag{3.19}$$

Given the existence of the limit (3.5), the proposed inequality $\langle \omega_0 \rangle^{(\beta)+} > 0$ will follow if we can prove that

$$\mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = -1) \;\le\; \epsilon \tag{3.20}$$

with $\epsilon < 1/2$ *independent of* Λ.

At this point we resort to (3.18):

$$\mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = -1) \;=\; \sum_{\Gamma \in \mathcal{C}_\Lambda^+} \mathbb{1}_{\mathrm{odd}}(\Gamma)\, \frac{\prod_{\gamma \in \Gamma} W^{(\beta)}(\gamma)}{\Theta_\Lambda^{(\beta)+}} \tag{3.21}$$

where $\mathbb{1}_{\mathrm{odd}}(\Gamma)$ assigns the value one to families Γ having an odd number of contours around the origin, and 0 otherwise. This function is majorized by the function that takes value one if Γ has *some* contour around the origin, and zero otherwise. Therefore,

$$\begin{aligned} &\mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = -1) \\ &\le\; \frac{1}{\Theta_\Lambda^{(\beta)+}} \sum_{\gamma_0 \,:\, \mathrm{Int}\,\gamma_0 \ni 0} W^{(\beta)}(\gamma_0) \sum_{\substack{n \ge 1 \,;\, \{\gamma_0, \gamma_1, \ldots, \gamma_n\} \,:\, \\ \{\gamma_0, \gamma_1, \ldots, \gamma_n\} \in \mathcal{C}_\Lambda^+}} \prod_{i=1}^n W^{(\beta)}(\gamma_i) \;. \end{aligned} \tag{3.22}$$

Sub-step 2.2. The contour-removal operation

Now, a seemingly obvious but crucial observation is applied:

$$\{\gamma_0, \gamma_1, \ldots, \gamma_n\} \in \mathcal{C}_\Lambda^+ \implies \{\gamma_1, \ldots, \gamma_n\} \in \mathcal{C}_\Lambda^+ \,. \tag{3.23}$$

This is a manifestation of the fact that the map (3.10) is a bijection. Given a family of contours defining a configuration $\omega \in \Omega_\Lambda^+$, the family obtained by omitting one of them is also the family of contours of a (different) configuration in Ω_Λ^+. In fact, there is an algorithm —*contour-removal operation*— to obtain the new configuration: Take the configuration ω and flip all the spins in the interior of γ_0. This makes γ_0 disappear, but leaves intact the other contours (the configurations immediately outside and inside the contours in the interior of γ_0 change, but the contours remain present.). The existence of this simple contour-removal operation is again something particular of the symmetric nearest-neighbor Ising model. Things are not so immediate in other cases.

From (3.22) and (3.23) we obtain

$$\mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = -1) \;\leq \sum_{\gamma_0 \,:\, \mathrm{Int}\,\gamma_0 \ni 0} W^{(\beta)}(\gamma_0) \,. \tag{3.24}$$

Sub-step 2.3. Energy versus entropy

By (3.14):

$$W^{(\beta)}(\gamma_0) \;=\; W^{(\beta)}(|\gamma_0|) \;=\; \exp\left[-\beta 2J\,|\gamma_0|\right] \,, \tag{3.25}$$

hence

$$\mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = -1) \;\leq \sum_{\ell \,\geq\, \ell_{\min}} N_\ell \, e^{-\beta 2J\ell} \tag{3.26}$$

with

$$N_\ell \;=\; |\{\gamma : \mathrm{Int}\,\gamma \ni 0, |\gamma| = \ell\}| \,, \tag{3.27}$$

and $\ell_{\min} = 2d$ is the length (area) of the smallest possible contour (that is, of a cube with unit sides). The right-hand side of (3.26) has the typical energy-vs-entropy form. The "entropy", though, grows at most linearly with ℓ:

$$N_\ell \;\leq\; \ell \, c_d^\ell \tag{3.28}$$

where c_d is a geometrical constant that depends only on the dimension d. The factor ℓ is a bound on the number of positions for the intersections of the contour with the 1st axis. After fixing the position of this first segment, there are at most c_d^ℓ ways to assemble the contour for a certain number c_d. In $d = 2$, with the "chop-off" prescription, each contour is a closed path that can be assembled by choosing sequentially one of the (at most) three available directions for the next segment. Hence $c_d = 3$. In higher dimensions, or in $d = 2$ without chopping off the 4-segment intersections, we can resort to the Königsberg bridge problem to conclude that there is a number $\alpha_d < \infty$ such that for each contour there is a path that visits all the plaquettes without visiting each one more than α_d times. Hence we have at most $(2d-1)^{\alpha_d \ell} =: c_d^\ell$ ways to assemble the contour. [Note that we are neglecting the fact that contours have no dangling ends.]

We conclude that

$$\mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = -1) \;\leq\; \sum_{\ell \geq 2d} \ell\, c_d^\ell\, e^{-\beta 2J\ell} \tag{3.29}$$

which converges for

$$\beta \;>\; \frac{\log c_d}{2J} \;=:\; \beta_{\mathrm{PP}} \tag{3.30}$$

("PP" = Peierls Percolation, see below). Moreover, the series tends to zero as β tends to infinity, hence there exists a temperature β_{PM} ("PM" = Peierls Magnetization), such that

$$\sum_{\ell \geq 2d} \ell\, c_d^\ell\, e^{-\beta_{\mathrm{PM}} 2J\ell} \;=\; \frac{1}{2}\,. \tag{3.31}$$

By (3.19), (3.29) and Step 0:

$$\langle \omega_0 \rangle^{(\beta)+} \;>\; 0 \quad \text{for } \beta_{\mathrm{PM}} < \beta < \infty\,. \tag{3.32}$$

[I am using the fact that $\mathrm{Prob}_\Lambda^{(\beta)+}(\omega_0 = -1)$ is a non-increasing function of β. This follows from (3.19) and the second Griffiths inequality.]

We notice that, as c_d increases with d, both β_{PP} and β_{PMP} *increase* with d. However, the 2nd Griffiths inequality implies that the actual critical β of the model —i.e. the temperature at which the phases $\langle\,\cdot\,\rangle^+$ and $\langle\,\cdot\,\rangle^-$ cease to be different— *decreases* with d. The reason why

the Peierls bound to the critical temperature becomes poorer as the dimension grows is discussed next.

In an analogous manner one analyzes the probabilities for the "−"-phase. By symmetry the *same* $\beta_{\rm PP}$ and $\beta_{\rm PM}$ are obtained. In particular this implies that for $\beta_{\rm PM} < \beta < \infty$ the "+" and "−" phases have different magnetization, and hence are different. In fact, as we now discuss, the argument also shows that the phases remain different at least for $\beta > \beta_{\rm PP}$.

3.2.3 The sea-with-islands picture

A somehow less known aspect of the Peierls argument is that it also provides a good description of the typical configurations for the "+" and "−" phases. Indeed, looking back to (3.22)–(3.26) we see that what really has been proved is that

$$\sum_{\ell \geq 2d} \text{Prob}^{(\beta)+}(C_\ell) \;\leq\; \sum_{\ell \geq 2d} \ell \, c_d^\ell \, e^{-\beta 2J\ell} \tag{3.33}$$

with

$$C_\ell \;=\; \{\omega : \exists \gamma \in \Gamma(\omega) \text{ with } |\omega| = \ell \text{ and } \text{Int}\,\gamma \ni 0\} \;. \tag{3.34}$$

Hence the convergence of the last series, that is, the condition $\beta > \beta_{\rm PP}$ implies that the events have summable probabilities. By the first Borel-Cantelli lemma (see eg. [43, Section 2.7]), this means that

$$\text{Prob}^{(\beta)+}(\text{infinitely many } C_\ell \text{ happen}) \;=\; 0 \;. \tag{3.35}$$

In words, at such low temperatures, typical configurations of the measure $\langle \cdot \rangle^{(\beta)+}$ exhibit at most a *finite* number of contours around the origin or, by translation invariance, around any other site of the lattice. Thus, every site is always contained in some finite *external* contour and the minority spins (in this case spins "−") do not percolate. Moreover, as β grows, the lower-order terms dominate the sums (3.33), hence large (finite!) contours appear with increasingly small (but nonzero!) probability.

The fact that the Peierls argument is rather an estimation of the absence of minority percolation explains why it yields bounds for the

critical β that increase with d: The range of temperatures where minority spins percolate increases with d, because more connections are involved and hence percolation becomes easier. The widespread belief is that already for $d \geq 3$ minority percolation stops at a temperature strictly smaller than the critical one. This fact, however, has been rigorously proven only at dimensions large enough [2].

3.3 Cluster expansions

3.3.1 The importance of ratios of partition functions

The cluster expansion is the technique *par excellence* to deal with gases of hard-core objects, that is, with ensembles of objects interacting only by volume exclusion. For us, it will be a means to extract the most from contour arguments. Already the right-hand side of (3.22) gives the clue of what is needed. This formula can be written in terms of ratios of contour partition functions:

$$\mathrm{Prob}_{\Lambda}^{(\beta)+}(\omega_0 = -1) \;\leq\; \sum_{\gamma_0 \,:\, \mathrm{Int}\,\gamma_0 \ni 0} W^{(\beta)}(\gamma_0)\, \Theta_{\mathrm{Int}\,\gamma_0}^{(\beta)+}\, \frac{\Theta_{\Lambda \setminus \mathrm{Int}\,\gamma_0}^{(\beta)+}}{\Theta_{\Lambda}^{(\beta)+}} \;. \qquad (3.36)$$

More generally, the probability of all cylindrical events can be written, using the inclusion-exclusion principle, as a combination of probabilities of cylinders of the form $\{\omega : \omega_A = +_A\}$, where "$+_A$" indicates the configuration in Ω_A equal to $+1$ at each site of A (see [10, pp. 17–18] for the complete argument). These probabilities take the form

$$\begin{aligned} \mathrm{Prob}_{\Lambda}^{(\beta)+}(\omega_A = +_A) &= e^{\beta E^{+}_{\mathrm{Int}\,A}}\, \frac{Z_{\Lambda \setminus A}^{(\beta)+}}{Z_{\Lambda}^{(\beta)+}} \\ &= \frac{\Theta_{\Lambda \setminus A}^{(\beta)+}}{\Theta_{\Lambda}^{(\beta)+}} \;. \end{aligned} \qquad (3.37)$$

Thus, the contour ensembles can provide us with detailed information of the spin measure $\langle \cdot \rangle^{(\beta)+}$ if we have an efficient way to compute *ratios* of partition functions of the form $\Theta_{\Lambda}^{(\beta)+}/\Theta_{\Delta}^{(\beta)+}$. This is precisely what the cluster expansion is designed to do.

While rather general approaches to the cluster-expansion technology have been devised [29, 10], we shall present here the simplest possible formulation able to deal both with the Ising contours introduced above and the contours to be used in Chapters 4 and 5.

3.3.2 General contour ensembles

We consider objects γ which we call *contours*, but are also called *polymers* [23] or *animals* [10]. Each γ has associated a *finite nonempty* set denoted $\mathrm{supp}\,\gamma$ and called the *support* of γ, contained in a set $\mathcal{L}$ isomorphic to $\mathbb{Z}^d$. For the Peierls contours discussed above, $\mathrm{supp}\,\gamma$ is the set of (centers of the) segments or plaquettes forming γ. For the contours of Chapters 4 and 5 the supports are in fact finite sets of sites in $\mathbb{Z}^d$.

Two contours γ_1 and γ_2 are said *compatible* if $\mathrm{supp}\,\gamma_1 \cap \mathrm{supp}\,\gamma_2 = \emptyset$. A *compatible family* of contours is a (possibly infinite) family of pairwise compatible contours. Two (not necessarily compatible) families of contours are *mutually compatible* if each contour of one family is compatible with all the contours of the other one. A *cluster* is a collection of contours that can not be decomposed into two mutually compatible subfamilies. A cluster may contain the same contour more than once.

We then consider the set $\mathcal{C}$ formed by all *compatible families* Γ of contours. The empty family is always an element of every contour ensemble. We shall call each Γ a *contour-family* (Dobrushin calls it a *herd*). Note that the compatibility requirement implies that each $\Gamma \in \mathcal{C}$ is a countable family. In addition, for each finite $\Lambda \subset \mathcal{L}$ let me denote $\mathcal{C}_\Lambda$ the set of compatible contour-families whose contours have support in Λ. We assume that $\mathcal{C}_\Lambda$ is a *finite* set for each finite $\Lambda \subset \mathcal{L}$.

