Metastability of the Ising model with random interaction coefficients

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Saeda Marello
aus
Udine, Italien

Bonn, November 2022

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Erstgutachter: Prof. Dr. Anton Bovier
Zweitgutachterin: Prof. Dr. Véronique Gayrard

Tag der Promotion: 13.02.2023
Erscheinungsjahr: 2023

Abstract

In this Ph.D. thesis we present results on metastability of random modifications of Ising spin systems which evolve with Glauber dynamics. Their Hamiltonians have random and possibly inhomogeneous interaction coefficients. We study these models at fixed temperature, with constant external magnetic field and in the large volume limit.

The stochastic modification adds a level of randomness to the models and makes the direct study of metastability usually very difficult. Therefore, our main strategy consists in studying the model with random interaction coefficients by comparison with the model where these coefficients are replaced by their expectations. Here the latter model is called annealed model.

In the first part of this thesis we summarise a joint paper with A. Bovier and E. Pulvirenti (2021). Therein we studied metastability of the Ising model on the dense Erdős–Rényi random graph with constant edge probability, also called randomly dilute Curie–Weiss model (RDCW), by comparing it with the well-known Curie–Weiss model. The main novelty in the proofs is the application of Talagrand's concentration inequality to characterise the randomness of a certain generalised partition function.

In the second part we prove a simple unpublished extension to more general models of the generalised partition function methods used in the first part.

The third part contains a summary of a joint paper with A. Bovier and F. den Hollander (2022), in which we studied in detail metastability of an Ising model with random interaction coefficients having a product structure. The model we analysed is the annealed version of the Ising model on a Chung–Lu-like random graph with i.i.d. weights which have finite support (ICL). We provided detailed information on the metastable regime and analysed the mean metastable exit time, proving sharp asymptotic estimates and characterising its randomness up to leading order.

In the last part we give an overview on how the results of the first part were extended to a wide class of spin systems with more general random interaction coefficients, in a recent joint work with A. Bovier, F. den Hollander, E. Pulvirenti and M. Slowik. This class of models includes Ising models on various inhomogeneous dense random graphs (e.g. ICL) and randomly diluted spin models. In addition to estimates on the tails of the random mean metastable exit times (showed also in the first part for RDCW), we provided estimates on their moments and conditions on metastability, always in comparison with the annealed model. The methods used include McDiarmid's inequality and novel localisation techniques developed by Schlichting and Slowik (2019).

To Elena Pulvirenti,
for having helped me and for having taught me
so much.

Acknowledgements

I am thankful to my supervisor and coauthor, Anton Bovier. I thank him for giving me the opportunity to be one of his Ph.D. students, for suggesting the first project, for his intuition, for his availability and for what he taught me. Thanks also for speaking Italian with me, which made me feel more at home, and for suggesting the SSF swimming pool.

I would like to thank all my other coauthors Elena Pulvirenti, Frank den Hollander and Martin Slowik for involving me in their projects, for devoting many hours to them, for their patience and their support. In addition, I thank both Frank and Elena for their warm hospitality and for the time they found for me during the four weeks I visited them in the Netherlands. Thanks also to the Global Math Program for financing the visit.

Some more words must be dedicated to Elena. She joined my first project right after it started and never stopped giving mathematical and human support since then. She has been a special mix of colleague, experienced collaborator, mentor and friend.

I want to thank also all the members of the stochastics group that I met in these years for the nice working environment, in particular Anton's group, Mensa and "common-room lunch" groups and my nice and patient office-mates, Katharina and Francis. I want to thank especially Katharina for four years of sharing office and Ph.D. experiences She was always very kind and extremely helpful: I am indebted with her, among other things, for the countless translations from German. Moreover, thanks to the secretaries Mei-Ling Wang, essential for finding my apartment and always very willing to help, and Karen Bingel, always extremely helpful and kind.

Thanks to the members of the doctoral examination committee for their availability and time, and to the BIGS and the SFB 1060 for their financial support.

If these years where only work, I would have not written this thesis. Therefore, I thank here (too) briefly the people with whom I shared time (also) outside working hours, in particular Anna L., Chiara, Gioele, Joy, Lina, Pilar and Rohit. Among them Anna deserves a special acknowledgement. I feel blessed I met her: I cannot imagine how all would have gone without meeting her and I am extremely grateful for what she did and for the moments we had together.

An acknowledgement is also due to all the people who kept in touch, supported me and helped me, despite large physical distance. *Grazie* first of all to my family: they always cared and believed in me. A special mention is due to my brother for his patience and time. *Grazie* also to my "old friends" from Italy, whose support and company I benefited a lot, especially during lockdowns, holidays and when I asked for help. I would like to mention among them Alice, Elisa, Lina and Sara for the moments and thoughts we shared in our long conversations. Furthermore, I am grateful to those who came to visit me in Bonn.

Finally, I want to thank also many people I have not explicitly mentioned here, for company, conversations, advices, for reaching out or simply for a word, a "hi!" or a smile.

Grazie a tutti !

Contents

Chapter 1

Introduction

This Ph.D. thesis contains the study of metastable behaviour of some particular spin systems, which are random modifications of the Ising model and evolve with Glauber dynamics. It is the result of joint works with Anton Bovier, Frank den Hollander, Elena Pulvirenti and Martin Slowik.

In this chapter we give a general introduction of the topics with dealt with, present the models we studied and discuss our results. More precisely, we introduce the concept of metastability in Section 1.1 and devote a separate section (1.2) to the approach we use throughout this thesis: the well-established potential-theoretic approach to metastability. Then we move our attention to the models: we provide a brief introduction of the Ising and Curie–Weiss models in Section 1.3, which gives us the foundations for presenting our models, quantities of interest and results in Section 1.4. An overview on the literature surrounding our work and some useful techniques are given respectively in Sections 1.5 and 1.6. We conclude this introductory chapter in Section 1.7 with an outline of the reminder of the thesis.

1.1 Metastability

1.1.1 Basic introduction

Metastability is a phenomenon occurring in dynamical systems presenting multiple equilibria, namely multiple states or sets to which the dynamics is attracted. When the system reaches a local equilibrium it stabilises there, meaning it stays there for a very long time and it appears stable as if it reached its global equilibrium. The peculiar feature of metastability is that the system, after a long enough time, moves suddenly to a more stable equilibrium. We stress that metastability is strictly linked with some randomness in the dynamics of the system, for instance small random perturbations, and it cannot occur in purely deterministic systems, in which exiting stable states is impossible.

This phenomenon of metastability appears often in nature. Examples range from physics (e.g. supercooled liquids, supersaturated solutions, ferromagnets below Curie temperature, ...) to chemistry, economy and climatology. In the following we provide a few more details about some of these phenomena.

Before diving more into the features of metastable behaviour, it is important to have in mind that the dynamics of the systems we study is usually defined, according to physics laws, in terms of some energy of the configurations, in a way which favours low energy configurations. Despite both motion towards higher and lower energy configurations are possible, the probability of the

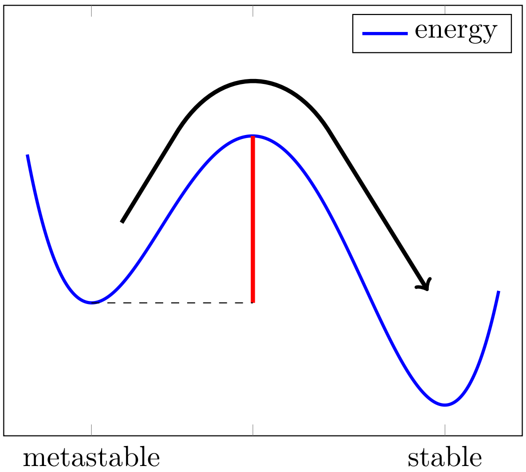


Figure 1.1: One dimensional double well energy landscape.

system decreasing energy is much larger than the one of increasing. Therefore, the minimal points of the energy are the equilibria, the attractors of the dynamics, and the smaller their energy the more stable they are. In this framework, local equilibria will be called *metastable states*, while the global equilibrium *stable state*.

In the following it is useful to have in mind Figure 1.1, which helps to understand various basic features of the metastable behaviour. It represents the paradigmatic one dimensional double well energy landscape of a metastable system, with two equilibria: the one on the left is the metastable state and the one on the right is the (more) stable state. We stress that this picture is much simpler than the landscapes usually studied, which are multidimensional, often have more than two equilibria and might present various degeneracies or delicate cases.

When in an equilibrium state, what makes it difficult for a metastable system to move to a more stable equilibrium (lower energy state) is the presence of an *energy barrier*. It consists in the increment of energy that the system has to attain before reaching the more stable equilibrium, namely it is the difference between a (suitably defined) "energy peak" and the starting energy. For instance, in Figure 1.1 the energy barrier is depicted in red. Since in every step it is much more likely to move towards lower energy configurations than to higher energy ones, the time it takes to overcome the energy barrier and to move to a more stable equilibrium, is very large. To go more into detail, we can divide a path towards a more stable state into two parts, separated by the maximum energy point in the chosen path. Having in mind the dynamics definition, one can see that the second part of the path, from the peak of energy to the more stable state, should be considerably faster than the first part of the path. Indeed, as we shall see later, the leading order term of the mean time to go from a local equilibrium to a more stable one is essentially given by the time to reach the energy peak. In practice, this is often reflected in a long time in which the systems attempts many times to overcome the energy barrier, going unsuccessfully back near the local equilibrium, followed by a very quick transition to a more stable equilibrium after reaching a critical configuration.

We stress again that the one dimensional picture in Figure 1.1 is a very simple case: for instance, while in a one dimensional landscape all paths go through the same maximum, in higher dimensions there can be many paths with different maxima. Therefore, in general one should carefully analyse the paths to define the relevant critical points with high energy that must be be attained by the system for the transition to occur. In multiple dimensions those

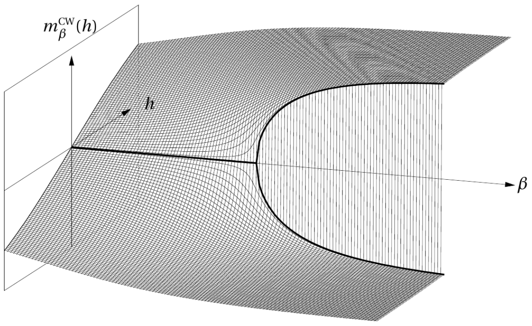


Figure 1.2: Magnetisation of the global equilibrium in the Curie–Weiss model, parameters β, h. [Image from [44]]

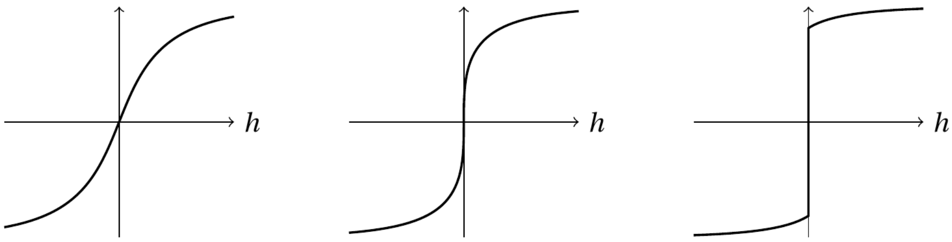


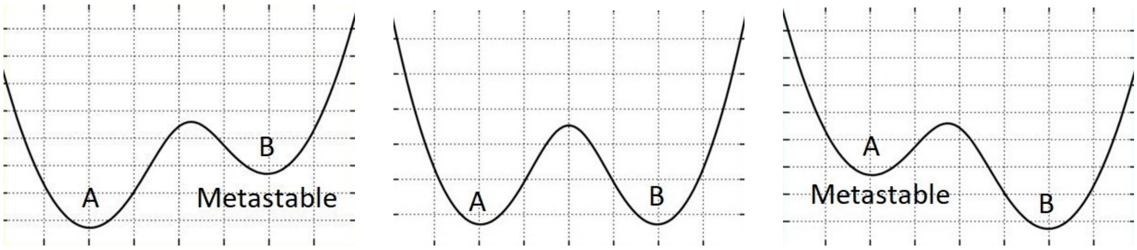
Figure 1.3: Three sections of Figure 1.2 with fixed increasing β. [Image from [44]]

critical points are not maxima but saddle points of the energy.

From this heuristic description one can already notice that a feature of metastable systems is the presence of *multiple time scales*: for each metastable state there is a time scale in which the system is mostly trapped around it, and a faster one, in which the systems jumps to a lower energy metastable state. Only in the case in which all the energy barriers have the same height there is a time scale in which the system jumps among metastable states.

We mentioned examples, described heuristically the metastable behaviour and said that it occurs when the system presents multiple equilibria, but when does this phenomenon occur? In statistical physics it occurs in systems which are close to a first order phase transition. The not expert reader could see this as follows. There is a set of parameters, for example thermodynamical quantities (temperature, pressure, ...), depending on which the energy of the system, namely its behaviour, changes. A first order phase transition occurs when there is a first order discontinuity in the global equilibrium landscape, considered as a function of the parameters. An example can be found in Figure 1.2. It is the graph of the global equilibrium magnetisation in the Curie–Weiss model, where the parameters are $h \in \mathbb{R}$ and $\beta > 0$ and a first order discontinuity occurs for $h = 0$ only (see sections in Figure 1.3). On the parameter region in which the discontinuity occurs the system has multiple global equilibria that coexist (two in Figure 1.2, see Figure 1.4b for an example of energy in this case). The relevant point for us is that when looking closer at the energy landscape near the phase transition one notices the presence of local equilibria (not visible in the global equilibria landscape) which smoothly reach the same energy of the global equilibria, while approaching the discontinuity region (Figures 1.4a and 1.4c).

Understanding this mechanism leads us to two important remarks. First of all, observe that even physical systems that present metastable behaviour are not always metastable. Indeed, a



(a) One global minimum.
In Curie–Weiss $h < 0$ small, $\beta > 1$.

(b) Two global minima.
In Curie–Weiss at phase transition, $h = 0$, $\beta > 1$.

(c) One global minimum.
In Curie–Weiss $h > 0$ small, $\beta > 1$.

Figure 1.4: Double well one dimensional energy landscape, close to phase transition.

system does not exhibit metastable behaviour per-se but it does when its parameters are in a specific set, in which case we say that the system is in the *metastable regime*. Moreover, after noticing that a modification of the parameters usually yields a change in the energy landscape, and consequently in the stability of the equilibria and in the height of the energy barrier, we could further appreciate the power of knowing the energy landscape and in particular the metastable regime. In fact, by changing suitably the parameters within and out of the metastable regime one could ideally encourage or dissuade systems from staying long time in specific equilibria.

The presence of local minima close to the phase transition explains the delay which occurs in nature in the change of phase of certain systems. For instance, in the phenomenon of supercooled water a delay in change from liquid to solid state of water occurs. The water, initially liquid at temperature above zero Celsius, is cooled down below zero. Then, despite being at negative temperature, the water stays in liquid state for long time and becomes all of a sudden solid, after the creation of some critical, sufficiently large, “ice agglomerates”. The picture described above helps to explain the phenomenon as follows (see again Figure 1.4 for reference, think of the temperature decreasing from 1.4a to 1.4c). Before being cooled the water is in its liquid global equilibrium phase, then, while cooling, it reaches the phase transition. With further decreasing temperature the liquid phase becomes the local equilibrium in which the water is trapped for a long time, until it reaches the critical configurations corresponding to the energy peak and then goes very quickly from that to the global equilibrium. In case some thermodynamical quantity is suddenly changed, for instance by shaking the water, the energy landscape might be subsequently modified (e.g. to a landscape with only one equilibrium) and the water may then reach the solid state immediately, without waiting long time.

Something similar happens in ferromagnetic materials (that will be described in more detail in Section 1.3.1) below the so-called Curie temperature within the phenomenon called hysteresis (which is present in many other areas of science [23]). A ferromagnet immersed in an external magnetic field acquires a magnetisation which has the same sign as the field. Below the Curie temperature, there is a first order discontinuity in the acquired magnetisation, when the magnetic field vanishes (as in the right most graph in Figure 1.3). Similarly to the previous example, after the external magnetic field is decreased from positive to negative the magnetisation does not jump immediately to the negative “phase”, but stays for long time very close to the positive value which was the global equilibrium for positive external field, and all of a sudden it goes very quickly to the negative value. Also in this case the system is trapped in a local equilibrium and, after long time, it reaches a critical high energy configuration from which it easily moves to the global equilibrium.

We have already encountered two of the main quantities of interest in the study of metastability: the metastable regime and the equilibria. More precisely, one is interested both in local minima (the equilibria) and in local saddle points (maxima in one dimension) which are relevant to the dynamics, as we have seen when talking about energy barrier. Furthermore, in metastability one is particularly interested in quantitative estimates of the transition between different equilibria. In particular, the interest focuses on the time it takes for such transitions to occur and on the typical (critical) configurations that the system has to attain in the path among equilibria. In this thesis, we do not deal with paths between equilibria, as the approach we use does not allow so, but we focus on precise estimates on metastable transition times from local equilibria to more stable ones.

1.1.2 Approaches to metastability

Different approaches to study metastability have been developed in the last decades. Freidlin and Wentzell initiated in the late 1960's the pathwise approach to metastability, developing a theory which allows to control the metastable behaviour of dynamical system through large deviation techniques. Cassandro, Galves, Olivieri and Vares in [26] started to apply the methods of Freidlin and Wentzell to the study of metastability of interactive particle systems. The pathwise approach uses large deviations to estimate the exponential leading order term of the transition times and provides very detailed information on the metastable behaviour of the system: it allows to characterise with much detail the typical trajectories of the transition between metastable to stable states, often using combinatorial techniques, which make this approach very model dependent and often complicated. A general reference for this approach and applications is the book [64] by Olivieri and Vares.

In the 1980's another approach was developed by Davies. He characterises metastability of a Markov process in terms of spectral properties of its generator, in particular on the presence of a cluster of eigenvalues which are much smaller than the others. The metastable exit times are large and can be expressed in terms of inverses of eigenvalues. This approach is rarely used because it is usually very difficult to verify its technical assumptions.

A third approach was developed in the early 2000's, by Bovier, Eckhoff, Gayrard and Klein ([12, 13]). Called *potential-theoretic approach*, it derives its name from the central role of the tools from potential theory it uses. A classical reference for its theory and applications is the monograph [18] by Bovier and den Hollander, published in 2015, where also many references for the other approaches can be found. In the last years this approach found plenty of applications, mainly in the context of reversible Markov processes modelling interacting particle systems, but also in non-reversible cases (see for instance Gaudillière and Landim [46], the recent work by Seo [70] and further references in [18, Section 7]).

In this thesis we study metastability using only the potential-theoretic approach, enlarging the already large number of applications of its methods. We postpone the introduction and more details to Section 1.2, which is entirely devoted to it.

1.1.3 Quantitative estimates and definitions

We mention here some key facts from the historical survey on metastability made by Bovier and den Hollander in [18]. The study of quantitative aspects of metastability started at the end of the nineteenth century with formulas on rates of chemical reactions. Molecules have to acquire a certain activation energy E before reacting and the *Arrhenius law*

$$R = A\exp(-E/(kT)) \tag{1.1}$$

gives the reaction rate, in terms of E, of the so-called amplitude A and of the absolute temperature T, where k is the Boltzmann constant (approximately $k = 1.38065 \times 10^{-23} J/K$). Equation 1.1 is a formula by van 't Hoff, refined by Arrhenius who added the amplitude factor. The inverse of R can be seen as the average reaction time in a system at temperature T with activation energy E. The Arrhenius law is considered to be universal for systems where an event occurs only after an energy barrier is exceeded.

Kramer provided in 1940 an explicit formula for the activation energy and the amplitude in the Arrhenius law in a one dimensional diffusive system with a double well potential W. The so-called Kramer's formula (see [18, Eq. (2.1.2)]) gives the mean transition time from a local minimum a of the potential W to a global minimum b via a local maximum s: the exponent (the activation energy in Arrhenius law) is the difference of potential between the starting point a and the local maximum s. The prefactor is computed explicitly and depends on the second derivatives of the potential in a and s.

Kramer's formula was later refined and extended to further models. Its extension to multiple dimensions is known as Eyring-Kramer's formula, despite the fact that Eyring and Kramer gave only heuristic arguments. Freidlin and Wentzell [43] derived rigorously the exponent in the 1980's, while the prefactor was computed only later by Bovier, Eckhoff, Gayrard and Klein in [14]. For a more detailed historical summary on the matter see for instance Berglund [4].

We proceed stating two formal definitions of metastability in the context of Markov processes. In order to do that we need to introduce the following notation.

Let $(X(t))_{t\geq 0}$ be a continuous time Markov chain with countable state space S and generator L. Let $\mathbb{P}_\nu$ denote the probability distribution of X conditioned on starting with initial distribution ν and let $\mathbb{E}_\nu$ be the corresponding mean. With abuse of notation we will write $\mathbb{P}_x$ and $\mathbb{E}_x$ when the initial distribution is non-zero only on the state $x \in S$. The first return time to a subset $\mathcal{A} \subset S$ is denoted by

$$\tau_{\mathcal{A}} \equiv \inf \{t > 0 : X(t) \in \mathcal{A}, X(t_-) \notin \mathcal{A}\}. \tag{1.2}$$

Remark 1.1.1 (Continuous vs discrete time). *In this thesis we deal not only with continuous time Markov processes (Chapters 4, 5) but also with discrete time Markov chains (2). Since the space we consider is discrete, the continuous time processes are piece-wise constant right continuous and are characterised by the discrete time jump process and the exponential times between two consecutive jumps. (See Stroock [72].) In this introduction we define quantities and state results for continuous time Markov processes only (except in Section 1.6.2 where we deal with discrete time Markov chains), in order to avoid repetitions. Indeed, for the quantities we deal with, switching from continuous to discrete time is usually natural and straightforward: it is sufficient to use the proper definitions in the right context and to replace the index $t > 0$ with $n \in \mathbb{N}$ and the transition rates with the transition probabilities. However, if relevant changes between the two settings occur we will notify the reader.*

The following two definitions are taken from the book [18, Chapter 8], but their origin can be traced back to the earlier paper [13], where a slightly stronger formulation of (1.3) is provided.

Definition 1.1.2 (Metastability and metastable points). *If $|S| < \infty$, a Markov process $(X(t))_{t\geq 0}$ on S is called* ρ-metastable *with respect to the* set of metastable points $\mathcal{M} \subset S$, *if*

$$|S| \frac{\sup_{x\in\mathcal{M}} \mathbb{P}_x \left(\tau_{\mathcal{M}\setminus\{x\}} < \tau_x\right)}{\inf_{y\notin\mathcal{M}} \mathbb{P}_y \left(\tau_{\mathcal{M}} < \tau_y\right)} \leq \rho \ll 1. \tag{1.3}$$

For later purposes we define here the valley of a metastable Markov process with set of metastable points $\mathcal{M}$.

Definition 1.1.3 (Valley around a metastable point). *The* valley *around a metastable point $m \in \mathcal{M}$ is defined by*

$$A(m) = \left\{ \sigma \in S \colon \mathbb{P}_\sigma(\tau_m = \tau_{\mathcal{M}}) = \sup_{n \in \mathcal{M}} \mathbb{P}_\sigma(\tau_n = \tau_{\mathcal{M}}) \right\}. \tag{1.4}$$

Schlichting and Slowik [68] extended Definition 1.1.2 from metastable points to metastable sets, allowing them to prove general results which allow easier control on some norms useful to estimate mean metastable exit times with the potential-theoretic approach. We refer to [68, Remark 1.2] for detailed comments on the extension. We shall see an example of application of their results in Chapter 5.

We quote here their [68, Definition 1.1] for completeness.

Definition 1.1.4 (Metastability and metastable sets). *For fixed $\rho > 0$ and $K \in \mathbb{N}$, let $\{\mathcal{M}_1, \dots, \mathcal{M}_K\}$ be a set of subsets of S such that $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset$ for all $1 \leq i \neq j \leq K$. The Markov process $(X(t))_{t \geq 0}$ with invariant measure μ is called ρ-*metastable *with respect to the* set of metastable sets $\{\mathcal{M}_1, \dots, \mathcal{M}_K\}$, *if*

$$K \frac{\max_{j \in \{1,\dots,K\}} \mathbb{P}_{\mu|\mathcal{M}_j}\left(\tau_{\mathcal{M}\setminus\mathcal{M}_j} < \tau_{\mathcal{M}_j}\right)}{\min_{\mathcal{X} \subset S \setminus \mathcal{M}} \mathbb{P}_{\mu|\mathcal{X}}\left(\tau_{\mathcal{M}} < \tau_{\mathcal{X}}\right)} \leq \rho \ll 1, \tag{1.5}$$

where $\mathcal{M} = \bigcup_{i=1}^K \mathcal{M}_i$ and $\mu|\mathcal{X}$ denotes, for any non-empty set $\mathcal{X} \subseteq S$, the measure μ conditioned on the set $\mathcal{X}$, namely $\mu|\mathcal{X}(x) = \frac{\mu(x)}{\mu(\mathcal{X})}$ for any $x \in \mathcal{X}$.

This definition, as explained in detail in [68, Remark 1.2], is an extension of Definition 1.1.2 and its main advantage is, as we shall see in Chapter 5, an easier control on some norms useful to estimate mean hitting times using the potential theoretic approach to metastability.

The following definition of valley around a metastable set, which can be found in [68], is an extension of Definition 1.1.3 to sets.

Definition 1.1.5. *The* valley *around a metastable set $\mathcal{M}_i$ is defined by*

$$\mathcal{M}_i \cup \left\{ \sigma \in \mathcal{M}^c \colon \mathbb{P}_\sigma\left[\tau_{\mathcal{M}_i} < \tau_{\mathcal{M}\setminus\mathcal{M}_i}\right] \geq \max_{j \neq i} \mathbb{P}_\sigma\left[\tau_{\mathcal{M}_j} < \tau_{\mathcal{M}\setminus\mathcal{M}_j}\right] \right\}, \tag{1.6}$$

where $\mathcal{M} = \bigcup_{i=1}^K \mathcal{M}_i$ is the union of all the metastable sets.

Both definitions of metastability provide, roughly speaking, information on the returning to the starting point in the following different cases: starting from a metastable set (or point) the probability of moving to a different metastable set (or point) before going back to the start is very small, much smaller that the probability of being attracted to some metastable set before going back to the start, when starting outside any of them. These definitions encode the fact that metastable sets are attractors of the dynamics, and that moving among of them is very unlikely.

1.2 Potential-theoretic approach to metastability

The potential-theoretic approach to metastability was initiated in 2001 by Bovier, Eckhoff, Gayrard and Klein in [12] and [13], and further extended from Markov chains on a discrete space to continuous diffusions in [14] by the same authors and in [17] by Bovier, Gayrard and Klein. It is based on the analogy between electrical networks and finite state Markov chains. This link is thoroughly analysed in [39] by Doyle and Snell and reviewed by Gaudillière in the lecture notes [45]. The main idea is to identify an electrical network (meaning an undirected weighted graph with positive weights, called conductances) with a finite state Markov chain whose transition graph is the undirected graph of the electrical network and whose transition probabilities are proportional to the conductances. Thus, notions from potential theory as resistance, current, equilibrium potential, capacity and their properties can be used in the context of Markov chains.

The problem addressed in [12, 13] was the lack of precise estimates on transition times between stable points involving the passage through a neighbourhood of an unstable critical point. The authors provide estimates on mean metastable transition times up to order constant, in a multidimensional setting, largely improving previous results which were known either up to leading exponential order or only for specific models in one or two dimensions.

In [12] estimates on transition times between local maxima of the invariant measure are provided for Markov chains having as state space the intersection of some connected subset of $\mathbb{R}^d$ and a lattice of spacing $O(1/N)$ in $\mathbb{R}^d$. In addition, they prove asymptotic (in the large volume limit, $N \to \infty$) exponential distribution of the transition times divided by their mean and give, as an example of application, a precise analysis of the metastable behaviour of the random field Curie–Weiss model, with random field taking values in a finite set. The main tools used are the Dirichlet principle from potential theory and renewal estimates.

In [13], after providing a definition of metastable points, the authors extend the results of [12] to the setting of Markov chains with general countable state space and provide different proofs with more tools from potential theory. They first prove asymptotic estimates up to order one of mean exit times from metastable points, in terms of the invariant measure and of certain transition probabilities. The latter can be written in terms of capacities, which can be estimated by well-known dual variational principles (as we shall see more in detail later). Moreover, they prove that each simple eigenvalue of the generator matrix can be associated to a metastable point and it is essentially equal to the inverse of the mean exit time from that point. Finally, they prove the exponential distribution of the rescaled metastable exit times.

The first detailed application of the new methods initiated in [12, 13] is provided by Bovier and Manzo [21] for reversible Markov chains with finite state space, in the zero temperature limit. In this setting they compute accurately the capacities, obtaining estimates on mean hitting times which are precise up to a multiplicative error tending to zero exponentially fast. Their results sharpen largely the ones obtained earlier on the same models with large deviation techniques using the pathwise approach.

Since then, this approach has been successful for studying metastability in different models. After introducing some notions from potential theory, later in this introduction we will focus on models and well-known results which are relevant for this work. We refer again to the monograph [18] for a more extensive introduction on metastability and the potential-theoretic approach applied to reversible Markov processes.

1.2.1 Dirichlet forms, equilibrium potential, last-exit biased distribution, capacity

We will define here the main tools of potential theory that are needed in this work. Let X be a Markov chain with countable state space S, generator L, transition rates $p(\cdot,\cdot)$ and reversible invariant measure μ. Recall that the first return time to any subset $\mathcal{A} \subset S$ is denoted by $\tau_{\mathcal{A}}$ and defined in (1.2).

Consider the following Dirichlet problem

$$\begin{cases} (Lf)(\sigma) = 0, & \sigma \in S \setminus (\mathcal{A} \cup \mathcal{B}) \\ f(\sigma) = \mathbb{1}_{\mathcal{A}}(\sigma), & \sigma \in \mathcal{A} \cup \mathcal{B}, \end{cases} \tag{1.7}$$

where $\mathbb{1}_{\mathcal{A}}$ denotes the indicator function of the set $\mathcal{A}$. The first equation corresponds physically to the Kirchhoff law. The solution of this problem is called *equilibrium potential* and is denoted by $h_{\mathcal{A},\mathcal{B}}$. It can be shown that

$$h_{\mathcal{A},\mathcal{B}}(\sigma) = \begin{cases} \mathbb{P}_\sigma(\tau_{\mathcal{A}} < \tau_{\mathcal{B}}), & \sigma \in S \setminus (\mathcal{A} \cup \mathcal{B}) \\ \mathbb{1}_{\mathcal{A}}(\sigma), & \sigma \in \mathcal{A} \cup \mathcal{B}. \end{cases} \tag{1.8}$$

As in [18, (7.1.39)] the capacity of two disjoint sets $\mathcal{A}, \mathcal{B} \subset S$ is defined by

$$\operatorname{cap}(\mathcal{A},\mathcal{B}) = \sum_{\sigma \in \mathcal{A}} \mu(\sigma) \mathbb{P}_\sigma \left(\tau_{\mathcal{B}} < \tau_{\mathcal{A}}\right). \tag{1.9}$$

The capacity is an essential quantity in potential theory as we shall see in the next sections. The main reason is the existence of variational characterisations of the capacity that are very useful for finding upper and lower bounds. As explained in [45], the capacity $\operatorname{cap}(\mathcal{A},\mathcal{B})$ is, in terms of electrical networks, the current associated with the equilibrium potential $h_{\mathcal{A},\mathcal{B}}$.