A *contour ensemble* is a family $\mathcal{C}$ as above plus an assignment of weights

$$\Gamma \;\to\; \prod_{\gamma\in\Gamma} w(\gamma)\;. \tag{3.38}$$

If $w(\gamma) \geq 0$, (3.38) makes each $\mathcal{C}_\Lambda$ a probability space. However, in the present formalization, the contour weights $\omega(\gamma)$ are allowed to be *complex numbers* because, in any case, the cluster expansion is required to be absolutely convergent. Thus, the technique exceeds a

mere probabilistic setting. In this section, except for formula (3.56), periodicity is not assumed.

3.3.3 The fundamental theorem and its corollaries

The theory is based on an expansion for the contour free-energies

$$\log \Theta_\Lambda \;:=\; \log\Big[1 + \sum_{\substack{\Gamma\in\mathcal{C}_\Lambda \\ \Gamma\neq\emptyset}} \prod_{\gamma\in\Gamma} w(\gamma)\Big] \,. \tag{3.39}$$

Logarithms are tricky functions, The naive expansion of (3.39) gives a series with alternating signs. A simple control via term-wise absolute values is out of question, because the number of terms grows too fast. Convergence can only be established if cancelations between different terms are taken into account. This is a laborious task. The end result is the following:

Theorem 3.1 *If*

$$\epsilon := \sup_{x\in\mathcal{L}} \sum_{\gamma\,:\,\mathrm{supp}\,\gamma\ni x} |w(\gamma)|\; e^{|\mathrm{supp}\,\gamma|} \;<\; 1 \,, \tag{3.40}$$

then, there exists an absolutely convergent expansion of the form

$$\log \Theta_\Lambda \;=\; \sum_{n\geq 1} \frac{1}{n!} \sum_{\gamma_1\,:\,\mathrm{supp}\,\gamma_1\subset\Lambda} \cdots \sum_{\gamma_n\,:\,\mathrm{supp}\,\gamma_n\subset\Lambda} w^T(\gamma_1,\ldots,\gamma_n) \,, \tag{3.41}$$

with

$$w^T(\gamma_1,\ldots,\gamma_n) \;=\; \varphi^T(\gamma_1,\ldots,\gamma_n) \prod_{i=1}^{n} w(\gamma_i) \tag{3.42}$$

where φ is a factor independent of the weights w and symmetric under permutations of $\{1,\ldots,n\}$ that satisfies

$$\varphi^T(\gamma_1,\ldots,\gamma_n) \;=\; 0 \quad \textit{if } \{\gamma_1,\ldots,\gamma_n\} \textit{ is not a cluster.} \tag{3.43}$$

Moreover, for any $x\in\mathcal{L}$ we have the bound

$$\sup_{x\in\mathcal{L}} \frac{1}{n!} \sum_{\substack{\gamma_1\,:\,\mathrm{supp}\,\gamma_1\subset\Lambda\,,\\ \mathrm{supp}\,\gamma_1\ni x}} \sum_{\gamma_2\,:\,\mathrm{supp}\,\gamma_2\subset\Lambda} \cdots \sum_{\gamma_n\,:\,\mathrm{supp}\,\gamma_n\subset\Lambda} \Big|w^T(\gamma_1,\ldots,\gamma_n)\Big| \;\leq\; \mathrm{O}(\epsilon^n) \,, \tag{3.44}$$

where $\mathrm{O}(\epsilon^n)$ *is* Λ*-independent. This last formula is valid for any finite* ϵ *in (3.40), not necessarily smaller than one.*

A very clear presentation of this theorem is given, for instance, in [34, Section 3]. It relies on an identity involving certain tree graphs whose edges join contours with intersecting supports. These trees are summed using Cayley's formula (which counts the trees with fixed incidence numbers), a procedure that introduces exponential factors at each "vertex" of the tree. This is the genesis of the exponential that shows up in (3.40).

The version of the expansion due to Kotecký and Preiss [29] yields a convergence condition that is more generous than (3.40). This version can be interpreted as a resummation of the series (3.41) which leads to resummed w^T that are no longer multilinear in the original weights $w(\gamma_i)$.

A particularly important expression is

$$T_\Lambda(A) \;:=\; \sum_{n\geq 1} \frac{1}{n!} \sum_{\substack{\gamma_1\,:\,\mathrm{supp}\,\gamma_1\in\Lambda\,,\\ \mathrm{supp}\,\gamma_1\cap A\neq\emptyset}} \sum_{\gamma_2\,:\,\mathrm{supp}\,\gamma_2\in\Lambda} \cdots \sum_{\gamma_n\,:\,\mathrm{supp}\,\gamma_n\in\Lambda} w^T(\gamma_1,\ldots,\gamma_n) \tag{3.45}$$

for finite $A \subset \mathcal{L}$. ["T" stands for "touch".] Its importance derives from the fact that if $\Delta \supset \Lambda$ and the series $\log \Theta_\Delta$ and $\log \Theta_\Gamma$ converge absolutely,

$$\log \Theta_\Delta - \log \Theta_\Lambda \;=\; T_\Delta(\Delta \setminus \Lambda)\;. \tag{3.46}$$

Here and in the sequel we denote $\lim_{\Lambda\to\mathcal{L}}$ to symbolize that for every exhausting sequence of volumes Λ_n there exists a $\lim_{\Lambda_n\to\infty}$ which is independent of the sequence.

Corollary 3.2 *In the regime (3.40),*

$$\lim_{\Lambda\to\mathcal{L}} T_\Lambda(A) \;=:\; T(A) \tag{3.47}$$

exists for all finite $A \subset \mathcal{L}$ *and is given by the absolutely convergent series*

$$T(A) \;:=\; \sum_{n\geq 1} \frac{1}{n!} \sum_{\gamma_1\,:\,\mathrm{supp}\,\gamma_1\cap A\neq\emptyset} \sum_{\gamma_2} \cdots \sum_{\gamma_n} w^T(\gamma_1,\ldots,\gamma_n)\;. \tag{3.48}$$

Moreover,

$$\left.\begin{matrix} |T_\Lambda(A)| \\ |T(A)| \end{matrix}\right\} \;\leq\; |A|\;\delta \tag{3.49}$$

with

$$\delta \;:=\; \sup_{x\in\mathcal{L}} \sum_{n\geq 1} \frac{1}{n!} \sum_{\substack{\gamma_1 : \\ \operatorname{supp}\gamma_1 \ni x}} \sum_{\gamma_2} \cdots \sum_{\gamma_n} \left| w^T(\gamma_1,\ldots,\gamma_n) \right| \tag{3.50}$$

$$\;=\; \mathrm{O}(\epsilon)\;. \tag{3.51}$$

Proof. The series (3.48) converges absolutely by (3.44). Moreover, the series formed by its absolute values:

(i) Majorizes the analogous series but with contours supported in Λ. This proves (3.47) by dominated convergence and, thus, the existence and finiteness of δ defined by (3.50).

(ii) Is bounded above by $|A|\;\delta$. This proves (3.49). ∎

For the sake of consistency with established notation, let me denote

$$f_x^{\mathcal{C}} \;:=\; \sum_{n\geq 1} \sum_{\substack{\{\gamma_1,\ldots,\gamma_n\}: \\ \operatorname{sup}\gamma_1\cup\cdots\cup\operatorname{sup}\gamma_n \ni x}} \frac{w^T(\gamma_1,\ldots,\gamma_n)}{|\operatorname{sup}\gamma_1 \cup \cdots \cup \operatorname{sup}\gamma_n|} \tag{3.52}$$

$$\;=\; \mathrm{O}(\epsilon)\;. \tag{3.53}$$

Corollary 3.3 *In the regime (3.40), we have the following bound on finite-volume effects: For every finite* $\Lambda \subset \mathbb{Z}^d$,

$$\left| \log \Theta_\Lambda + \sum_{x\in\Lambda} f_x^{\mathcal{C}} \right| \;\leq\; \delta\;|\partial\Lambda|\;. \tag{3.54}$$

This is proven using the previous corollary for the series $T_\Lambda(\partial\Lambda)$.

The next corollary is the only result of this section on cluster expansions where we assume periodicity. Let us call the contour ensemble *periodic* if

(i) Translations act on the collection of contours, that is, for each contour γ and $a \in \mathcal{L}$ there is another contour $\tau_a\gamma$ with $\operatorname{supp}\tau_a\gamma = \tau_a \operatorname{supp}\gamma$.

(ii) The assignment (3.38) is *periodic*:

$$w(\tau_a \gamma) \;=\; w(\gamma) \quad \forall \gamma \in \mathcal{C} \;;\; \forall a \in F_b \;, \tag{3.55}$$

for some fundamental cell F_b [see (2.5)].

Corollary 3.4 *For a periodic contour ensemble in the regime (3.40):*

(i) The contour free-energy density, that is, the limit

$$f^{\mathcal{C}} \;:=\; -\lim_{\Lambda \to \mathbb{Z}^d} \frac{1}{|\Lambda|} \log \Theta_\Lambda \tag{3.56}$$

exists and is given by the absolutely convergent expansion

$$\begin{aligned} f^{\mathcal{C}} \;&=\; \frac{1}{|F_b|} \sum_{x \in F_b} f_x^{\mathcal{C}} \\ &=\; \mathrm{O}(\epsilon) \;. \end{aligned} \tag{3.57}$$

(ii) For every $\Lambda \subset \mathbb{Z}^d$ that is a disjoint union of translates of F_b,

$$\Big| \log \Theta_\Lambda + |\Lambda| \; f^{\mathcal{C}} \Big| \;\leq\; \delta \; |\partial \Lambda| \;. \tag{3.58}$$

This is basically a rewriting of the previous corollary.

3.3.4 Contour "probabilities"

The original goal of controlling rates of partition functions is accomplished via (3.46):

$$\frac{\Theta_{\Lambda \setminus A}}{\Theta_\Lambda} \;=\; e^{T_\Lambda(A)} \;. \tag{3.59}$$

By Corollary 3.2 this ratio has a limit as $\Lambda \to \mathcal{L}$. For the Ising model (ensemble $\mathcal{C}^+$ of Peierls contours), this observation implies that if β has a real part large enough so that (3.40) is satisfied, each of the limits

$$\lim_{\Lambda \to \mathbb{Z}^d} \mathrm{Prob}_\Lambda^{(\beta)+}(\omega_A = \sigma_A) \;=:\; \mathrm{Prob}^{(\beta)+}(\omega_A = \sigma_A) \tag{3.60}$$

exists. Moreover, given the exponential dependence of the weights in β, one concludes that these limits are *analytic* in β, for β in this region. This example shows that the cluster-expansion technology can be used actually to show the *existence* of the infinite-volume measures of spin systems. While for systems like the Ising model this is nothing new (except for the additional insight on analyticity), the construction may be the *only* way available to prove the existence of infinite-volume expectations if the weights are complex valued, as in quantum statistical mechanical systems [8, 5].

In general, the basic objects evaluated via contour ensembles are the "probabilities" of occurrence of a fixed family of contours:

$$\begin{aligned} &\text{"Prob"}_{\mathcal{C}_\Lambda}(\{\Gamma \in \mathcal{C}_\Lambda \,:\, \gamma_1, \ldots, \gamma_n \in \Gamma\}) \\ &\quad = \; w(\gamma_1) \cdots w(\gamma_n) \, \frac{\Theta_{\Lambda \setminus (\operatorname{supp}\gamma_1 \cup \cdots \operatorname{supp}\gamma_n)}}{\Theta_\Lambda} \\ &\quad = \; w(\gamma_1) \cdots w(\gamma_n) \, \exp[T_\Lambda(\operatorname{supp}\gamma_1 \cup \cdots \operatorname{supp}\gamma_n)] \; . \end{aligned} \tag{3.61}$$

The quotation marks are a reminder that these can be complex-valued "probabilities". Corollary 3.2 implies the existence of the $\Lambda \to \mathbb{Z}^d$ limit of this expression, which, by (3.49) satisfies the bound

$$\begin{aligned} &|\text{"Prob"}_{\mathcal{C}}(\{\Gamma \in \mathcal{C} \,:\, \gamma_1, \ldots, \gamma_n \in \Gamma\})| \\ &\qquad \leq \; |w(\gamma_1)| \, e^{|\operatorname{supp}\gamma_1|\,\delta} \cdots |w(\gamma_n)| \, e^{|\operatorname{supp}\gamma_n|\,\delta} \; . \end{aligned} \tag{3.62}$$

For the Ising case, this inequality is weaker than the bound behind (3.24). The latter corresponds to replacing each $\exp[|\operatorname{supp}\gamma_i|\,\delta]$ by 1. But in general non-symmetric cases, or with signed or complex-valued weights, the previous bound is the best one can expect.