Another relevant quantity which will arise when characterising capacities is the *Dirichlet form* of two functions $f, g \colon S \to [0,1]$, defined by

$$\mathcal{E}(f,g) = \frac{1}{2} \sum_{\sigma \in S} \sum_{\sigma' \in S} \mu(\sigma) p(\sigma,\sigma')[f(\sigma') - f(\sigma)][g(\sigma') - g(\sigma)]. \tag{1.10}$$

A crucial distribution in the potential-theoretic approach is the so-called *last-exit biased distribution* $\nu_{\mathcal{A},\mathcal{B}}$ of a set $\mathcal{A} \subset S_N$ with respect to a disjoint set $\mathcal{B} \subset S_N$ and it is defined by

$$\nu_{\mathcal{A},\mathcal{B}}(\sigma) = \frac{\mu(\sigma)\mathbb{P}_\sigma\left(\tau_{\mathcal{B}} < \tau_{\mathcal{A}}\right)}{\sum_{\sigma \in \mathcal{A}} \mu(\sigma)\mathbb{P}_\sigma\left(\tau_{\mathcal{B}} < \tau_{\mathcal{A}}\right)}, \qquad \sigma \in \mathcal{A}. \tag{1.11}$$

Notice that the normalising factor at the denominator is exactly the capacity defined in (1.9).

1.2.2 Variational characterisations of capacities

We state here some of the well-known variational principles characterising the capacity. These are extremely important as they allow to easily obtain upper and lower bounds on capacities, and obtaining sharp bounds reduces to the task of finding proper test functions. In addition to references mentioned above, a summary on the interpretation of Dirichlet and Thomson principles in terms of electrical networks can be found the appendix of [49], by den Hollander and Jansen. Furthermore, Slowik's Ph.D. thesis [71] contains a detailed introduction to the variational principles with statements in the discrete state space setting.

Lemma 1.2.1 (Dirichlet principle)**.** *For any non-empty disjoint $\mathcal{A}, \mathcal{B}$ subsets of S,*

$$\operatorname{cap}(\mathcal{A}, \mathcal{B}) = \min_{f \in \mathcal{H}_{\mathcal{A},\mathcal{B}}} \mathcal{E}(f, f) = \min_{f \in \mathcal{H}_{\mathcal{A},\mathcal{B}}} \frac{1}{2} \sum_{\sigma, \sigma' \in S} \mu(\sigma) p(\sigma, \sigma')[f(\sigma) - f(\sigma')]^2, \tag{1.12}$$

where

$$\mathcal{H}_{\mathcal{A},\mathcal{B}} = \{f : S \to [0, 1] \ s.t. \ f|_{\mathcal{A}} = 1, f|_{\mathcal{B}} = 0\} \tag{1.13}$$

and the minimum is attained for $h_{\mathcal{A},\mathcal{B}}$, the equilibrium potential defined in (1.8)*.*

See Bovier and den Hollander [18, Section 7.3.1 and (7.1.29)] for further details.

Definition 1.2.2. *([71, Definition 1.17]) Given a space S and the transition rates $p(\cdot, \cdot)$ of a Markov chain on S, an* anti-symmetric flow from $\mathcal{A}$ to $\mathcal{B}$*, disjoint subsets of S, is a real valued function $\phi \colon S \times S \to \mathbb{R}$ such that the following conditions hold*

1. *for any $x, y \in S$, if $p(x, y) = 0$, then $\phi(x, y) = 0$;*

2. *for any $x, y \in S$, $\phi(x, y) = -\phi(y, x)$;*

3. *the* Kirchhoff law *holds for every $x \in S \setminus (\mathcal{A} \cup \mathcal{B})$, i.e.*

$$\sum_{y \in S} \phi(x, y) = 0 \tag{1.14}$$

   (this condition is often called divergence free*, as the divergence of a flow ϕ at x is the left hand side of* (1.14)*);*

4. *the total flow out of $\mathcal{A}$ equals the total flow into $\mathcal{B}$, i.e.*

$$\sum_{a \in \mathcal{A}, x \in S \setminus \mathcal{A}} \phi(a, x) = \sum_{b \in \mathcal{B}, x \in S \setminus \mathcal{B}} \phi(x, b). \tag{1.15}$$

A flow from $\mathcal{A}$ to $\mathcal{B}$ is said to be unit *if the quantity in* (1.15) *equals* 1*.*

We are ready to state the Thomson principle in the version of Slowik [71].

Lemma 1.2.3 (Thomson principle)**.** *For any non-empty disjoint $\mathcal{A}, \mathcal{B}$ subsets of S,*

$$\operatorname{cap}(\mathcal{A}, \mathcal{B}) = \sup_{\phi \in \mathcal{U}_{\mathcal{A},\mathcal{B}}} \frac{1}{\mathcal{D}(\phi)}, \tag{1.16}$$

where $\mathcal{U}_{\mathcal{A},\mathcal{B}}$ is the space of all unit anti-symmetric flows $\phi \colon S \times S \to \mathbb{R}$ from $\mathcal{A}$ to $\mathcal{B}$, and $\mathcal{D}(\phi)$ is the norm of ϕ defined by

$$\mathcal{D}(\phi) = \frac{1}{2} \sum_{\sigma, \sigma' \in S} \mathbb{1}(p(\sigma, \sigma') \neq 0) \frac{\phi(\sigma, \sigma')^2}{\mu(\sigma) p(\sigma, \sigma')}. \tag{1.17}$$

The supremum is attained for the so-called harmonic flow *$\phi_{\mathcal{A},\mathcal{B}}$ defined by*

$$\phi_{\mathcal{A},\mathcal{B}}(\sigma, \sigma') = \frac{[h_{\mathcal{A},\mathcal{B}}(\sigma') - h_{\mathcal{A},\mathcal{B}}(\sigma)]_+ \, \mu(\sigma) p(\sigma, \sigma')}{\operatorname{cap}(\mathcal{A}, \mathcal{B})}, \tag{1.18}$$

for $\sigma, \sigma' \in S$, where $[\cdot]_+$ denotes the positive part.

Dirichlet and Thomson principles are widely used throughout this thesis. Another important characterisation of the capacity is given by Berman-Konsowa principle, proved by Berman and Konsowa in [5] for discrete time Markov chains on a finite state space and then extended by den Hollander and Jansen in [49] to continuous time Markov processes on Polish spaces.

In the literature not only anti-symmetric flows are used in this context but also *non-negative* flows, as we shall see in the statement of Berman-Konsowa principle. Here we prefer to state well-known results as they are most commonly found in the literature, therefore we define also the notion of non-negative flows.

Definition 1.2.4. *([71, Definition 1.21]) Given a space S and the transition rates $p(\cdot,\cdot)$ of a Markov chain on S, a* non-negative flow from $\mathcal{A}$ to $\mathcal{B}$, *disjoint subsets of S, is a non-negative function $\psi\colon S\times S\to[0,\infty)$ such that the following conditions hold*

1. *for any $x,y\in S$, if $p(x,y)=0$, then $\psi(x,y)=0$;*
2. *for any $x,y\in S$, if $\psi(x,y)>0$, then $\psi(y,x)=0$;*
3. *the* Kirchhoff law *holds for every $x\in S\setminus(\mathcal{A}\cup\mathcal{B})$*

$$\sum_{y\in S\setminus\{x\}}\psi(x,y)=\sum_{z\in S\setminus\{x\}}\psi(z,x), \tag{1.19}$$

namely the flow into and out of any point (not in $\mathcal{A}$ nor in $\mathcal{B}$) is the same;

4. *the total flow out of $\mathcal{A}$ equals the total flow into $\mathcal{B}$, i.e.* (1.15) *holds.*

As above, a flow from $\mathcal{A}$ to $\mathcal{B}$ is said to be unit *if the quantity in* (1.15) *equals* 1. *Moreover, a non-negative flow from $\mathcal{A}$ to $\mathcal{B}$ is called* loop-free *when any finite path $x_1,\dots,x_n$ of elements of S, with $x_1\in\mathcal{A},x_n\in\mathcal{B}$, satisfying $\psi(x_i,x_{i+1})>0$ for all $i\in\{1,\dots,n-1\}$ is self-avoiding (namely no element of the sequence $x_1,\dots,x_n$ is repeated).*

Remark 1.2.5. *Given a non-negative unit flow ψ from $\mathcal{A}$ to $\mathcal{B}$ one can construct the corresponding anti-symmetric flow ϕ from $\mathcal{A}$ to $\mathcal{B}$ by setting $\phi(x,y)=\psi(x,y)-\psi(y,x)$ and vice versa, by setting $\psi(x,y)=[\phi(x,y)]_+$. The following results do not hold for both flows but equivalent versions for non-negative flows can be found, for instance by slightly modifying some constants.*

Given a non-negative loop-free unit flow ψ from A to B, we can define a probability distribution on the set of self-avoiding paths $\gamma=(x_1,\dots,x_n)$ from $\mathcal{A}$ to $\mathcal{B}$ by

$$\mathbb{P}^{\psi}(\gamma)=\left(\sum_{y\in S}\psi(x_1,y)\right)\prod_{i=1}^{n-1}\frac{\psi(x_i,x_{i+1})\mathbb{1}(x_i\notin\mathcal{B})}{\sum_{y\in S}\psi(x_i,y)}. \tag{1.20}$$

Lemma 1.2.6 (Berman-Konsowa principle)**.** *[71, Proposition 1.11] Given two disjoint subsets $\mathcal{A},\mathcal{B}$ of the state space S, it holds that*

$$\operatorname{cap}(\mathcal{A},\mathcal{B})=\sup_{\psi\in\bar{\mathcal{U}}_{\mathcal{A},\mathcal{B}}}\sum_{\gamma}\mathbb{P}^{\psi}(\gamma)\left(\sum_{(x,y)\in\gamma}\frac{\psi(x,y)}{\mu(x)p(x,y)}\right)^{-1}, \tag{1.21}$$

where $\bar{\mathcal{U}}_{\mathcal{A},\mathcal{B}}$ denotes the set of all the non-negative loop-free unit flows from A to B, and the first sum is over all the positive self-avoiding paths γ from $\mathcal{A}$ to $\mathcal{B}$.

In this thesis we do not use the Berman-Konsowa principle directly. However, in [19] we use a lower bound for capacities derived from that principle by Bovier and den Hollander in [18, Lemma 9.4]. For completeness we state here the bound, after defining one more quantity. A *defective loop-free non-negative unit flow* with defect $\delta\colon S \to \mathbb{R}$ is a function $\psi\colon S \times S \to \mathbb{R}$ satisfying all the conditions in Definition 1.2.4, with (1.19) replaced by

$$\sum_{y \in S\setminus\{x\}} \psi(x,y) = \sum_{z \in S\setminus\{x\}} \psi(z,x) + \delta(x). \tag{1.22}$$

We are ready to state [18, Lemma 9.4], for S finite.

Lemma 1.2.7 (Berman-Konsowa defective flow bound)**.** *Given two disjoint subsets $\mathcal{A}, \mathcal{B}$ of the state space S, for any defective loop-free non-negative unit flow ψ from $\mathcal{A}$ to $\mathcal{B}$ with defect function δ, it holds that*

$$\mathrm{cap}(A,B) \geq \prod_{i=1}^{M} \left(1 + \left[\max_{x \in A_i} \frac{\delta(x)}{\sum_{y \in S} \psi(x,y)}\right]_+\right)^{-1} \sum_{\gamma} \mathbb{P}^{\psi}(\gamma) \left[\left(\sum_{(x,y)\in\gamma} \frac{\psi(x,y)}{\mu(x)p(x,y)}\right)^{-1}\right], \tag{1.23}$$

where $[\cdot]_+$ denotes the positive part, the sum is over all the positive self-avoiding paths γ from $\mathcal{A}$ to $\mathcal{B}$, $\mathbb{P}^{\psi}$ is defined in (1.20) *and $(A_i)_{1\leq i\leq M}$ is a partition of S depending on ψ defined in [18, Eq. (7.3.31)].*

1.2.3 Main formulas

We state here the two main formulas for metastable exit times, that are the core of the potential-theoretic approach to metastability.

Let X be a metastable Markov chain with metastable set $\mathcal{M}$ and invariant measure μ. Let, for $x \in S$, $\mathcal{M}(x) = \{\sigma \in S\colon \mu(x) \leq \mu(\sigma)\}$ and $A(x)$ be the valley around x as defined in Definition 1.1.3. Then for every $x \in \mathcal{M}$

$$\mathbb{E}_x[\tau_{\mathcal{M}(x)}] = [1 + o(1)] \frac{\mu(A(x))}{\mathrm{cap}(x, \mathcal{M}(x))}, \tag{1.24}$$

follows from [18, Theorem 8.15].

Furthermore, for any two non-empty and disjoint $\mathcal{A}, \mathcal{B}$ subsets of S, the following holds as in [18, Eq. (7.1.41)],

$$\mathbb{E}_{\nu_{\mathcal{A},\mathcal{B}}}[\tau_{\mathcal{B}}] = \sum_{\sigma\in\mathcal{A}} \nu_{\mathcal{A},\mathcal{B}}(\sigma)\mathbb{E}_{\sigma}[\tau_{\mathcal{B}}] = \frac{\sum_{\sigma\in S} \mu(\sigma) h_{\mathcal{A},\mathcal{B}}(\sigma)}{\mathrm{cap}(\mathcal{A},\mathcal{B})}, \tag{1.25}$$

where $\nu_{\mathcal{A},\mathcal{B}}$ is defined in (1.11) and $h_{\mathcal{A},\mathcal{B}}$ is the equilibrium potential defined in (1.8). The numerator will be often referred to as *harmonic sum* or $\|h_{\mathcal{A},\mathcal{B}}\|_{\mu}$.

Despite one would like to estimate mean hitting times starting from a specific configuration within certain sets, for general spaces (1.24) is rarely used, because computing the capacity of a single configuration at the denominator of (1.24) is often very difficult. An exceptional case are nearest neighbours random walks on a lattice in $\varepsilon\mathbb{Z}^d$ with ε positive and usually equal to $1/N$ for $N \in \mathbb{N}$. In this case one can use the approximation techniques explained in [18, Chapter 10], and also used in [19] as mentioned later in Chapter 4.

Regardless of being less precise than (1.24), what is often used for general spaces is (1.25), which characterises mean hitting times starting from a set $\mathcal{A}$ with the specific distribution $\nu_{\mathcal{A},\mathcal{B}}$.

The advantage of (1.25) compared to (1.24) is that it does not involve capacities of singletons but only of sets. As mentioned above, estimates on capacities are usually obtained exploiting variational principles (as the ones in Section 1.2.2) and examples of those estimates are frequent in this thesis. Therefore, obtaining estimates on the denominator of (1.25) is usually not a problem. The challenging part when applying (1.25) consists in estimating the harmonic sum. In [22] we use the same techniques as in [7] which are computationally involved and much model dependent (see Chapter 2). In [20] a general technique deriving from the work by Schlichting and Slowik [68] is used. For further details see Chapter 5.

1.3 Ising and Curie–Weiss models

In this section we introduce the concept of ferromagnetism and two well-known spin models, the Ising and the Curie–Weiss models, which will serve as foundation for presenting in Section 1.4 the models we study in this thesis.

1.3.1 Ferromagnetism

All materials acquire a magnetic moment when immersed in a magnetic field. Most of them lose their magnetisation when the external field vanishes: these are called either paramagnetic, when the acquired field is aligned with the external one, or diamagnetic, when it is opposite [67].

Ferromagnetic (and anti-ferromagnetic) materials are an exception. They acquire as well a magnetisation when immersed in a magnetic field but, below a specific temperature called the Curie temperature (named after Pierre Curie's studies in 1895), different for every material, part of it remains when the external field is removed. This internal acquired magnetisation is often called spontaneous magnetisation. Above the Curie temperature, ferromagnetic materials lose their peculiar property and behave as paramagnetic ones. This change in behaviour can be seen in Figure 1.3 which shows the acquired magnetisation of a ferromagnetic material as a function of the external magnetic field h at different temperatures (β is the inverse temperature of the system, as we will see later): from left to right the paramagnetic, critical and ferromagnetic cases, with temperatures above, equal and below the Curie temperature, respectively. More details can be found for instance in Baxter [3]. The difference between ferromagnetic and anti-ferromagnetic behaviour will be shorty explained later.

Similarly to the change of phase in supercooled water, the change in sign of the spontaneous magnetisation in ferromagnets is not immediate. Indeed, if the external field is decreased to zero and then reversed in sign, the acquired magnetisation of a ferromagnet decreases continuously, changing sign only after having been immersed in the negative field for some time. This phenomenon is called hysteresis and the graph of the magnetisation close to vanishing external field is called hysteresis loop due to its peculiar shape (for more details on hysteresis see for instance Brokate and Sprekels [23]).

1.3.2 Origin and definition of the Ising model

Pierre Weiss suggested in 1907 [77] that a ferromagnet could be thought as a set of elementary magnets or interacting magnetic dipoles which can rotate ideally with any angle, and that the spontaneous magnetisation is motivated by the spontaneous alignment of those dipoles. Based on Weiss's theory, Wilhelm Lenz suggested in 1920 [57] a model assuming that the dipoles could take only two possible opposite directions (namely up and down, or as we will see later positive or negative). Lenz did not specify the interactions among the dipoles [61].

Ernst Ising, Lenz's doctoral student, investigated in his Ph.D. thesis (1924, [51, 52]) the model Lenz suggested with few additional assumptions: the dipoles interact only with their nearest neighbours and they are placed on a one dimensional lattice (namely on $\mathbb{Z}$). Starting from Peierls' paper [66], Lenz's model with Ising's hypothesis of the nearest neighbour interaction is known as the *Ising model*.

The Ising model models a spin system. A spin in physics is the angular momentum of a subatomic particle [75] and it provides the essential information to describe the magnetic dipoles of Weiss's theory. Here spins are assumed to be quantised and have only two possible values $+1$ and -1, following Lenz's suggestion. Moreover, they are placed at the vertices of a graph so to provide the nearest neighbour structure necessary to describe the interaction assumed by Ising. Therefore, one might simply think of the Ising model as a model of particles, or any objects, which are placed on a grid, can have only two values ($+1$ or -1) and interact only with their nearest neighbours on the grid.

Having in mind the up and down position of the dipoles, we say that a spin "flips" when it changes its value and that two spins are "aligned" when they have the same value.

Let $G = (V, E)$ be a graph with vertex set V and edge set E, and assume that a spin is placed on every vertex. A *configuration* specifies the values of every spin, namely it is an element $\sigma = (\sigma_i)_{i\in V}$ of the *configuration space* in $S = \{-1, +1\}^V$, where each σ_i is the value of the spin at vertex $i \in V$. As usual in statistical mechanics, given an Hamiltonian function $H\colon S \to \mathbb{R}$, the system is described by the *Gibbs measure* which is defined, for any configuration $\sigma \in S$, by

$$\mu(\sigma) = \frac{\mathrm{e}^{-\beta H(\sigma)}}{Z}, \tag{1.26}$$

where $Z = \sum_{\sigma\in S} \mathrm{e}^{-\beta H(\sigma)}$ is the normalising factor, called *partition function*, and $\beta = \frac{1}{kT} \in (0, \infty)$ is called *inverse temperature* as T is the absolute temperature and k is the Boltzmann constant (approximately $k = 1.38065 \cdot 10^{-23} J/K$). The Gibbs measure corresponds to the Maxwell-Boltzmann distribution of statistical mechanics, which postulates that the value $\mu(\sigma)$ defined in (1.26) is the probability that a configuration with energy $H(\sigma)$ is present in the system at temperature T.

The Hamiltonian H of the Ising model is defined for each configuration $\sigma \in \{-1, +1\}^V$ by

$$H(\sigma) = -J \sum_{i,j\colon (i,j)\in E} \sigma_i\sigma_j - h \sum_{i\in V} \sigma_i, \tag{1.27}$$

where $J > 0$ represents the interaction strength and $h \in \mathbb{R}$ the external magnetic field.

The first term of the Hamiltonian H is an interaction term, to which only nearest neighbouring spins contribute. The parameter J is set positive in order to model the ferromagnetic behaviour in which the configurations with aligned neighbouring spins are more likely. Taking J negative would lead to a description of a system where pairs of neighbouring spins with opposite signs are favoured: this is referred to as anti-ferromagnetic behaviour.

The second term of H encodes the influence of the external magnetic field on the system. Notice that here and in the following the parameter h is independent of $i \in V$, meaning that the external magnetic field influences equally every spin.

We refer to Selinger [69] and Friedli and Velenik [44] for a more detailed introduction of the Ising model at stationarity.

1.3.3 History and importance of the Ising model

Ising proved that no ferromagnetic behaviour, namely no first order phase transition, occurs in the Ising model on the one dimensional lattice ($\mathbb{Z}$) and inferred that the same would happen in three dimensions. Later Ising was proved wrong: the Ising model presents ferromagnetic behaviour in $\mathbb{Z}^2$ and $\mathbb{Z}^3$ [66]. (For further mathematical history on the Ising model we refer to Brush [24] and for a more recent overview to Duminil-Copin [40].)

Only few years after Ising's work, Heisenberg showed that the energy function of ferromagnets is more complicated that the one assumed by Ising and depends also on the speed of the spin flips [66]. Therefore, the Ising model lost its physical interest and had been considered for years interesting only for a purely mathematical point of view. However, in the second half of the last century the Ising model became very relevant in modern physics and other sciences.

The importance of the Ising model nowadays is essentially threefold. First of all it is one of the few models in statistical mechanics which is solvable (in dimensions one and two). In addition, despite being not very realistic in describing ferromagnetism, the model describes a system which exhibits a phase transition. Moreover, it can be naturally seen as a model of cooperative phenomena which makes it suitable for many applications. For a deeper insight on interpretations and importance of the Ising model from the 1920's to the 1970's we refer to a series of three papers ([61, 62, 63]) by Niss, where he reviews how the perception of the model changed in those decades.

1.3.4 The Curie–Weiss model

Pierre Weiss introduced in [77] the *mean-field* assumption, trying to explain ferromagnetic behaviour, which was earlier studied by Pierre Curie, who identified the Curie temperature and the Curie's law. Weiss's assumption led to what is nowadays known as the Curie–Weiss model. It models a system of $N \in \mathbb{N}$ spins, immersed in a uniform external magnetic field $h \in \mathbb{R}$, where all spins interact with each other with uniform interaction strength, namely all in the same way. The homogeneity of the interaction among all the spins is known as the *mean-field* property of this model.

Let

$$S_N = \{-1, 1\}^N \tag{1.28}$$

be the configuration space. The Hamiltonian of the Curie–Weiss model is defined, for any $\sigma \in S_N$, by

$$H_N(\sigma) = -\frac{1}{N} \sum_{i,j:\, 1 \le i < j \le N} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i. \tag{1.29}$$

Notice that the Curie–Weiss model can be viewed as the Ising model on the complete graph, i.e. the graph with edges connecting each pair of vertices (with a rescaling $1/N$ so for the Hamiltonian not to take values larger than order N). The Gibbs measure is defined as in (1.26) by

$$\mu_N(\sigma) = \frac{\mathrm{e}^{-\beta H_N(\sigma)}}{Z_N}, \tag{1.30}$$

where $Z_N = \sum_{\sigma \in S_N} \mathrm{e}^{-\beta H_N(\sigma)}$ is the partition function.

The mean-field property of the Curie–Weiss model makes it much easier to treat than most of the models in statistical mechanics. Indeed, the homogeneity in the interaction allows one to rewrite the Hamiltonian as

$$H_N(\sigma) = -Nm_N(\sigma)^2 - Nhm_N(\sigma), \tag{1.31}$$

where the function $m_N : S_N \to [-1, 1]$ is called *magnetisation* and is defined, for any $\sigma \in S_N$, by

$$m_N(\sigma) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i. \tag{1.32}$$

Thus, H_N does not depend on the value of each spin but only on their mean. Therefore, as one usually tries to do in statistical mechanics the focus moves from the microscopical configurations S_N to the following mesoscopic subset of the interval $[-1, 1]$,

$$\Gamma_N = \left\{ -1, -1 + \frac{2}{N}, \ldots, 1 - \frac{2}{N}, 1 \right\}, \tag{1.33}$$

which is the set of all possible magnetisations and it is more treatable than S_N. The Gibbs measure on the set Γ_N turns out to be, for any $m \in \Gamma_N$,

$$Q_N(m) = \mu_N(m_N^{-1}(m)) = \frac{\mathrm{e}^{-\beta N F_N(m)}}{Z_N}, \tag{1.34}$$

where $Z_N = \sum_{\sigma \in S_N} \mathrm{e}^{-\beta H_N(\sigma)}$ is the partition function and the free energy $F_N : \Gamma_N \to \mathbb{R}$ is defined by

$$F_N(m) = -\frac{1}{2} m^2 - hm + \frac{1}{\beta} \left[-\frac{1}{N} \log \binom{N}{\frac{(1+m)}{2} N} \right], \tag{1.35}$$

where the first two terms come from the Hamiltonian (see (1.31)) and the last one is the entropy term.

We refer to Bovier [11] or Friedli and Velenik [44] for an analysis of the Curie–Weiss model at stationarity. The metastable regime and behaviour for this model with Glauber dynamics is also very well known: we give a brief summary on the topic in Bovier, den Hollander and Marello [19] and we refer for instance to Bovier and den Hollander [18, Chapter 13] for further details.

1.4 Models of interest and results

In this thesis we focus on metastability of reversible Markov chains which model spin systems. The models we study are stochastic modifications of the Ising and Curie–Weiss models. More precisely, we study spin systems of (large) size N at fixed inverse temperature $\beta > 0$, immersed in a constant external magnetic field $h \in \mathbb{R}$. As for the Curie–Weiss model a configuration is a sequence of $N \in \mathbb{N}$ spins, namely an element $\sigma = (\sigma_i)_{i \in \{1,\ldots,N\}}$ of the configuration space S_N defined in (1.28). For every $\sigma \in S_N$ we define the Hamiltonian $H_N : S_N \to \mathbb{R}$ as follows

$$H_N(\sigma) = -\frac{1}{N} \sum_{i,j\colon 1 \le i < j \le N} J_{ij} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i, \tag{1.36}$$

where $h \in \mathbb{R}$ is fixed and $(J_{ij})_{1 \le i < j}$ is a triangular sequence of random variables which will be specified later. The J_{ij}'s are usually called *interaction* (or *coupling*) *coefficients*. Moreover, J_{ij} as a function of the natural numbers i and j is also called *pair potential*.

One can immediately see that taking $J_{ij} \equiv 1$ for all $i, j \in \mathbb{N}$, the Hamiltonian in (1.36) turns out to be equal to one of the Curie–Weiss model (see (1.29)). Moreover, if all the variables $(J_{ij})_{ij}$ have only the two values 0 and $J > 0$, then (1.36) is the Hamiltonian of an Ising model on the graph $G = (V, E)$ with vertex set $V = \{1, \dots N\}$ and edge (i, j) present in the edge set E if and only if $J_{ij} = J$. Similarly to the Curie–Weiss model, there is a rescaling of $1/N$ in order for the Hamiltonian not to take values larger than order N, for instance in case of dense graphs.

It is important to notice that, if each J_{ij} is, for instance, a Bernoulli random variable with mean $p_{ij} \in [0, 1]$, then we are dealing with an Ising model on a *random graph* in which the edge connecting two vertices i and j is present in the graph with probability p_{ij}. This connection with random graphs makes these models more interesting and increases the possible number of applications. In this thesis the notions needed from random graphs theory are simply the definitions of Erdős–Rényi [41] and Chung-Lu [28] random graphs. However, if the reader is interested in exploring the theory, a classical reference is the book by van der Hofstad [47], where also further references on those two graphs can be found.

1.4.1 The three models

Depending on the choice of the random variables $(J_{ij})_{ij}$ in (1.36), the model will of course be different. In this thesis we present three articles. In each of them we studied metastability of a different model. All three models have the Hamiltonian defined in (1.36) and differ in the choice of the coupling coefficients as follows:

1. in Bovier, Marello and Pulvirenti [22] $(J_{ij})_{ij}$ is a sequence of i.i.d. Bernoulli random variables with fixed mean $p \in (0, 1)$: the resulting model is the so-called randomly dilute Curie–Weiss model, which can be seen as an Ising model on the Erdős–Rényi random graph with fixed edge probability p (discussed in Chapter 2);

2. in Bovier, den Hollander and Marello [19] $(J_{ij})_{ij} = (J(i)J(j))_{ij}$, where $(J(i))_{i\in\mathbb{N}}$ is a sequence of i.i.d. non-negative random variables with finite support (discussed in Chapter 4);

3. in Bovier, den Hollander, Marello, Pulvirenti and Slowik [20] $(J_{ij})_{ij}$ is a sequence of random variables which are uniformly bounded and independent conditionally to a given σ-algebra (discussed in Chapter 5).

1.4.2 Glauber dynamics and metastable exit times

We study metastability of Markov processes $(X_N(t))_{t\geq 0}$ with discrete state space $S_N = \{-1, +1\}^N$ (see (1.28)), for large N. As the initial distribution will be specified later, for completing the definition of the Markov process we are left to define the transition rates of those Markov processes. We use the Metropolis transition rates $p_N(\cdot, \cdot)$ defined as follows for any two $\sigma, \sigma' \in S$

$$p_N(\sigma, \sigma') = \begin{cases} \exp\left(-\beta\left[H_N(\sigma') - H_N(\sigma)\right]_+\right), & \text{if } \sigma \sim \sigma', \\ 0, & \text{otherwise,} \end{cases} \tag{1.37}$$

where $\sigma \sim \sigma'$ means that σ' is obtained from σ by a flip of a single spin and $[\cdot]_+$ denotes the positive part. This is a so-called *Glauber dynamics*: the only transitions occur between configurations which differ exactly on one spin, namely only single spin flips are allowed. The Markov process $(X_N(t))_{t\geq 0}$ has the Gibbs measure defined in (1.30) as reversible invariant measure. We denote with $\mathbb{P}_\nu^N$ the law of $(X_N(t))_{t\geq 0}$ with initial distribution ν and with $\mathbb{E}_\nu^N$ the mean with respect to $\mathbb{P}_\nu^N$. Moreover, we abbreviate $\mathbb{P}_\sigma^N$ and $\mathbb{E}_\sigma^N$ when initial distribution is not zero only on the configuration $\sigma \in S_N$.

As we explained in Remark 1.1.1 here we define all quantities only in continuous time, despite in case 1 (Chapter 2, [22]) we deal with discrete time Markov chains.

The main quantity of our interest is the hitting time $\tau_{\mathcal{B}_N}^N$ of a set $\mathcal{B}_N \subset S_N$ of the Markov process $(X_N(t))_{t\geq 0}$ starting from a set $\mathcal{A}_N \subset S_N$, disjoint from $\mathcal{B}_N$. This hitting time is a metastable transition time (often called also metastable exit time) if $(X_N(t))_{t\geq 0}$ is metastable and $\mathcal{A}_N$ and $\mathcal{B}_N$ are metastable sets for the process. We mainly focus on the mean of $\tau_{\mathcal{B}_N}^N$ starting from a specific configuration $\sigma \in \mathcal{A}_N$ or from the last-exit biased distribution $\nu_{\mathcal{A}_N,\mathcal{B}_N}^N$ of $\mathcal{A}_N$ (as defined in (1.11)): $\mathbb{E}_\sigma^N[\tau_{\mathcal{B}_N}^N]$ or $\mathbb{E}_{\nu_{\mathcal{A}_N,\mathcal{B}_N}^N}^N[\tau_{\mathcal{B}_N}^N]$. The reasons why we focus only on these two conditions are embedded in the approach we use to study to metastability, potential-theoretic approach (see Section 1.2.3 for a discussion).

It is important to stress that by introducing some randomness (in the coupling coefficients) in the Hamiltonian we add *levels of randomness* in the picture. Indeed, not only the process we study (the Markov process $(X_N(t))_{t\geq 0}$ on S_N) is random, but also the transition rates p_N defining the process are themselves random variables, bringing an additional level of randomness or possibly more than one, as might happen in case 3. Notice that, as a consequence, mean hitting times as $\mathbb{E}_{\nu_{\mathcal{A}_N,\mathcal{B}_N}^N}^N[\tau_{\mathcal{B}_N}^N]$ are random variables: the mean $\mathbb{E}_{\nu_{\mathcal{A}_N,\mathcal{B}_N}^N}^N$ averages out only one level of randomness, the Markov process on S_N, but leaves untouched the randomness on the transition rates, which comes from the random coupling coefficients $(J_{ij})_{ij}$.

Remark 1.4.1 (Notational remark)**.** *We warn the reader that in some of the papers presented in this thesis the quantities $\mathbb{P}, \mathbb{E}, \mathbb{P}_\nu^N, \mathbb{E}_\nu^N$ are denoted in a slightly different way. However, the notation should be consistent within each chapter of this work.*

1.4.3 Results

In this section we summarise our results about metastability of the models introduced in the previous sections, having in mind the definition of the three cases given in Section 1.4.1. More details will be given in following chapters.

Let us denote with $\mathbb{P}$ the (conditional in case 3) law of $(J_{ij})_{ij}$ and with $\mathbb{E}$ the corresponding mean.

Results on first and third models: a comparison with the averaged model

Clearly, the model with $(J_{ij})_{ij}$ as in 3 (paper [20]) is an extension of the model with $(J_{ij})_{ij}$ as in 1 (paper [22]). In both cases we study metastability of the system by comparing the model with Hamiltonian 1.36 with the averaged model, assuming metastability of the latter. More precisely, the averaged model (called also *annealed model* in [20]) is the model defined by the Hamiltonian in 1.36 in which we replace the random interaction coefficients $(J_{ij})_{ij}$ with their (conditional) mean, i.e. the Hamiltonian $\widetilde{H}_N : S_N \to \mathbb{R}$ defined, for $\sigma \in S_N$, by

$$\widetilde{H}_N(\sigma) = \mathbb{E}(H_N(\sigma)) = -\frac{1}{N} \sum_{i,j:\, 1\leq i<j\leq N} \mathbb{E}(J_{ij})\sigma_i\sigma_j - h\sum_{i=1}^N \sigma_i. \tag{1.38}$$

Notice that in the model with Hamiltonian $\widetilde{H}_N$ there is one level of randomness less than in the one with Hamiltonian H_N. The model with Hamiltonian H_N will be often called *quenched* model, when it is compared with the annealed model.

After taking two specific sets $\mathcal{A}_N$ and $\mathcal{B}_N$ which are metastable for the annealed model with size N, we provide in both [22] and [20], estimates on the tail distribution of the mean hitting time $\mathbb{E}^N_{\nu^N_{\mathcal{A}_N,\mathcal{B}_N}}[\tau^N_{\mathcal{B}_N}]$ divided by the corresponding quantity of the annealed (or averaged) model. In both papers we prove that, asymptotically in N, the random ratio is of order constant times a random factor which behaves like an exponential of a sub-Gaussian random variable. In formulas our results have the following form

$$\lim_{N\to\infty} \mathbb{P}\left(\mathrm{e}^{-t-C_N} \le \frac{\mathbb{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tau^N_{\mathcal{B}_N}\right]}{\tilde{\mathbb{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tilde{\tau}^N_{\mathcal{B}_N}\right]} \le \mathrm{e}^{+t+K_N}\right) \ge 1 - k_1\,\mathrm{e}^{-k_2 t^2}, \tag{1.39}$$

where in both papers C_N and K_N are explicit values, related with the variance of the coefficients $(J_{ij})_{ij}$ and of order at most constant in N. In [20] K_N is smaller than in [22], while it is not clear which C_N provides a better bound because in [22] it has a quite complicated expression. However, a clear improvement in [20] consists in the fact that k_1 and k_2 are explicit constants, while they are not provided in [22].

Both papers use the potential-theoretic approach to metastability (mainly Section 1.2.2 and (1.25)), but there are mainly two differences in the strategy. First, Talagrand's concentration inequality is used in [22] to deal with the randomness of the interaction coefficients, while McDiarmid's inequality is used in [20]. The latter gives the explicit constants k_1 and k_2 and allows for more generality because, contrary to Talagrand's, it does not require convexity checks. The second difference is in the techniques to estimate the harmonic sum in (1.25): [22] uses the same model dependent and computationally long techniques used by Bianchi, Bovier and Ioffe in [7], while [20] exploits definitions and results by Schlichting and Slowik [68] which yield more general estimates with less technicalities.

We have already pointed out that [20] contains results on more general models and with more explicit constants than in [22]. Moreover, in [20] we provided two additional results, not present in [22]. We proved that if the annealed model is metastable with respect to some sets, then also the quenched model is metastable with respect to the same set. Furthermore, definitely in the size N of the system, we proved in [20, Theorem 2.13(ii)] order constant upper and lower bounds on the (conditional) moments of the mean metastable transition time $\mathbb{E}^N_{\nu^N_{\mathcal{A}_N,\mathcal{B}_N}}[\tau^N_{\mathcal{B}_N}]$, divided by the corresponding quantity for the annealed model.

Results on the second model: direct estimates and metastable regime

In paper [19] (for the model with $(J_{ij})_{ij}$ as in 2 in Section 1.4.1), we study directly the metastability of the system and provide explicit estimates, that hold with high probability, on metastable exit times and metastable regimes, without comparing it to an averaged model. We found out that, interestingly, the critical value of the external magnetic field might be not increasing in β. Moreover, we found explicit sharp estimates on the metastable exit time and an explicit characterisation, up to order 1, of the randomness of the exponent its leading order term.

Despite the fact the model might be interesting per-se as a model with random interaction coefficients, our biggest motivation to study its metastability in [19] was the following. It is the annealed model of the Ising model on a Chung-Lu like random graph, i.e. the model where $(J_{ij})_{ij}$ is a sequence of independent Bernoulli random variables with (random) mean $J(i)J(j)$. As a consequence, one can apply the results in [20] (see the general $(J_{ij})_{ij}$ in 3) to the Ising model on a Chung-Lu like random graph, because [19] provides explicit details on the metastability of its annealed model.

1.5 Context and literature

1.5.1 Disordered spin systems at equilibrium

The literature studying equilibrium properties of spin systems having randomness in the interaction coefficients is extremely large. Here we give a brief partial review on work which is related with this thesis, focusing mainly on the Ising model on various different graphs.

Bovier and Gayrard proved in [15] that the free energy and the law of the mean magnetisation of the randomly dilute Curie–Weiss model (see case 1) converge to the ones of the undiluted Curie–Weiss model for any fixed $\beta > 0$ in the thermodynamic limit, when the dilution parameter p times the size of the system N diverges with N. Their paper was of inspiration for comparing the metastable behaviour of those two models in [22]. In [15] the results are proved also for coupling coefficients satisfying Bernstein's condition or normally distributed ones.

Recently, Kabluchko, Löwe and Schubert [53], inspired by [15], proved a central limit theorem for the magnetisation of the same model, referred there as Ising model on dense Erdős–Rényi random graphs, for high temperatures and zero external field. The same authors provided also further results for the sparse case in [54] and [55]. In those three papers the reader can find references for Ising spin systems on sparse random graphs at equilibrium. An additional overview on the topic, until year 2013, and results on the ferromagnetic Ising model on power-law random graphs are present in Dommers' PhD thesis [34].

Moreover, Ising models on dense regular graphs (without external magnetic field) have been studied very recently from a statistical point of view by Xu and Mukherjee [79].

At equilibrium, the model we considered in [19] (see case 2) was studied by Tindemans and Capel [76] and Dommers, Giardinà, Giberti, van der Hofstad and Prioriello [36]. To the best of our knowledge, these are the only two works in the literature which consider that model.

1.5.2 Metastability of the random field Curie–Weiss model with Glauber dynamics

As a first attempt to approach metastability on random modifications of spin models, attention was drawn to the random field Curie–Weiss model (RFCW). It is a spin model with Hamiltonian as the one of the Curie–Weiss model (1.29) where the external field h is a random variable and might depend on the spin it is applied to (thus, the term $h\sum_{i=1}^{N}\sigma_i$ becomes $\sum_{i=1}^{N}h_i\sigma_i$).

Mathieu and Picco [58] and Fontes, Mathieu and Picco [42] were the first to study metastability of the RFCW, obtaining exponential asymptotics for metastable exit times, for large volumes at fixed temperature, when the random field takes only the values 1 and -1. In the case when the external random field can take a finite number of values, the metastable behaviour was fully analysed by Bovier, Eckhoff, Gayrard and Klein in [12], as an application of their new established potential-theoretic approach. Using the same approach, Bianchi, Bovier and Ioffe [7] provided sharp estimates on the mean metastable exit time of the RFCW with continuous distribution of the random field. Their methods for estimating the harmonic sum were largely used in our paper [22] (see Chapter 2). Later, the same three authors proved in [8] the exponential law of the metastable exit time using coupling methods.

1.5.3 Metastability of the Ising model on sparse random graphs

Studying disorder at the level of the interaction coefficients turns out to be more complicated as the Hamiltonian strongly depends on all the realisations of the random coefficients and various techniques used for the RFCW cannon be used in this context.

However, in the last two decades various dynamical properties of Ising model on random graphs with Glauber dynamics have been determined. The first results in this context were given by Bianchi in her Ph.D. thesis [6], where she studied relaxation times of the Ising model on regular random graphs with Glauber dynamics, for low temperatures. Later, Mossel and Sly studied mixing times of the Ising model on sparse Erdős-Rényi random graphs [59] and on random regular graphs [60], both for high temperatures.

Studies on metastability of the Ising model on $\mathbb{Z}^d$ with dilution on the edges (called dilute Ising model) have been carried out around the same time: see for instance Bodineau, Graham and Wouts [9], Wouts [78] and references therein. Later, Dommers [35] and Dommers, den Hollander, Jovanovski and Nardi [38] focused on random regular graphs and the configuration model, respectively, and computed metastable transition times for finite volume and low temperature. In both papers the pathwise approach to metastability is used.

Recently, Can, van der Hofstad and Kumagai [25] extended the results in [60] by studying mixing times the Ising model on random regular graphs for low temperatures.

All these results are about the Ising model on sparse random graphs.

1.5.4 Metastability of the Ising model on dense random graphs and further disordered spin systems

To the best of our knowledge, the papers included in this thesis are the first ones dealing with metastability of the Ising model on dense random graphs, with one exception: the earlier work by den Hollander and Jovanovski [50]. They studied metastability on the exact same model we analyse in [22], the randomly dilute Curie–Weiss model (RDCW), at fixed temperature and large volume, before us, using mainly the pathwise approach to metastability. They proved that the leading order term of the mean transition time is the same as the one in the Curie–Weiss and the random correction term is at most polynomial in the size of the system. Despite the fact that we improved in [22] the precision of the correction term to order constant, their result remains interesting because it is proved with more generality in the initial distribution: their estimate is independent from the initial configuration within the starting set, while ours holds only starting from the last-exit biased distribution.

Acharyya and Štefankovič [1] were the first to study dynamical properties of Glauber spin dynamics on dense random graphs. They considered a homogeneous ferromagnetic Ising model on converging dense graphons. They studied phase transitions and estimated mixing times at fixed temperature, but they considered random graphs only in a very special case. Moreover, they did not study metastability.

Therefore, our results in [20] (see Chapter 5) are the first on the metastable behaviour of the Ising model on general dense random graphs (where the presence of edges is independent) and of a much wider class of spin systems where the interaction coefficients are bounded and (conditionally) independent.

1.6 Some useful techniques

In this section we introduce some relevant notions and techniques which were used in the papers presented in this work.

1.6.1 Coarse-graining

Coarse-graining, as the word suggests, is a technique which consists in considering only part of the very fine available details of a system, especially when this is sufficient to have enough information on the system. In practice the grained quantity is the configuration space: one looks for a not injective function, say m, of the configuration space whose image provides sufficient information to characterise the quantities of interest, usually the Hamiltonian of the system.

In statistical mechanics the grained quantity is the configuration space S and it is reduced to some smaller Γ which carries all the relevant information about the system. More precisely, one looks for a not injective, surjective function $m\colon S \to \Gamma$ such that the Hamiltonian $H(\cdot)\colon S \to \mathbb{R}$ characterising the system could be written as $H(\sigma) = E(m(\sigma))$ for any $\sigma \in S$, with some $E\colon \Gamma \to \mathbb{R}$. This means that the Hamiltonian does not depend on all the information carried by the configuration σ but it is a function only of $m(\sigma)$.

This technique is extremely useful when dealing with spin systems, because the analysis is much simplified if the configuration space is reduced for instance from $S_N = \{-1, 1\}^N$ to a discrete smaller subset of $\mathbb{R}^d$, where the critical points of the Hamiltonian can be found by differentiation. The simplest example of our interest is the Curie–Weiss model in which the function m is the *magnetisation*, a one-dimensional function which maps every configuration $\sigma \in S_N$ to the average of its spins in $\Gamma_N \subset [-1, 1]$ (see (1.32)).

Coarse-graining is very effectively used in more complicated settings, where finitely distributed i.i.d. random variables depending on a single spin are present in the Hamiltonian. This is the case in the random field Curie–Weiss model (with finitely distributed field) (see [12]) or in the model with "product" coupling disorder studied in [19] (Chapter 4) where i.i.d. random variables, say $(V(i))_{i\in\mathbb{N}}$ having values in a finite set $I = \{a_1, \dots, a_k\}$ of cardinality k, appear in the form $\sum_{i=1}^{N} V(i)\sigma_i$. In these cases the strategy is to first randomly partition the set $\{1, \dots, N\}$ into k sets $(A_{\ell,N})_{\ell\in\{1,\dots,k\}}$ according to the realisations of the random variables, namely

$$A_{\ell,N} = \{i \in \{1, \dots, N\}\colon V(i) = a_\ell\}. \tag{1.40}$$

The key point is noticing that one can rewrite

$$\sum_{i=1}^{N} V(i)\sigma_i = \sum_{\ell=1}^{k} a_\ell \sum_{i\in A_{\ell,N}} \sigma_i. \tag{1.41}$$

Notice that the right hand side is a sum of a finite number of terms (even in the limit $N \to \infty$) depending on σ only through a function of the spins. Thus, one defines the random magnetisation m_N associated to $\sigma \in S_N$ as the k-dimensional vector of averages of spins grouped according to the partition in (1.40), as in (4.5) below. Finally, using (1.41) it is possible to rewrite the Hamiltonian of the system as a function of the magnetisation m_N. An example of application of this method and consequences of this procedure are explained in Section 4.3.1.

As it is shown in Bianchi, Bovier and Ioffe [7] for the random field Curie–Weiss model (with continuously distributed field) coarse-graining can be used, in the same spirit, also when the random variables $(V(i))_{i\in\mathbb{N}}$ take values in an infinite bounded set $I \subset \mathbb{R}$. In this case the set I is partitioned in n sets $(I_\ell)_{\ell\in\{1,\dots,n\}}$ of Lebesgue measure vanishing in n, where n is taken to infinity later. The sets $(I_\ell)_{\ell\in\{1,\dots,n\}}$ play the role of the values $\{a_1, \dots, a_k\}$ above. Then, for any fixed N the set $\{1, \dots, N\}$ is partitioned in n sets according to the realisations of the random variables, with respect to the sets $(I_\ell)_{\ell\in\{1,\dots,n\}}$, as follows

$$A_{\ell,N} = \{i \in \{1, \dots, N\}\colon V(i) \in I_\ell\}. \tag{1.42}$$

As above, one defines the random magnetisation m_N associated to $\sigma \in S_N$ as the n-dimensional vector of the averages of the spins with indexes in that partition. Since I_ℓ is not a singleton we cannot obtain something as concise as (1.41). Instead, defining $\bar{a}_\ell = \frac{1}{|A_{\ell,N}|}\sum_{i\in A_{\ell,N}} V(i)$ as reference value for $V(i)$, it turns out that

$$\sum_{i=1}^{N} V(i)\sigma_i = \sum_{\ell=1}^{n} \bar{a}_\ell \sum_{i\in A_{\ell,N}} \sigma_i + \sum_{\ell=1}^{n}\sum_{i\in A_{\ell,N}} (V(i)-\bar{a}_\ell)\sigma_i. \tag{1.43}$$

Here the first term on the right hand side, similarly to (1.41), is a finite sum depending on σ only through its magnetisation while the second term is an error term, which turns out to be treatable despite its dependence on the whole configuration σ (see [7] or [18, Chapter 15] for further details).

Coarse-graining is particularly useful in processes which are lumpable with respect to the function m, as we shall see in the next section.

1.6.2 Lumping

We provide here a slightly modified, but equivalent, version of the definition of lumpable process given by Kemeny and Snell in [56, Section 6.3] and state the main properties. Let be $X = (X(n))_{n\in\mathbb{N}}$ be a Markov chain with state space S, with initial distribution π. and let Γ be a finite set. Let $m : S \to \Gamma$ be a surjective and not injective function and notice that m induces a natural partition on $S = \bigcup_{\gamma\in\Gamma} S[\gamma]$, where we denote $S[\gamma] = m^{-1}(\gamma)$ (this notation will be used throughout the thesis). Then, we define the *lumped process* $Y = (Y(n))_{n\in\mathbb{N}}$ with state space Γ as $Y(n) = m(X(n))$, with initial distribution $\pi_Y(\gamma) = \pi(S[\gamma])$, for any $\gamma \in \Gamma$. Using this notation we give the following definition.

Definition 1.6.1 (Lumpable process, discrete time)**.** *Let S be a finite state space. Then, we say that X is* lumpable *with respect to the function m if the process $Y = (Y(n))_{n\in\mathbb{N}} = (m(X(n)))_{n\in\mathbb{N}}$ is a Markov process for every initial distribution π of X and the transition probabilities of Y do not depend on π.*

Proposition 1.6.2. *As in [56, Theorem 6.3.2], a necessary and sufficient condition for X to be lumpable with respect to m is that for any $y_0, y_1 \in \Gamma$, $n \in \mathbb{N}$ and any $x_0 \in S[y_0]$, the quantity*

$$\sum_{x\in S[y_1]} \mathbb{P}\left(X(n+1) = x | X(n) = x_0\right) \tag{1.44}$$

is independent of $x_0 \in S[y_0]$. Then, the transition probability $\mathbb{P}(Y(n+1) = y_1|Y(n) = y_0)$ is equal to the quantity in (1.44)*.*

Definition 1.6.1 and Proposition 1.6.2 hold true also for continuous time Markov chains on a finite state space, replacing $n \in \mathbb{N}$ and $n+1$ with $s > 0$ and t such that $t > s$, and the transition probabilities with the transition rates (recall Remark 1.1.1). We refer to [18, Section 9.3]) for more general cases.

Applications to potential theory

If a Markov process X is lumpable with respect to a function m then mean hitting times, equilibrium potentials and capacities of X are equal to the ones of the lumped process $m(X)$. More precisely, the following holds.

Proposition 1.6.3. *Let X be a Markov process on a finite state space S with invariant measure $\mu^{(X)}$, lumpable with respect to a function $m\colon S \to \Gamma$. Let $Y = m(X)$ the lumped process, with invariant measure $\mu^{(Y)}(\cdot) = \mu^{(X)}(m^{-1}(\cdot))$. Then, for all $b \in \Gamma$ and $\sigma \in S \setminus m^{-1}(b)$*

$$\mathbb{E}_\sigma^{(X)}[\tau_{m^{-1}(b)}^{(X)}] = \mathbb{E}_{m(\sigma)}^{(Y)}[\tau_b^{(Y)}] = [1 + o(1)]\, \frac{\mu^{(Y)}(A(m(\sigma)))}{\mathrm{cap}^{(Y)}(m(\sigma), b)}, \tag{1.45}$$

where $A(x)$ is the valley around $x \in \Gamma$ as in (1.24). Moreover, for any two non-empty disjoint subsets $\mathcal{A}, \mathcal{B}$ of S, such that there exist $a, b \in \Gamma$ satisfying $\mathcal{A} = m^{-1}(a)$ and $\mathcal{B} = m^{-1}(b)$, the following hold for all $\sigma \in S$ and $x \in m^{-1}(\sigma)$

$$h_{\mathcal{A},\mathcal{B}}^{(X)}(\sigma) = h_{a,b}^{(Y)}(x) \tag{1.46}$$

and

$$\mathrm{cap}^{(X)}(\mathcal{A}, \mathcal{B}) = \mathrm{cap}^{(Y)}(a, b). \tag{1.47}$$

The superscripts indicate with respect to which process the quantities are defined.

Proof. In (1.45) the first equality follows from Proposition 1.6.2, while the second one directly from (1.24). Equalities (1.46) (1.47) are proved for instance in [18, Theorem 9.7]. □

These properties are very helpful in the study of metastability in case the metastable sets can be written as counter-images of elements in Γ. Equations (1.46) and (1.47) were used in [22] on the Curie–Weiss model, while (1.45) and (1.47) were used in [19] where the model is lumpable with respect to the k-dimensional magnetisation m_N mentioned in Section 1.6.1.

Equation (1.45) is very useful because, assuming that we are able to estimate capacities and invariant measure of the lumped model, it provides mean metastable exit times starting from any initial configuration in a metastable set. Estimating those quantities in the lumped model is usually easier than in the original one because of the reduction to a smaller and simpler state space, for example a subset of $\mathbb{R}^d$. This allows one to use, for instance, the approximation techniques explained in [18, Chapter 10] (which we used in [19] as mentioned in Chapter 4) and obtain accurate asymptotic estimates.

When the Markov chain is not lumpable, (1.45) is not necessarily true. Thus, in order to obtain sharp estimates on mean metastable exit times, one uses either (1.24) or (1.25). The first one is hardly useful because computing capacities of single configuration is usually not feasible. Therefore, the second one is the only tool left: we use it in [22] and [20], where the models are not lumpable.

1.7 Outline

The remainder of this thesis is structured as follows. We summarise in three separate chapters our three papers where we analysed metastability of the models introduced in Section 1.4.1.In Chapter 2 our work on metastability of the randomly dilute Curie–Weiss model published in [22] is summarised. The following Chapter 3 contains a short unpublished generalisation of some results from [22]. In Chapter 4 we present the results published in [19] on metastability of a spin model having particular coupling coefficients with product structure. In Chapter 5 we summarise our preprint [20] on metastability of a general class of spin systems with (conditionally) independent coupling coefficients.

Finally, a summary (Chapter 6) will precede three appendices containing the publications and preprints summarised in Chapters 2, 4 and 5: [22] is in Appendix A, [19] in Appendix B and [20] in Appendix C.

Chapter 2

Randomly dilute Curie–Weiss model

We summarise here the content of the joint work with Anton Bovier and Elena Pulvirenti, contained in the paper [22] which was published in March 2021 on the Electronic Journal of Probability.

2.1 Summary

We considered a discrete time Markov chain $\sigma(n)_{n\in\mathbb{N}}$ with state space $S_N = \{-1,+1\}^N$, where we discarded the N dependence from the notation. We assumed it evolves with a Glauber dynamics (defined in Section 1.4.2), in particular with the following Metropolis transition probabilities

$$p_N(\sigma,\sigma') = \begin{cases} \frac{1}{N}\exp(-\beta[H_N(\sigma') - H_N(\sigma)]_+), & \text{if } \sigma\sim\sigma', \\ 1-\sum_{\eta\neq\sigma} p_N(\sigma,\eta), & \text{if } \sigma=\sigma', \\ 0, & \text{else,} \end{cases} \tag{2.1}$$

for $\sigma,\sigma'\in S_N$, where, as usual, $\sigma\sim\sigma'$ means that σ' is obtained from σ by a single spin flip and $\beta>0$ is fixed. The Hamiltonian H_N is defined for fixed $p\in(0,1)$ and for any $\sigma\in S_N$ by

$$H_N(\sigma) = -\frac{1}{Np}\sum_{i,j\in[N],i<j} J_{ij}\,\sigma_i\sigma_j - h\sum_{i\in[N]}\sigma_i, \tag{2.2}$$

where $(J_{ij})_{i<j\in\mathbb{N}}$ is a triangular sequence of i.i.d. Bernoulli random variables with mean p. We denoted with $\mathbb{P}_J$ the law of the triangular sequence $(J_{ij})_{i,j}$ and with $\mathbb{E}$ the mean with respect to it. Notice that, because of the normalisation by p, the mean of the random Hamiltonian H_N is equal to the deterministic Hamiltonian of the Curie–Weiss model $\widetilde{H}_N$ (1.29), namely $\mathbb{E}(H_N(\cdot)) = \widetilde{H}_N(\cdot)$.

Remark 2.1.1 (Notational remark)**.** *In the paper both quantities with superscript CW and $\sim$ referred to the Curie–Weiss model. In this section, in order to be consistent with the rest of the thesis and to uniform the notation, we replace the superscript CW with $\sim$.*

The model defined above is the *randomly dilute Curie–Weiss model* (RDCW). As the name suggests, the model is a result of a dilution of the Curie–Weiss model, by means of the random variables $(J_{ij})_{ij}$. With this dilution the strength of the interaction remains unaltered (up to the $1/p$ rescaling), while the interaction graph is changed from the complete graph to the Erdős–Rényi random graph with uniform edge probability p. By interaction graph here we mean the graph which sets the nearest neighbour interaction as in the Ising model (see Sections 1.3.2 and 1.4), namely the one having vertices $\{1,\ldots,N\}$ and edge (i,j) present in the graph if and only if $J_{ij} \neq 0$. Notice that, since p is fixed and independent on N, the resulting Erdős–Rényi random graph is dense because the expected degree of each vertex is Np.

The RDCW was studied at equilibrium by Bovier and Gayrard in [15]: they proved the convergence of free energy and law of the mean magnetisation of the randomly dilute Curie–Weiss model to the ones of the Curie–Weiss model (CW). In our paper we studied metastability of the RDCW, improving, with special initial conditions, the earlier results by den Hollander and Jovanovski [50]. We studied in particular the tail behaviour of some hitting time of the RDCW compared to the corresponding one in the CW. In order to be more precise we need to introduce some notation.

Let m_-, m_+, m^* be the critical points of the limiting free energy $\lim_{N\to\infty} F_N$ (see (1.35)) of the Curie–Weiss model (local minimum, global minimum and local maximum, respectively) within its well known metastable regime, i.e. $\beta \geq 1$, $0 \leq h \leq \bar{h}_c(\beta)$ where $\bar{h}_c$ is defined explicitly for instance in [19, Eq. (A.13)]. Moreover, let $m_-(N), m_+(N), m^*(N)$ be the closest points to m_-, m_+, m^* in Γ_N (defined in (1.33)) and m_N be the magnetisation function defined in (1.32). The starting set for our dynamics is the set of configurations with magnetisation $m_+(N)$, namely the set $S_N[m_+(N)] = m_N(m_+(N))^{-1}$, with the notation introduced in Section 1.6.2. The initial distribution is the last-exit biased distribution $\nu^N_{S_N[m_-(N)],S_N[m_+(N)]}$ defined similarly to (1.11) for the RDCW dynamics. We abbreviated it with $\nu^N_{m_-,m_+}$.

Our object of interest was the (random) mean hitting time $\mathbb{E}_{\nu^N_{m_-,m_+}}\left[\tau_{S_N[m_+(N)]}\right]$ from the set $S_N[m_-(N)]$ (metastable set of CW) to the set $S_N[m_+(N)]$ (the stable set of CW), where $\mathbb{E}$ is the mean with respect to the law of the Markov chain $\sigma(n)$. To be more precise, our main result in [22] is an asymptotic estimate on the tails of that hitting time divided by the corresponding (deterministic) quantity in the Curie–Weiss model as follows. We quote here [22, Theorem 1.4], with a slight modification in the denominator of (2.3), for clarity and to avoid introducing additional notation. The statement is equivalent because of lumpability of the Curie–Weiss model.

Theorem 2.1.2. *For $\beta > 1$, $h > 0$ small enough and for $s > 0$, there exist absolute constants $k_1, k_2 > 0$ and $C_1(p,\beta) < C_2(p,\beta,h)$ independent of N, such that*

$$\lim_{N\uparrow\infty} \mathbb{P}_J\left(\mathrm{e}^{-2\beta(1+h)-\alpha+\kappa}\mathrm{e}^{-s}(1+o(1)) \leq \frac{\mathbb{E}_{\nu^N_{m_-,m_+}}\left[\tau_{S_N[m_+(N)]}\right]}{\widetilde{\mathbb{E}}_{\tilde{\nu}^N_{m_-,m_+}}\left[\tilde{\tau}_{S_N[m_+(N)]}\right]} \leq \mathrm{e}^{2\beta(1+h)+2\alpha}\mathrm{e}^{s}(1+o(1))\right) \geq 1 - k_1\mathrm{e}^{-k_2 s^2} \tag{2.3}$$

where

$$\alpha = \frac{\beta^2(1-p)}{4p} \tag{2.4}$$

and $\kappa > \alpha$ is explicit (see [22, Eq. (1.27)]).