In the present general discussion I adopted the word "contours" —rather than polymers or animals— to emphasize the fact that each $\operatorname{supp}\gamma$ can be associated to a (possibly "thick") closed boundary. More precisely, each $\operatorname{supp}\gamma$ divides $\mathbb{Z}^d$ into disjoint components: one with infinite cardinality — the *exterior*, $\operatorname{Ext}\gamma$,— and the remaining one involving only finitely-many sites. The union of these finite-cardinality components is called the *interior* of γ and denoted $\operatorname{Int}\gamma$. In this case, it makes sense to evaluate

$$\text{"Prob"}_{\mathcal{C}}(\{\Gamma \,:\, \exists \gamma \in \Gamma \text{ with } \; |\operatorname{supp}\gamma| = \ell \text{ and } \; \operatorname{Int}\gamma \ni 0\})$$

$$\leq \ \ell \sup_{x\in\mathcal{L}} \text{“Prob”}_{\mathcal{C}}(\{\Gamma \,:\, \exists\gamma\in\Gamma \text{ with } |\operatorname{supp}\gamma| = \ell \text{ and } \operatorname{supp}\gamma \ni x\}) \,. \tag{3.63}$$

The factor ℓ counts, as in the Peierls argument, the number of sites where the contour can intercept the 1st axis. Thus, from (3.62) and (3.50),

$$\begin{aligned} &\text{“Prob”}_{\mathcal{C}}(\{\Gamma \,:\, \exists\gamma\in\Gamma \text{ with } \operatorname{Int}\gamma \ni 0\}) \\ &\quad\leq \sup_{x\in\mathcal{L}} \sum_{\gamma\,:\,\operatorname{supp}\gamma\ni x} |\operatorname{supp}\gamma| \ |w(\gamma)| \ e^{|\operatorname{supp}\gamma|\,\delta} \\ &\quad\leq \frac{\epsilon}{(1-\delta)\,e} \,, \end{aligned} \tag{3.64}$$

where we used the fact that $\max_{\ell\geq 0} \ell e^{-(1-\delta)\ell} = 1/[(1-\delta)e]$ and assumed that ϵ is small enough so $\delta < 1$ [recall that $\delta = \mathrm{O}(\epsilon)$].

When applied to the $\mathcal{C}^+$ ensemble of the Ising model, this implies that for an observable g with $g(\omega) = g(\omega_B)$

$$\Big|\langle g\rangle^{(\beta)\,+} - g(+)\Big| \ \leq \ \|g\|_\infty \ |B| \ \mathrm{O}(\epsilon) \tag{3.65}$$

for (real part of) β large enough. This is an explicit estimation of “closeness” between the low-temperature “+”-phase and the deterministic zero-temperature distribution δ_+. Estimations like this remain valid even for complex-valued weights.

On the other hand, if the contour ensemble has real non-negative weights, i.e. it is an honest probability space, we can combine (3.64) with the first Borel-Cantelli lemma to conclude, as in Section 3.2.3, that typical configurations exhibit only a *finite* number of contours around each site.

3.3.5 Mixing properties

Let us now consider *families* of contour ensembles parameterized by a certain $\tau > 0$ —the *Peierls constant*— satisfying the *Peierls bound*:

$$\Big|w^{(\tau)}(\gamma)\Big| \ \leq \ e^{-\tau\,|\operatorname{supp}\gamma|} \,. \tag{3.66}$$

Under this condition, if $\tau_{\mathrm{c}} = \tau_{\mathrm{c}}(\epsilon)$ is the solution of

$$\sup_{x\in\mathcal{L}} \sum_{\gamma\,:\,\operatorname{supp}\gamma\ni x} e^{-(\tau_{\mathrm{c}}-1)|\operatorname{supp}\gamma|} \ = \ \epsilon \,, \tag{3.67}$$

then for a given $\epsilon < 1$ the fundamental Theorem 3.1 holds for all models with

$$\tau \geq \tau_{\rm c} \ . \tag{3.68}$$

In particular, the contour "probabilities" enjoy the properties discussed above.

If we combine (3.67) with the multilinear dependence (3.42) on the weights and the bound (3.44), we obtain

$$\begin{aligned} \sup_{x\in\mathcal{L}} \frac{1}{n!} \sum_{\gamma_1\,:\,\mathrm{supp}\,\gamma_1 \ni x} \sum_{\gamma_2} \cdots \sum_{\gamma_n} \Big| w^{(\tau)\,T}(\gamma_1,\ldots,\gamma_n)\Big| \\ \times\ \mathbb{1}[|\mathrm{supp}\,\gamma_1 \cup \cdots \cup \mathrm{supp}\,\gamma_n| \geq D] \ \leq\ e^{-(\tau-\tau_{\rm c})D}\,\mathrm{O}(\epsilon^n) \ . \end{aligned} \tag{3.69}$$

As a consequence, the sum

$$\begin{aligned} T^{(\tau)}_\Lambda(A,B) \ :=& \\ \sum_{n\geq 1} \frac{1}{n!} \sum_{\gamma_1\,:\,\mathrm{supp}\,\gamma_1\subset\Lambda} \cdots \sum_{\gamma_n\,:\,\mathrm{supp}\,\gamma_n\subset\Lambda} & w^{(\tau)\,T}(\gamma_1,\ldots,\gamma_n) \\ \times\ \mathbb{1}\Big[\begin{array}{l} (\mathrm{supp}\,\gamma_1 \cup \cdots \cup \mathrm{supp}\,\gamma_n) \cap A \neq \emptyset \ , \\ (\mathrm{supp}\,\gamma_1 \cup \cdots \cup \mathrm{supp}\,\gamma_n) \cap B \neq \emptyset \end{array} \Big] & \end{aligned} \tag{3.70}$$

converges for $\epsilon < 1$ and satisfies the bound

$$|T^{(\tau)}_\Lambda(A,B)| \ \leq\ |A|\ |B|\ \mathrm{O}(\epsilon)\ \exp[-(\tau-\tau_{\rm c})\mathrm{dist}(A,B)] \ . \tag{3.71}$$

As before,

$$\lim_{\Lambda\to\mathcal{L}} T^{(\tau)}_\Lambda(A,B) \ =:\ T^{(\tau)}(A,B) \tag{3.72}$$

exists.

The relevance of (3.71) is explained in the following proposition, where we omit the superscript "(τ)" to simplify the notation.

Proposition 3.5 *Consider contour ensembles satisfying the Peierls bound (3.66) and a value $\tau > \tau_{\rm c}$, and let γ_1 and γ_2 be two compatible contours. Then, for Λ large enough*

$$\begin{aligned} \Big| \text{"Prob"}_{\mathcal{C}_\Lambda}(\{\Gamma\ :\ \gamma_1,\gamma_2\in\Gamma\}) \\ -\ \text{"Prob"}_{\mathcal{C}_\Lambda}(\{\Gamma\ :\ \gamma_1\in\Gamma\})\ \text{"Prob"}_{\mathcal{C}_\Lambda}(\{\Gamma\ :\ \gamma_2\in\Gamma\})\Big| \\ \leq\ |w(\gamma_1)|\ e^{|\mathrm{supp}\,\gamma_1|\,\delta}\ |w(\gamma_2)|\ e^{|\mathrm{supp}\,\gamma_2|\,\delta}\,\mathrm{O}\Big[e^{-(\tau-\tau_{\rm c})\,\mathrm{dist}(\mathrm{supp}\,\gamma_1,\mathrm{supp}\,\gamma_2)}\Big] \ . \end{aligned} \tag{3.73}$$

Proof. By (3.62)

$$\begin{aligned} \text{LHS} &= |w(\gamma_1)|\ |w(\gamma_2)|\ \left| e^{T_\Lambda(\operatorname{supp}\gamma_1 \cup \operatorname{supp}\gamma_2)} - e^{T_\Lambda(\operatorname{supp}\gamma_1)}\, e^{T_\Lambda(\operatorname{supp}\gamma_2)} \right| \\ &= |w(\gamma_1)|\ |w(\gamma_2)|\ e^{|T_\Lambda(\operatorname{supp}\gamma_1)|}\, e^{|T_\Lambda(\operatorname{supp}\gamma_2)|}\ \left| e^{-T_\Lambda(\operatorname{supp}\gamma_1\,,\,\operatorname{supp}\gamma_2)} - 1 \right| . \end{aligned} \tag{3.74}$$

Now use Corollary 3.2 and inequality (3.71). ∎

Of course the mixing property (3.73) implies, via the limit $\Lambda \to \mathbb{Z}^d$, a similar inequality for the "probabilities" on the whole of $\mathcal{C}$. Analogous results, with analogous proofs, hold when γ_1 and γ_2 are replaced by mutually compatible families of contours $\{\gamma_1^{(1)}, \dots, \gamma_{n_1}^{(1)}\}$ and $\{\gamma_1^{(2)}, \dots, \gamma_{n_2}^{(2)}\}$.

When applied to the ensemble $\mathcal{C}^+$ for the Ising model, the previous proposition shows that the +-phase is exponentially mixing (at least at low temperature) and gives a (rough) estimation of the correlation length.

3.3.6 Summary of properties

For future quick reference, let me summarize the properties proven for contour ensembles in the regime (3.40), and advance their implications when the ensemble is associated to the η-phase of a spin system.

(P1) Existence of the free-energy density (for periodic ensembles, see Corollary 3.4). The resulting expression is a series expansion for the free energy density of the associated spin system, which is useful, for instance, to exhibit analyticity properties.

(P2) Existence of the infinite-volume limit of the contour probabilities [see (3.60)]. This implies the existence of the η-Gibbs distribution of the associated spin system.

(P3) Absence of infinite sequence of nested contours (for contours with a notion of interior and non-negative weights). This follows from (3.64) and Borel-Cantelli. For an associated spin system this implies the validity of the sea-with-islands picture for the η-phase: Each

site is surrounded at most by finitely many contours and (groups of) spins not configured as in η do not percolate.

(P4) The probability of occurrence of contours around a site is $O(\epsilon)$ (for contours with a notion of interior). Again, this follows from (3.64). For an associated spin system this means that each expectation $\langle g \rangle^\eta$ is $O(\epsilon)$-close to the value $g(\eta)$. [This $O(\epsilon)$ depends on g.]

(P5) Exponential mixing [for ensembles satisfying the Peierls bound (3.66), see Proposition 3.5]. It implies the same property for the associated η-phase of the spin system, which, in turns, implies that the η-Gibbs distribution is *extremal*.

Chapter 4

Pirogov-Sinai theory

Let us now discuss how to extend the Peierls argument to more general situations. In principle, our goal is simply stated: Given a system we should find contour ensembles such that:

1. For each rigid ground-state configuration η, there is a one-to-one correspondence between configurations in Ω_Λ^η and families of contours with support in Λ.

2. We can apply the cluster expansion technology of Section 3.3 to the contour ensemble(s).

Both aspects will turn out to be rather subtle, and usually impossible to satisfy simultaneously, as we shall illustrate with a number of examples. We shall see that each rigid η will require its own contour ensemble, and that the cluster expansion will be applicable at the expense of loosing the one-to-one correspondence. For the Ising model, the contour ensembles $\mathcal{C}^+$ and $\mathcal{C}^-$ coincide, so we can think of them as a single ensemble for which both requirements above are cleanly satisfied. This is a consequence partially of the flipping symmetry and partially of the nearest-neighbor character of the interaction. We can not expect being so lucky in the general case. In a sense, the Ising model may be misleading in its simplicity.

4.1 The definition of contours

4.1.1 Aspects to consider

Let me motivate, via examples, the definition of contours to be presented below. The discussion here will be a little imprecise. Precision will come in Section 4.1.2

Aspect 1: The need for labels

Let me go back to the Blume-Capel model (2.44). We can proceed by analogy with the Ising case and draw segments (plaquettes) perpendicular to pairs of misaligned nearest neighbors. If one of the rigid ground-state configurations —all-"+1", all-"0" or all-"−1"— is chosen as external condition, the resulting contours are closed circuits (surfaces) separating regions where spins are aligned in one of the three energy-minimizing possibilities. However, it is obvious that to reconstruct univocally the configuration we need to specify what type of misaligned pairs are separated by each contour. In fact, it is enough to add to each contour labels that indicate which configuration is present immediately outside the contour, and immediately inside each component of its interior.