This means that the ratio in (2.3) is asymptotically bounded from above and below by quantities of order constant times the exponential of a sub-Gaussian random variable.

Definition and characterisations of sub-Gaussian random variables can be found for instance in Boucheron, Lugosi and Massart [10]. The most relevant for us is the following. A real random variable X, with distribution $\mathbb{P}$ and expectation equal to 0, is *sub-Gaussian* if there exists $v > 0$ such that for every $t > 0$

$$\max\{\mathbb{P}(X \geq t), \mathbb{P}(X \leq -t)\} \leq \mathrm{e}^{-\frac{t^2}{2v}}. \tag{2.5}$$

2.2 Discussion

Our paper is not the first work focusing on this: den Hollander and Jovanovski proved first in [50, Theorem 1.4] that the same mean hitting time that we studied is, with high probability, proportional to the Curie–Weiss one times a random unknown prefactor which is at most polynomial in N. They used the pathwise approach to metastability and coupled the randomly dilute Curie–Weiss model with two versions of the standard Curie-Weiss model (modifying suitably the magnetic field). As a consequence of their detailed computations, the exponential distribution of the hitting time follows. In [22] we gave a more precise estimate of the prefactor, losing in generality because our results hold only for the specific initial last-exit biased distribution, while their results hold uniformly in any initial configuration in the starting set.

2.3 Techniques and summary of the proofs

Remark 2.3.1 (Notational remark)**.** *In this chapter we use notation which is consistent with the introduction and of this thesis. The notation in the paper [22] differs only in the free energy: here we use F_N, F, there $f_{N,\beta}$ and f_β were used instead. Moreover, here we discard the dependency on β also from the notation of the invariant measure μ_N, on the mesoscopic measure $\mathcal{Q}_N$ and on the partition function Z_N.*

The model defined above presents more difficulties compared to the well studied random field Curie-Weiss (RFCW) model. Despite each random variable in the Hamiltonian has finite support, our model is not lumpable, not even approximately (as happens for the RFCW with compact support distributions). Indeed, the Hamiltonian cannot be written, for instance via coarse-graining, in terms of a function of the configurations because it depends on the values of every single spin and the model cannot be reduced. Therefore, as explained at the end of Section 1.6.2 we used (1.25) to write mean hitting times as a ratio of harmonic sum and capacity. Since dealing directly with these quantities is difficult, our strategy consists in studying separately how close they are to the corresponding well known quantities of the averaged model, the Curie–Weiss model.

Our paper is divided in three parts. In a preliminary section, [22, Section 2], we compared RDCW and CW at equilibrium by estimating what can be seen as a generalised partition function having as Hamiltonian $H_N(\cdot) - \widetilde{H}_N(\cdot)$. Then, using those results together with well known techniques, we gave estimates on the capacity in [22, Section 3] and the harmonic sum ([22, Section 4]). These estimates were formulated similarly to the one on the mean hitting time in Theorem 2.1.2 and, together with (1.25), they yielded our main result.

2.3.1 Some estimates at equilibrium

This part, summarises the most innovative part of the paper, i.e. [22, Section 2].

The strategy to estimate capacity and harmonic sum was to rewrite the random quantities of interest of the RDCW (capacities, Dirichlet form, invariant measure, transition rates, ...) with two terms, one containing the corresponding well-known CW quantities and the (random) error term. Thus, we preliminary focused on estimates on the error term. The quantity we would like to estimate is $\Delta_N(\sigma) := H_N(\sigma) - \mathbb{E}(H_N(\sigma)) = H_N(\sigma) - \widetilde{H}_N(\sigma)$ or, more precisely, the following sum

$$\mathcal{Z}_{N,g} = \sum_{m\in\Gamma_N} g(m) \sum_{\sigma\in S_N[m]} \mathrm{e}^{-\beta\Delta_N(\sigma)}, \tag{2.6}$$

for some $g : \Gamma_N \to [0,\infty)$ deterministic function which might very well depend on N. We recall that Γ_N is defined in (1.33).

Remark 2.3.2. *The reader might ask why we are not considering the more general quantity* $\sum_{\sigma\in S_N} g(\sigma)\,\mathrm{e}^{-\beta\Delta_N(\sigma)}$ *instead, with* $g : S_N \to [0,\infty)$*. The reason is that, since the* $\mathbb{E}(H_N)$ *is the Hamiltonian of the Curie–Weiss model and it depends only on the magnetisation, we used only the more specific* $\mathcal{Z}_{N,g}$ *defined above, for which the proof of our results is slightly simpler that the general case. However, the techniques used in [19] can be extended to that more general case, with little effort: this is unpublished work carried out by the author of this thesis (see Chapter 3).*

In [22] we proved that $\mathcal{Z}_{N,g}$ can be written as a product of the deterministic quantity $\exp(\mathbb{E}[\log \mathcal{Z}_{N,g}])$, which we bounded asymptotically with quantities of order constant, and the random variable $\exp(\log \mathcal{Z}_{N,g} - \mathbb{E}[\log \mathcal{Z}_{N,g}])$, which we characterised to be an exponential of a sub-Gaussian random variable. More in detail, we proved in [22, Proposition 2.1] that for any $t > 0$

$$\mathbb{P}_J\left(|\log \mathcal{Z}_{N,g} - \mathbb{E}\left[\log \mathcal{Z}_{N,g}\right]| \geq t\right) \leq c_1 \exp\Big(-\gamma t^2\Big), \tag{2.7}$$

and in [22, Lemmas 2.2 and 2.3]

$$\kappa + \log G_g + o\left(1\right) \leq \mathbb{E}\left[\log \mathcal{Z}_{N,g}\right] \leq \alpha + \log G_g + o\left(1\right), \tag{2.8}$$

where α and κ are the same as in Theorem 2.1.2, and here we abbreviate

$$G_g := \sum_{m\in\Gamma_N} g(m) \exp\left(-N I_N(m)\right), \tag{2.9}$$

where $I_N(m)$ is exactly the entropy term appearing in the square brackets in the Curie–Weiss free energy (1.35).

The bound (2.7) was proved directly using Talagrand's concentration inequality (see (2.16)). The upper bound in (2.8) was proved using Jensen's inequality and estimates of $\mathbb{E}\mathcal{Z}_{N,g}$ obtained via Taylor expansion. The lower bound in (2.8) is obtained in a more involved way, following a technique used by Talagrand in [73, Theorem 2.2.1], exploiting Paley-Zygmund inequality (see (2.17)), an upper bound on $\mathbb{E}\mathcal{Z}^2_{N,g}$, obtained via Taylor expansion, and again Talagrand's concentration inequality.

Finally, combining (2.7) and (2.8) we obtained [22, Corollary 2.5] which states that, for any $s > 0$,

$$\mathbb{P}_J\left(\mathrm{e}^{-s+\kappa} G_g\left(1 + o(1)\right) \leq \mathcal{Z}_{N,g} \leq \mathrm{e}^{s+\alpha} G_g\left(1 + o(1)\right)\right) \leq 1 - k_1 \exp\left(-k_2 t^2\right), \tag{2.10}$$

where $k_1, k_2 > 0$ are absolute constants independent of s and N coming from Talagrand's concentration inequality. This result, largely used in the paper, allowed us to quantitatively compare the mesoscopic measures $\mathcal{Q}_N = \mu_N(S_N[m])$ of the randomly dilute Curie–Weiss model and $\widetilde{\mathcal{Q}}_N = \widetilde{\mu}_N(S_N[m])$ of the Curie–Weiss model in [22, Corollary 2.7].

2.3.2 Capacity estimates

In [22, Section 3] we obtained estimates on the capacity $\text{cap}(S_N[m_1], S_N[m_2])$, for any $m_1 \neq m_2 \in \Gamma_N$. We quote here upper and lower bonds ([22, Theorems 1.5 and 1.6]). For any $s > 0$, there exist absolute constants $k_1, k_2 > 0$ such that

$$\mathbb{P}_J\left(\frac{Z_N\ \text{cap}\left(S_N[m_1], S_N[m_2]\right)}{\widetilde{Z}_N\ \widetilde{\text{cap}}\left(S_N[m_1], S_N[m_2]\right)} \leq \mathrm{e}^{s+2\beta(1+h)+\alpha}(1+o(1))\right) \geq 1 - k_1\mathrm{e}^{-k_2 s^2}, \tag{2.11}$$

and

$$\mathbb{P}_J\left(\frac{Z_N\ \text{cap}\left(S_N[m_1], S_N[m_2]\right)}{\widetilde{Z}_N\ \widetilde{\text{cap}}\left(S_N[m_1], S_N[m_2]\right)} \geq \mathrm{e}^{-(s+2\beta(1+h)+\alpha)}(1+o(1))\right) \geq 1 - k_1\mathrm{e}^{-k_2 s^2}, \tag{2.12}$$

asymptotically as $N \to \infty$, where α is defined in (2.4).

As usual within the potential-theoretic approach to metastability the well-established strategy to estimate capacities is via the variational principles presented in Section 1.2.2.

For the upper bound we used the Dirichlet principle (1.12), rough estimates of the ratio of the transition probabilities of RDCW and CW, and (2.10) for a suitably chosen g. The term I_N in G_g (see (2.9)) provided the needed entropy part to reconstruct the CW free energy.

For the lower bound we applied the Thomson principle (1.16) using as test flow the optimal flow of the Curie–Weiss model. Moreover, rough estimates of the ratio of the transition probabilities of RDCW and the CW, together with (2.10), allowed us to obtain the desired lower bound in terms of the CW capacity.

2.3.3 Harmonic sum estimate

The last section of the paper is devoted to estimating the "harmonic sum"

$$\sum_{\sigma\in S_N} \mu_N(\sigma) h^N_{m_-,m_+}(\sigma) \tag{2.13}$$

which is the numerator of (1.25). $\mu_N(\cdot)$ is the invariant measure of the process $(\sigma(n))_{n\in\mathbb{N}}$ and $h^N_{m_-,m_+}$ is a short notation for the harmonic function $h_{S_N[m_-],S_N[m_+]}$ of $(\sigma(n))_{n\in\mathbb{N}}$ (see (1.8)). We quote here upper and lower bonds on ([22, Theorems 1.5 and 1.6]). For any $s > 0$, there exist absolute constants $k_1, k_2 > 0$ such that

$$\mathbb{P}_J\left(\sum_{\sigma\in S_N} \mu_N(\sigma) h^N_{m_-,m_+}(\sigma) \leq \mathrm{e}^{\alpha+s}\ \frac{\exp\left(-\beta N F(m_-)\right)}{Z_N\sqrt{\left(1-m_-^2\right)\beta F''(m_-)}}\ (1+o(1))\right) \geq 1-k_1\mathrm{e}^{-k_2 s^2}, \tag{2.14}$$

and

$$\mathbb{P}_J\left(\sum_{\sigma\in S_N} \mu_N(\sigma) h^N_{m_-,m_+}(\sigma) \geq \mathrm{e}^{\kappa-s}\ \frac{\exp\left(-\beta N F(m_-)\right)}{Z_N\sqrt{\left(1-m_-^2\right)\beta F''(m_-)}}\ (1+o(1))\right) \geq 1-k_1\mathrm{e}^{-k_2 s^2}, \tag{2.15}$$

asymptotically as $N \to \infty$. α and κ are defined in Theorem 2.1.2, and $F = \lim_{N\to\infty} F_N$ the limit of the CW free energy F_N (1.35).

The proof is an involved computation following [7, Section 6]. We summarise here the main idea. We first partitioned the magnetisation space Γ_N in the three sets $U_{\delta,N}(m_-), U_{\delta,N}(m_+), U^c_{\delta,N}$, according to the value of the free energy F. $U_{\delta,N}(m_-)$ and $U_{\delta,N}(m_+)$ are neighbourhoods of $m_-(N)$ and $m_+(N)$, respectively, in which the free energy F is bounded from above by $F(m_-)+\delta$, with $\delta > 0$ chosen sufficiently small. (For a more details of the construction of these sets we refer to [22, Section 4.1].) Then we partitioned the configuration space S_N according to the magnetisation m_N of the configurations, following the partition of Γ_N. Thus, we studied separately the contribution of each set of the S_N partition to the sum in (2.13) and proved that the only relevant contribution to the harmonic sum is given by the configurations with magnetisation in $U_{\delta,N}(m_-)$. Estimating the relevant contribution on $U_{\delta,N}(m_-)$ was done using the saddle point method (see, for instance, de Bruijn [29][7, Chp 5.7]), after approximating the RDCW mesoscopic measure $\mathcal{Q}_N$ with the explicit expression of CW one $\widetilde{\mathcal{Q}}_N$, using results in Section 2.3.1. Proving that the contribution of the third set is negligible was simple: it was sufficient to bound the equilibrium potential by 1 and use that the invariant measure is small there, by definition of the sets. The most challenging part was to prove that also the contribution on $U_{\delta,N}(m_+)$ is negligible. For that we followed the strategy of [7, Proof of Proposition 6.3], where one finds a function which is super-harmonic in a specific interval and uses Doob's optimal stopping theorem to obtain bounds on some transition probabilities.

This strategy was simpler in our case. Indeed, while we used the one-dimensional Curie–Weiss magnetisation m_N, in [7] the magnetisation was n-dimensional due to coarse graining, with n taken to infinity at the end, implying that many quantities were n-dimensional.

2.3.4 Two inequalities

We state here for completeness two inequalities we used in our proofs.

Talagrand concentration inequality, in the version of Tao [74, Theorem 2.1.13], states that if $G : \mathbb{R}^M \to \mathbb{R}$ is a 1-Lipschitz and convex function, and $X = (X_1, \dots, X_M)$, a sequence of $M \in \mathbb{N}$ independent random variables uniformly bounded by $K > 0$, then, for any $t \geq 0$,

$$\mathbb{P}\Big(|G(X) - \mathbb{E}G(X)| \geq tK\Big) \leq c_1 \exp\big(-c_2 t^2\big), \tag{2.16}$$

with positive absolute constants c_1, c_2.

Paley–Sygmund inequality states that for any non negative random variable X, with law P and expectation $\mathbb{E}$, and any $\eta \in (0,1)$

$$\mathbb{P}\big(X \geq \eta\,\mathbb{E}X\big) \geq (1-\eta)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}. \tag{2.17}$$

2.4 Contribution

My contribution to this paper consisted in helping carrying out the proofs and writing them down, thanks to the other authors' ideas. I largely contributed in simplifying and adapting the techniques used by Bianchi, Bovier and Ioffe [7] to our case, which lead to the proofs in Section 4 of the paper concerning estimates on the harmonic sum. Moreover, I created the figure and gave a major contribution in structuring the paper.

Chapter 3

Some generalisations

In this chapter we present some *unpublished* results which generalise Sections 2 and 3 [22], which were presented in Chapter 2. This work has been carried out by the author of this thesis, with the supervision and some ideas by Anton Bovier and Frank den Hollander.

3.1 Setting

For $N \in \mathbb{N}$, consider a discrete time Markov chain $(\sigma_N(n))_{n\in\mathbb{N}}$ on $S_N = \{-1,1\}^N$, which evolves with Glauber dynamics, defined by the Metropolis transition probabilities p_N in (2.1). We consider a more general Hamiltonian H_N than the one in [22] (Chapter 2): we take H_N as in (1.36) with $h \geq 0$ fixed and $(J_{ij})_{ij}$ sequence of independent random variables, defined on an abstract probability space $(\Omega, \mathbb{P}, \mathcal{F})$. We assume that the variables $(J_{ij})_{ij}$ are absolutely uniformly bounded by a constant $k_J > 0$. The invariant measure of $(\sigma_N(n))_{n\in\mathbb{N}}$ is denoted by μ_N and the expectation with respect to $\mathbb{P}$ by $\mathbb{E}$. Moreover, we denote with $\tilde{H}_N = \mathbb{E}(H_N)$ the averaged Hamiltonian, as defined in (1.38). Let $(\tilde{\sigma}_N(n))_{n\in\mathbb{N}}$ be the *annealed* Markov chain, with Metropolis transition probabilities. We will denote with a superscript $\sim$ the quantities referring to the process $(\tilde{\sigma}_N(n))_{n\in\mathbb{N}}$.

Remark 3.1.1. *For consistency with [22] we consider discrete time Markov chains, but our results hold with no modification also for continuous time processes.*

3.2 Estimate on a generalised partition function

In this section we extend (2.10) or [22, Corollary 2.5] to the more general setting defined above, namely to models with random interaction coefficients (not necessarily identically distributed). This generalisation was announced in Remark 2.3.2.

For any deterministic function $g\colon S_N \to [0,\infty)$, we are interested in studying the random quantity

$$\sum_{\sigma\in S_N} g(\sigma)\, \mathrm{e}^{-\beta(H_N(\sigma)-\mathbb{E}[H_N(\sigma)])}, \tag{3.1}$$

which can be seen as a *generalised partition function* with Hamiltonian $H_N(\sigma) - \mathbb{E}[H_N(\sigma)]$. In the same spirit of our results in [22] and [20], we prove that it is equal to the same quantity in which that Hamiltonian is replaced by its expectation times a prefactor. We characterise the prefactor to be, asymptotically, bounded by constants times the exponential of a sub-Gaussian random variable (see (2.5) for a definition). More precisely, we prove the following.

Proposition 3.2.1. *There exist absolute constants* $k_1, k_2 > 0$ *and* $c_1, c_2 > 0$ *such that, for any deterministic function* $g\colon S_N \to [0,\infty)$*, with* $\sum_{\sigma\in S_N} g(\sigma) < \infty$*, and for any* $s > 0$

$$\lim_{N\to\infty} \mathbb{P}\left(\mathrm{e}^{-s+\kappa_N}[1+o(1)] \le \frac{\sum_{\sigma\in S_N} g(\sigma)\,\mathrm{e}^{-\beta(H_N(\sigma)-\mathbb{E}[H_N(\sigma)])}}{\sum_{\sigma\in S_N} g(\sigma)} \le \mathrm{e}^{s+\alpha_N}[1+o(1)] \right) \ge 1 - k_1 \mathrm{e}^{-k_2 s^2}, \tag{3.2}$$

where

$$\alpha_N = \frac{\beta^2}{2N^2} \sum_{1\le i<j\le N} \mathbb{V}\mathrm{ar}(J_{ij}), \tag{3.3}$$

$$\kappa_N = \alpha_N + \max_{\eta\in(0,1)} \left\{ \log\eta - \beta \left[\max_{1\le i<j\le N} |J_{ij} - \mathbb{E}J_{ij}| \right] \frac{\sqrt{2\alpha_N + \log\left(\frac{c_1}{(1-\eta)^2}\right)}}{\sqrt{2c_2}} \right\}. \tag{3.4}$$

where $c_1, c_2 > 0$ *are the constants appearing in Lemma 3.2.4 below.*

Notice that α_N in (3.3) is exactly the one defined in (5.3) (i.e. in [20]) and it is an extension of the quantity α defined in (2.4) (i.e. in [22]) for not identically distributed random variables $(J_{ij})_{ij}$.

Remark 3.2.2. *We neglect the* N*-dependence in the* g *notation, despite the function* g *depends on* N *in the domain and possibly also in its values.*

Remark 3.2.3. (3.2) *also holds replacing* β *with* $-\beta$*.*

Notice that an immediate application of Proposition 3.2.1 is the comparison between Z_N and the annealed correspondent $\widetilde{Z}_N$, taking $g(\sigma) = \mathrm{e}^{-\beta\mathbb{E}[H_N(\sigma)]}$.

Before starting the proof we introduce some more notation. As in [22] we abbreviate

$$\Delta_N(\sigma) = H_N(\sigma) - \mathbb{E}[H_N(\sigma)]. \tag{3.5}$$

Moreover, for consistency with the notation in Section 2.3.1, we name the numerator in (3.2)

$$\mathcal{Z}_{N,g} = \sum_{\sigma\in S_N} g(\sigma)\,\mathrm{e}^{-\beta\Delta_N(\sigma)} \tag{3.6}$$

and the denominator in (3.2)

$$G_g = \sum_{\sigma\in S_N} g(\sigma). \tag{3.7}$$

To prove Proposition 3.2.1 we go exactly along the lines of [22, Section 2]. Here we indicate only some steps and the slight modifications we made to extend those proofs. In order to follow the notation in [22, Section 2] the reader could replace $\frac{1}{N}\log\mathcal{Z}_{N,g}$ by $F_{N,g}$ and $\frac{1}{N}\mathbb{E}[\log\mathcal{Z}_{N,g}]$ by $p_{N,g}$. However, here we decided not to use $F_{N,g}$ and $p_{N,g}$, in favour of more clarity in our arguments.

Proof of Proposition 3.2.1. After rewriting

$$\mathcal{Z}_{N,g} = \mathrm{e}^{\mathbb{E}[\log \mathcal{Z}_{N,g}]}\mathrm{e}^{(\log \mathcal{Z}_{N,g} - \mathbb{E}[\log \mathcal{Z}_{N,g}])}, \tag{3.8}$$

the proof proceeds in two steps. First, we prove concentration of $\log \mathcal{Z}_{N,g}$ i.e. we characterise the random variable $(\log \mathcal{Z}_{N,g} - \mathbb{E}[\log \mathcal{Z}_{N,g}])$ as a sub-Gaussian random variable (Lemma 3.2.4 below). Second, we derive upper and lower bounds on $\mathbb{E}[\log \mathcal{Z}_{N,g}])$ in terms of $\log G_g$ (Lemma 3.2.5 below). Thus, combining these two steps we obtain estimates on the ratio $\frac{\mathcal{Z}_{N,g}}{G_g}$, concluding the proof of Proposition 3.2.1. □

The following lemma extends [22, Proposition 2.1]

Lemma 3.2.4. *For $N \in \mathbb{N}$, there exist absolute constants $c_1, c_2 > 0$ such that for any $t > 0$*

$$\mathbb{P}\left(|\log \mathcal{Z}_{N,g} - \mathbb{E}[\log \mathcal{Z}_{N,g}]| \geq t\right) \leq c_1 \exp\left(-c_2 \frac{2t^2}{4\beta^2 k_J^2}\right). \tag{3.9}$$

Sketch of the proof. Use Talagrand's concentration inequality as in the proof of [22, Proposition 2.1], with $G = \log \mathcal{Z}_{N,g}\left(\frac{\beta}{\sqrt{2}}\right)^{-1}$ as a 1-Lipschitz function of the random variables $\{(J_{ij} - \mathbb{E}J_{ij})\colon 1 \leq i < j \leq N\}$. Moreover, here $|J_{ij} - \mathbb{E}J_{ij}| \leq 2k_J =: K$, while K was equal to 1 in [22]. □

The following lemma extends [22, Lemmas 2.2 and 2.3] and our proof is a slight generalisation of those proofs.

Lemma 3.2.5. *As $N \to \infty$,*

$$\kappa_N + \log G_g + o(1) \leq \mathbb{E}[\log \mathcal{Z}_{N,g}] \leq \alpha_N + \log G_g + o(1), \tag{3.10}$$

where α_N and κ_N are defined in (3.3) *and* (3.4)*, respectively.*

Proof. Step 1. $\mathbb{E}[\mathcal{Z}_{N,g}]$.

We first calculate

$$\mathbb{E}\left[\mathrm{e}^{-\beta\Delta_N(\sigma)}\right] = [1 + o(1)] \exp\left[\frac{\beta^2}{2N^2} \sum_{1 \leq i < j \leq N} \mathbb{V}\mathrm{ar}[J_{ij}]\right] = [1 + o(1)]\, \mathrm{e}^{\alpha_N}, \tag{3.11}$$

which motivates the definition of α_N in (3.3). Here, we used the independence of the J_{ij}'s and followed the argument in [22, Proof of Lemma 2.2]. In particular, we used that, for any random variable $\hat{J}$ with $\mathbb{E}[\hat{J}] = 0$, $\mathbb{E}[\hat{J}^2] > 0$ and all higher moments finite, we have $\mathbb{E}[\mathrm{e}^{x\hat{J}}] = \exp[\frac{x^2}{2}\mathbb{E}[\hat{J}^2] + o(x^2)]$, $x \to 0$. Here we set $\hat{J} = J_{ij} - \mathbb{E}[J_{ij}]$ and $x = \frac{\beta}{N}\sigma_i\sigma_j$, because $\{J_{ij} - \mathbb{E}[J_{ij}]\}_{i,j}$ are independent with mean 0 and variance $\mathbb{V}\mathrm{ar}[J_{ij}]$.

From (3.11), we obtain

$$\mathbb{E}[\mathcal{Z}_{N,g}] = \sum_{\sigma \in S_N} g(\sigma)\, \mathbb{E}\left[\mathrm{e}^{-\beta\Delta_N(\sigma)}\right] = [1 + o(1)]\, \mathrm{e}^{\alpha_N} G_g. \tag{3.12}$$

Step 2. Estimate on $\mathbb{E}[\mathcal{Z}_{N,g}^2]$ in terms of $\mathbb{E}[\mathcal{Z}_{N,g}]^2$.

We first estimate the covariance for any $\sigma, \sigma' \in S_N$

$$\begin{aligned} \mathbb{E}\left[\mathrm{e}^{-\beta\Delta_N(\sigma)}\mathrm{e}^{-\beta\Delta_N(\sigma')}\right] &= \prod_{1\le i<j\le N} \exp\left[\frac{\beta^2}{2N^2}\left[2+2\sigma_i\sigma_j\sigma'_i\sigma'_j\right] \mathbb{V}\mathrm{ar}[J_{ij}] + o\left(\frac{1}{N^2}\right)\right] \\ &\le [1+o(1)] \prod_{1\le i<j\le N} \exp\left[4\frac{\beta^2}{2N^2} \mathbb{V}\mathrm{ar}[J_{ij}]\right] \\ &= [1+o(1)]\,\mathbb{E}\left[\mathrm{e}^{-\beta\Delta_N(\sigma)}\right]\,\mathbb{E}\left[\mathrm{e}^{-\beta\Delta_N(\sigma')}\right]\,\mathrm{e}^{2\alpha_N}, \end{aligned} \tag{3.13}$$

where we used (3.11) and the bound $[2+2\sigma_i\sigma_j\sigma'_i\sigma'_j] \le 4$, in addition to the same argument used in (3.11), now with $x = \frac{\beta}{N}[\sigma_i\sigma_j + \sigma'_i\sigma'_j]$. Thus,

$$\mathbb{E}[\mathcal{Z}_{N,g}^2] = \sum_{\sigma\in S_N} g(\sigma) \sum_{\sigma'\in S_N} g(\sigma')\,\mathbb{E}\left[\mathrm{e}^{-\beta\Delta_N(\sigma)}\mathrm{e}^{-\beta\Delta_N(\sigma')}\right] \le [1+o(1)]\,\mathbb{E}[\mathcal{Z}_{N,g}]^2\mathrm{e}^{2\alpha_N}. \tag{3.14}$$

The upper bound in (3.10) follows from Jensen's inequality and (3.12). The lower bound in (3.10) follows as in the proof of [22, Lemma 2.3] after replacing p by $\frac{1}{2k_J}$ (namely replacing the $\frac{2p^2}{\beta^2}$ in the exponent of [22, Eq. (2.22) and (2.37)] with the $\frac{2}{4\beta^2k_J^2}$ in (3.9)), α by α_N, and using (3.14). The motivation for the definition of κ_N in (3.4) is given in the proof of [22, Lemma 2.3] as well. □

3.3 Capacity estimates

As a consequence of the Proposition 3.2.1, in this section we extend to our more general setting the capacity estimates stated in [22, Theorems 1.5 and 1.6] or in Section 2.3.2 and proved in [22, Section 3].

Proposition 3.3.1. *For any two disjoint $\mathcal{A}, \mathcal{B} \subset S_N$, asymptotically as $N \to \infty$*

$$\mathbb{P}\left(\frac{Z_N\ \mathrm{cap}_N(\mathcal{A},\mathcal{B})}{\widetilde{Z}_N\ \widetilde{\mathrm{cap}}_N(\mathcal{A},\mathcal{B})} \le \mathrm{e}^{s+2\beta(\max_{1\le i<j\le N}|\mathbb{E}J_{ij}|+h)+\alpha_N}(1+o(1))\right) \ge 1 - k_1\mathrm{e}^{-k_2s^2}, \tag{3.15}$$

and

$$\mathbb{P}\left(\frac{Z_N\ \mathrm{cap}_N(\mathcal{A},\mathcal{B})}{\widetilde{Z}_N\ \widetilde{\mathrm{cap}}_N(\mathcal{A},\mathcal{B})} \ge \mathrm{e}^{-\left(s+2\beta(\max_{1\le i<j\le N}|\mathbb{E}J_{ij}|+h)+\kappa_N\right)}(1+o(1))\right) \ge 1 - k_1\mathrm{e}^{-k_2s^2}, \tag{3.16}$$

where α_N and κ_N are defined in (3.3) *and* (3.4)*, respectively.*

3.3.1 Upper bound

Proof of (3.15). This proof is similar to [22, Section 3.1].

We have the uniform bound

$$\frac{\mathrm{e}^{-\beta[H_N(\sigma^{(k)})-H_N(\sigma)]_+}}{\mathrm{e}^{-\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}} = \frac{\mathrm{e}^{-\beta\left[\sigma_k\left(\frac{2}{N}\sum_{i:\ i\ne k} J_{ik}\sigma_i+2h\right)\right]_+}}{\mathrm{e}^{-\beta\left[\sigma_k\left(\frac{2}{N}\sum_{i:\ i\ne k} \mathbb{E}J_{ik}\sigma_i+2h\right)\right]_+}} \le \mathrm{e}^{2\beta(\max_{1\le i<j\le N}|\mathbb{E}J_{ij}|+h)}. \tag{3.17}$$

We apply the Dirichlet principle (1.12) and replace the term $E(m(\cdot))$ in [22, (3.4)] with $\mathbb{E}(H_N(\cdot))$. Moreover, using (3.17) and the fact that the dynamics is single spin flip (see (2.1)), we obtain the bound

$$\begin{aligned}
&Z_N \operatorname{cap}_N(\mathcal{A},\mathcal{B}) = \min_{f\in\mathcal{H}_{\mathcal{A},\mathcal{B}}} Z_N \mathcal{E}_N(f,f) \\
&= \min_{f\in\mathcal{H}_{\mathcal{A},\mathcal{B}}} \sum_{\sigma\in S_N} \mathrm{e}^{-\beta H_N(\sigma)} \sum_{k=1}^{N} \mathrm{e}^{-\beta[H_N(\sigma^{(k)})-H_N(\sigma)]_+}[f(\sigma^{(k)})-f(\sigma)]^2 \\
&\le \min_{f\in\mathcal{H}_{\mathcal{A},\mathcal{B}}} \sum_{\sigma\in S_N} \mathrm{e}^{-\beta H_N(\sigma)} \sum_{k=1}^{N} \frac{\mathrm{e}^{-\beta[H_N(\sigma^{(k)})-H_N(\sigma)]_+}}{\mathrm{e}^{-\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}} \mathrm{e}^{-\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}[f(\sigma^{(k)})-f(\sigma)]^2 \\
&\le \min_{f\in\mathcal{H}_{\mathcal{A},\mathcal{B}}} \mathrm{e}^{2\beta(\max_{1\le i<j\le N}|\mathbb{E}J_{ij}|+h)} \sum_{\sigma\in S_N} \mathrm{e}^{-\beta H_N(\sigma)} \sum_{k=1}^{N} \mathrm{e}^{-\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}[f(\sigma^{(k)})-f(\sigma)]^2,
\end{aligned} \tag{3.18}$$

where $\mathcal{H}_{\mathcal{A},\mathcal{B}}$ is defined in (1.13) and $\sigma^{(k)}$ is the configuration in S_N obtained from σ by flipping the spin σ_k, i.e. $\sigma_i^{(k)} = \sigma_i$ for all $i \neq k$, and $\sigma_k^{(k)} = -\sigma_k$.

Now, let $\tilde{f}$ be the minimizer of the Dirichlet principle for $\widetilde{\operatorname{cap}}_N(\mathcal{A},\mathcal{B})$. After recalling that $\tilde{H}_N(\cdot) = \mathbb{E}[H_N(\cdot)]$, we apply Proposition 3.2.1 to the right-hand side of (3.18) with

$$g(\sigma) = \mathrm{e}^{-\beta\mathbb{E}[H_N(\sigma)]} \sum_{k=1}^{N} \mathrm{e}^{-\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}[\tilde{f}(\sigma^{(k)})-\tilde{f}(\sigma)]^2, \tag{3.19}$$

and use the Dirichlet principle for $\widetilde{\operatorname{cap}}_N(\mathcal{A},\mathcal{B})$ to obtain that there exist absolute constants $k_1, k_2 > 0$ such that for any $s > 0$

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{Z_N \operatorname{cap}_N(\mathcal{A},\mathcal{B})}{\widetilde{Z}_N \widetilde{\operatorname{cap}}_N(\mathcal{A},\mathcal{B})} \le [1+o(1)]\,\mathrm{e}^{2\beta(\max_{1\le i<j\le N}|\mathbb{E}J_{ij}|+h)+s+\alpha_N} \right) \ge 1 - k_1\mathrm{e}^{-k_2 s^2}. \tag{3.20}$$

□

3.3.2 Lower bound

Proof of (3.16). This proof follows the same idea of [22, Section 3.2]. In order to recover $\widetilde{\operatorname{cap}}_N(\mathcal{A},\mathcal{B})$ from $\operatorname{cap}_N(\mathcal{A},\mathcal{B})$, we use the Thomson principle (1.16) for $\operatorname{cap}_N(\mathcal{A},\mathcal{B})$ using the harmonic flow of $\widetilde{\operatorname{cap}}_N(\mathcal{A},\mathcal{B})$ as test flow.

Recalling (1.17), denote with $\mathcal{D}_N(\psi)$ the norm of ψ for the process $(\sigma_N(n))_{n\in\mathbb{N}}$ and with $\tilde{\mathcal{D}}_N(\psi)$ the norm of ψ for the $(\tilde{\sigma}_N(n))_{n\in\mathbb{N}}$.

We use the Thomson principle (1.16) for $\operatorname{cap}_N(\mathcal{A},\mathcal{B})$, choosing as test flow $\tilde{\psi} \in \mathcal{U}_{\mathcal{A},\mathcal{B}}$ the harmonic flow for $\widetilde{\operatorname{cap}}_N(\mathcal{A},\mathcal{B})$ (see (1.18)). Thus, using (2.1) and the definition of the Gibbs invariant measure we write

$$\begin{aligned}\frac{1}{Z_N\,\mathrm{cap}_N(\mathcal{A},\mathcal{B})} &\le \frac{1}{Z_N}\mathcal{D}_N(\tilde{\psi}) = \frac{1}{2}\sum_{\sigma\in S_N}\sum_{k=1}^{N}\tilde{\psi}(\sigma,\sigma^{(k)})^2\,\mathrm{e}^{\beta[H_N(\sigma^{(k)})-H_N(\sigma)]_+}\,\mathrm{e}^{\beta\,H_N(\sigma)}\\ &= \frac{1}{2}\sum_{\sigma\in S_N}\mathrm{e}^{\beta[H_N(\sigma)-\mathbb{E}[H_N(\sigma)]]}\,\mathrm{e}^{\beta\,\mathbb{E}[H_N(\sigma)]}\\ &\quad\times\sum_{k=1}^{N}\frac{\mathrm{e}^{\beta[H_N(\sigma^{(k)})-H_N(\sigma)]_+}}{\mathrm{e}^{\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}}\,\tilde{\psi}(\sigma,\sigma^{(k)})^2\,\mathrm{e}^{\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}\\ &\le \frac{\mathrm{e}^{2\beta(\max_{1\le i<j\le N}|J_{ij}|+h)}}{2}\sum_{\sigma\in S_N}\mathrm{e}^{\beta\Delta_N(\sigma)}\,\mathrm{e}^{\beta\,\mathbb{E}[H_N(\sigma)]}\sum_{k=1}^{N}\tilde{\psi}(\sigma,\sigma^{(k)})^2\,\mathrm{e}^{\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+},\end{aligned} \tag{3.21}$$

because

$$\begin{aligned}\frac{\mathrm{e}^{\beta[H_N(\sigma^{(k)})-H_N(\sigma)]_+}}{\mathrm{e}^{-\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}} &\le \mathrm{e}^{\beta[H_N(\sigma^{(k)})-H_N(\sigma)]_+} = \mathrm{e}^{\beta\left[\sigma_k\left(\frac{2}{N}\sum_{i:\,i\neq k}J_{ik}\sigma_i+2h\right)\right]_+}\\ &\le \mathrm{e}^{2\beta(\max_{1\le i<j\le N}|J_{ij}|+h)} = \mathrm{e}^{2\beta(k_J+h)}.\end{aligned} \tag{3.22}$$

Recall that, since $\tilde{H}_N(\cdot)=\mathbb{E}[H_N(\cdot)]$, by the Thompson principle (1.16) for $\widetilde{\mathrm{cap}}_N(\mathcal{A},\mathcal{B})$ and our choice of $\tilde{\psi}$

$$\frac{1}{\widetilde{Z}_N\widetilde{\mathrm{cap}}_N(\mathcal{A},\mathcal{B})} = \frac{1}{\widetilde{Z}_N}\tilde{\mathcal{D}}_N(\tilde{\psi})\sum_{\sigma\in S_N}\mathrm{e}^{\beta\,\mathbb{E}[H_N(\sigma)]}\sum_{k=1}^{N}\tilde{\psi}(\sigma,\sigma^{(k)})^2\,\mathrm{e}^{\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}. \tag{3.23}$$

Now, after changing the sign of β (see Remark 3.2.3), we can apply (3.2) to the right-hand side of (3.21) with

$$g(\sigma) = \mathrm{e}^{\beta\mathbb{E}[H_N(\sigma)]}\sum_{k=1}^{N}\tilde{\psi}(\sigma,\sigma^{(k)})^2\,\mathrm{e}^{\beta[\mathbb{E}[H_N(\sigma^{(k)})]-\mathbb{E}[H_N(\sigma)]]_+}, \tag{3.24}$$

implying $\frac{1}{\widetilde{Z}_N\widetilde{\mathrm{cap}}_N(\mathcal{A},\mathcal{B})}=\sum_{\sigma\in S_N}g(\sigma)$. Thus, we obtain

$$\lim_{N\to\infty}\mathbb{P}\left(\widetilde{Z}_N\widetilde{\mathrm{cap}}_N(\mathcal{A},\mathcal{B})\sum_{\sigma\in S_N}g(\sigma)\,\mathrm{e}^{\beta[H_N(\sigma)-\mathbb{E}[H_N(\sigma)]]}\le[1+o(1)]\,\mathrm{e}^{s+\alpha_N}\right)\ge 1-k_1\mathrm{e}^{-k_2s^2}. \tag{3.25}$$

Using (3.21) and (3.23) we get

$$\lim_{N\to\infty}\mathbb{P}\left(\frac{\widetilde{Z}_N\widetilde{\mathrm{cap}}_N(\mathcal{A},\mathcal{B})}{Z_N\,\mathrm{cap}_N(\mathcal{A},\mathcal{B})}\le[1+o(1)]\,\mathrm{e}^{2\beta(k_J+h)+s+\alpha_N}\right)\ge 1-k_1\mathrm{e}^{-k_2s^2}, \tag{3.26}$$

concluding the proof of (3.16). □

Chapter 4

Curie–Weiss model with product coupling disorder

In this chapter we introduce and summarise the joint work with Anton Bovier and Frank den Hollander, which appears in the paper [19], submitted in July 2021 and published in the journal Communications in Mathematical Physics in February 2022.

Remark 4.0.1 (Notational remark). *In this chapter we use notation which is consistent with the introduction and the other chapters of this thesis. The notation in [19] is slightly different. We provide here a list of the main differences, showing on the right the notation we use here and on the left the one used in the paper:* $N \mapsto n, \sigma = (\sigma_i)_i \mapsto \sigma = (\sigma(i))_i$ *(spins),* $\sigma(t) \to \sigma_t$ *(Markov process),* $p_N(\cdot,\cdot) \to r_n(\cdot,\cdot)$*.*

4.1 Setting and motivation

We studied metastable regime, metastable sets and random mean metastable exit times for the following random modification of the Curie–Weiss model. For $N \in \mathbb{N}$ we considered the continuous time Markov process $(\sigma(t))_{t\geq 0}$ on $S_N = \{-1,1\}^N$ which evolves with Glauber dynamics with Metropolis transition rates as in (1.37) and has Hamiltonian H_N defined in (4.1). We denoted with $\mathbb{P}_\sigma$ and $\mathbb{E}_\sigma$ law and expectation of $(\sigma(t))_{t\geq 0}$ starting at $\sigma(0) = \sigma$, discarding the N dependence from the notation.

In the Hamiltonian we introduced a coupling disorder, replacing the identically 1 interaction coefficients in the Hamiltonian of the Curie–Weiss model (1.29) with a product of i.i.d. random variables with finite support. This means that the Hamiltonian H_N of our target model writes, $\sigma \in S_N$,

$$H_N(\sigma) = -\frac{1}{N} \sum_{1\leq i<j\leq N} J(i)J(j)\,\sigma_i\sigma_j - h\sum_{i=1}^{N}\sigma_i, \tag{4.1}$$

where $h \geq 0$ is fixed and $(J(i))_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables which have finite support $\{a_1,\dots,a_k\} \subset [0,\infty)$ of cardinality $k \in \mathbb{N}$ and law $\mathcal{P}$ defined through the values $\{\omega_1,\dots,\omega_k\}$ by $\mathcal{P}(J(i) = a_\ell) = \omega_\ell$, for any $\ell \in \{1,\dots,k\}$ and $i \in \mathbb{N}$. We assumed w.l.o.g. that $a_1 < a_2 < \cdots < a_k$.

The interest in studying metastability on this model is motivated by the fact that the Hamiltonian in (4.1) is the mean (with respect to $\mathcal{P}$) of the Hamiltonian of the Ising model on a

Chung–Lu-like random graph, defined by the Hamiltonian in (1.36) where the coupling coefficients $(J_{ij})_{ij}$ are a sequence of independent Bernoulli random variables with mean $J(i)J(j)$. Therefore, paper [19] provides details on the metastable behaviour of the annealed version of the Ising model on a Chung–Lu-like random graph. Thus, thanks to [19], the latter became an example of model to which apply the results appeared in paper [20] (see Chapter 5). This example is particularly relevant because it is the first, to the best of our knowledge, in which the annealed model has an Hamiltonian with random interactions ($(J(i))_{i\in\mathbb{N}}$ in (4.1) are random variables) and its metastability has been studied in detail.

4.2 Results

4.2.1 Metastable regime

As we explained in Section 1.1.1 an important quantity of interest in the study of metastability is the regime of parameters in which the system exhibits metastable behaviour, the metastable regime. In [19] the parameters are $(\beta, h) \in (0,\infty) \times [0,\infty)$ and the metastable regime has the form $\beta \in (\beta_c, \infty), h \in [0, h_c(\beta))$. We recall that the metastable regime of the Curie–Weiss model is $\beta \in (1,\infty), h \in [0, \bar{h}_c(\beta))$, where $\bar{h}_c$ is an increasing function of β defined explicitly for instance in [19, Eq. (A.13)].

We found the value of the critical inverse temperature $\beta_c = \left[\sum_{\ell=1}^{k} a_\ell^2 \omega_\ell\right]^{-1}$ and the following limits of the continuous critical external magnetic field $h_c : (\beta_c, \infty) \to \mathbb{R}$:

$$\lim_{\beta\downarrow\beta_c} h_c(\beta) = 0, \qquad \lim_{\beta\to\infty} h_c(\beta) = \min_{\ell\in[k]^*} \left(\sum_{\ell'=\ell}^{k} a_\ell \, a_{\ell'} \, \omega_{\ell'} - \sum_{\ell'=1}^{\ell-1} a_\ell \, a_{\ell'} \, \omega_{\ell'} \right), \tag{4.2}$$

where $\min_{\ell\in[k]^*}$ is the minimum over all $\ell \in \{1,\dots,k\}$ such that the quantity in brackets is positive. Moreover, in [20, Lemma 3.6] we proved the upper bound

$$\sup_{\beta\in(\beta_c,\infty)} h_c(\beta) < \left(\max_{\ell\in\{1,\dots,k\}} a_\ell \right) \sum_{\ell=1}^{k} a_\ell \, \omega_\ell. \tag{4.3}$$

Contrary to what we expected and what holds in the Curie–Weiss model, $h_c(\cdot)$ is not necessarily an increasing function. In other words this is equivalent to the presence of a so-called “re-entrant” phase transition, meaning that there exist three values $\beta_c \le \beta_1 < \beta_2 < \beta_3$ and a $h > 0$ such that the pairs (β_1, h) and (β_3, h) are in the metastable regime, while (β_2, h) is not. Finding analytical proof of this fact was too complicated, but we found an explicit numerical example of that (see [19, Appendix C] for details).

Noteworthy, similarly to what stated in [12, Lemma 7.4] for the random field Curie–Weiss model, the metastable landscape has an underlying 1-dimensional structure, meaning that the critical points of the $k-$dimensional free energy, which are relevant for the metastable behaviour, correspond to the critical points of a 1-dimensional function. The reduction to a 1-dimensional problem explained below in the proof summary motivates this fact at an intuitive level.

Furthermore, we found numerical evidence of the presence of multiple metastable states, for some choice of parameters and law of the coupling coefficients (see [19, Appendix B]). Hence, we conjectured that for any finite k, and any $\ell \in \{0, 1, \dots, k\}$ there exist a pair (β, h) and a law $\mathcal{P}$ such that the free energy has exactly $2\ell + 1$ critical points. Thus, due to the 1-dimensional structure, if $\ell \ge 1$, there would be $\ell + 1$ metastable magnetisations separated by ℓ saddle points.

4.2.2 Metastable exit time

We summarise here in words the main results concerning metastable exit times which can be found in [19, Theorems 1.1, 1.2, 1.3]. Recall Remark 4.0.1.

In [19, Theorem 1.1], we found an explicit expression, precise to order 1, of the random mean metastable exit time, namely the time it take to the process $\sigma(t)$, starting in a configuration within a metastable set, to hit a metastable set with lower free energy (i.e. more stable). Exactly as in [18, Theorem 10.9], it turns out to be exponential in N with a prefactor of order constant. The exponent is proportional to the inverse temperature β times the (random) free energy gap between the metastable starting set and the saddle point. Our result holds the same hypotheses of [18, Theorem 10.9], in the metastable regime, uniformly in the starting configuration within the metastable set and with $\mathcal{P}$-probability tending to 1.

Since the Hamiltonian is random, also the free energy appearing in the exponent of the mean metastable exit time is random. In [19, Theorem 1.3] we gave a characterisation of that random free energy gap. It turns out to be equal, in distribution, to its deterministic limit plus a random correction term, which is $\frac{1}{\sqrt{N}}$ times a centered normal random variable. The variance of this Gaussian random variable can be explicitly computed and depends in a complicated way on the values $\{a_1, \dots, a_k\}$ and $\{\omega_1, \dots, \omega_k\}$.

Furthermore, with $\mathcal{P}$-probability tending to 1, as in [18, Theorem 10.11], the considered metastable exit time divided by its mean turns out to be exponentially distributed ([19, Theorem 1.2]).

4.3 Main ideas of the proofs

4.3.1 Preliminary model reduction

The product structure of the random coupling coefficients simplifies very much the model, because the randomness depends on single spins and not on pairs of spins (or edges). Indeed, one can rewrite the Hamiltonian in (4.1) (up to add the diagonal term which is a constant shift) as follows, for $\sigma \in S_N$

$$H_N(\sigma) = -\frac{1}{2N}\left(\sum_{i=1}^{N} J(i)\sigma_i\right)^2 - h\sum_{i=1}^{N}\sigma_i. \tag{4.4}$$

Hence, a model reduction is possible through a finite coarse-graining as explained in Section 1.6.1. We defined the random level sets $A_{\ell,N}$, $\ell \in \{1, \dots, k\}$ as in (1.40) (replacing $V(i)$ with $J(i)$). Then we set the level magnetisations $m_N(\sigma) = (m_{\ell,N}(\sigma))_{\ell\in\{1,\dots,k\}}$ to be the average of spins in $A_{\ell,N}$, namely for $\ell \in \{1, \dots, k\}$ and $\sigma \in S_N$

$$m_{\ell,N}(\sigma) = \frac{1}{|A_{\ell,N}|}\sum_{i\in A_{\ell,N}}\sigma_i. \tag{4.5}$$

Thus, $m_N(\sigma)$ takes values in the set

$$\Gamma_N = \bigtimes_{\ell\in\{1,\dots,k\}}\Gamma_{\ell,N}, \qquad \Gamma_{\ell,N} = \left\{-1, -1+\tfrac{2}{|A_{\ell,N}|}, \dots, 1-\tfrac{2}{|A_{\ell,N}|}, 1\right\}. \tag{4.6}$$

Using this notation and abbreviating $\omega_{\ell,N} = \frac{|A_{\ell,N}|}{N}$, one can rewrite the Hamiltonian as follows

$$H_N(\sigma) = -N\left[\frac{1}{2}\left(\sum_{\ell=1}^{k} a_\ell\,\omega_{\ell,N}\,m_{\ell,N}(\sigma)\right)^2 + h\sum_{\ell=1}^{k}\omega_{\ell,N}\,m_{\ell,N}(\sigma)\right] = NE_N(m_N(\sigma)). \tag{4.7}$$

Then it is clear that H_N depends on the configurations $\sigma \in S_N$ only through the magnetisation $m_N(\sigma)$

Remark 4.3.1. *In [18, Section 14.4] a very similar coarse-graining is carried out for the random field Curie–Weiss model. The only two differences are in the notation and in the rescaling in the definition of* $m_{\ell,N}$ (4.5)*, which is* $\frac{1}{N}$ *in [18]. The choice of the rescaling affects of course the set* Γ_N*, but consistency with the choice yields the same results.*

From (4.7) and the fact that, for any two $\sigma, \sigma' \in S_N$ satisfying $m_N(\sigma) = m_N(\sigma')$, the sets $m_N^{-1}(\sigma)$ and $m_N^{-1}(\sigma')$ have the same cardinality, one proves that the process $(\sigma(t))_{t\geq 0}$ is *lumpable* (see Section 1.6.2) with respect to the map $m_N : S_N \to \Gamma_N$. Hence, we studied first the lumped process, namely the Markov process on Γ_N with transition probabilities

$$\bar{r}_N(m, m') = \mathrm{e}^{-\beta N[E_N(m')-E_N(m)]_+} \sum_{\ell=1}^{k} |A_{\ell,N}| \left[\frac{1-m_\ell}{2}\,\mathbb{1}(m' = m^{\ell,+}) + \frac{1+m_\ell}{2}\,\mathbb{1}(m' = m^{\ell,-})\right] \tag{4.8}$$

for any $m \in \Gamma_N$ and $m' \in \Gamma_N$ equal to m in all component but one, say $\ell \in \{1, \dots, k\}$, in which $m'_\ell = m_\ell \pm \frac{2}{|A_{N,\ell}|}$. All other transitions have zero probability to occur. Then, after estimating capacities and invariant measure at the magnetisation level, we proved our results on S_N using the properties presented in Section 1.6.2.

For later reference we add here few more details on the lumped process. With the notation $S_N[\cdot]$ introduced in Section 1.6.2, its invariant measure is the Gibbs measure $\mathcal{Q}_N(m) = \mu_N(S_N[m]) = \frac{1}{Z_N}\mathrm{e}^{-\beta N F_N(m)}$, $m \in \Gamma_N$, where $F_N(m)$ is the free energy given by

$$F_N(m) = -\frac{1}{2}\left(\sum_{\ell=1}^{k} a_\ell\,\omega_{\ell,N}\,m_\ell\right)^2 - h\sum_{\ell=1}^{k}\omega_{\ell,N}\,m_\ell - \frac{1}{\beta}\frac{1}{N}\log\left[\prod_{\ell=1}^{k}\binom{|A_{\ell,N}|}{\frac{1+m_\ell}{2}|A_{\ell,N}|}\right]. \tag{4.9}$$

4.3.2 Metastable regime

Since we were interested in the large volume setting and motivated by [18, Theorem 10.6], we studied the metastable regime at the limit $N \to \infty$. Differentiating the limit of the free energy (4.9), we obtained the following system of equations for the k-dimensional critical points $m = (m_\ell)_{\ell\in\{1,\dots,k\}}$

$$m_\ell = \tanh\left(\beta\left[a_\ell\left(\sum_{\ell'=1}^{k} a_{\ell'}\,\omega_{\ell'}\,m_{\ell'}\right) + h\right]\right), \qquad \ell \in \{1, \dots, k\}. \tag{4.10}$$

Looking at (4.10), we noticed that a further reduction is possible. Indeed, any solution $m \in [-1, 1]^k$ of the system of equations in (4.10) is characterised by the 1-dimensional quantity

$$K = K(m) = \sum_{\ell=1}^{k} a_\ell\,\omega_\ell\,m_\ell. \tag{4.11}$$

Thus, (4.10) reduces to the 1-dimensional fixed point equation

$$K = T_{\beta,h}(K), \qquad T_{\beta,h}(K) = \sum_{\ell=1}^{k} a_\ell\,\omega_\ell\,\tanh(\beta[a_\ell K + h]). \tag{4.12}$$

Therefore, since (4.12) characterises the critical points of the free-energy, we obtained our results on the metastable regime by studying the conditions for $T_{\beta,h}(K)$ to have more than three fixed points not tangent to the diagonal.

4.3.3 Mean metastable exit time

We recall that, by lumpability and model reduction described in Section 4.3.1, we focused on the Markov chain on Γ_N with transition rates in (4.8). In [19, Hypothesis 1] we imposed on free energy the same assumptions in [18, Section 10.1]. Thus, we used the techniques explained in [18, Chapter 10] which have already been applied in a similar way in [18, Chapter 14] to the random field Curie–Weiss model. Our contribution in this part was to find that a slightly modification of the matrix $\mathbb{B}_N$ was needed to adapt the methods of [18, Chapter 10] to the random interactions and to the continuous time setting.

We summarise here the main idea of the proof on the mean metastable exit time, which follows the line of proof of [18, Theorem 10.9]. We used the potential-theoretic approach to metastability, in particular (1.24). For estimates of the numerator in (1.24) we used Taylor expansions, [18, Lemma 10.12 and Eq. (10.2.33)]. In [19, Section 5] we provided estimates of the capacity at the denominator: we used the Dirichlet principle (1.12) for the upper bound and the Berman-Konsova principle (1.21) for the lower bound, as explained in [18, Sections 10.2, 10.3]. Matching upper and lower bounds were obtained via the construction explained in [18, Sections 10.2] and implemented in [19, Section 4]. The procedure consists in two steps: first define a suitable function $\tilde{g}$ to be used as test function in the Dirichlet principle and to build the test flow for the Berman-Konsova principle, second construct a linearised approximated dynamics, whose Dirichlet form $\tilde{\mathcal{E}}(\cdot,\cdot)$ in $\tilde{g}$ could be explicitly computed. That will be used as an approximation of the original $\mathcal{E}(\tilde{g},\tilde{g})$ in the Dirichlet principle.

The exponential law follows directly applying [18, Theorem 8.45].

The characterisation of the random exponent followed from a computation which uses mainly the Central Limit Theorem for the quantity $\omega_{N,\ell}$ (defined below (4.6)) and the equations characterising the critical points of the free energy.

4.4 Contribution

I carried out the detailed computation for the proofs of the main theorems, finding that few details of well-known had to be adapted to study metastability in our model of interest. Moreover, I largely contributed in finding the metastable regime. Since the number of metastable points and the properties of the critical curve β_c were not clear to us, as they depend in a complicated way on all the parameters involved I used the software Mathematica® to produce some plots, some of which appeared in [19, Appendices B,C]. These allowed us to understand better the behaviour of the model and to have numerical evidence of two facts: the presence of multiple metastable states for some choices of (β, h) and the re-entrant phase transition for some choices of the distribution of $J(\cdot)$. Finally, I contributed in typing and structuring the paper.

Chapter 5

Ising model with inhomogeneous coupling disorder

In this chapter we summarise the content of the preprint [20]: *Metastability of Glauber dynamics with inhomogeneous coupling disorder*, joint work with Anton Bovier, Frank den Hollander, Elena Pulvirenti and Martin Slowik. The preprint was submitted to `arxiv.org` in September 2022.

5.1 Setting and summary

We fixed an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and fix $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. The model we considered is a continuous time Markov process $(\Sigma_N(t))_{t\geq 0}$ with state space $S_N = \{-1,+1\}^N$, modelling a spin system of size N, evolving with Glauber dynamics (defined in Section 1.4.2), determined by the Hamiltonian H_N in (1.36). We took the coupling coefficients $(J_{ij})_{ij}$ in H_N to be uniformly bounded random variables (i.e. there exists $k_J > 0$ such that $|J_{ij}| \leq k_J$ for all $i, j \in \mathbb{N}$) that are independent conditionally to the given σ-algebra $\mathcal{G}$. We denoted with $\mathbb{P}_\mathcal{G}$ a regular conditional distribution for the sequence $(J_{ij})_{ij}$ with respect to $\mathcal{G}$, and with $\mathbb{E}_\mathcal{G}$ and $\mathbb{V}\mathrm{ar}_\mathcal{G}$ the expectation and the variance with respect to $\mathbb{P}_\mathcal{G}$.

We proved two theorems which compare the metastable behaviour of this model (called quenched) with the one of the model (called annealed or averaged) with Hamiltonian $\widetilde{H}_N$ in which the coupling coefficients $(J_{ij})_{ij}$ of the Hamiltonian H_N are replaced by their conditional mean $\mathbb{E}_\mathcal{G} J_{ij}$ (see (1.38)). Our strategy is based on the potential-theoretic approach to metastability. In addition, we used the McDiarmid's inequality and results and techniques from Schlichting and Slowik [68].

5.2 Examples

Our results are very general. They apply to a wide class of Ising spin systems on (possibly inhomogeneous) random graphs, in which the presence of the edges is independent. Of particular interest, due to the rescaling of the interaction term, is the case of dense random graphs.

A first example is the randomly dilute Curie–Weiss model, which is an Ising model on an Erdős–Rényi random graph with edge probability p. It is obtained by taking $(J_{ij})_{ij}$ distributed as i.i.d. Bernoulli random variables with fixed mean $p \in (0, 1)$. Its metastability was studied already in [22] (see Chapter 2), which we extended, and earlier in [50].

Taking instead the coupling coefficients $(J_{ij})_{ij}$ distributed as independent Bernoulli random variables with (random) mean $J(i)J(j)$, with $(J(i))_{i\in\mathbb{N}}$ i.i.d. random variables which have finite support, leads to the Ising model on a Chung–Lu-like [28] random graph. The metastability of that model can be effectively studied using our results and the information about the metastable behaviour of its annealed model which we studied, for this very purpose, in [19] and explained in Chapter 4.

Our general model includes many randomly diluted models. With our results their metastable behaviour can be studied by comparing it with the one of the corresponding undiluted models. An example in this category is the randomly diluted Hopfield model. It is obtained by taking as $(J_{ij})_{ij}$ the usual coefficients of the Hopfield model multiplied (namely diluted) by i.i.d. Bernoulli random variables with mean p, and as $\mathcal{G}$ the σ-algebra generated by the coupling coefficients of the Hopfield model. In spite of the fact that the metastable behaviour of the undiluted Hopfield model is not known yet (except for a small regime of parameters), this example is already relevant as it motivates why we chose to study the general case of conditional averaging (see discussion in Section 5.5).

5.3 Results

In the first theorem ([20, Theorem 2.10]) we provided conditions for metastability. More precisely, we proved that if, for some $k_1 > 0$, the annealed model is $\mathrm{e}^{-k_1 N}$-metastable with respect to a finite number of sets (in the sense of Definition 1.1.4 by Slowik and Schlichting [68]), definitely in N with probability 1, then the quenched model is e^{-cN}-metastable for any $c \in (0, k_1)$, definitely in N with probability 1, with respect to the *same* metastable sets of the annealed model.

In order to give estimates on metastable exit times, we required a *non-degeneracy* hypothesis ([20, Assumption 2.11]), similar to the one in [13, Definition 1.2], as in [68, Theorem 1.7], and assumed in addition the annealed model to be $\mathrm{e}^{-k_1 N}$-metastable definitely in N with probability one.