Moreover, contours with different labels have, in general, different weights. Indeed, a $(+,-)$ nearest-neighbor pair raises the energy four times more than either a $(+,0)$ or a $(-,0)$ pair. This dependence of the weights on the set of labels is typical of situations in which, unlike the Ising model, there is no symmetry operation connecting the different ground-state configurations.

Aspect 2: The need for thickness

The other feature of Ising contours, namely their being "thin" ethereal circuits living in the dual lattice, is obviously specific to the nearest-neighbor character of the interaction. For interactions of larger range one needs to regard simultaneously groups of spins within the range, to determine the different contributions to the energy. A simple example

could be a range-2 Ising model

$$H \;=\; -J \sum_{\substack{(x,y)\,:\\ \mathrm{dist}(x,y)\le 2}} \sigma_x\sigma_y \;. \tag{4.1}$$

This model has the same all-"+1" and all-"−1" rigid ground-state configurations as the nearest-neighbor version. But if we want to associate excess energy to contours we have to consider them formed by plaquettes of sides 2 where the spins are not aligned. Note that these "thick" contours can have in (one of the components of) the interior the same rigid configuration as in the exterior. Thus, labels are still necessary, or some other means to identify the configurations in the exterior and in the different components of the interior. The flipping symmetry of the model implies that every contour has a flipped version with the same energy. Therefore, the symmetry extends to the partition function

$$Z^+_\Lambda \;=\; Z^-_\Lambda \;. \tag{4.2}$$

Another model lacking a similar symmetry and requiring "thick" contours is the Fisher antiferromagnet (I have tuned out the parameters):

$$H \;=\; \sum_{<x,y>} \omega_x\omega_y - \sum_{\substack{(x,y)\,:\\ \mathrm{dist}(x,y)=\sqrt{2}}} \omega_x\omega_y - 2d\sum_x \omega_x \;. \tag{4.3}$$

This model has three rigid ground-state configurations: The all-"+1" and the two alternating "+−" configurations (in physics nomenclature, ferromagnetic and Neél ordered ground states.). While in the previous model the thickness in the contours was needed for a correct account of the energy, in this model the plaquettes are needed to determine boundaries of excitations. Looking just to a nearest neighbor pair, one can not in general decide whether the pair is part of an excitation (except if both spins are "−") or not. One needs to look to plaquettes of side (at least) 2. In $d = 2$, the energy minimizing plaquettes are

$$\eta_+ = \begin{pmatrix} + & + \\ + & + \end{pmatrix} \;,\; \eta_{+-} = \begin{pmatrix} + & - \\ - & + \end{pmatrix} \;,\; \eta_{-+} = \begin{pmatrix} - & + \\ + & - \end{pmatrix} \;, \tag{4.4}$$

other plaquette configurations correspond to excitations.

Aspect 3: The need for reference configurations that are not ground-state

Consider the Ising model with (2.17) with $h > 0$. Such model has a single rigid ground-state configuration, namely the all-"+1". Having accepted the possibility of "thick" contours, we could be tempted to define as part of them all sites that are "−". Many such contours will then have very thick supports, and the dominant contribution to their energy will come from the misalignment of the spins with the magnetic field. That is,

$$w(\omega) \simeq e^{-\beta h|\operatorname{supp}\gamma|} . \tag{4.5}$$

This means that the estimations made via cluster-expansion methods will get worse as $h \to 0$. This is certainly not desirable, because the vicinity of $h = 0$ is the most interesting region of the model (coexistence of phases!). It is therefore preferable to come back to the original policy of associating contours with pairs of misaligned spins, so the exponential damping is proportional to J. The contribution of h is then taken into account by the difference in energy densities [see (3.9)]. In other words, it is convenient to consider two possible external configurations —all-"+1" and all-"−1"— even when the latter is not a ground state configuration.

This example shows that, specially if one wants to study the effect of varying parameters (phase diagrams), one must include external conditions that are *not* ground-state in the region under study (they usually are ground-state in some other region). So contours will, in general, be defined considering some set of *reference configurations* not necessarily associated to energy minimizers. The condition to be discussed in Section 4.2.2 will fix univocally (albeit not in a constructive manner) what configurations should be chosen.

4.1.2 Definition of contours for general models

After the preceding detailed motivation, let us plunge into the right general definition of contour, as presented by Pirogov and Sinai in their original work [35, 36]. From now on, the rest of these notes are a minor adaptation of parts of [8].

We start with a set of periodic *reference configurations* $\mathcal{K} = \{\eta_1, \ldots, \eta_P\}$ and a number r. This number will represent the range of the interaction, but, except for this r, the definition of contour makes no reference to a particular interaction. (When studying phase diagrams, one chooses a definition suited to all the interactions being considered). We fix *sampling plaquettes* $W_a(x) = \{y \in \mathbb{Z}^d : |x_i - y_i| \leq a \text{ for } 1 \leq i \leq d\}$. The size a must be strictly larger than (i) the periods of the reference configurations $\eta_1, \ldots, \eta_P$, and (ii) the given number r (in practice, the range of the interaction).

Condition (i) implies the following *extension property*:

$$\begin{aligned}&\textit{If } \omega \textit{ coincides with the configuration } \eta_p \textit{ on a plaquette } W_a(x)\\ &\textit{and with } \eta_q \textit{ on } W_a(y) \textit{ with } \operatorname{dist}(x,y) \leq 1\textit{, then } \eta_p = \eta_q.\end{aligned} \tag{4.6}$$

Two sets, A and B, in $\mathbb{Z}^d$ are said to be connected if $\operatorname{dist}(A, B) \leq 1$ in lattice units. A subset M of a set $A \subset \mathbb{Z}^d$ is called a *component* of A if M is a maximal connected subset of A, i.e., M is connected and $M \subset M' \subset A$, $M \neq M'$ imply that M' cannot be connected.

Contours are constructed out of "incorrect" plaquettes. A site x is said to be p-correct, for a configuration ω, if the latter coincides with η_p on *every* sampling plaquette that contains x. The set of sites that are not p-correct for any p, $1 \leq p \leq P$, are referred to as "incorrect". If all the reference configurations η_i are translation-invariant, then we define the *defect set* of ω as the set of plaquettes for which at least one site is "incorrect". If some of the η_i are only periodic, it is necessary to consider regions that respect this periodicity, so formulas like (3.58) can be directly applied. To this end, we first choose a set F_b which is a fundamental set for *all* the reference configurations η_i (this is possible because there are only finitely many) and pave $\mathbb{Z}^d$ with disjoint translates of this F_b. This pavement is fixed once and for all. At this point we can choose:

Choice 1) We "collapse" the systems so that each translate of F_b becomes a point. This amounts to rewriting the interaction in a suitable (equivalent) manner. The new system has all reference configurations translation-invariant and we define the defect set of each ω as above.

Choice 2) We agree to work only with sets that are union of the tiles of the pavement. In particular, we define the defect set of ω as

the smallest covering of its set of incorrect points by tiles of the pavement.

Both choices are equivalent. In fact, all the formulas that follow look identical in both cases if, for the second choice we agree on the following

- (i) All volumes Λ, Δ, etc, are understood to be formed by disjoint union of F_b-tiles.
- (ii) The cardinality of sets of sites is measured in units of fundamental cells:

$$|\Lambda| \ := \ \frac{\mathrm{card}\,\Lambda}{\mathrm{card} F_b} \ . \tag{4.7}$$

A *contour* of a configuration ω is a pair $\gamma = (\mathrm{supp}\,\gamma, \omega_{\mathrm{supp}\,\gamma})$ where $\mathrm{supp}\,\gamma$ —the *support* of γ— is a component of the defect set of ω. To abbreviate the long formulas of next chapter, let us agree on denoting

$$|\gamma| \ := \ |\,\mathrm{supp}\,\gamma| \ . \tag{4.8}$$

[Where the right-hand side is understood as in (4.7).] An alternative, more physically sounding, name for these contours is *excitations*, and plaquettes belonging to a contour are said to be *excited* .

Henceforth, we shall only consider *finite* contours (i.e., $|\,\mathrm{supp}\,\gamma| < \infty$). For each such contour, γ, the space $\mathbb{Z}^d \setminus \mathrm{supp}\,\gamma$ is divided into a finite number of components. By the extension property (4.6), we can extend the configuration on a single plaquette in a component to a unique configuration of $\mathcal{K}$ in that component. In this way we can label each connected component of $\mathbb{Z}^d \setminus \mathrm{supp}\,\gamma$ by a particular reference configuration. Thus, we obtain the unique configuration ω^γ that has γ as its *only* contour. We shall refer to such a configuration as a one-contour configuration. The only infinite component of $\mathbb{Z}^d \setminus \gamma$ is called the *exterior* of the contour, $\mathrm{Ext}\,\gamma$, and the union of the other components constitute the *interior*, $\mathrm{Int}\,\gamma$. The union of components of $\mathrm{Int}\,\gamma$ labeled by a reference configuration η_q is called the q-interior, $\mathrm{Int}_q\,\gamma$. The contour is called a *p-contour* if its exterior is labeled by the configuration $\eta_p \in \mathcal{K}$. When necessary, we shall write γ^p to explicitly indicate that γ is a p-contour.

Each configuration defines a unique family of contours from which it can be reconstructed, but *not* all families of contours correspond to configurations. The situation is more complicated than for the Peierls contours of Section 3.2. Besides having disjoint supports, nested contours must have matching internal and external labels. A family of contours which corresponds to a configuration will be called *admissible*.

A contour γ of a configuration ω is called an *exterior contour* of ω if its support is not contained in the interior of any other contour of ω, i.e., if $\gamma \subset \mathrm{Ext}(\gamma')$ holds, for any other contour γ' of ω.

4.2 The Peierls condition

4.2.1 Contour energies

Consider an interaction $\{\Phi_B^{\mathrm{cl}}\}$ of range not exceeding r. The functions Φ_B can be complex-valued. Let us see how the energy of a configuration can be written in terms of contours. Following the pattern set by the Peierls argument, we should fix an external configuration η_p and assign to the contours the excess energy with respect to it. Let us first consider one-contour configurations.

Let ω^γ be a one-contour configuration which has the p-contour γ as its only contour. To compute the energy cost of γ, relative to its exterior configuration η_p, we write

$$H_\Lambda^{\eta_p}(\omega^\gamma) \;=\; \sum_{B\cap(\mathrm{Int}\,\gamma\cup\mathrm{supp}\,\gamma)\neq\emptyset} [\Phi_B(\omega^\gamma)-\Phi_B(\eta_p)] + \sum_{B\cap\Lambda\neq\emptyset} \Phi_B(\eta_p)\ , \quad (4.9)$$

where Λ is large enough to contain $\mathrm{supp}\,\gamma$. It is convenient to use the decomposition [44]

$$\begin{aligned}\sum_{B\cap(\mathrm{Int}\,\gamma\cup\mathrm{supp}\,\gamma)\neq\emptyset} [\Phi_B(\omega^\gamma)-\Phi_B(\eta_p)] \;=\; \\ \sum_B \frac{|B\cap\mathrm{supp}\,\gamma|}{|B|}\,[\Phi_B(\omega^\gamma)-\Phi_B(\eta_p)] \\ +\;\sum_B \frac{|B\cap\mathrm{Int}\,\gamma|}{|B|}\,[\Phi_B(\omega^\gamma)-\Phi_B(\eta_p)]\end{aligned}$$

$$+ \sum_{B\cap(\mathrm{Int}\,\gamma\cup\mathrm{supp}\,\gamma)\neq\emptyset} \frac{|B\cap \mathrm{Ext}\,\gamma|}{|B|}\,[\Phi_B(\omega^\gamma)-\Phi_B(\eta_p)]\;, \tag{4.10}$$

and, analogously,

$$\sum_{B\cap\Lambda\neq\emptyset}\Phi_B(\eta_p) \;=\; e(\eta_p)\,|\Lambda| \;+\; \sum_{B\cap\Lambda\neq\emptyset}\frac{\left|B\cap(\mathbb{Z}^d\setminus\Lambda)\right|}{|B|}\,\Phi_B(\eta_p)\;. \tag{4.11}$$

In this fashion, (4.9) becomes

$$H_\Lambda^{\eta_p}(\omega^\gamma) \;=\; E(\gamma)+\sum_{u=1}^{P}\left[e(\eta_u)-e(\eta_p)\right]\,|\mathrm{Int}_u\,\gamma|+e(\eta_p)\,|\Lambda|+R_\Lambda(\eta_p)\;, \tag{4.12}$$

where

$$E(\gamma) \;=\; \sum_B \frac{|B\cap\mathrm{supp}\,\gamma|}{|B|}\left[\Phi_B(\omega^\gamma)-\Phi_B(\eta_p)\right] \tag{4.13}$$

is the *contour energy* of γ relative to the energy of its exterior configuration, and $R_\Lambda(\eta_p)$ is a *boundary* term which is independent of the configuration ω. This term will be omitted from our analysis. More precisely, we shall work with

$$Z_p(\Lambda) \;:=\; e^{\beta R_\Lambda^p}\,Z_\Lambda^{\eta_p}\;. \tag{4.14}$$

[The notation is purposely similar, it is not worthwhile to keep track of the difference between both objects.] The free energy density obtained with either of both partitions is the same, but the left-hand side is what enters in the argument.

In obtaining (4.13) we have profited from having chosen the plaquette size *a larger* than the range r, so that

$$(\omega^\gamma)_B=(\eta_u)_B,\;\;\text{if}\;\;B\cap\mathrm{Int}_u\,\gamma\neq\emptyset, \tag{4.15}$$

for any B with $\Phi_B\neq 0$, and, since γ is a p-contour,

$$(\omega^\gamma)_B=(\eta_p)_B\;\;\text{if}\;\;B\cap\mathrm{Ext}\,\gamma\neq\emptyset. \tag{4.16}$$

Hence the latter bonds do not contribute to the contour energy $E(\gamma)$.

If all the reference configurations $\eta_p \in \mathcal{K}$ are rigid ground-state configurations, they all have the same specific energy and hence, from (4.12), the energy cost of a contour γ is simply given by $E(\gamma)$.

The reader can now convince him/herself that if a configuration ω is characterized by a family of contours $\Gamma = \Gamma(\omega)$, the corresponding energy can be written as

$$H_\Lambda^{\eta_p}(\omega) - S_\Lambda^p \;=\; e(\eta_p)\,|\Lambda| + E(\Gamma) + \sum_{u=1}^{P} \Big[e(\eta_u) - e(\eta_p)\Big]\,|I_u| \;, \quad (4.17)$$

where

$$E(\Gamma) \;=\; \sum_{\gamma\in\Gamma} E(\gamma), \quad (4.18)$$

and I_u is the set of sites in Λ that are either u-correct or belong to a u-contour.

Let me remark that the contours γ may extend outside Λ. This happens if ω has some incorrect site *on* the internal boundary of Λ. In this case all those plaquettes which contain this site, but extend outside Λ, also belong to a contour. Hence, in general, the contours are contained in the larger set formed by the plaquettes that touch Λ:

$$\widehat{\Lambda} \;:=\; \bigcup \{W_a(x) \,:\, W_a(x) \cap \Lambda \neq \emptyset\} \;. \quad (4.19)$$

This means that in $E(\gamma)$ one may be counting bonds $B \subset \mathbb{Z}^d \setminus \Lambda$ that are not counted in $H_\Lambda(\omega|\eta_p)$. However, the identity (4.17) remains valid, because these bonds do not contribute to the energy of a contour, [see sentence following (4.16)].

4.2.2 The Peierls condition

The very least we need to reproduce the Peierls argument in the present general setting is that the energy of each contour be proportional to the cardinal of its support. This requirement is called *Peierls condition*:

Definition 4.1 *An interaction Φ satisfies the Peierls condition, respect to a finite family of reference configurations $\{\eta_1, \ldots, \eta_P\}$, with Peierls constant J if*

$$\mathrm{Re}\, E(\gamma) \;\geq\; J|\gamma| \;, \quad (4.20)$$

where $E(\gamma)$ is the contour energy defined through (4.13).

In many examples, the Peierls condition is verified very easily, almost by inspection. These are cases in which one can show that the excess energy of each excited plaquette is nonzero, *irrespective* of the particular configuration on the plaquettes surrounding it. The situation could, conceivably be more difficult. For instance, certain configurations of an excited plaquette may have no excess energy there, but they can not be extended without paying some extra energy somewhere (tilings). In those cases, to prove the Peierls condition one must show that this energy cost is payed within a bounded distance of the excited plaquette for all possible extensions. This fact has been established by Holsztynski and Slawny [25] for general m-potentials (Definition 2.2). They showed:

Proposition 4.2 *If Φ is periodic and $\mathcal{M}(\Phi)$ is a non-empty finite set, then the system with reference configurations $\mathcal{K} = \mathcal{M}(\Phi)$ satisfies the Peierls condition.*

[Recall that $\mathcal{M}(\Phi)$ is the set of configuration minimizing all functions Φ_B.]

The case in which the reference configurations are not ground-state usually corresponds to an interaction with some parameters $\underline{\mu}$ (like the field h for the Ising model) introduced to break the degeneracy of the ground-state configurations. One is interested in studying stability of phases when the parameters are varied from a certain value $\underline{\mu}_0$ ($h = 0$ in the Ising model). The condition imposed on the size of the sampling plaquettes, namely $a > r$, simplifies the situation, since it permits us to resort to some perturbative results (discussed for instance in [42, pages 1126–1127]) which can be summarized in the following statement:

Proposition 4.3 *Consider a family of interactions $\{\Phi_{\underline{\mu}} B\}$ differentiable in $\underline{\mu}$. Assume that, for some value $\underline{\mu}_0$ of the parameters, the interaction $\{\Phi_{\underline{\mu}_0} B\}$ has a finite number of periodic ground states $\mathcal{K} = \{\eta_1, \ldots, \eta_P\}$ and that it satisfies the Peierls condition, having them as reference configurations, with Peierls constant J_0. Then, for $\alpha > 0$ small enough, there exist open neighborhoods $\mathcal{O}_\alpha \ni \underline{\mu}_0$ such that all the interactions $\{\Phi_{\underline{\mu}} B\}$ with $\underline{\mu} \in \mathcal{O}_\alpha$ satisfy the Peierls condition with Peierls constant $J_0 - \alpha$ and for the same set of reference configurations $\mathcal{K}$.*

The violation of the Peierls condition for a given set of reference configurations may be an indication that we are forgetting some other important configurations [45]. Indeed, if for instance a ground-state configuration has been left out, plaquettes so configured will not raise the energy and will lead to arbitrarily thick contours with little energy cost. In some sense, this observation determines a criterion to decide the set of reference configurations: choose exactly what you need to satisfy the Peierls condition. This is the criterion referred to at the end of the paragraph "Aspect 3" above. It is not a constructive criterion, and there may be situations in which it can not be fulfilled at all, but it is the best we've got.

4.3 The general scenario

4.3.1 Procedure to fall into the cluster-expansion framework

As illustrated extensively in Chapter 3, contour arguments are decided at the level of partition functions. Let us consider an interaction satisfying the Peierls condition. With the decomposition (4.17), we can write $Z_p(\Lambda)$ [defined in (4.14)] in the form

$$Z_p(\Lambda) \;=\; e^{-\beta e(\eta_p)|\Lambda|} \sum_{\substack{\{\gamma_k\}\subset\widehat{\Lambda} \\ \text{admissible}}} \Big[\prod_k w(\gamma_k)\Big]\Big[\prod_{u=1}^{P} e^{-\beta\,[e(\eta_u)-e(\eta_p)]\,|I_u|}\Big]\,, \tag{4.21}$$

where the exterior contours of each compatible family are p-contours. The contour weights,

$$w(\gamma) \;=\; e^{-\beta E(\gamma)}\,, \tag{4.22}$$

are parameterized by β (omitted from the notation for simplicity) and satisfy the Peierls bound

$$|w(\gamma)| \;\le\; e^{-\beta J}\,. \tag{4.23}$$

This bound encourages us to apply the cluster-expansion technology of Section 3.3, except that we now have a problem: The requirement of admissibility is a highly nonlocal condition (two far away contours can be rendered inadmissible by a mismatch of labels). This puts

the expression out of reach of cluster-expansion methods, which only work under very local conditions, like volume exclusion. In particular, we do not have a simple "contour-removal operation" granting that, once a contour is taken out, the remaining contours correspond to another configuration. There exists, however, a procedure, introduced by Minlos and Sinai [31, 32], to eliminate the inconvenient admissibility condition in (4.21). First, (4.21) is resummed (all sums here are finite!) over the contours in the interior of the exterior contours,

$$Z_p(\Lambda) \;=\; e^{-\beta e(\eta_p)|\Lambda|} \sum_{\substack{\{\gamma_k^p\}\subset\widehat{V}\\ \text{exterior}\\ \text{non}-\text{intersecting}}} \prod_k \Big[w(\gamma_k^p) \prod_{u=1}^{P} Z_u\Big(\mathrm{Int}_u\,\gamma_k^p\Big) e^{\beta e(\eta_p)\left|\mathrm{Int}_u\,\gamma_k^p\right|}\Big] \,. \tag{4.24}$$

Then, in the right-hand side the different partition functions are multiplied and divided by $Z_p\Big(\mathrm{Int}_u\,\gamma_k^p\Big)$ to obtain

$$Z_p(\Lambda) e^{\beta e(\eta_p)|\Lambda|} \;=\; \sum_{\substack{\{\gamma_k^p\}\subset\widehat{V}\\ \text{exterior}\\ \text{non}-\text{intersecting}}} \prod_k \Big[W(\gamma_k^p) \prod_{u=1}^{P} Z_p\Big(\mathrm{Int}_u\,\gamma_k^p\Big) e^{\beta e(\eta_p)\left|\mathrm{Int}_u\,\gamma_k^p\right|}\Big] \,. \tag{4.25}$$

with the new weights

$$W(\gamma^p) \;:=\; w(\gamma^p) \prod_{u=1}^{P} \frac{Z_u\Big(\mathrm{Int}_u\,\gamma^p\Big)}{Z_p\Big(\mathrm{Int}_u\,\gamma^p\Big)} \,. \tag{4.26}$$

One can now repeat the same procedure for each factor,

$$Z_u\Big(\mathrm{Int}_u\,\gamma_k^p\Big) e^{\beta e(\eta_p)\left|\mathrm{Int}_u\,\gamma_k^p\right|} \tag{4.27}$$

in (4.25). This iteration finally yields the expression

$$Z_p(\Lambda) \;=\; e^{-\beta e(\eta_p)|\Lambda|} \sum_{\substack{\{\gamma_k^p\}\subset\widehat{V}\\ \text{non}-\text{intersecting}}} \prod_k W(\gamma_k^p) \tag{4.28}$$

$$:=\; e^{-\beta e(\eta_p)|\Lambda|}\, \Theta_\Lambda^p \,. \tag{4.29}$$

As we've gone through external contours, the final condition only involves non-intersection.

Let us pause to analyze what has been done. The partition function has finally been written in terms of non-intersecting contours, but at the cost of introducing an *artificial* contour ensemble: All contours in (4.28) are p-contours. The contour families obtained do not correspond, in general, to any spin configuration. In other words, in the absence of a clean contour-removal operation the procedure advises us to sacrifice the one-one correspondence with configurations and to adopt, instead, an unphysical ensemble of contours. Note that, nevertheless, the *exterior* contours are the same as for the original physical spin system, because the resummation procedure started from the physical exterior contours down. Let me summarize the procedure so far:

- P different contour ensembles are introduced, one for each reference configuration. The p-contour ensemble $\mathcal{C}^p$—corresponding to the reference configuration η_p— is formed by all families of non-intersecting p-contours with weights (4.26).

- For each finite $\Lambda \subset \mathbb{Z}^d$, the ensemble $\mathcal{C}^p_\Lambda$ has the same distribution of *external* contours as the measure $\langle \cdot \rangle^{(\beta)\,\eta_p}$. Therefore, *if* the cluster-expansion technology is applicable, it will yield information on probabilities of events determined by exterior contours. This, however, is enough to conclude about the existence and mixing properties of the η_p-phase of the spin system.

There is a big "if" in the previous statement: At this point we do not know whether cluster expansions will converge for these artificial contour ensembles. Whereas the original weights $w(\gamma^p)$ satisfied the Peierls condition (4.23), there is no a priori bound on the new weights $W(\gamma^p)$, defined in (4.26). This is an *essential* obstruction and it arises for a good reason. As we shall see, the convergence of the cluster expansion is associated to a stable η_p-phase.