Let $\tilde{\mu}_N$ be the invariant measure of the annealed model. We fixed as starting set $\mathcal{A}_N$, one of the metastable sets (satisfying the non-degeneracy assumption), and as target set $\mathcal{B}_N$, the union of all metastable sets which have measure $\tilde{\mu}_N$ not smaller than $\tilde{\mu}_N(\mathcal{A}_N)$. Here, following the notation in [20], we denote with P^N_σ the law of $(\Sigma_N(t))_{t\geq 0}$ starting in $\sigma \in S_N$, and with $E^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}$ the mean with initial distribution $\nu_{\mathcal{A}_N,\mathcal{B}_N}$. Note that the same quantities were denoted instead by $\mathbb{P}^N_\sigma$ and $\mathbb{E}^N_{\nu^N_{\mathcal{A}_N,\mathcal{B}_N}}$ in Section 1.4. Moreover, quantities with a superscript $\sim$ will refer to the annealed model (namely they are defined via the Hamiltonian $\tilde{H}_N$).

For the reasons explained in Section 1.2.3, in our results the starting distribution of $(\Sigma_N(t))_{t\geq 0}$ is always the last-exit biased distribution

$$\nu_{\mathcal{A}_N,\mathcal{B}_N}(\sigma) = \frac{\mu_N(\sigma) P^N_\sigma\left(\tau^N_{\mathcal{B}_N} < \tau^N_{\mathcal{A}_N}\right)}{\sum_{\sigma\in\mathcal{A}_N} \mu_N(\sigma) P^N_\sigma\left(\tau^N_{\mathcal{B}_N} < \tau^N_{\mathcal{A}_N}\right)}, \qquad \sigma \in \mathcal{A}. \tag{5.1}$$

as it is defined in (1.11).

Therefore, the mean hitting time $E^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}[\tau^N_{\mathcal{B}_N}]$ studied in our second theorem is the mean metastable transition time from the set $\mathcal{A}_N$ (with distribution $\nu_{\mathcal{A}_N,\mathcal{B}_N}$) to one of the more stable metastable sets.

We are ready to present in detail the results of [20, Theorem 2.13]. In the first part of the theorem we provided the following estimates on tails, which are formulated similarly to the ones in [22, Theorem 1.4]. For $t \geq 0$, $\mathbb{P}$-a.s.,

$$\lim_{N\to\infty} \mathbb{P}_{\mathcal{G}} \left(\mathrm{e}^{-t-\alpha_N} \leq \frac{E^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tau^N_{\mathcal{B}_N}\right]}{\tilde{E}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tilde{\tau}^N_{\mathcal{B}_N}\right]} \leq \mathrm{e}^{+t+2\alpha_N} \right) \geq 1 - 4\,\mathrm{e}^{-t^2/(2\beta k_J)^2}, \tag{5.2}$$

where

$$\alpha_N = \frac{\beta^2}{2N^2} \sum_{\substack{i,j=1\\ i<j}}^{N} \mathbb{V}\mathrm{ar}[J_{ij}]. \tag{5.3}$$

Notice that, in the case of $(J_{ij})_{ij}$ i.i.d. Bernoulli random variables with fixed mean p, namely for the randomly dilute Curie–Weiss model, α_N is exactly α defined in (2.4), as in [22, Theorem 1.4].

The result in (5.2) shows that the mean metastable exit time $E^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tau^N_{\mathcal{B}_N}\right]$ is mainly concentrated around $\tilde{E}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tilde{\tau}^N_{\mathcal{B}_N}\right]$ (the corresponding quantity for the annealed model) up to multiplicative constants and random prefactors which are concentrated as exponential of sub-Gaussian random variables (see (2.5) for a definition).

In the second part of [20, Theorem 2.13] we showed that the q-th conditional moment of the mean metastable exit time $E^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tau^N_{\mathcal{B}_N}\right]$ is $\mathbb{P}$-a.s. bounded both from above and below by constants times $\tilde{E}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tilde{\tau}^N_{\mathcal{B}_N}\right]$. More precisely, we proved that for any fixed $q \geq 1$ and $c \in (0,\infty)$,

$$\mathrm{e}^{-\alpha_N}\left(1 - \tfrac{c}{\sqrt{N}}\right) \leq \frac{\mathbb{E}_{\mathcal{G}}\left[E^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tau^N_{\mathcal{B}_N}\right]^q\right]^{1/q}(\omega)}{\tilde{E}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\left[\tilde{\tau}^N_{\mathcal{B}_N}\right](\omega)} \leq \mathrm{e}^{4q\alpha_N}\left(1 + \tfrac{c}{\sqrt{N}}\right) \tag{5.4}$$

holds, definitely in N for $\mathbb{P}$-almost every $\omega \in \Omega$, where α_N is defined in (5.3).

5.4 Summary of the proofs

5.4.1 Key idea: control set on the Hamiltonians

One of the main issues in our proofs was the control on the quantity $|H_N(\cdot) - \tilde{H}_N(\cdot)|$, which is random and might very well be huge for some realisations of the random coupling coefficients. Our strategy to overcome this problem consists in defining the set $\Xi(a_N)$ ([19, Eq. (3.1)]) as the subset of Ω in which that quantity is uniformly bounded by a positive real a_N. Then we proved fine estimates within $\Xi(a_N)$. On the complement $\Xi(a_N)^c$ we were able to obtain only rough estimates. However, choosing suitably the sequence $(a_N)_{N\in\mathbb{N}}$ (for instance to be of order $\sqrt{N}$) yields nice estimates within the set $\Xi(a_N)$ and makes $\mathbb{P}(\Xi(a_N)^c)$ vanish sufficiently fast as $N \to \infty$ for our rough estimates within $\Xi(a_N)^c$ to be precise enough in the limit.

5.4.2 Proof of metastability

By applying the Dirichlet principle, we proved the result on metastability [20, Theorem 2.10] on the intersection of the set $\Xi(a_N)$ (with some $(a_N)_{N\in\mathbb{N}}$ of order $\sqrt{N}$) and of the set where metastability of the annealed model holds. We proved the asymptotic result using Borel Cantelli's lemma together with the assumption of metastability definitely in N of the annealed model.

5.4.3 Mean metastable exit time estimates

Capacity and harmonic sum

The first step in the proof of [20, Theorem 2.13] consists in writing the mean hitting times as a ratio of harmonic sum and capacity, using (1.25), as usual within the potential-theoretic approach. Hence, we first focused separately on estimating those two quantities, then combined them, and finally took the limit $N \to \infty$: only there, at the very end, we used that the annealed model is assumed to be metastable definitely in N with probability 1.

Two types of estimates

We computed the following two kinds of estimates both for capacities and harmonic sums, more precisely in the next formulas, both for $Y_N = Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})$ (for any two disjoint $\mathcal{X}, \mathcal{Y} \subset S_N$) and $Y_N = Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}$. As we shall see later the estimates on the latter will hold only when the process $(\Sigma_N(t))_{t\geq 0}$ is metastable, with the ones on the capacities hold on the whole Ω.

We use the $\sim$ notation as mentioned above.

1. *Concentration estimates* of sub-Gaussian type of the logarithms, i.e. bounds of the following form, with explicit $c > 0$, for any $t \geq 0$

$$\mathbb{P}_{\mathcal{G}}\Big[\big|\log Y_N - \mathbb{E}_{\mathcal{G}}[\log Y_N]\big| > t\Big] \leq 2\,\mathrm{e}^{-ct^2} + o(N). \tag{5.5}$$

   The right most term (vanishing for $N \to \infty$) on the right hand side of (5.5) is not present for the capacity estimates.

2. *"Annealed estimates"*, namely estimates comparing conditional means of quantities of the quenched model with quantities of the annealed model. Two types of annealed estimates are needed:

   - Difference of logarithms (needed for the result on the tails (5.2)):

$$\mathbb{E}_{\mathcal{G}}[\log Y_N] - \log \widetilde{Y}_N. \tag{5.6}$$

     We proved that the quantity in (5.6) for the capacity is equal to α_N (5.3) in absolute value and with an error of order $\frac{1}{\sqrt{N}}$, while for the harmonic sum it is contained in $[0, \alpha_N]$ with an error of order $\frac{1}{N}$;

   - Ratios of powers (needed for the result on the moments (5.4)). For any $q \in [1, \infty)$ there exists a $c \in (0, \infty)$ such that

$$\mathrm{e}^{-\alpha_N}\Big(1 - \frac{c}{N^u}\Big) \leq \frac{\mathbb{E}_{\mathcal{G}}\Big[Y_N^q\Big]^{1/q}}{\widetilde{Y}_N} \leq \mathrm{e}^{q\alpha_N}\Big(1 + \frac{c}{N^u}\Big), \tag{5.7}$$

     where $u = \frac{1}{2}$ for the capacities and $u = 1$ for the harmonic sum, and the constant c is different in the two cases. α_N is defined in (5.3).

Combining estimates of these two types, which hold for any finite $N \in \mathbb{N}$ sufficiently large, and taking the limit $N \to \infty$ yields our final results.

Two main technical tools

McDiarmid's inequality. The main tool used to prove both concentration inequalities is the McDiarmid inequality. Its conditional version provides almost sure sub-Gaussian type concentration for functions of conditionally independent random variables which satisfy a bounded differences condition. The exact statement can be found in [20, Appendix A].

The fact that, contrary to Talagrand's concentration inequality used in [22], no convexity is required allowed us to prove concentration directly on logarithms of capacities, whose convexity is hard to prove. In addition, remarkably, the constants in this inequality are given explicitly, which is an improvement compared to [22].

Invariant measure and transition rates. In [20, Lemma 4.2] we proved two results. First, the conditional expectation of invariant Gibbs measure is equal to e^{α_N} times the invariant Gibbs measure of the annealed model. Moreover, the conditional expectation of the transition rates of the quenched model is equal to e^{α_N} times the invariant Gibbs measure and the transition rates of the annealed model. These results hold up to an error vanishing in N. In the proofs we used conditional independence and boundedness of $(J_{ij})_{ij}$, Taylor expansions and Jensen's inequality ($\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X])$ for integrable real random variables X and convex real functions φ).

Capacity estimates

The capacity estimates we provided do not require any assumption on metastability and hold for any pair of disjoint sets $\mathcal{X}, \mathcal{Y} \subset S_N$. Moreover, in the proofs we did not use the set $\Xi(a_N)$.

We proved concentration (5.5) by applying McDiarmid's inequality, using the hypothesis of conditional independence of the sequence $(J_{ij})_{ij}$. The required bounded differences for $\log\left(Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\right)$ were proved using boundedness of the coupling coefficients and the Dirichlet principle (1.12).

To prove the annealed estimates ((5.6),(5.7)) we used [20, Lemma 4.2], Dirichlet and Thomson (1.16) principles, together with Jensen's inequality. For (5.7) we used also Minkowski's inequality ($\mathbb{E}[(X+Y)^q]^{1/q} \leq \mathbb{E}[X^q]^{1/q} + \mathbb{E}[Y^q]^{1/q}$, for any $q \in [1, \infty)$ and any two q-integrable random variables X, Y).

Harmonic sum estimates

We stress that all the estimates we proved for finite N for the harmonic sum hold, contrary to the generality of the capacity estimates, only in the subset of Ω in which the annealed model is metastable for that N.

The core of our harmonic sum estimates and *main novel technique* of this paper (together with the idea of $\Xi(a_N)$ and the use of McDiarmid inequality) is [20, Proposition 5.2], which strongly relies on definitions and results of Schichting and Slowik's [68]. In this proposition we proved that, on the event $\Xi(a_N)$, the harmonic sum of the quenched model localises on the metastable valley of $\mathcal{A}_N$ (see Definition 1.1.5), if the annealed model is metastable, a_N is sublinear in N and N is large enough. In order to prove this we followed mostly the line of proof of [68, Lemma 3.3] which uses reversibility, [68, Lemma 3.1] and the non-degeneracy assumption. This is the only point in which we used that assumption.

Similarly to the capacity case, to show concentration (5.5) we proved first bounded differences, using the localisation of the harmonics sum on $\Xi(a_N)$ for some a_N of order $\sqrt{N}$, in addition to boundedness of the coupling coefficients. We concluded the proof by McDiarmid inequality using conditional independence of $(J_{ij})_{ij}$.

To prove the annealed estimates ((5.6),(5.7)), in addition to the localisation of the harmonic sum on $\Xi(a_N)$ for some a_N of order $\sqrt{N}$, we applied [20, Lemma 4.2], together with Jensen's and Minkowski's inequalities.

5.5 Discussion

This paper represents a large development in the study of sharp asympotics on metastability of systems in inhomogeneous random environments. Before our paper this field was largely unexplored mainly due to technical difficulties in the computation of harmonic sums. These problems are related to the fact that those models are not lumpable because the inhomogeneity involves pairs of spins and thus the Hamiltonians depend highly on every spin and on the realisation of the random environment. Dealing instead with inhomogeneity on single spins, as in the random field Curie–Weiss model or the "separable" model studied in [19] (see Chapter 4, (4.4)), is much easier via coarse-graining techniques (see Section 1.6.1).

The generality of our theorems is remarkable because in the literature most results for metastability in random environments are given for specific models. The breakthrough to achieve this generality was given by the results by Schlichting and Slowik in [68], which allow one easier and much model independent estimates of the harmonic sum which was the bottleneck in the application of the general potential-theoretic approach (in particular formula (1.25).

We conclude our discussion explaining why we dealt with conditional independence and expectations, instead of the simple ones as in [22]. We made this choice for our results to hold for a wide class of models, including randomly diluted models. Looking at the example of the randomly diluted Hopfield model (RDH) helps visualising why. One would like to compare RDH with the undiluted Hopfield model, which has *random* interaction coefficients which are *not independent*. Conditioning to a non-trivial σ-algebra in our setting had two consequences. Defining the annealed model using a conditional mean on the random couplings allowed us to average only partially on the randomness of interaction coefficients. Thus, we obtained an annealed model which is "less random" than the quenched but not necessarily fully deterministic, as it was in [22], where the annealed model was the Curie–Weiss model. Moreover, assuming conditional independence of the interaction coefficients allowed us to include models whose dilution terms are independent but the coupling coefficients of their annealed version are not.

5.6 Contribution

I largely contributed in merging two independent original projects which dealt with the randomly dilute Hopfield model and with the Ising model on inhomogeneous random graphs, respectively, and in obtaining the current general formulation which includes both models. Moreover, I gave a contribution in structuring of the paper, writing introduction, explanations and remarks. I also helped in checking, simplifying and writing hypotheses and proofs, especially for metastability and harmonic sum estimates.

Chapter 6

Summary

This Ph.D. thesis contains results on metastability of a large class of Ising-like spin models with random interaction coefficients, obtained using the potential-theoretic approach.

We started this thesis with a basic introduction to the phenomenon of metastability, which included two formal definitions and a brief overview of the main approaches used to study it. Subsequently, we focused on the potential-theoretic approach to metastability, established by Bovier, Eckhoff, Gayrard and Klein starting with the paper [12]. We gave a sketch of the first developments in the approach and provided essential elements of potential theory. The most relevant quantities are capacity and equilibrium potential. We presented the main formulas we used in our research, including some well-known variational principles characterising capacities, which represent a core element in the potential-theoretic approach.

We proceeded by introducing the Ising and Curie–Weiss models. After providing a brief review on the origin of the models for describing ferromagnetism, we dived more into detail of their mathematical formulation. This set the basic ground for presenting our models of interest and our results. Afterwards, we mentioned some previous related work which constitutes the context of our research. We concluded the introductory chapter presenting two relevant techniques, coarse-graining and lumping, which were essential at various levels in this work.

In this thesis we summarised the work contained in three papers. The first two papers have been published, while the third one has appeared only on `arxiv.org` as a preprint and it has not been peer-reviewed yet:

1. [22] *Metastability for the dilute Curie–Weiss model with Glauber dynamics*, joint work with Anton Bovier and Elena Pulvirenti;
2. [19] *Metastability for Glauber dynamics on the complete graph with coupling disorder*, joint work with Anton Bovier and Frank den Hollander;
3. [20] *Metastability of Glauber dynamics with inhomogeneous coupling disorder*, joint work with Anton Bovier, Frank den Hollander, Elena Pulvirenti and Martin Slowik.

Furthermore, we added an unpublished slight extension of some techniques used in [22].

In our papers we studied metastability of random modifications of Ising models. More in detail, we studied reversible Markov processes which model the evolution of spin systems of size N, which are immersed in a constant external magnetic field at fixed temperature. These Markov processes evolve with Glauber dynamics (with Metropolis transition rates), which allows only one spin to flip at a time, and have a Gibbs measure as invariant measure. Our results hold for systems of large size.

The Hamiltonian which characterises our models of interest is a random modification of the Ising model Hamiltonian. We modified only the spin interaction. We took random, possibly inhomogeneous, interaction coefficients $(J_{ij})_{ij}$, with a $\frac{1}{N}$ rescaling as in the Curie–Weiss model, to avoid the Hamiltonian to attained values larger than order N. The only earlier work in which metastability of these models was studied is den Hollander and Jovanovski's [50].

A peculiar feature of these models is the presence of more than one level of randomness. The first one, present also in the Curie–Weiss and Ising models, is the randomness of the Markov process which is by definition random. An additional level is added by having random coefficients in the Hamiltonian, which characterises the transition rates of the Markov process. Depending on the definition of the random coefficients, more randomness could be added, e.g. using random coefficients having random mean. Such cases are included in [20].

The Hamiltonian depends highly on all the spins in a configuration because the randomness is placed in its interaction terms (namely "on the edges", if we think in terms of the interaction graph). Hence, in most cases coarse graining techniques cannot be applied and, even if the random coefficients have finite support, the models we consider are usually not lumpable. For these reasons well-established techniques, which rely on model reduction, cannot be used and studying metastability of those models directly is very difficult.

To overcome these problems, our main strategy in [22] and [20] consisted in studying those models by comparison with simpler models, meaning models which have less randomness in their Hamiltonian. Indeed, we compare the metastable behaviour of models which have interaction coefficients $(J_{ij})_{ij}$ with the one of models with interaction coefficients $(\mathbb{E}[J_{ij}])_{ij}$, where $\mathbb{E}$ is some (possibly conditional) mean. Our results show how close these two behaviours are for large systems.

We conclude this summary with brief reviews of the results presented in the central chapters of this thesis.

Review of [22] and partial extension

As summarised in Chapter 2, in the first paper [22] we studied the metastable behaviour of the randomly dilute Curie–Weiss model (RDCW), which can be seen as the Ising model on the Erdős–Rényi random graph, with fixed edge probability $p \in (0, 1)$. In this case the interaction coefficients are i.i.d. Bernoulli random variables with fixed mean p, multiplied by $\frac{1}{p}$. We compared this model with the well-known Curie–Weiss model (CW), whose Hamiltonian is the mean of RDCW. Assuming metastability of the Curie–Weiss model, we provided asymptotic estimates on the mean time it takes to the RDCW, starting in the metastable set of the CW with last-exit biased distribution, to hit the stable set of CW. This quantity can be written as a product of the corresponding well-known deterministic mean hitting time of CW times a prefactor. We proved that, asymptotically as $N \to \infty$, this prefactor is bounded by terms of order constant times the exponential of a sub-Gaussian random variable.

Den Hollander and Jovanovski studied in [50] the same model and quantities, using the pathwise approach to metastability. They proved a weaker result on the random prefactor: they showed that it is at most polynomial in N. However, their result is uniform in the initial configuration within the starting set.

Following the potential-theoretic approach, to obtain our estimates we first wrote the mean hitting times as a ratio of harmonic sum and capacity. The main novelty in our paper consists in writing capacity and harmonic sum in terms of the corresponding deterministic quantities of the Curie–Weiss model and in being able to isolate and characterise the random correction term.

In order to do that we estimated a generalised partition function, involving the difference of the Hamiltonian of RDCW and CW. Using Talagrand's concentration inequality and a second moment method, established as well by Talagrand, we proved that the generalised partition function is a product of an exponential of a sub-Gaussian random variable and an order constant factor.

We concluded our proofs using this fact together with well-known methods. Upper and lower bounds on capacities were obtained via Dirichlet and Thomson principles, using as test function and test flow the optimal ones of the Curie–Weiss model. To estimate harmonic sums we used the technique by Bianchi, Bovier and Ioffe [7], which consists in proving that the most relevant terms of the harmonic sum are concentrates around the local minimum of the Curie–Weiss model free energy.

The unpublished work presented in Chapter 3 extends the results on the generalised partition function to a more general setting. This extension yields, together with Dirichlet and Thompson principles, to bounds on general bounds on capacities similar to the ones obtained in [22]. However, it is important to stress that these results are of no help to estimate harmonic sums. Hence, more sophisticated techniques are needed to fully extend [22], as was done later in [20].

Review of [19]

In our second paper [19], summarised in Chapter 4, we studied metastability of the averaged (or annealed) model of the Ising model on a Chung–Lu-like random graph. The interaction coefficients of the latter model are independent Bernoulli random variables with mean the product of weights on vertices. Thus, in [19] we considered a model with product interaction coefficients. More in detail, we choose $J_{ij} = J(i)J(j)$, where $(J(i))_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with finite support.

The interest in studying metastability of this model comes from the project of extending the results of the first paper [22] to the Ising model on further random graphs. Such an extension, carried out later in [20], allows one to have information about metastability of the Ising model on a random graph from knowledge on the metastable behaviour of its averaged model (the model obtained averaging its Hamiltonian). Thus, [19] makes the Ising model on a Chung–Lu-like random graph a new example of model to which the results in [20] can be effectively applied. Notice that this example is more interesting that the one of the Ising model on Erdős–Rényi random graph because of the inhomogeneity of the interactions and the double randomness of the averaged model.

In [19] we proved results on metastable regime, mean metastable exit times and distribution of the metastable exit times.

Studying the mean metastable exit times and their distribution was done with well-known techniques which explained in the monograph by Bovier and den Hollander [18, Chapters 9, 10, 14] and were used already by Bovier, Eckhoff, Gayrard and Klein [12, 13]. Besides providing a new example of their application, the contribution of this part of the paper consists in adapting those methods for use on models with random interactions and in continuous time.

Besides this, which is standard, we managed to characterise asymptotically, up to order constant, the leading randomness of the exponent in the mean metastable exit time. We proved that it is equal, in distribution, to its deterministic limit plus $\frac{1}{\sqrt{N}}$ times a centered normal random variable, whose variance can be explicitly computed.

The most interesting results in [19] are related to the metastable regime. We found the critical inverse temperature, and limit and upper bound of the critical external magnetic field. Moreover, we showed numerically that, in some cases, the system presents multiple metastable states, up to $k+1$, where k is the cardinality of the support of $J(\cdot)$. Furthermore, we found numerical evidence of a re-entrant phase transition. This means that the critical external magnetic field can be not increasing, which, for instance, does not happen in the Curie–Weiss model.

These results show that the metastable regime and its analysis become more complicated already when a simple product structure disorder is added to the system. As a consequence, we infer that the development of new more accurate and computationally efficient techniques might be needed to analyse in detail more complicated systems.

Review of [20]

As summarised in Chapter 5, in our third paper [20] we studied metastability of a large class of Ising-like models, extending the results obtained on the Ising model on an Erdős–Rényi random graph in our first paper [22]. In [20] the interaction coefficients are assumed to be independent, conditionally to a given σ-algebra $\mathcal{G}$, and uniformly bounded. The models considered include Ising models on (dense) random graphs, with independent edge probabilities, like the Ising models on Chung–Lu-like random graphs and on the Erdős–Rényi random graph, and randomly diluted models, like the randomly diluted Hopfield model.

Our results compare the metastable behaviour of a model, called quenched model, with the one of its averaged (or annealed) model, asymptotically in the size N of the system. The annealed model is the model obtained by averaging (conditionally with respect to $\mathcal{G}$) the interaction coefficients in the Hamiltonian of the quenched model.

We proved three results. First, if the annealed model is metastable with respect to some sets, then the quenched model is metastable with respect to the same sets. In addition, we wrote the mean metastable exit times of the quenched model as the same quantity of the annealed model times a prefactor, which we proved to be asymptotically of order constant times the exponential of a sub-Gaussian random variable. These estimates are similar to the ones in [22] for the particular case RDCW, with the advantage of explicit constants. Furthermore, we estimated upper and lower bounds of order constant on the conditional moments of the mean metastable exit of the quenched model divided by the mean metastable exit time of the annealed model.

We stress that in [20] we use the definition of metastability given by Schlichting and Slowik in [68] and continue this review by mentioning the techniques used in our proofs.

After applying the standard formula to write the mean metastable exit times as harmonic sum divided by capacity, our strategy consisted in finding two types of estimates on both those quantities. We proved concentration estimates and so-called annealed estimates (comparing conditional mean of a quantity of the quenched model with the same quantity of the annealed model).

In this paper, in addition to well-established variational principles for capacities and standard inequalities, we used the following three innovative techniques.

The main tool used to prove concentration inequalities is a conditional version of McDiarmid's inequality, which has milder hypothesis and higher precision than the Talagrand's inequality used in [22] for similar purposes.

Moreover, for results on metastability and estimates on the harmonic sum we used the following scheme to overcome difficulties on controlling the random difference of the annealed and quenched model Hamiltonians. We first proved fine results for finite large N on a specific control set, in which that difference is bounded by an N-dependent sequence. For diverging sequences, the probability of being in the control set tends exponentially fast to 1 in limit

$N \to \infty$. Thus, choosing suitably that sequence, very rough estimates on the complement of the control set turn out to be sufficient to obtain sharp results in the limit.

Furthermore, to estimate the harmonic sum we widely used the results and model-independent techniques by Schlichting and Slowik [68]. The key point is the localisation of the harmonic sum on the valley of the starting metastable set. The generality and precision of our results in [20] show how powerful their methods are for finding estimates on the harmonic sum, which were the main issue in studying metastability of non-lumpable spin system.

Appendix A

Publication: Metastability for the dilute Curie-Weiss model with Glauber dynamics

This appendix reproduces exactly the content of the paper [22] with title "Metastability for the dilute Curie-Weiss model with Glauber dynamics", authored by Anton Bovier, Saeda Marello and Elena Pulvirenti, and published in *Electronic Journal of Probability, Institute of Mathematical Statistics and Bernoulli Society, 2021, 26, 1-38*, `https://doi.org/10.1214/21-EJP610`.
This paper was summarised in Chapter 2.

A.1 Introduction and main results

The randomly dilute Curie–Weiss model (RDCW) is a classical model of a disordered ferromagnet and was studied, e.g. in Bovier and Gayrard [15]. It generalises the standard Curie–Weiss model (CW) in that the fixed interactions between each pair of spins is replaced by independent, identically distributed, random ferromagnetic couplings between any pair of spins. In Bovier and Gayrard [15] it is proven that the RDCW free energy converges, in the thermodynamic limit, to that of the CW model, under some assumptions on the coupling distribution. Their result relies on the fact that the RDCW Hamiltonian can be approximated by that of the CW model up to a small perturbation which can be uniformly bounded in high probability. In the last decade the RDCW model have gained again some attention and various results at equilibrium have been proven, both in the annealed and quenched case. De Sanctis and Guerra [31] give an exact expression of the free energy first in the high temperature and low connectivity regime, and then at zero temperature. The control of the fluctuations of the magnetisation in the high temperature limit is addressed by De Sanctis [30], while recently Kabluchko, Löwe and Schubert [53] prove a quenched Central Limit Theorem for the magnetisation in the high temperature regime.

One of the features which make these random systems with "bond disorder" very appealing is their deep connection with the theory of *random graphs*, which attracted great interest in the last years due to their application to real-world networks. Indeed, if the random couplings are chosen as i.i.d. Bernoulli random variables with mean p, one can view the model as a spin system on an Erdős–Rényi random graph with *fixed* edge probability p, which makes it a dense graph. There has been an extensive study of the Ising model at equilibrium on different kinds of random graphs, e.g. in Dembo, Montanari [32] and Dommers, Giardinà, van der Hofstad [37], where several thermodynamic quantities were analysed when the graph size tends to infinity. These results were all obtained for sparse graphs which have a locally tree-like structure. We refer to van der Hofstad [48] for a general overview of these results.

In contrast to the substantial body of literature on the equilibrium properties of the RDCW model, much less is known about its dynamical properties. The present paper focuses on the phenomenon of *metastability* for the RDCW model where, for simplicity, the couplings are Bernoulli distributed with *fixed* parameter $p \in (0,1)$, independent of the number of vertices N, and the system evolves according to a Glauber dynamics. In particular, we give a precise estimate of the mean transition time from a certain probability distribution on the *metastable state* (called the last-exit biased distribution) to the *stable state*, when the external magnetic field is small enough and positive and when N tends to infinity. We obtain asymptotic bounds on the probability of the event that the average time is close to the CW one times some constants of order 1 which depend on the parameters of the system.

In the context of metastability for interacting particle systems on random graphs, progress has been made for the case of the random regular graph, analysed by Dommers [35] and for the configuration model, studied by Dommers, den Hollander, Jovanovski, and Nardi [38], both subject to Glauber dynamics, in the limit as the temperature tends to zero and the number of vertices is fixed. Both are dealing with sparse random graphs. In [50] den Hollander and Jovanovski investigate the same model considered in the present paper and obtain estimates on the average crossover time for fixed temperature in the thermodynamic limit. They show that, with high probability, the exponential term is the same as in the CW model, while the multiplicative term is polynomial in N. Their analysis relies on coupling arguments and on the *pathwise approach* to metastability. This method uses large deviations techniques in path space and focuses on properties of typical paths in the spirit of Freidlin-Wentzell theory. We refer to the classical book by Olivieri and Vares [65] for an overview on this method.

In contrast, in the present paper, we use the *potential theoretic approach* initiated by Bovier, Eckhoff, Gayrard and Klein in a series of papers [12, 13, 14] (see the monograph of Bovier and den Hollander [18] for an in-depth review of this as well as other approaches). This method gives less information on the evolution of the system, but leads to more precise estimates of the metastable transition time. It has been successfully applied to a large variety of systems such as the random field CW model, where the external magnetic field is given by i.i.d. random variables, first by Bovier, Eckhoff, Gayrard and Klein in [12] and later by Bianchi, Bovier and Ioffe in [7]. Furthermore, inspired by the results of Bovier and Gayrard [15], namely that the equilibrium properties of the RDCW model are very close to those of the CW model, we observe that, using Talagrand's concentration inequality, the mesoscopic measure can be expressed in terms of that of CW.

Before stating our results we give a precise definition of the model.