4.3.2 Stable boundary conditions

To motivate the definitions to come, let us reflect on what is expected. In general, we expect that some of the configurations η_p gives rise to stable phases and others do not. Moreover, the stable ones should exhibit the sea-with-island picture. If so, for η_p stable we expect that the

new weights W of the ensemble $\mathcal{C}^p$ be exponentially decreasing with $|\omega^p|$. On the other hand, if η_q is not stable we expect that whenever we force the η_q external conditions for an arbitrarily large Λ, typically there will be a large contour, close to the boundary of Λ, where the system flips into one of the stable η_p's. Therefore, we expect that the new weights of $\mathcal{C}^q$ favor *large* contours.

In other words, we expect that for stable η_p the additional factors $Z_u(\Delta)/Z_p(\Delta)$, contributing to the new weights, either decrease or at least do not grow much with $|\Delta|$, while for η_q not stable these factors are expected to grow exponentially fast with $|\Delta|$ to beat the Peierls bound of the original weight w. Therefore, stability can be detected by imposing a suitable cutoff over the growth of these quotients of partition functions that can not possibly be satisfied by non-stable η_q's. A convenient form of this cutoff has been proposed by Zahradník [44] (I use the constants adopted by Borgs and Imbrie [3]).

Definition 4.4

(i) A region $\Lambda \subset \mathbb{Z}^d$ is p-stable if $Z_p(\Lambda) \neq 0$ and

$$\left| \frac{Z_u(\Lambda)}{Z_p(\Lambda)} \right| \leq \exp[4\,|\partial\Lambda|] \tag{4.30}$$

is satisfied for all u.

(ii) A p-contour γ^p is stable if each $\mathrm{Int}_u(\gamma^p)$ is p-stable, for $1 \leq u \leq P$.

For β large enough, the weights $W(\gamma^p)$ of *stable* contours γ^p satisfy the crucial condition (3.40) and we can apply the cluster expansion technology of Section 3.3. Hence we can define the *truncated contour partition functions* [44]:

$$\begin{aligned} Z'_p(\Lambda) &:= e^{-\beta e(\eta_p)|\Lambda|} \sum_{\substack{\{\gamma_k^p\} \subset \widehat{V} \\ \text{stable} \\ \text{non-intersecting}}} \prod_k W(\gamma_k^p) \\ &:= e^{-\beta e(\eta_p)|\Lambda|}\, \Theta'^p_\Lambda \,. \end{aligned} \tag{4.31}$$

By Corollary 3.4 we have that for β large the *truncated contour free energies*

$$f'_p \;:=\; -\lim_{\Lambda\to\mathbb{Z}^d} \frac{1}{|\Lambda|} \log Z'_p(\Lambda) \tag{4.32}$$

exist, and are of the form

$$f'_p \;=\; \beta\, e(\eta_p) + f^{\mathcal{C}'_p}\,, \tag{4.33}$$

where $f^{\mathcal{C}'_p}$ is given by the cluster expansion (3.57), with w replaced by the new weights W and only accepting stable contours. Physicists would like a factor $1/\beta$ in the right-hand side of (4.32)–(4.33). For the expressions in the proof of the main theorem (Chapter 5) such factor would be a bit of a nuisance, so I omit it. In most references on Pirogov-Sinai, the inverse temperature is absorbed as part of the interaction, what amounts to set $\beta = 1$. I prefer to exhibit β explicitly and to juggle between physical tradition and notational convenience.

The key observation of Zahradník [44], is that these truncated free energies f'_p are the objects promised at the end of Chapter 2: Its minimization decides which phases are stable.

4.4 The main result

4.4.1 Criterion for the stability of a phase

Let

$$a_p(\beta) \;:=\; \mathrm{Re}\, f'_p(\beta) - \min_{\eta_u\in\mathcal{K}} \mathrm{Re}\, f'_u(\beta)\,. \tag{4.34}$$

Then the main result of Pirogov-Sinai theory, in the formulation due to Zahradník [44] is the following

Theorem 4.5 *Consider a finite-range complex-valued interaction satisfying the Peierls condition with respect to a finite number of periodic reference configurations $\{\eta_1,\dots,\eta_P\}$, and consider β large enough. Then, $a_p(\beta) = 0$ implies that for such β the η_p-phase is stable. More explicitly:*

(i) The truncated free-energy density f'_p coincides with β times the actual free-energy density.

(ii)

$$\lim_{\Lambda\to\mathbb{Z}^d} \langle g\rangle_\Lambda^{(\beta)\,\eta_p} \;=:\; \langle g\rangle^{(\beta)\,\eta_p} \tag{4.35}$$

exists for each local observable g. Moreover

$$\left|\langle g\rangle^{(\beta)\,\eta_p} - g(\eta_p)\right| \;\le\; \|g\|_\infty\, |B|\; \mathrm{O}(e^{-\beta\widetilde{J}}) \tag{4.36}$$

for some constant $\widetilde{J} > 0$.

(iii) The η_p-phase is exponentially mixing: If $g_1(\omega) = g_1(\omega_{A_1})$ and $g_2(\omega) = g_2(\omega_{A_2})$:

$$\begin{aligned}&\left|\langle g_1\, g_2\rangle^{(\beta)\,\eta_p} - \langle g_1\rangle^{(\beta)\,\eta_p}\langle g_2\rangle^{(\beta)\,\eta_p}\right| \\ &\qquad\le\; \|g_1\|_\infty\, \|g_2\|_\infty\, |A_1|\; |A_2|\; \mathrm{O}(e^{-\mathrm{dist}(A_1,A_2)/\xi(\beta)})\;,\end{aligned} \tag{4.37}$$

for some $\xi(\beta) > 0$.

(iv) If the interaction is real valued, then the sea-with-island picture is valid for $\langle g\rangle^{(\beta)\,\eta_p}$:

- *Infinite sequences of nested contours have probability zero.*
- *Defects and other reference configurations do not percolate.*
- *The probability of occurrence of contours around a given site is $\mathrm{O}(e^{-\beta\widetilde{J}})$.*

4.4.2 Application: Stability of zero-temperature phases

In the large β limit the lowest-order terms dominate the truncated contour free-energy densities f'_p. These are terms involving clusters formed by a single contour with the lowest possible energy cost. Such contours are always stable because they have empty interior. If at that order some f'_q are already larger than the others, the corresponding boundary conditions can be disregarded as unstable, and the analysis proceeds with clusters with higher-order contributions. Such a scheme was in fact systematized by Slawny [39] on the basis of the original Pirogov-Sinai theory.

As an illustration, let me formalize the discussion on the stability of the rigid configurations of the Blume-Capel model (2.44). The excitations (2.45)–(2.47) define precisely the contours with the lowest-possible energy cost, thus

$$f'_0 = -2e^{-2d\beta} + \mathrm{O}(e^{-(4d-2)\beta}) \tag{4.38}$$

$$f'_+ = f'_- = -e^{-2d\beta} + \mathrm{O}(e^{-4d\beta}) \; . \tag{4.39}$$

We see that $f'_0 < f'_+, f'_-$ for β large enough, which proves that the all-"0" configuration is stable. The discussion of Section 2.3.3 was perfectly accurate, the only step missing was the identification of the truncated f'_p as the "free-energy densities" vaguely introduced there. Rigorously speaking, nowhere in this notes the unstability of the other two configurations is proven. Such a result is the object of the studies of "completeness of the phase diagram" [44], not discussed here.

On the other hand, the addition of the "g"- term of (2.48) changes the truncated free energies to

$$f'_0 = -2e^{-(2d-g)\beta} + \mathrm{O}(e^{-[(4d-2)-2g]\beta}) \tag{4.40}$$

$$f'_+ = f'_- = -\beta g - e^{-(2d+g)\beta} + \mathrm{O}(e^{-4d\beta}) \; . \tag{4.41}$$

Hence, for g small enough, the value $\beta^{(3)}(g)$ where the three phases are stable is the solution of

$$-\beta g - e^{-(2d+g)\beta} + 2e^{-(2d-g)\beta} = \mathrm{O}(e^{-[(4d-2)-2g]\beta}) \; . \tag{4.42}$$

I leave to the reader the analogous verification that for the Fisher's antiferromagnet (4.3) the ferromagnetic phase [η_+ in (4.4)] is stable, and the determination of the tuning of the field [coefficient of the last term in (4.4)] that keeps the three phases stable at very low temperatures.

Chapter 5

Proof of the main result

In this chapter I closely follow the excellent presentation of Borgs and Imbrie [3]. The proof is just a transcription of the proof of Theorem 3.1 in [3], except that I do not aim at the detailed description provided by their formulas (3.10)–(3.12). Hence I can allow myself some small simplifications.

5.1 The key lemma

Given the discussion at the end of Section 4.3.2 and the bounty offered by the cluster-expansion technology (see summary in Section 3.3.6), we conclude that Theorem 4.5 is a consequence of the following key lemma:

Lemma 5.1 *The following statements are equivalent:*

(i) $a_p = 0$.

(ii) All regions Λ *are p-stable.*

Proof. We first prove that (ii)$\Longrightarrow$(i), assuming that (i)$\Longrightarrow$(ii) holds. For this purpose, we consider a boundary condition η_v for which $a_v = 0$. For each Λ, we have that $Z_p(\Lambda) = Z'_p(\Lambda)$, by assumption, and $Z_v(\Lambda) = Z'_v(\Lambda)$ holds because (i)$\Longrightarrow$(ii). Therefore

$$\exp\Big[4\ |\partial\Lambda|\Big]\ \geq\ \left|\frac{Z_v(\Lambda)}{Z_p(\Lambda)}\right|\ =\ \exp\Big[a_p\ |\Lambda| + \mathrm{O}(e^{-\beta\widetilde{J}})\,|\partial\Lambda|\Big]\,. \qquad (5.1)$$

The leftmost inequality expresses p-stability of the region Λ, while the equality on the right is a consequence of the bound (3.58) on finite-volume effects for the contour ensemble. If $a_p > 0$ the equation in (5.1) leads to a contradiction for regions Λ with diverging volume-to-surface-area ratio. Thus we conclude that $a_p = 0$.

Proof of (i)$\Longrightarrow$(ii). In order to understand the steps and definitions that follow, it is useful to inspect the ratio of partition functions corresponding to different boundary conditions. From (3.58) we have that, for any $\eta_v, \eta_q \in \mathcal{K}$,

$$\left|\frac{Z'_v(\Lambda)}{Z'_p(\Lambda)}\right| \;=\; \exp\Big[-(a_v - a_p)\;|\Lambda| + \mathrm{O}(e^{-\beta\widetilde{J}})\,|\partial\Lambda|\Big]\;. \tag{5.2}$$

From (5.2) we conclude that if $a_p = 0$ then

$$\left|\frac{Z'_v(\Lambda)}{Z'_p(\Lambda)}\right| \;\leq\; \exp(|\partial\Lambda|), \tag{5.3}$$

for large β. Hence, in this case, the proof would be complete if the truncated partition functions in (5.3) could be replaced by the untruncated ones and

$$\left|\frac{Z_v(\Lambda)}{Z_p(\Lambda)}\right| \;\leq\; \exp(\mathrm{const}|\partial\Lambda|). \tag{5.4}$$

More generally, for regions Λ for which

$$a_p\,|\Lambda| \;\leq\; |\partial\Lambda|\;, \tag{5.5}$$

we have from (5.2) that, for large β,

$$\left|\frac{Z'_v(\Lambda)}{Z'_p(\Lambda)}\right| \;\leq\; \exp(2\,|\partial\Lambda|). \tag{5.6}$$

As a first step, we would like to show that the primes in (5.3) and (5.6) can be removed for regions satisfying (5.5). If condition (5.5) were inherited by subregions of Λ then we could prove inductively, from (5.6), that $Z_p(\Lambda) = Z'_p(\Lambda)$. However, it is not true that the bound (5.5) remains valid for arbitrary subregions of Λ. Therefore it is convenient to resort to a sufficient condition that has this hereditarity feature. For this purpose, we introduce the notion of *small regions* and *small contours*, adopting the definitions of [3].

Definition 5.2

(i) A region Λ is q-small if

$$a_q \operatorname{diam} \widehat{\Lambda} \;\leq\; 1 \,. \tag{5.7}$$

[where $\widehat{\Lambda}$ is defined in (4.19)].