A.1.1 Glauber dynamics for the RDCW model

Let $[N] = \{1, ..., N\}$, $N \in \mathbb{N}$, be a set of vertices. To each vertex $i \in [N]$ an Ising spin σ_i with values in $\{-1, +1\}$ is associated. We denote by $\sigma = \{\sigma_i : i \in [N], \sigma_i \in \{-1, +1\}\}$ a spin configuration and we define the state space $\mathcal{S}_N = \{-1, +1\}^N$ to be the set of all such configurations σ. We fix a probability $p \in (0,1)$. Then the *randomly dilute Curie–Weiss* model

(RDCW) has the following *random Hamiltonian* $H_N : \mathcal{S}_N \to \mathbb{R}$

$$H_N(\sigma) = -\frac{1}{Np} \sum_{1 \le i < j \le N} J_{ij}\sigma_i\sigma_j - h \sum_{i \in [N]} \sigma_i, \tag{A.1}$$

where $h \in \mathbb{R}$ represents an external constant magnetic field, while J_{ij}/Np is a ferromagnetic random coupling. In particular, $\{J_{ij}\}_{i,j\in[N]}$ is a sequence of i.i.d. random variables with $J_{ij} \sim \text{Ber}(p)$ and $J_{ij} = J_{ji}$.

Let us denote by $\mathbb{P}_J$ the joint probability distribution of the the random couplings J_{ij} with $i, j \in [N]$ and by $\mathbb{E}$ the corresponding mean value.

The RDCW model can be seen as the Ising model on the Erdős–Rényi random graph with vertex set $[N]$, edge set E and edge probability $p \in (0, 1)$ (see van der Hofstad [47] for a general overview on random graphs). In this picture the Hamiltonian can also be written as

$$H_N(\sigma) = -\frac{1}{Np} \sum_{\{i,j\} \in E} \sigma_i\sigma_j - h \sum_{i \in [N]} \sigma_i. \tag{A.2}$$

The Gibbs measure associated to the random Hamiltonian H_N is

$$\mu_{\beta,N}(\sigma) = \frac{\mathrm{e}^{-\beta H_N(\sigma)}}{Z_{\beta,N}}, \quad \sigma \in \mathcal{S}_N, \tag{A.3}$$

where $\beta \in (0, \infty)$ is the inverse temperature and the partition function is defined as

$$Z_{\beta,N} = \sum_{\sigma \in \mathcal{S}_N} \mathrm{e}^{-\beta H_N(\sigma)}. \tag{A.4}$$

The Gibbs measure $\mu_{\beta,N}$ is the unique invariant (and reversible) measure for the (discrete time) Glauber dynamics on $\mathcal{S}_N$ with Metropolis transition probabilities

$$p_N(\sigma, \sigma') = \begin{cases} \frac{1}{N} \exp(-\beta[H_N(\sigma') - H_N(\sigma)]_+), & \text{if } \sigma \sim \sigma', \\ 1 - \sum_{\eta \ne \sigma} p_N(\sigma, \eta), & \text{if } \sigma = \sigma', \\ 0, & \text{else,} \end{cases} \tag{A.5}$$

where $\sigma \sim \sigma'$ means $||\sigma - \sigma'|| = 2$ with $|| \cdot ||$ the ℓ_1-norm on $\mathcal{S}_N$, i.e. $\sigma \sim \sigma'$ if and only if σ' is obtained from σ by a single spin flip. We denote this Markov chain by $\{\sigma(t)\}_{t\ge0}$ and write $\mathbb{P}_\nu$ for the law of the process $\sigma(t)$ with initial distribution ν conditioned on the realisation of the random couplings. Analogously, $\mathbb{E}_\nu$ is the quenched expectation w.r.t. the Markov chain with initial distribution ν. Moreover, we set $\mathbb{P}_\sigma = \mathbb{P}_{\delta_\sigma}$. For any subset $A \subset \mathcal{S}_N$ we define the hitting time of A as

$$\tau_A = \inf\{t > 0 : \sigma_t \in A\}. \tag{A.6}$$

Notice that H_N, $\mu_{\beta,N}$ and p_N are random variables, with respect to the random realisation of the random variables $\{J_{ij}\}_{i,j\in[N]}$. In this paper the results involving these random variables hold pointwise, namely for every realisation of $\{J_{ij}\}_{i,j\in[N]}$, unless we specify it differently, as in our main theorems.

A.1.2 The Curie–Weiss model

Before stating the main results, we recall some results for the mean-field Curie–Weiss (CW) model (see e.g. Bovier and den Hollander [18, Section 13] and Bovier, Eckhoff, Gayrard and Klein [12]). The CW Hamiltonian $\tilde{H}_N$ can be obtained taking the mean value of (A.1) (namely, the first equality in (A.8) below). A simplifying feature of the CW model is that its Hamiltonian depends on the configuration $\sigma \in \mathcal{S}_N$ only through the empirical magnetisation $m_N : \mathcal{S}_N \to \Gamma_N$ defined as

$$m_N(\sigma) = \frac{1}{N}\sum_{i=1}^{N} \sigma_i \in \Gamma_N = \left\{-1, -1 + \tfrac{2}{N}, ..., 1 - \tfrac{2}{N}, 1\right\}. \tag{A.7}$$

From now on we will drop the dependency on N from the magnetisation. Then we can write

$$\tilde{H}_N(\sigma) = -\frac{1}{N}\sum_{1\le i<j\le N} \sigma_i\sigma_j - h\sum_{i\in[N]} \sigma_i = -N\left(\tfrac{1}{2}m(\sigma)^2 + hm(\sigma)\right) \tag{A.8}$$

and we can define, for any $m \in \Gamma_N$,

$$E(m) = -\tfrac{1}{2}m^2 - hm, \tag{A.9}$$

obtaining

$$\tilde{H}_N(\sigma) = NE(m(\sigma)). \tag{A.10}$$

The associated Gibbs measure is

$$\tilde{\mu}_{\beta,N}(\sigma) = \frac{\mathrm{e}^{-\beta NE(m(\sigma))}}{\tilde{Z}_{\beta,N}}, \quad \sigma \in \mathcal{S}_N, \tag{A.11}$$

where $\tilde{Z}_{\beta,N} = \sum_{\sigma\in\mathcal{S}_N} \mathrm{e}^{-\beta\tilde{H}_N(\sigma)}$ is the normalising partition function.

We denote the law of $m(\sigma)$ under the Gibbs measure by

$$\tilde{\mathcal{Q}}_{\beta,N} = \tilde{\mu}_{\beta,N} \circ m^{-1}. \tag{A.12}$$

Then

$$\tilde{\mathcal{Q}}_{\beta,N}(m) = \frac{\mathrm{e}^{-\beta NE(m)}}{\tilde{Z}_{\beta,N}}\sum_{\sigma\in\mathcal{S}_N} \mathbf{1}_{m(\sigma)=m} = \frac{\mathrm{e}^{-\beta NE(m)}}{\tilde{Z}_{\beta,N}}\binom{N}{\frac{1+m}{2}N} = \frac{\mathrm{e}^{-\beta Nf_{\beta,N}(m)}}{\tilde{Z}_{\beta,N}}, \tag{A.13}$$

where

$$f_{\beta,N}(m) = E(m) + \beta^{-1}I_N(m) = -\frac{m^2}{2} - hm + \beta^{-1}I_N(m) \tag{A.14}$$

is the finite volume *free energy*, while the *entropy* of the system is given by the following combinatorial coefficient

$$I_N(m) = -\frac{1}{N}\log\binom{N}{\frac{1+m}{2}N} \tag{A.15}$$

and it has the following properties: as $N \to \infty$,

$$I_N(m) \to I(m) \equiv \frac{1-m}{2}\log\frac{1-m}{2} + \frac{1+m}{2}\log\frac{1+m}{2}, \tag{A.16}$$

more precisely,

$$I_N(m) - I(m) = \frac{1}{2N}\ln\frac{1-m^2}{4} + \frac{\ln N + \ln(2\pi)}{2N} + O\left(\frac{1}{N^2}\right). \tag{A.17}$$

As reference see for example Bovier, Eckhoff, Gayrard and Klein [12, (7.18)].

Notice that the previous definitions imply

$$\tilde{\mu}_{\beta,N}(\sigma) = \tilde{\mathcal{Q}}_{\beta,N}(m(\sigma))\, \mathrm{e}^{NI_N(m(\sigma))}. \tag{A.18}$$

We use the notation $f_\beta(m) = \lim_{N\to\infty} f_{\beta,N}(m)$. We refer to Bovier and den Hollander [18, (13.2.6)] for more details on the following result.

Lemma A.1.1. *For $m \in (-1,1)$,*

$$\mathrm{e}^{-\beta N f_{\beta,N}(m)} = \mathrm{e}^{-\beta N f_\beta(m)}(1+o(1))\sqrt{\frac{2}{\pi N(1-m^2)}} \tag{A.19}$$

and for $m \in \{1,-1\}$, $f_{\beta,N}(m) = f_\beta(m)$.

Remark A.1.2. *Comparing our definitions and the literature (e.g. Bovier and den Hollander [18, Section 13.1]), one notices that the Gibbs measure is often defined with an additional factor 2^{-N}, corresponding to the reference measure. More precisely, the Gibbs measure would be $\tilde{\mu}_{\beta,N}(\sigma) = \frac{1}{\tilde{Z}_{\beta,N}}\mathrm{e}^{-\beta N E(m(\sigma))}2^{-N}$, where the partition function would be defined by $\sum_{\sigma\in\mathcal{S}_N}\mathrm{e}^{-\beta\tilde{H}_N(\sigma)}2^{-N}$. We preferred to discard the 2^{-N} from our definitions. Therefore, for consistency, our definition of I_N differs from the classical one by a factor 2^{-N} inside the logarithm, yielding a difference of* $\log(2)$ *in the limit in* (A.16) *with respect to Bovier and den Hollander [18, (13.1.14)] or Bovier, Eckhoff, Gayrard and Klein [12, (7.17)].*

We consider the Glauber dynamics associated to the CW Hamiltonian in analogy with (A.5) and with transition probabilities $\tilde{p}_N(\sigma,\sigma')$. A particular feature of this model is that the image process $m(t) \equiv m(\sigma(t))$ of the Markov process $\sigma(t)$ under the map m is again a Markov process on Γ_N, with transition probabilities

$$\tilde{r}_N(m,m') = \begin{cases} \exp(-\beta N[E(m')-E(m)]_+)\frac{(1-m)}{2} & \text{if } m' = m+\frac{2}{N}, \\ \exp(-\beta N[E(m')-E(m)]_+)\frac{(1+m)}{2} & \text{if } m' = m-\frac{2}{N}, \\ 0 & \text{else.} \end{cases} \tag{A.20}$$

The equilibrium CW model displays a phase transition. Namely, there is a critical value of the inverse temperature $\beta_c = 1$ such that, in the regime $\beta > \beta_c$, $h > 0$ and small, the free energy $f_\beta(m)$ is a double-well function with local minimisers m_-, m_+ and saddle point m^*. They are the solutions of equation $m = \tanh(\beta(m+h))$. Since $f_\beta(m_-) > f_\beta(m_+)$, the phase with m_- represents the metastable state, while m_+ represents the stable state for the system. Define $m_-(N), m^*(N), m_+(N)$ as the closest points in Γ_N to m_-, m^*, m_+ respectively, with respect to the Euclidean distance on $\mathbb{R}$. $\{m_-(N), m_+(N)\}$ form a metastable set in the sense of Definition 8.2 of Bovier and den Hollander [18]. Let $\mathbb{E}^{\mathrm{CW}}_{m_-(N)}$ be the expectation w.r.t. the Markov process $m(t)$ with transition probabilities $\tilde{r}_N$ and starting at $m_-(N)$. Then the following theorem holds.

Theorem A.1.3. *For $\beta > 1$ and $h > 0$ small enough, as $N \to \infty$,*

$$\begin{aligned} \mathbb{E}^{CW}_{m_-(N)}[\tau_{m_+(N)}] &= \exp\Big(\beta N\left[f_\beta(m^*) - f_\beta(m_-)\right]\Big) \\ &\quad \times \frac{\pi}{1+m^*}\sqrt{\frac{1-m^{*2}}{1-m_-^2}}\,\frac{N(1+o(1))}{\beta\sqrt{f''_\beta(m_-)\left(-f''_\beta(m^*)\right)}}. \end{aligned} \tag{A.21}$$

As a reference see Bovier and den Hollander [18, Theorem 13.1]. The difference of sign in the denominator with respect to our statement is due to the fact that their result holds for $h < 0$, while ours for $h > 0$.

We conclude this section by giving the explicit formula of the capacity for the CW model. The definition of *capacity* is given in (A.31), while its relation with the mean hitting time is given by the key relation (A.30). Let us denote, for any subset U of Γ_N, the set of configurations with magnetisation in U by

$$\mathcal{S}_N[U] = \{\sigma \in \mathcal{S}_N : m(\sigma) \in U\} \tag{A.22}$$

and for simplicity, for any $m \in \Gamma_N$, the set of configurations with given magnetisation m by $\mathcal{S}_N[m]$. Notice that $\mathcal{S}_N[m]$ has cardinality $\mathrm{e}^{-NI_N(m)}$, where $I_N(m)$ is defined in (A.15).

Then, the following formula,

$$\mathrm{cap}^{\mathrm{CW}}(\mathcal{S}_N[m_-(N)], \mathcal{S}_N[m_+(N)]) = \frac{1}{\tilde{Z}_{\beta,N}} \mathrm{e}^{-\beta N f_\beta(m^*)} \frac{\sqrt{\beta(-f''_\beta(m^*))}}{\pi N} \sqrt{\frac{1+m^*}{1-m^*}}(1+o(1)), \tag{A.23}$$

follows from standard arguments (see e.g. techniques used in the proof of Bovier and den Hollander [18, Theorem 13.1]).

A.1.3 Main results

For any $A, B \subset \mathcal{S}_N$ disjoint, we define the so-called *last-exit biased distribution* on A for the transition from A to B as

$$\nu_{A,B}(\sigma) = \frac{\mu_{\beta,N}(\sigma)\mathbb{P}_\sigma(\tau_B < \tau_A)}{\sum_{\sigma \in A} \mu_{\beta,N}(\sigma)\mathbb{P}_\sigma(\tau_B < \tau_A)}, \qquad \sigma \in A. \tag{A.24}$$

Since we are going to use $\nu_{A,B}$ on the sets $\mathcal{S}_N[m_-(N)], \mathcal{S}_N[m_+(N)]$ defined above, we introduce the following simplified notation

$$\nu^N_{m_-,m_+} = \nu_{\mathcal{S}_N[m_-(N)],\mathcal{S}_N[m_+(N)]}. \tag{A.25}$$

The following theorem gives a description of the dynamical properties of the RDCW model in the *metastable regime* where h is positive and small enough, $\beta > \beta_c = 1$ (β_c is the critical inverse temperature for the RDCW model) and N is going to infinity. We provide an estimate on the mean time it takes to the system, starting with initial distribution $\nu^N_{m_-,m_+}$, to reach $\mathcal{S}_N[m_+(N)]$. More precisely, we estimate, in the limit as $N \to \infty$, its ratio with the mean metastable exit time for the CW model to go from $m_-(N)$ to $m_+(N)$, providing constant upper and lower bounds independent of N. Because of the random interaction, the result is given in the form of tail bounds.

After recalling that notation $\mathbb{P}_J$ and $\mathbb{E}_\nu$ was introduced in Section A.1.1, while $\mathbb{E}^{\mathrm{CW}}_{m_-(N)}$ was introduced in Section A.1.2, we are ready to formulate our main theorem.

Theorem A.1.4 (Mean metastable exit time)**.** *For $\beta > 1$, $h > 0$ small enough and for $s > 0$, there exist absolute constants $k_1, k_2 > 0$ and $C_1(p,\beta) < C_2(p,\beta,h)$ independent of N, such that*

$$\lim_{N\uparrow\infty} \mathbb{P}_J \left(C_1 \mathrm{e}^{-s}(1+o(1)) \le \frac{\mathbb{E}_{\nu^N_{m_-,m_+}}\left[\tau_{\mathcal{S}_N[m_+(N)]}\right]}{\mathbb{E}^{CW}_{m_-(N)}\left[\tau_{m_+(N)}\right]} \le C_2 \mathrm{e}^{s}(1+o(1)) \right) \ge 1 - k_1 \mathrm{e}^{-k_2 s^2}. \tag{A.26}$$

The quantities C_1 and C_2 in the previous theorem can be explicitly written. Set

$$\alpha = \frac{\beta^2(1-p)}{4p}, \qquad \kappa = \alpha + \max_{\eta\in(0,1)} \left\{ \log\eta - \frac{\beta\sqrt{2\alpha + \log\left(\frac{c_1}{(1-\eta)^2}\right)}}{p\sqrt{2c_2}} \right\}, \tag{A.27}$$

where $c_1, c_2 > 0$ are absolute constants coming from Theorem A.2.8. It is easy to see that $\kappa < \alpha$. With this notation

$$C_1 = C_1(\beta, h, p) = \mathrm{e}^{-2\beta(1+h)-\alpha+\kappa}, \tag{A.28}$$

$$C_2 = C_2(\beta, h, p) = \mathrm{e}^{2\beta(1+h)+2\alpha}. \tag{A.29}$$

A.1.4 Proof of the main theorem

The proof of Theorem A.1.4 is based on the *potential theoretic approach* to metastability, which turns out to be a rather powerful tool to analyse the main object we are interested in, i.e. the mean hitting time of $\mathcal{S}_N[m_+(N)]$ for the system with initial distribution $\nu^N_{m_-,m_+}$. The general ideas of this approach were first introduced in a series of papers by Bovier, Eckhoff, Gayrard and Klein [12, 13, 14]. We refer to Bovier and den Hollander [18] for an overview on this method.

The crucial formula in the study of metastability is given by the following relation linking mean hitting time and *capacity* of two sets $A, B \in \mathcal{S}_N$, which can be found in Bovier and den Hollander [18, Eq. (7.1.41)]

$$\mathbb{E}_{\nu_{A,B}}[\tau_B] = \sum_{\sigma\in A} \nu_{A,B}(\sigma)\mathbb{E}_\sigma[\tau_B] = \frac{1}{\mathrm{cap}(A,B)} \sum_{\sigma'\in\mathcal{S}_N} \mu_{\beta,N}(\sigma')h_{AB}(\sigma'), \tag{A.30}$$

where the capacity, as in Bovier and den Hollander [18, (7.1.39)], is defined by

$$\mathrm{cap}(A,B) = \sum_{\sigma\in A} \mu_{\beta,N}(\sigma)\mathbb{P}_\sigma(\tau_B < \tau_A). \tag{A.31}$$

The function h_{AB} is called *harmonic function* and has the following probabilistic interpretation

$$h_{AB}(\sigma) = \begin{cases} \mathbb{P}_\sigma(\tau_A < \tau_B) & \sigma \in \mathcal{S}_N \setminus (A\cup B), \\ \mathbb{1}_A(\sigma) & \sigma \in A \cup B. \end{cases} \tag{A.32}$$

We refer to Bovier and den Hollander [18, Section 7.1.2] for further details on the latter quantities.

By (A.30), in order to estimate mean hitting times one needs estimates both on the capacity and on the harmonic function.

We prove bounds on the capacity of two sets $\mathcal{S}_N[m_1], \mathcal{S}_N[m_2]$, stated in the two following theorems.

Theorem A.1.5. *For any $m_1 \neq m_2 \in \Gamma_N$ and any $s > 0$, there exist absolute constants $k_1, k_2 > 0$ such that*

$$\mathbb{P}_J\left(\frac{Z_{\beta,N}\ \mathrm{cap}\left(\mathcal{S}_N[m_1], \mathcal{S}_N[m_2]\right)}{\tilde{Z}_{\beta,N}\ \mathrm{cap}^{CW}\left(\mathcal{S}_N[m_1], \mathcal{S}_N[m_2]\right)} \le \mathrm{e}^{s+2\beta(1+h)+\alpha}(1+o(1)) \right) \ge 1 - k_1\mathrm{e}^{-k_2 s^2}, \tag{A.33}$$

asymptotically as $N \to \infty$, where α is defined in (A.27).

Theorem A.1.6. *For any $m_1 \neq m_2 \in \Gamma_N$ and any $s > 0$, there exist absolute constants $k_1, k_2 > 0$ such that*

$$\mathbb{P}_J\left(\frac{Z_{\beta,N}\ \mathrm{cap}\left(\mathcal{S}_N[m_1],\mathcal{S}_N[m_2]\right)}{\tilde{Z}_{\beta,N}\ \mathrm{cap}^{CW}\left(\mathcal{S}_N[m_1],\mathcal{S}_N[m_2]\right)} \geq \mathrm{e}^{-(s+2\beta(1+h)+\alpha)}(1+o(1))\right) \geq 1-k_1\mathrm{e}^{-k_2 s^2}, \tag{A.34}$$

asymptotically as $N \to \infty$, where α is defined in (A.27)*.*

We state asymptotic upper and lower bounds on the sum over the harmonic function in the numerator of (A.30) in the following proposition. We used the simplified notation

$$h^N_{m_-,m_+} = h_{\mathcal{S}_N[m_-(N)],\mathcal{S}_N[m_+(N)]}. \tag{A.35}$$

Theorem A.1.7. *For any $s > 0$, there exist absolute constants $k_1, k_2 > 0$ such that*

$$\mathbb{P}_J\left(\sum_{\sigma\in\mathcal{S}_N}\mu_{\beta,N}(\sigma)h^N_{m_-,m_+}(\sigma) \leq \mathrm{e}^{\alpha+s}\,\frac{\exp\left(-\beta N f_\beta(m_-)\right)}{Z_{\beta,N}\sqrt{\left(1-m_-^2\right)\beta f''_\beta(m_-)}}\left(1+o(1)\right)\right) \geq 1-k_1\mathrm{e}^{-k_2 s^2}, \tag{A.36}$$

$$\mathbb{P}_J\left(\sum_{\sigma\in\mathcal{S}_N}\mu_{\beta,N}(\sigma)h^N_{m_-,m_+}(\sigma) \geq \mathrm{e}^{\kappa-s}\,\frac{\exp\left(-\beta N f_\beta(m_-)\right)}{Z_{\beta,N}\sqrt{\left(1-m_-^2\right)\beta f''_\beta(m_-)}}\left(1+o(1)\right)\right) \geq 1-k_1\mathrm{e}^{-k_2 s^2}, \tag{A.37}$$

asymptotically as $N \to \infty$, and where α and κ are defined in (A.27)*.*

We conclude this section using Theorems A.1.5-A.1.7, to prove the main theorem. First, we introduce the following notation which will be extensively used:

$$A \overset{P(s)}{\gtreqless} B \qquad \text{is equivalent to} \qquad \mathbb{P}_J(A \gtreqless B) \geq 1-k_1\mathrm{e}^{-k_2 s^2}, \tag{A.38}$$

for all $s > 0$ and for some absolute constants $k_1, k_2 > 0$, whose values might change along the paper.

Proof of Theorem A.1.4 . We prove here only the upper bound, as the lower bound follows similarly. More precisely, we prove

$$\frac{\mathbb{E}_{\nu^N_{m_-,m_+}}\left[\tau_{\mathcal{S}_N[m_+(N)]}\right]}{\mathbb{E}^{\mathrm{CW}}_{m_-(N)}\left[\tau_{m_+(N)}\right]} \overset{P(s)}{\leq} C_2\mathrm{e}^s. \tag{A.39}$$

We start from (A.30), which in our case reads

$$\mathbb{E}_{\nu^N_{m_-,m_+}}\left[\tau_{\mathcal{S}_N[m_+(N)]}\right] = \frac{\sum_{\sigma\in\mathcal{S}_N}\mu_{\beta,N}(\sigma)h^N_{m_-,m_+}(\sigma)}{\mathrm{cap}\left(\mathcal{S}_N[m_-(N)],\mathcal{S}_N[m_+(N)]\right)}. \tag{A.40}$$

From (A.36) we obtain

$$\mathbb{E}_{\nu^N_{m_-,m_+}}\left[\tau_{\mathcal{S}_N[m_+(N)]}\right] \overset{P(s)}{\leq} \frac{\mathrm{e}^{\alpha+s}\ \exp\left(-\beta N f_\beta(m_-)\right)\left(1+o(1)\right)}{Z_{\beta,N}\ \mathrm{cap}(\mathcal{S}_N[m_-(N)],\mathcal{S}_N[m_+(N)])\sqrt{(1-m_-^2)\beta f''_\beta(m_-)}}. \tag{A.41}$$

Via the lower bound on the capacity from Theorem A.1.6, we obtain

$$\begin{aligned}\mathbb{E}_{\nu^N_{m_-,m_+}}\left[\tau_{\mathcal{S}_N[m_+(N)]}\right] &\overset{P(s)}{\leq} \mathrm{e}^{2s+2\beta(1+h)+2\alpha}\sqrt{\frac{1-m^*}{1+m^*}}\,\frac{\pi\, N\,\exp\left(\beta N\left[f_\beta(m^*)-f_\beta(m_-)\right]\right)}{\beta\sqrt{(1-m_-^2)f''_\beta(m_-)(-f''_\beta(m^*))}}(1+o(1))\\ &= \mathrm{e}^{2s+2\beta(1+h)+2\alpha}\,\mathbb{E}^{\mathrm{CW}}_{m_-(N)}[\tau_{m_+(N)}],\end{aligned} \tag{A.42}$$

where we used (A.23) and Theorem A.1.3. □

A.1.5 Outline

The remainder of this paper is organised as follows. In Section A.2 we use the powerful *Talagrand's concentration inequality* to obtain bounds on the equilibrium measure of the RDCW model. These bounds allow us to write the RDCW *mesoscopic* measure in terms of the deterministic CW one, times a random factor which is the exponential of a sub-Gaussian random variable. In Section A.3 we give the proof of Theorems A.1.5 and A.1.6 via two dual variational principles, the Dirichlet and the Thomson principles, which are the building blocks of the potential theoretic approach to metastability. In obtaining upper and lower bounds on the capacity, the main strategy is to use the results of Section A.2 in order to recover the capacity of the CW model. In Section A.4 we prove Theorem A.1.7, i.e. we compute the asymptotics of the numerator in the formula for the mean hitting time using estimates on the harmonic function.

A.2 Equilibrium analysis via Talagrand's concentration inequality

In this section we prove that the equilibrium mesoscopic measure of the RDCW model is in fact very close to that of the CW model. This is done in two steps. First, we prove that the difference between the *random free energy* at fixed magnetisation and its average can be controlled via *Talagrand's concentration inequality*. Second, we find upper and lower bounds on the aforementioned average by estimating first and second moments of the partition function of the RDCW model at fixed magnetisation.

A.2.1 Mesoscopic measure and closeness to the CW model

We start by analysing the equilibrium measure of the RDCW model. The aim is to express the equilibrium measure $\mu_{\beta,N}$, defined in (A.3), in terms of the empirical magnetisation in order to obtain a *mesoscopic* description, as we did for the CW model in Section A.1.2. Let us define the measure $\mathcal{Q}_{\beta,N}$ on Γ_N, and let the partition function be its normalisation

$$\mathcal{Q}_{\beta,N}(\cdot) = \mu_{\beta,N}\circ m^{-1}(\cdot) = \sum_{\sigma\in\mathcal{S}_N[\cdot]}\mu_{\beta,N}(\sigma), \qquad Z_{\beta,N} = \sum_{m\in\Gamma_N}\mathcal{Q}_{\beta,N}(m). \tag{A.43}$$

A priori the Hamiltonian of the RDCW model is not only depending on m, but it depends of course on the whole spin configuration. Nonetheless, we will see later in this section that the mesoscopic measure $\mathcal{Q}_{\beta,N}$ can be written in terms of the mesoscopic measure $\tilde{\mathcal{Q}}_{\beta,N}$ of the standard CW model.

$$\mathbb{E}[H_N(\sigma)] = -\frac{1}{Np}\sum_{i<j}\mathbb{E}[J_{ij}]\sigma_i\sigma_j - h\sum_i\sigma_i = -\frac{p}{Np}\sum_{i<j}\sigma_i\sigma_j - h\sum_i\sigma_i = \tilde{H}_N(\sigma). \tag{A.44}$$

Therefore, we can split the Hamiltonian into the mean-field part and the remaining random part obtaining

$$H_N(\sigma) = \mathbb{E}[H_N(\sigma)] + \Delta_{N,p}(\sigma), \tag{A.45}$$

where, introducing the notation $\hat{J}_{ij} = J_{ij} - p$,

$$\Delta_{N,p}(\sigma) = H_N(\sigma) - \tilde{H}_N(\sigma) = -\frac{1}{Np}\sum_{i<j} \hat{J}_{ij}\sigma_i\sigma_j. \tag{A.46}$$

Note that $\Delta_{N,p}$ is a random variable with zero mean. In order to simplify the notation, we drop from now on the dependence on N and p, from $\Delta_{N,p}$. Next, we write the mesoscopic measure as

$$\mathcal{Q}_{\beta,N}(m) = \frac{1}{Z_{\beta,N}} \mathrm{e}^{-\beta N E(m)} \cdot \sum_{\sigma\in\mathcal{S}_N[m]} \mathrm{e}^{-\beta\Delta(\sigma)}, \tag{A.47}$$

where $E(m)$ is defined in (A.8).

We will now focus on proving bounds for functions of $\sum_{\sigma\in\mathcal{S}_N[m]} \mathrm{e}^{-\beta\Delta(\sigma)}$ more general than $\mathcal{Q}_{\beta,N}(m)$. These results will be fundamental to prove our main theorem in the following sections. We will come back to $\mathcal{Q}_{\beta,N}$ at the end of this section, proving its closeness to the CW correspondent $\tilde{\mathcal{Q}}_{\beta,N}$ as a consequence of those general results.

Let us introduce the following notation, where we drop the dependence on β for simplicity

$$\mathcal{Z}_{N,g} = \sum_{m\in\Gamma_N} g(m) \sum_{\sigma\in\mathcal{S}_N[m]} \mathrm{e}^{-\beta\Delta(\sigma)} = \exp\left(Np_{N,g}\right)\exp\left(N\left[F_{N,g} - p_{N,g}\right]\right), \tag{A.48}$$

$$F_{N,g} = \frac{1}{N}\log \mathcal{Z}_{N,g}, \tag{A.49}$$

$$p_{N,g} = \mathbb{E}(F_{N,g}), \tag{A.50}$$

where $g : \Gamma_N \to [0,\infty)$ is a function which may depend on N.

We are interested in finding precise estimates on $\mathcal{Z}_{N,g}$ by writing it in terms of the entropic exponential term $\mathrm{e}^{-NI_N(m)}$ times some random factor which takes into account the randomness of the couplings. We notice that $\mathcal{Z}_{N,g}$ is the product of a deterministic factor $\mathrm{e}^{Np_{N,g}}$ and a random factor $\mathrm{e}^{N(F_{N,g}-p_{N,g})}$.

We first characterise the random variable $N(F_{N,g} - p_{N,g})$ in the following Proposition.