(ii) A contour γ is q-small if

$$a_q \operatorname{diam} \operatorname{supp} \gamma \;\leq\; 1 \,; \tag{5.8}$$

otherwise the contour is called q-large.

It is clear that smallness is inherited by subregions. Moreover, the bound (5.5) is valid for q-small regions, because

$$\begin{aligned} a_q \; |\Lambda| \;&\leq\; a_q \operatorname{diam} \Lambda \, |\partial\Lambda| \\ &\leq\; |\partial\Lambda| \,. \end{aligned} \tag{5.9}$$

In particular, all contours inside a q-small region are q-small contours. [It is for the sake of this property that we used $\widehat{\Lambda}$ in (5.7)]. The hypothesis that $a_p = 0$ implies that *all* regions are p-small. As a consequence, the proof of the implication (i)$\Longrightarrow$(ii) is completed by proving the following lemma [44].

5.2 The inductive proof

Lemma 5.3 *For all q, q-small regions are q-stable. As a consequence, all q-contours contained in q-small regions are stable.*

Proof. The proof is by induction on the diameters of the regions.

Let us assume that, *for all* u, $a_u \operatorname{diam} \Lambda \leq 1$ implies that $Z_u(\widetilde{\Lambda}) \neq 0$ and

$$\left| \frac{Z_v(\widetilde{\Lambda})}{Z_u(\widetilde{\Lambda})} \right| \;\leq\; \exp(4|\partial\Lambda|), \tag{5.10}$$

for all v and for all regions $\widetilde{\Lambda}$, contained in Λ, with diameter less than or equal to m. We pick some $\eta_q \in \mathcal{K}$ and some q-small region $\widehat{\Lambda}$ of

diameter $m+1$, and prove the bound (5.10), with $u = q$. All contours γ^q in this region are q-small, hence their interiors are q-small and of diameter strictly smaller than $m+1$. By the inductive hypothesis such interiors satisfy (5.10), and hence these contours are stable, yielding

$$Z'_q(\Lambda) \;=\; Z_q(\Lambda)\;. \tag{5.11}$$

We remark that if $a_v = 0$ then the proof is complete. This is because if $a_v = 0$ *all* regions are v-small, and, consequently, all v-contours in $\widehat{\Lambda}$ are stable. This implies that $Z'_v(\Lambda) \;=\; Z_v(\Lambda)$, which, along with (5.11) and (5.6), implies that (4.30) is true.

Let us now consider a boundary condition η_v for which $a_v \neq 0$. To estimate $Z_v(\Lambda)/Z_q(\Lambda)$ we write the numerator as in (4.24) and resum the contribution of *v-small exterior contours*. This yields

$$\frac{Z_v(\Lambda)}{Z_q(\Lambda)} \;=\; \sum_{\substack{\{\gamma_k^v\}\subset\widehat{\Lambda}\\ v-\text{large}\\ \text{exterior}}} \frac{Z_v^{\text{small}}(\text{Ext})}{Z_q(\Lambda)} \prod_k w(\gamma_k^v)\, e^{-\beta e(\eta_v)|\gamma_k^v|}\; Z\Big(\text{Int}\,\gamma_k^v\Big)\;. \tag{5.12}$$

Here "Ext" is the region outside the v-large exterior contours $\{\gamma_k^v\}$, the label "small" indicates a restriction to configurations where all the exterior contours are v-small, and $Z(\text{Int}\,\gamma^v) := \prod_{\tilde{v}} Z_{\tilde{v}}(\text{Int}_{\tilde{v}}(\gamma^v))$. If we multiply and divide the right-hand side of (5.12) by

$$Z_q(\text{Int}) \;:=\; \prod_k \prod_{\tilde{v}} Z_q(\text{Int}_{\tilde{v}}(\gamma_k^v))\;, \tag{5.13}$$

we obtain

$$\frac{Z_v(\Lambda)}{Z_q(\Lambda)} \;=\; \sum_{\substack{\{\gamma_k^v\}\subset\widehat{\Lambda}\\ v-\text{large}\\ \text{exterior}}} \frac{Z_v^{\text{small}}(\text{Ext})\, Z_q(\text{Int})\, e^{-\beta e(\eta_v)|\gamma_k^v|}}{Z_q(\Lambda)} \prod_k Y(\gamma_k^v)\;, \tag{5.14}$$

with

$$Y(\gamma_k^v) \;:=\; w(\gamma_k^v) \prod_{\tilde{v}} \frac{Z_{\tilde{v}}(\text{Int}_{\tilde{v}}\,\gamma_k^v)}{Z_q(\text{Int}_{\tilde{v}}\,\gamma_k^v)}\;. \tag{5.15}$$

We observe that, by the inductive hypothesis,

$$Z_v'^{\,\text{small}}(\text{Ext}) \;=\; Z_v^{\text{small}}(\text{Ext}) \tag{5.16}$$

Identities (5.11) and (5.16) allow us to apply the finite-volume bound (3.58) to all the factors in (5.14), except $\prod_k Y(\gamma_k^v)$. We then obtain

$$\begin{aligned}\left|\frac{Z_v^{\text{small}}(\text{Ext})\, Z_q(\text{Int})\, e^{-\beta e(\eta_v)\sum_k|\gamma_k^v|}}{Z_q(\Lambda)}\right| \\ \leq \quad \exp\Big[-\text{Re}\,(f_v'^{\text{small}} - f_q')\,|\Lambda\setminus\text{Int}\,| \;+\; 2\,|\partial\Lambda|\Big]\,\prod_k e^{(2d+1)|\gamma_k^v|}\;.\end{aligned} \tag{5.17}$$

We have used the geometrical bound $|\partial\,\text{Ext}\,|+|\partial\,\text{Int}\,| \leq |\partial\Lambda|+2d\sum_k|\gamma_k^v|$. We now resort to the q-smallness of Λ, inequality (5.9), to bound

$$\begin{aligned}-\text{Re}\,(f_v'^{\text{small}} - f_q')\,|\Lambda\setminus\text{Int}\,| \;&=\; (-a_v^{\text{small}} + a_q)\,|\Lambda\setminus\text{Int}\,| \\ &\leq\; -a_v^{\text{small}}\,|\Lambda\setminus\text{Int}\,| + |\partial\Lambda|\;.\end{aligned} \tag{5.18}$$

Furthermore, the Peierls condition (4.23) and the inductive hypothesis (5.10) for $u = q$ (combined with the bound $|\partial\,\text{Int}\,\gamma_k^v| \leq 2d|\gamma_k^v|$) imply that

$$|Y(\gamma_k^v)| \;\leq\; e^{-\beta J|\gamma_k^v|}\,e^{8d|\gamma_k^v|}\;. \tag{5.19}$$

Substituting (5.17), (5.18) and (5.19) in (5.14), we get the bound

$$\begin{aligned}\left|\frac{Z_v(\Lambda)}{Z_q(\Lambda)}\right| \;&\leq\; e^{3|\partial\Lambda|}\sum_{\substack{\{\gamma_k^v\}\subset\widehat{\Lambda}\\ v-\text{large}\\ \text{exterior}}} e^{-a_v^{\text{small}}|\Lambda\setminus\text{Int}\,|}\,\prod_k e^{-\beta J|\gamma_k^v|}\,e^{(10d+1)|\gamma_k^v|} \\ &=:\; e^{3|\partial\Lambda|}\sum_{\substack{\{\gamma_k^v\}\subset\widehat{\Lambda}\\ v-\text{large}\\ \text{exterior}}} e^{-a_v^{\text{small}}|\Lambda\setminus\text{Int}\,|}\,\prod_k w^*(\gamma_k^v)\;.\end{aligned} \tag{5.20}$$

To show that $e^{4|\partial\Lambda|}$ is an upper bound for (5.20), and hence complete the proof of the lemma, it is convenient to follow [44] and consider the quantity

$$\widetilde{Z}_v^{\text{large}}(\Lambda) \;:=\; \sum_{\substack{\{\gamma_k^v\}\subset\widehat{\Lambda}\\ v-\text{large}\\ \text{non}-\text{intersecting}}}\;\prod_k w^*(\gamma_k)\,e^{2d|\gamma_k|}\;. \tag{5.21}$$

[Note that $w^*(\gamma) \geq 0$.] This quantity can be interpreted as the partition function of an ensemble of contours having weights

$$\widetilde{w}(\gamma) \;:=\; w^*(\gamma)\, e^{2d|\gamma|} \tag{5.22}$$

and confined to a volume $\widehat{\Lambda}$. It is evident that, for β large enough, the contour weights $\widetilde{w}(\gamma)$ satisfy condition (3.40), and hence the cluster expansion converges. Moreover, if $\widetilde{f}_v^{\,\mathrm{large}}$ is the corresponding contour free-energy density, it follows from (3.58) that

$$[\widetilde{Z}_v^{\mathrm{large}}(\Lambda)]^{-1} \;\leq\; e^{\widetilde{f}_v^{\mathrm{large}}|\Lambda|}\, \exp\Big[\mathrm{O}(e^{-\beta\widetilde{J}})\; |\partial\Lambda|\Big]\;, \tag{5.23}$$

We claim that

$$a_v^{\mathrm{small}} \;\geq\; -\widetilde{f}_v^{\mathrm{large}}\;. \tag{5.24}$$

Indeed, for every v-large contour

$$a_v \operatorname{diam}\gamma \;>\; 1\;. \tag{5.25}$$

Hence, by (3.57)

$$a_v^{\mathrm{small}} \;=\; a_v + \mathrm{O}\Big(e^{-\beta\widetilde{J}/a_v}\Big)\;. \tag{5.26}$$

By the same argument,

$$\widetilde{f}_v^{\mathrm{large}} \;=\; \mathrm{O}\Big(e^{-\beta\widetilde{J}/a_v}\Big)\;. \tag{5.27}$$

Hence

$$a_v^{\mathrm{small}} + \widetilde{f}_v^{\mathrm{large}} \;\geq\; a_v + \mathrm{O}\Big(e^{-\beta\widetilde{J}/a_v}\Big) \tag{5.28}$$

which is non-negative for β large and λ small, proving (5.24).

The rest of the argument can be presented as a lemma of independent interest. The lemma shows that (5.24) causes the sum in (5.20) to yield at most a contribution exponential in the boundary. By substitution of the bound (5.30), shown below, into the right-hand side of (5.20), we obtain the bound (5.10). This completes the inductive proof. ∎

5.3 The final lemma

Lemma 5.4 *Consider weights $w^*(\gamma)$ satisfying a Peierls bound*

$$0 \;\le\; w^*(\gamma) \;\le\; e^{-\beta J} \tag{5.29}$$

and let $\widetilde{f}$ denote the contour free-energy density for the weights $\widetilde{w}(\gamma) = w^*(\gamma)\, e^{2d|\gamma|}$ (well defined if β is small enough). Then, for $g \ge -\widetilde{f}$,

$$\sum_{\substack{\{\gamma_k\}\subset\Lambda \\ \text{exterior}}} e^{-g|\Lambda\setminus \mathrm{Int}\,|} \prod_k w^*(\gamma_k) \;\le\; \exp\Big[\mathrm{O}(e^{-\beta\widetilde{J}})|\partial\Lambda|\Big] \,. \tag{5.30}$$

Proof. (This is Lemma 3.2 of [3], and the proof given there.) Multiply and divide the left-hand side of (5.30) by $\widetilde{Z}_v(\mathrm{Int})$. Using the analogue of (5.23) for the region $\mathrm{Int} := \bigcup_{\tilde{v},k} \mathrm{Int}_{\tilde{v}}\, \gamma_k$, and the bound $|\partial\, \mathrm{Int}\,| \le 2d \sum_k |\gamma_k|$ we obtain

$$[\widetilde{Z}(\mathrm{Int})]^{-1} \;\le\; e^{\widetilde{f}|\,\mathrm{Int}\,|} \prod_k e^{2d\sum_k |\gamma_k|} \tag{5.31}$$

for β small enough. Thus the left-hand side of (5.30) satisfies

$$\mathrm{LHS} \;\le\; \sum_{\substack{\{\gamma_k\}\subset\widehat{\Lambda} \\ \text{exterior}}} e^{-g|\Lambda\setminus\mathrm{Int}\,|}\, e^{\widetilde{f}|\,\mathrm{Int}\,|}\, \widetilde{Z}(\mathrm{Int}) \prod_k \widetilde{w}(\gamma_k) \,, \tag{5.32}$$

and, since $-g \le \widetilde{f}$, we have that

$$\begin{aligned} \mathrm{LHS} \;&\le\; e^{\tilde{f}|\Lambda|} \sum_{\substack{\{\gamma_k\}\subset\widehat{\Lambda} \\ \text{exterior}}} \widetilde{Z}(\mathrm{Int}) \prod_k \widetilde{w}(\gamma_k) \\ &\le\; e^{\tilde{f}|\Lambda|}\, \widetilde{Z}(\Lambda) \,. \end{aligned} \tag{5.33}$$

Hence, by (3.58) for the region Λ,

$$\mathrm{LHS} \;\le\; \exp\Big[\mathrm{O}(e^{-\beta\widetilde{J}})\; |\partial\Lambda|\Big] \,. \;\blacksquare \tag{5.34}$$

5.4 Last remarks

Expectations and probabilities

Let me briefly comment how Lemma 5.1 and the resummation procedure of Section 4.3.1 are used to show that expectations of the form $\langle g \rangle_\Lambda^{(\beta)\,\eta_p}$ (g local) admit, if $a_p = 0$ and β is large enough, an absolutely convergent cluster expansion with a well defined infinite-volume limit. If g is a local observable, for instance $g(\omega) = g(\omega_A)$, then

$$g(\omega) \;=\; g(\Gamma_A(\omega))\;, \tag{5.35}$$

where $\Gamma_A(\omega)$ represents all the contours of $\Gamma(\omega)$ whose support intersects A. Such family has a well defined exterior configuration, say η_q, that I will make explicit by writing Γ_A^q. It may happen that $\Gamma_A = \emptyset$, in which case A must be contained in the set C_u of u-correct sites for some $1 \le u \le P$. Therefore, we have

$$\begin{aligned}\langle g \rangle_\Lambda^{(\beta)\,\eta_p} \;=\; & \sum_{q=1}^{p} \Bigg\{ g(\eta_q)\, \exp\Big(\sum_{B \subset A} \Phi_B(\eta_q) \Big)\, \frac{Z_{p,q}(\Lambda \setminus A)}{Z_p(\Lambda)} \\ & + \sum_{\Gamma_A^q \text{ admissible}} \widetilde{g}(\Gamma_A^q) \Big[\prod_{\gamma \in \Gamma_A^q} w(\gamma) \Big] \frac{Z_{p,q}(\Lambda \setminus \operatorname{supp} \Gamma_A)}{Z_p(\Lambda)} \Bigg\}\;.\end{aligned} \tag{5.36}$$

Here $Z_{p,q}(\Lambda \setminus D)$ is the partition function in $\Lambda \setminus D$ with boundary conditions η_p in the exterior of Λ and η_q in the interior of D (D is assumed to be strictly contained in Λ); $\operatorname{supp} \Gamma_A$ is the union of the supports of the contours of the admissible family Γ_A, and $\widetilde{g}$ is just g expressed as function of contour configurations.

Via the resummation described in Section 4.3.1 all the partition functions involved can be written purely in terms of non-intersecting p-contours with new weights W [each term $Z_{p,q}(\Lambda \setminus D)$ with $p \neq q$ must have at least one contour surrounding D]. Whenever $a_p = 0$ Lemma 5.1 and the stability condition (4.30) imply that the new weights obey a Peierls bound and the different cluster expansions converge absolutely and uniformly in Λ. This guarantees the existence of the $\Lambda \to \mathbb{Z}^d$ limit. This expansion yields the other stability properties listed in Theorem 4.5. I leave the details to the reader.

On the definition of stable regions

The proof presented in this chapter clearly shows that there is considerable freedom in the definition of stable regions. On the one hand the constant "4" in (4.30) can be replaced by any other constant. Moreover, instead of brutally suppressing contours with large new weights W, one could devise some sort of smooth counter-weights to scale the weights down to convergence levels. In this manner one could, for instance, construct alternative truncated free-energy densities with certain desirable smoothness properties respect to some of the parameters of the interaction. Such possibility has indeed been exploited and put to good use [4].

Bibliography

[1] M. Aizenman. Translation invariance and instability of phase coexistence in the two dimensional Ising system. *Commun. Math. Phys.*, 73:83–94, 1980.

[2] M. Aizenman, J. Bricmont, and J. L. Lebowitz. Percolation of the minority spins in high-dimensional Ising models. *J. Stat. Phys.*, 49:859–865, 1987.

[3] C. Borgs and J. Z. Imbrie. A unified approach to phase diagrams in field theory and statistical mechanics. *Commun. in Math. Phys.*, 123:305–328, 1989.

[4] C. Borgs and R. Kotecký. A rigorous theory of finite-size scaling at first-order phase transitions. *J. Stat. Phys.*, 61:79–119, 1990.

[5] C. Borgs, R. Kotecký, and D. Ueltschi. Low temperature phase diagrams for quantum perturbations of classical spin systems. *Commun. Math. Phys.*, 181:409–46, 1996.

[6] D. C. Brydges. A short cluster in cluster expansions. In *Critical Phenomena, Random Systems, Gauge Theories*, pages 129–183. Elsevier, 1984.

[7] C. Cammarota. Decay of correlations for infintie range interactions in unbounded spin systems. *Commun. Math. Phys.*, 85:517–528, 1982.

[8] N. Datta, R. Fernández, and J. Fröhlich. Low-temperature phase diagrams of quantum lattice systems. I. Stability for quantum perturbations of classical systems with finitely-many ground-states. *J. Stat. Phys.*, 84:455–534, 1996.

[9] R. Dobrushin. Estimates of semiinvariants for the Ising model at low temperatures, 1994. Preprint 125, ESI, Wien.

[10] R. Dobrushin. Perturbation methods of the theory of Gibbsian fields. In *Ecole d'Eté de Probabilités de Saint-Flour XXIV – 1994*, pages 1–66. Springer-Verlag (Lecture Notes in Mathematics #1648), Berlin–Heidelberg–New York, 1996.

[11] R. L. Dobrushin. Existence of a phase transition in the two-dimensional and three-dimensional Ising models. *Soviet Phys. Doklady*, 10:111–113, 1965.

[12] R. L. Dobrushin. Gibbs states describing coexistence of phases for a three-dimensional Ising model. *Theor. Prob. Appl.*, 17:582–600, 1972.

[13] R. L. Dobrushin, J. Kolafa, and S. B. Shlosman. Phase diagram of the two-dimensional Ising antiferromagnet (computer-assisted proof). *Commun. Math. Phys.*, 102:89–103, 1985.

[14] R. L. Dobrushin and S. B. Shlosman. The problem of translation invariance of Gibbs states at low temperatures. *Sov. Sci. Rev. C Math. Phys.*, 5:53–196, 1985.

[15] R. Fernández, J. Fröhlich, and A. D. Sokal. *Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory.* Springer-Verlag, Berlin–Heidelberg–New York, 1992.

[16] M. E. Fisher. Scaling, universality and renormalization group theory. In F. J. W. Hahne, editor, *Critical Phenomena (Stellenbosch 1982)*, pages 1–139. Springer-Verlag (Lecture Notes in Physics #186), Berlin–Heidelberg–New York, 1983.

[17] J. Fröhlich and T. Spencer. The Berezinskii-Kosterlitz-Thouless transition (energy-entropy arguments and renormalization in defect gases). In J. Fröhlich, editor, *Scaling and Self-Similarity in Physics.* Birkhäuser, Boston–Basel–Stuttgart, 1983.

[18] H.-O. Georgii. *Gibbs Measures and Phase Transitions.* Walter de Gruyter (de Gruyter Studies in Mathematics, Vol. 9), Berlin–New York, 1988.

[19] J. Glimm and A. Jaffe. *Quantum Physics: A Functional Integral Point of View*. Springer-Verlag, New York, 1981 (2^{nd} edition 1987).

[20] J. Glimm, A. Jaffe, and T. Spencer. A convergent expansion about mean field theory. I. *Ann. Phys. (NY)*, 101:610–630, 1976.

[21] R. B. Griffiths. Peierls' proof of spontaneous magnetization of a two-dimensional Ising ferromagnet. *Phys. Rev.*, A136:437–439, 1964.

[22] R. B. Griffiths. Rigorous results and theorems. In C. Domb and M. S. Green, editors, *Phase Transitions and Critical Phenomena (Vol. 1)*. Academic Press, London–New York–San Francisco, 1972.

[23] C. Gruber and H. Kunz. General properties of polymer systems. *Commun. Math. Phys.*, 22:133–61, 1971.

[24] Y. Higuchi. On the absence of non-translationally invariant Gibbs states for the two-dimensional Ising system. In J. Fritz, J. L. Lebowitz, and D. Szasz, editors, *Random Fields: Rigorous Results in Statistical Mechanics and Quantum Field Theory (Esztergom 1979)*. North-Holland, Amsterdam, 1981.

[25] W. Holsztynski and J. Slawny. Peierls condition and the number of ground states. *Commun. Math. Phys.*, 61:177–190, 1978.

[26] R. B. Israel. *Convexity in the Theory of Lattice Gases*. Princeton Univ. Press, Princeton, N.J., 1979.

[27] R. Kotecký. Geometric representation of lattice models and large volume asymptotics. In G. R. Grimmett, editor, *Probability and Phase Transitions*, pages 153–76. Kluwer, 1994.

[28] R. Kotecký. Phase transitions of lattice models (Rennes lectures), 1995. Preprint CTS-05-97 (Center for Theoretical Study, the Institute of Advanced Studies at Charles University and the Academy of Sciences of the Czech Republic).

[29] R. Kotecký and D. Preiss. Cluster expansion for abstract polymer models. *Commun. in Math. Phys.*, 103:491–498, 1986.

[30] V. A. Malyshev. Cluster expansions in lattice models of statistical physics and quantum theory of fields. *Russian Mathematical Surveys*, 35:1–62, 1980.

[31] R. A. Minlos and Ya. G. Sinai. The phenomenon of phase separation at low temperatures in some lattice models of a gas. I. *Math. USSR-Sbornik*, 2:335–395, 1967.

[32] R. A. Minlos and Ya. G. Sinai. The phenomenon of phase separation at low temperatures in some lattice models of a gas. II. *Trans. Moscow Math. Soc.*, 19:121–196, 1968.

[33] R. Peierls. Ising's model of ferromagnetism. *Proc. Cambridge Philos. Soc.*, 32:477–481, 1936.

[34] C. E. Pfister. Large deviations and phase separation in the two-dimensional Ising model. *Helvetia Physica Acta*, 64:953–1054, 1991.

[35] S. A. Pirogov and Ya. G. Sinai. Phase diagrams of classical lattice systems. *Theor. Math. Phys.*, 25:1185–1192, 1976.

[36] S. A. Pirogov and Ya. G. Sinai. Phase diagrams of classical lattice systems. Continuation. *Theor. Math. Phys.*, 26:39–49, 1976.

[37] E. Seiler. *Gauge Theories as a Problem of Constructive Quantum Field Theory and Statistical Mechanics.* Lecture Notes in Physics #159. Springer-Verlag, Berlin–Heidelberg–New York, 1982.

[38] Ya. G. Sinai. *Theory of Phase Transitions: Rigorous Results.* Pergamon Press, Oxford–New York, 1982.

[39] J. Slawny. Low-temperature expansion for lattice systems with many ground states. *J. Stat. Phys.*, 20:711, 1979.

[40] J. Slawny. Low-temperature properties of classical lattice systems: phase transitions and phase diagrams. In C. Domb and J. L. Lebowitz, editors, *Phase Transitions and Critical Phenomena, Vol. 11.* Academic Press, London–New York, 1985.

[41] A. C. D. van Enter and R. Fernández. A remark on different norms and analyticity for many-particle interactions. *J. Stat. Phys.*, 56:965–972, 1989.

[42] 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.

[43] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.

[44] M. Zahradník. An alternate version of Pirogov-Sinai theory. *Commun. Math. Phys.*, 93:559–5581, 1984.

[45] M. Zahradník. Phase diagrams of lattice spin models. Cours de Troisième Cycle de la Physique en Suisse Romande (Lausanne), May 1987.

[46] M. Zahradník. A short course on the Pirogov-Sinai theory. File 96-300 in the mp_arc database (it can be retrieved by e-mail from mp_arc@math.utexas.edu, or visiting the site http://www.ma.utexas.edu), 1996.