Proposition A.2.1. *For any β, $t > 0$,*

$$\mathbb{P}_J\Big(|N(F_{N,g} - p_{N,g})| \geq t\Big) \leq c_1 \exp\Big(-\gamma t^2\Big), \tag{A.51}$$

where $\gamma \propto \frac{p^2}{\beta^2}$.

The previous result intuitively means that the *random* $F_{N,g}$ is in fact very well concentrated around its mean $p_{N,g}$.

As a second step we provide asymptotic bounds on the average of $F_{N,g}$, i.e. the *deterministic* term $p_{N,g}$.

Lemma A.2.2. *Asymptotically, as $N \to \infty$,*

$$p_{N,g} \leq \frac{\alpha}{N} + \frac{1}{N}\log\left(\sum_{m\in\Gamma_N} g(m)\exp\left(-NI_N(m)\right)\right) + o\left(\frac{1}{N}\right), \tag{A.52}$$

where $I_N(m)$ is defined in (A.15) *and α in* (A.27).

Lemma A.2.3. *Asymptotically, as $N \to \infty$,*

$$p_{N,g} \geq \frac{\kappa}{N} + \frac{1}{N} \log \left(\sum_{m \in \Gamma_N} g(m) \exp\left(-N I_N(m)\right) \right) + o\left(\frac{1}{N}\right), \tag{A.53}$$

where $I_N(m)$ is defined in (A.15) *and κ in* (A.27).

Proposition A.2.1 together with Lemmas A.2.2 and A.2.3 imply the following result.

Proposition A.2.4. *Asymptotically, as $N \to \infty$, we have*

$$\mathcal{Z}_{N,g} \leq \mathrm{e}^{\alpha} \left(\sum_{m \in \Gamma_N} g(m) \exp\left(-N I_N(m)\right) \right) \exp\left[N(F_{N,g} - p_{N,g})\right] (1 + o(1)), \tag{A.54}$$

and

$$\mathcal{Z}_{N,g} \geq \mathrm{e}^{\kappa} \left(\sum_{m \in \Gamma_N} g(m) \exp\left(-N I_N(m)\right) \right) \exp\left[N(F_{N,g} - p_{N,g})\right] (1 + o(1)), \tag{A.55}$$

where $\mathcal{Z}_{N,g}$ is defined in (A.48)*, α and κ in* (A.27)*, and $I_N(m)$ in* (A.15)*. Moreover, $N(F_{N,g} - p_{N,g})$ is a sub-Gaussian random variable with variance*

$$\mathbb{V}\mathrm{ar}\left[N(F_{N,g} - p_{N,g})\right] \leq \frac{c\,\beta^2}{p^2}, \tag{A.56}$$

where c is a positive constant.

We prove Proposition A.2.1 in Section A.2.2, and Lemmas A.2.2 and A.2.3 in Section A.2.3.
We are ready to state the main result of this section, as a corollary of Proposition A.2.1 and Proposition A.2.4.

Corollary A.2.5. *Asymptotically, as $N \to \infty$, using notation* (A.38)*, the following bounds hold for any $\beta > 0$ and any function $g : \Gamma_N \to [0, \infty)$*

$$\sum_{m \in \Gamma_N} g(m) \sum_{\sigma \in \mathcal{S}_N[m]} \mathrm{e}^{-\beta \Delta(\sigma)} \overset{P(s)}{\leq} \mathrm{e}^{s+\alpha} \left(\sum_{m \in \Gamma_N} g(m) \exp\left(-N I_N(m)\right) \right) (1 + o(1)), \tag{A.57}$$

$$\sum_{m \in \Gamma_N} g(m) \sum_{\sigma \in \mathcal{S}_N[m]} \mathrm{e}^{-\beta \Delta(\sigma)} \overset{P(s)}{\geq} \mathrm{e}^{-s+\kappa} \left(\sum_{m \in \Gamma_N} g(m) \exp\left(-N I_N(m)\right) \right) (1 + o(1)), \tag{A.58}$$

where α and κ are defined in (A.27)*, $I_N(m)$ in* (A.15) *and $\Delta(\sigma)$ in* (A.46).

Proof. By Proposition A.2.1 we obtain, for any fixed $s > 0$,

$$\exp\left[N(F_{N,g} - p_{N,g})\right] \overset{P(s)}{\leq} \mathrm{e}^{s} \quad \text{and} \quad \exp\left[N(F_{N,g} - p_{N,g})\right] \overset{P(s)}{\geq} \mathrm{e}^{-s}, \tag{A.59}$$

where $k_1, k_2 > 0$ are absolute constants.
To conclude the proof it is sufficient to use the definition of $\mathcal{Z}_{N,g}$ (A.48) and Proposition A.2.4.
□

Remark A.2.6. *The exact same statement of Corollary A.2.5 holds replacing* $e^{-\beta\Delta(\sigma)}$ *with* $e^{\beta\Delta(\sigma)}$. *The proof remains the same: the Lipschitz constant for the Talagrand concentration inequality (in Section A.2.2) is the same and the change of sign, being squared, disappears from* (A.72) *onwards.*

We conclude this section with an immediate application of Corollary A.2.5 which states the closeness of the random mesoscopic measure $\mathcal{Q}_{\beta,N}$ to the correspondent deterministic CW quantity $\tilde{\mathcal{Q}}_{\beta,N}$. This result will be widely used in Section A.4.

Corollary A.2.7. *Asymptotically, as* $N \to \infty$, *using notation* (A.38), *the following bounds hold for any fixed* $s > 0$ *and any function* $\bar{g} : \Gamma_N \to [0, \infty)$

$$\sum_{m\in\Gamma_N} \bar{g}(m)\, \mathcal{Q}_{\beta,N}(m) \overset{P(s)}{\leq} e^{s+\alpha} \frac{\tilde{Z}_{\beta,N}}{Z_{\beta,N}} \left(\sum_{m\in\Gamma_N} \bar{g}(m)\, \tilde{\mathcal{Q}}_{\beta,N}(m) \right) (1 + o(1))\,, \tag{A.60}$$

$$\begin{aligned}\sum_{m\in\Gamma_N} \bar{g}(m)\, \mathcal{Q}_{\beta,N}(m) \overset{P(s)}{\leq}\; & e^{s+\alpha} \frac{1}{Z_{\beta,N}} \left(\sum_{m\in\Gamma_N\setminus\{1,-1\}} \bar{g}(m) \exp\Big(- \beta N f_\beta(m)\Big) \sqrt{\frac{2}{\pi N(1-m^2)}} \right) (1 + o(1)) \\ & + e^{s+\alpha} \frac{1}{Z_{\beta,N}} \left(\sum_{m\in\{1,-1\}} \bar{g}(m) \exp\Big(- \beta N f_\beta(m)\Big) \right) (1 + o(1))\,,\end{aligned} \tag{A.61}$$

where α *and* κ *are defined in* (A.27).

Proof. Using (A.47) we obtain

$$\sum_{m\in\Gamma_N} \bar{g}(m)\, \mathcal{Q}_{\beta,N}(m) = \frac{1}{Z_{\beta,N}} \sum_{m\in\Gamma_N} \bar{g}(m)\, e^{-\beta N E(m)} \sum_{\sigma\in\mathcal{S}_N[m]} e^{-\beta\Delta(\sigma)}. \tag{A.62}$$

Now we can apply the upper bound in Corollary A.2.5, with $g(m) = \frac{1}{Z_{\beta,N}} \bar{g}(m)\, e^{-\beta N E(m)}$, to the right hand side of (A.62). We conclude the proof of (A.60) using the definition of $\tilde{\mathcal{Q}}_{\beta,N}$ (A.13) and (A.14).

(A.61) follows by (A.60) simply applying Lemma A.1.1. □

A.2.2 Sub-Gaussian bounds on the random term

Proposition A.2.1 follows from Talagrand's concentration inequality, which we cite for completeness in the version of Tao [74, Theorem 2.1.13].

Theorem A.2.8 (Talagrand concentration inequality)**.** *Let* $G : \mathbb{R}^M \to \mathbb{R}$ *be a* 1*-Lipschitz and convex function. Let* $M \in \mathbb{N}$, $X = (X_1, \ldots, X_M)$, *with* X_i *be independent r.v., uniformly bounded by* $K > 0$, *i.e.* $|X_i| \leq K$, *for every* $1 \leq i \leq M$. *Then, for any* $t \geq 0$,

$$\mathbb{P}\Big(|G(X) - \mathbb{E}G(X)| \geq tK\Big) \leq c_1 \exp\big(- c_2 t^2\big), \tag{A.63}$$

with positive absolute constants c_1, c_2.

Proof of Proposition A.2.1. We can apply Theorem A.2.8 to the free energies $F_{N,g}$ as a function of the N^2 coupling constants $\hat{J}_{ij}$. Indeed it is standard to see that $F_{N,g}$ is convex and Lipschitz continuous with constant $\frac{\beta}{Np\sqrt{2}}$ (see e.g. Talagrand [73, Corollary 2.2.5]). Thus, applying Theorem A.2.8 for $G = F_{N,g}\left(\frac{\beta}{Np\sqrt{2}}\right)^{-1}$ and $K = 1$, after defining $t' = t\frac{\beta}{Np\sqrt{2}}$ we obtain, for some positive constants c_1, c_2 and for any $t' \geq 0$,

$$\mathbb{P}_J\left(N|F_{N,g} - p_{N,g}| \geq t'\right) \leq c_1 \exp\left(-c_2\frac{2p^2}{\beta^2}t'^2\right), \tag{A.64}$$

concluding the proof of (A.51) and hence Proposition A.2.1. □

A.2.3 Asymptotic bounds on the deterministic term

In this section we prove first the upper bound on $p_{N,g}$ (Lemma A.2.2) and then the lower bound (Lemma A.2.3). The upper bound is obtained by estimates on the first moment of the random partition function $\mathcal{Z}_{N,g}$, while the lower bound is in the spirit of Talagrand [73, Theorem 2.2.1] and is more delicate. We will see that it involves also estimates on the second moment of the random partition function.

Proof of Lemma A.2.2. Observing that $\{\hat{J}_{ij}\}_{i,j\in[N]}$ defined in (A.46) are i.i.d. random variables such that $\mathbb{E}\hat{J}_{ij} = 0$, we easily obtain

$$\begin{aligned}\mathbb{E}[\mathcal{Z}_{N,g}] &= \sum_{m\in\Gamma_N} g(m) \sum_{\sigma\in\mathcal{S}_N[m]} \mathbb{E}\left(\exp\left[\frac{\beta}{Np}\sum_{i<j}\hat{J}_{ij}\sigma_i\sigma_j\right]\right) \\ &= \sum_{m\in\Gamma_N} g(m) \sum_{\sigma\in\mathcal{S}_N[m]} \prod_{i<j}\mathbb{E}\left(\exp\left[\frac{\beta}{Np}\hat{J}_{ij}\sigma_i\sigma_j\right]\right).\end{aligned} \tag{A.65}$$

In order to find estimates for (A.65), we first define

$$\Phi(x) := \mathbb{E}[\exp(x\hat{J}_{ij})], \tag{A.66}$$

which is a function independent of i, j, being $\{\hat{J}_{ij}\}_{i,j}$ i.i.d., with first and second derivatives

$$\Phi'(0) = \mathbb{E}\hat{J}_{ij} = 0, \tag{A.67}$$

$$\Phi''(0) = \mathbb{E}\hat{J}_{ij}^2 = p(1-p). \tag{A.68}$$

Performing a Taylor expansion of Φ we get

$$\Phi(x) = \Phi(0) + x\,\Phi'(0) + \frac{x^2}{2}\Phi''(0) + o(x^2) = 1 + \frac{x^2}{2}p(1-p) + o(x^2). \tag{A.69}$$

Thus, we can exponentiate $\Phi(x)$ to obtain

$$\Phi(x) = \exp\Big(\log\left(\Phi(x)\right)\Big) = \exp\left(\frac{x^2}{2}p(1-p) + o(x^2)\right), \tag{A.70}$$

where we used the expansion $\log(1+x) = x + o(x)$. Therefore, for any sequence of coefficients x_{ij}^2 which are independent of i, j and σ, we have the following

$$\begin{aligned}
&\sum_{m\in\Gamma_N} g(m) \sum_{\sigma\in\mathcal{S}_N[m]} \prod_{i<j} \mathbb{E}\Big[\exp(x_{ij}\hat{J}_{ij})\Big] = \sum_{m\in\Gamma_N} g(m) \sum_{\sigma\in\mathcal{S}_N[m]} \prod_{i<j} \Phi(x_{ij}) \\
&= \sum_{m\in\Gamma_N} g(m) \sum_{\sigma\in\mathcal{S}_N[m]} \prod_{i<j} \exp\left(\frac{x_{ij}^2}{2} p(1-p) + o(x_{ij}^2)\right) \\
&= \sum_{m\in\Gamma_N} g(m)\, \mathrm{e}^{-NI_N(m)} \exp\left(\frac{x_{ij}^2}{2} p(1-p) + o(x_{ij}^2)\right)^{N(N-1)/2} \\
&= \sum_{m\in\Gamma_N} g(m)\, \mathrm{e}^{-NI_N(m)} \exp\left(x_{ij}^2 p(1-p)\frac{N(N-1)}{4} + o\left(x_{ij}^2 N(N-1)\right)\right),
\end{aligned} \tag{A.71}$$

asymptotically, for $x_{ij} \to 0$, where the third equality holds only if x_{ij}^2 is independent of i, j and σ. Moreover, we used that the cardinality of $\mathcal{S}_N[m]$ is $\mathrm{e}^{-NI_N(m)}$, where $I_N(m)$ is defined in (A.15), and the cardinality of $\{(i,j) \in [N]^2 : i < j\}$ is $\frac{N(N-1)}{2}$.

We can apply (A.71) with $x_{ij} = \frac{\beta}{Np}\sigma_i\sigma_j$ because x_{ij}^2 is independent of i, j and σ. Indeed $x_{ij}^2 = \frac{\beta^2}{N^2p^2}$, being $\sigma_i, \sigma_j \in \{-1, +1\}$ for any $i, j \in [N]$ and $\sigma \in \mathcal{S}_N$. Thus, we get, asymptotically as $N \to \infty$,

$$\begin{aligned}
\mathbb{E}\left[\mathcal{Z}_{N,g}\right] &= \sum_{m\in\Gamma_N} g(m)\, \mathrm{e}^{-NI_N(m)} \exp\left(\frac{\beta^2(1-p)}{4p} + o(1)\right) \\
&= \exp\left(\alpha + o(1)\right) \sum_{m\in\Gamma_N} g(m) \exp\left(-NI_N(m)\right),
\end{aligned} \tag{A.72}$$

where α is defined in (A.27).

Therefore, by Jensen's inequality and (A.72), we have

$$\mathbb{E}\Big[\log \mathcal{Z}_{N,g}\Big] \le \log\Big(\mathbb{E}[\mathcal{Z}_{N,g}]\Big) = \alpha + o(1) + \log\left(\sum_{m\in\Gamma_N} g(m) \exp\left(-NI_N(m)\right)\right) \tag{A.73}$$

which proves the upper bound. □

Proof of Lemma A.2.3. A key ingredient in the proof is to control the upper bound on the second moment of $\mathcal{Z}_{N,g}$, i.e. prove that the following bound holds

$$\mathbb{E}\left[\mathcal{Z}_{N,g}^2\right] \le \mathrm{e}^{2\alpha}\, \mathbb{E}\left[\mathcal{Z}_{N,g}\right]^2 (1 + o(1)), \tag{A.74}$$

where α is defined in (A.27).

We estimate $\mathbb{E}\left[\mathcal{Z}_{N,g}^2\right]$ using the first two lines of (A.71) with $x_{ij} = \frac{\beta}{Np}\left(\sigma_i^{(1)}\sigma_j^{(1)} + \sigma_i^{(2)}\sigma_j^{(2)}\right)$, which hold also when x_{ij}^2 is not independent on i, j and σ,

$$\begin{aligned}
\mathbb{E}\left[\mathcal{Z}_{N,g}^2\right] &= \mathbb{E}\left[\sum_{m,m'\in\Gamma_N} g(m)g(m') \sum_{\substack{\sigma^{(1)}\in\mathcal{S}_N[m],\\ \sigma^{(2)}\in\mathcal{S}_N[m']}} \exp\left(\sum_{i<j}\frac{\beta}{Np}\hat{J}_{ij}\left(\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)}\right)\right)\right]\\
&= \sum_{m,m'\in\Gamma_N} g(m)g(m')\,\mathbb{E}\left[\sum_{\substack{\sigma^{(1)}\in\mathcal{S}_N[m],\\ \sigma^{(2)}\in\mathcal{S}_N[m']}} \exp\left(\sum_{i<j}\frac{\beta}{Np}\hat{J}_{ij}\left(\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)}\right)\right)\right]\\
&= \sum_{m,m'\in\Gamma_N} g(m)g(m') \sum_{\substack{\sigma^{(1)}\in\mathcal{S}_N[m],\\ \sigma^{(2)}\in\mathcal{S}_N[m']}} \prod_{i<j}\exp\left(\frac{1}{2}\frac{\beta^2}{N^2p^2}(\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)})^2p(1-p)+o\left(\frac{\beta^2}{N^2}\right)\right)\\
&\le \sum_{m,m'\in\Gamma_N} g(m)g(m') \sum_{\substack{\sigma^{(1)}\in\mathcal{S}_N[m],\\ \sigma^{(2)}\in\mathcal{S}_N[m']}} \prod_{i<j}\exp\left(\frac{\beta^2}{N^2p}2(1-p)+o\left(\frac{1}{N^2}\right)\right)\\
&= \sum_{m,m'\in\Gamma_N} g(m)g(m')\,\mathrm{e}^{-NI_N(m)}\,\mathrm{e}^{-NI_N(m')}\exp\left(\frac{N(N-1)}{2}\left[\frac{\beta^2}{N^2p}2(1-p)+o\left(\frac{1}{N^2}\right)\right]\right)\\
&= \sum_{m,m'\in\Gamma_N} g(m)g(m')\,\mathrm{e}^{-NI_N(m)}\mathrm{e}^{-NI_N(m')}\exp\left(\beta^2\frac{(1-p)}{p}+o(1)\right)\\
&= \exp\left(4\alpha+o(1)\right)\sum_{m\in\Gamma_N} g(m)\,\mathrm{e}^{-NI_N(m)}\sum_{m'\in\Gamma_N} g(m')\,\mathrm{e}^{-NI_N(m')}\\
&= \mathrm{e}^{2\alpha}\,\mathbb{E}\left[\mathcal{Z}_{N,g}\right]^2\left(1+o(1)\right),
\end{aligned} \tag{A.75}$$

where, similarly to the last steps in (A.71), we used that the cardinality of $\mathcal{S}_N[m]$ is $\mathrm{e}^{-NI_N(m)}$, the cardinality of $\{(i,j)\in[N]^2 : i<j\}$ is $\frac{N(N-1)}{2}$. Moreover, in the last line we used (A.72).

We recall the Paley–Zygmund inequality, which states that

$$\mathbb{P}_J(X \ge \eta\,\mathbb{E}X) \ge (1-\eta)^2\frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}, \tag{A.76}$$

for any non negative random variable X and any $\eta \in (0,1)$. Using (A.76) with $X = \mathcal{Z}_{N,g}$, (A.72) and (A.75) we get, asymptotically as $N \to \infty$,

$$\begin{aligned}
&\mathbb{P}_J\left(\frac{1}{N}\log\mathcal{Z}_{N,g} \ge \frac{1}{N}\log\left(\eta\,\mathbb{E}\mathcal{Z}_{N,g}\right)\right)\\
&= \mathbb{P}_J\left(\frac{1}{N}\log\mathcal{Z}_{N,g} \ge \frac{1}{N}\log\left(\mathbb{E}\mathcal{Z}_{N,g}\right)+\frac{1}{N}\log\eta\right)\\
&\ge \frac{(1-\eta)^2}{\exp\left(2\alpha+o(1)\right)}.
\end{aligned} \tag{A.77}$$

Moreover, using (A.64) together with (A.49) and the change of variables $t' = Nt''$, we obtain

$\forall\ t'' > 0$,

$$\mathbb{P}_J\left(\left|\frac{1}{N}\log\mathcal{Z}_{N,g} - p_{N,g}\right| \geq t''\right) \leq c_1 \exp\left(-\frac{2c_2N^2p^2t''^2}{\beta^2}\right). \tag{A.78}$$

Thus, taking the complementary event, we get

$$\mathbb{P}_J\left(-t'' \leq \frac{1}{N}\log\mathcal{Z}_{N,g} - p_{N,g} \leq t''\right) \geq 1 - c_1 \exp\left(-\frac{2c_2N^2p^2t''^2}{\beta^2}\right). \tag{A.79}$$

Now, using

$$\mathbb{P}_J\left(\frac{1}{N}\log\mathcal{Z}_{N,g} - p_{N,g} \leq t''\right) \geq \mathbb{P}_J\left(-t'' \leq \frac{1}{N}\log\mathcal{Z}_{N,g} - p_{N,g} \leq t''\right) \tag{A.80}$$

and the change of variable $t = \frac{Np\sqrt{2c_2}}{\beta}t''$ we obtain

$$\mathbb{P}_J\left(\frac{1}{N}\log\mathcal{Z}_{N,g} \leq p_{N,g} + \frac{t\beta}{Np\sqrt{2c_2}}\right) \geq 1 - c_1\exp(-t^2). \tag{A.81}$$

Next we prove that the intersection of the events in (A.77) and (A.81) is non empty. *Assuming*, for $\eta \in (0,1)$, that

$$\mathbb{P}_J\left(\frac{1}{N}\log\mathcal{Z}_{N,g} \leq p_{N,g} + \frac{t\beta}{Np\sqrt{2c_2}}\right) > 1 - \frac{(1-\eta)^2}{\exp\left(2\alpha + o(1)\right)} \tag{A.82}$$

and comparing (A.77) and (A.82), we notice that the sum of the probabilities of the two events

$$\left\{\frac{1}{N}\log\mathcal{Z}_{N,g} \leq p_{N,g} + \frac{t\beta}{Np\sqrt{2c_2}}\right\}, \tag{A.83}$$

and

$$\left\{\frac{1}{N}\log\mathcal{Z}_{N,g} \geq \frac{1}{N}\log\left(\mathbb{E}\mathcal{Z}_{N,g}\right) + \frac{1}{N}\log\eta\right\} \tag{A.84}$$

is strictly greater than 1. Therefore, they intersect in the not empty event

$$\left\{\frac{1}{N}\log\left(\mathbb{E}\mathcal{Z}_{N,g}\right) + \frac{1}{N}\log\eta \leq \frac{1}{N}\log\mathcal{Z}_{N,g} \leq p_{N,g} + \frac{t\beta}{Np\sqrt{2c_2}}\right\} \tag{A.85}$$

which is contained in the deterministic set

$$\left\{\frac{1}{N}\log\left(\mathbb{E}\mathcal{Z}_{N,g}\right) + \frac{1}{N}\log\eta \leq p_{N,g} + \frac{t\beta}{Np\sqrt{2c_2}}\right\}. \tag{A.86}$$

As a consequence, the latter set is non empty and, being deterministic,

$$p_{N,g} \geq \frac{1}{N}\log\left(\mathbb{E}\mathcal{Z}_{N,g}\right) + \frac{1}{N}\log\eta - \frac{t\beta}{Np\sqrt{2c_2}} \tag{A.87}$$

holds with probability 1.

It remains to choose a suitable $t > 0$ for assumption (A.82) to hold. A sufficient condition is, for every $\eta \in (0,1)$,

$$c_1\exp(-t^2) < \frac{(1-\eta)^2}{\exp\left(2\alpha + o(1)\right)}, \tag{A.88}$$

namely

$$t^2 > 2\alpha + \log\left(\frac{c_1}{(1-\eta)^2}\right) + o(1). \tag{A.89}$$

Therefore, by (A.87) and (A.89), using (A.72) we obtain, for every $\eta \in (0,1)$,

$$\begin{aligned} p_{N,g} &\geq \frac{1}{N}\log\left(\mathbb{E}\mathcal{Z}_{N,g}\right) + \frac{1}{N}\log\eta - \frac{\beta\sqrt{2\alpha + \log\left(\frac{c_1}{(1-\eta)^2}\right) + o(1)}}{Np\sqrt{2c_2}} \\ &= \frac{1}{N}\log\left(\sum_{m\in\Gamma_N} g(m)\exp\left(-NI_N(m)\right)\right) + \frac{\kappa_\eta}{N} + o\left(\frac{1}{N}\right), \end{aligned} \tag{A.90}$$

where

$$\kappa_\eta = \alpha + \log\eta - \frac{\beta\sqrt{2\alpha + \log\left(\frac{c_1}{(1-\eta)^2}\right)}}{p\sqrt{2c_2}}. \tag{A.91}$$

Notice that $\kappa_\eta < \alpha$. In order to obtain the best lower bound, namely the closer to the upper bound proven in Lemma A.2.2, we choose $\eta \in (0,1)$ s.t. $\alpha - \kappa_\eta$ is minimised and we conclude the proof. This choice motivates the maximum in the definition of κ, in (A.27). □

A.3 Capacity estimates

This section is entirely devoted to obtain upper and lower bounds on capacities between sets with a fixed magnetisation. These bounds are obtained via two dual variational principles, i.e. the *Dirichlet* and *Thomson principles* which are extensively discussed in Bovier and den Hollander [18, Sections 7.3.1, 7.3.2]. The result will be expressed in terms of the capacity for the Curie–Weiss model, see (A.23). In particular, we prove Theorem A.1.5 in Section A.3.1 and Theorem A.1.6 in Section A.3.2.

A.3.1 Asymptotics on capacity: upper bound

In this section we prove Theorem A.1.5, obtaining the upper bound on the capacity of the RDCW model in terms of the capacity of the CW model.

Proof of Theorem A.1.5. The main idea of the proof is to find an upper bound on the capacity via the following Dirichlet principle (see Bovier and den Hollander [18, Section 7.3.1 and (7.1.29)] for details)

$$\mathrm{cap}\left(\mathcal{S}_N[m_1], \mathcal{S}_N[m_2]\right) = \min_{f\in\mathcal{H}} \sum_{\sigma,\sigma'\in\mathcal{S}_N} \mu_{\beta,N}(\sigma)p_N(\sigma,\sigma')[f(\sigma)-f(\sigma')]^2, \tag{A.92}$$

where

$$\mathcal{H} = \left\{f : \mathcal{S}_N \to [0,1] \text{ s.t. } f|_{\mathcal{S}_N[m_1]} = 1, f|_{\mathcal{S}_N[m_2]} = 0\right\}. \tag{A.93}$$

Later it will be clear that we can restrict the previous variational principle over the functions on the space Γ_N, hence it is useful to define

$$\tilde{\mathcal{H}} = \left\{v : \Gamma_N \to [0,1] \text{ s.t. } v(m_1) = 1, v(m_2) = 0\right\}. \tag{A.94}$$

In order to simplify the notation we will often neglect the dependency on m_1, m_2 when this will not generate confusion.

From (A.92), in view of (A.5) and since $[f(\sigma)-f(\sigma')]$ vanishes for $\sigma=\sigma'$, we are left only with the terms such that $\sigma\sim\sigma'$ and obtain the following first equality in (A.95). The second equality in (A.95) follows by (A.10), (A.44), (A.45) and multiplying and dividing by $\exp\left(-\beta N\left[E(m(\sigma'))-E(m(\sigma))\right]_+\right)$. The inequality in (A.95) is obtained restricting the minimum on $\mathcal{H}$ to the minimum on $\{f\in\mathcal{H}: f(\eta)=f(\eta')\ \forall\eta,\eta'\in\mathcal{S}_N \text{ s.t. } m(\eta)=m(\eta')\}$ and noticing that the latter is in bijection with $\tilde{\mathcal{H}}$.

$$
\begin{aligned}
&Z_{\beta,N}\ \mathrm{cap}\left(\mathcal{S}_N[m_1],\mathcal{S}_N[m_2]\right)\\
&=\min_{f\in\mathcal{H}}\frac{1}{N}\sum_{\sigma,\sigma'\in\mathcal{S}_N}\mathbb{1}_{\sigma\sim\sigma'}\exp\left(-\beta H_N(\sigma)\right)\exp\left(-\beta\left[H_N(\sigma')-H_N(\sigma)\right]_+\right)\left[f(\sigma)-f(\sigma')\right]^2\\
&=\min_{f\in\mathcal{H}}\tilde{Z}_{\beta,N}\sum_{m,m'\in\Gamma_N}\sum_{\substack{\sigma\in\mathcal{S}_N[m],\\ \sigma'\in\mathcal{S}_N[m']}}\mathbb{1}_{\sigma\sim\sigma'}\frac{\exp\left(-\beta N E(m(\sigma))\right)}{\tilde{Z}_{\beta,N}N}\exp\left(-\beta N\left[E(m(\sigma'))-E(m(\sigma))\right]_+\right)\\
&\qquad\times\left[f(\sigma)-f(\sigma')\right]^2\exp\left(-\beta\Delta(\sigma)\right)\frac{\exp\left(-\beta\left[H_N(\sigma')-H_N(\sigma)\right]_+\right)}{\exp\left(-\beta N\left[E(m(\sigma'))-E(m(\sigma))\right]_+\right)}\\
&\le\min_{v\in\tilde{\mathcal{H}}}\tilde{Z}_{\beta,N}\sum_{m,m'\in\Gamma_N}\frac{\exp\left(-\beta N E(m)\right)}{\tilde{Z}_{\beta,N}N}\exp\left(-\beta N\left[E(m')-E(m)\right]_+\right)\left[v(m)-v(m')\right]^2\\
&\qquad\times\sum_{\sigma\in\mathcal{S}_N[m]}\exp\left(-\beta\Delta(\sigma)\right)\sum_{\sigma'\in\mathcal{S}_N[m']}\mathbb{1}_{\sigma\sim\sigma'}\frac{\exp\left(-\beta\left[H_N(\sigma')-H_N(\sigma)\right]_+\right)}{\exp\left(-\beta N\left[E(m')-E(m)\right]_+\right)}.
\end{aligned}
\tag{A.95}
$$

We turn now to the last sum in (A.95) and call this quantity $G(\sigma,m')$. If $\sigma\sim\sigma'$, then σ and σ' differ on a single vertex, say $\ell\in[N]$, i.e. $\forall i\in[N]\setminus\{\ell\}$, $\sigma_i=\sigma'_i$ and $\sigma_\ell=-\sigma'_\ell$. Thus, setting $m=m(\sigma)$ and recalling (A.46) and (A.8), we can write

$$
\Delta(\sigma')-\Delta(\sigma)=-\frac{2}{Np}\sum_{i:i\neq\ell}\hat{J}_{i\ell}\sigma'_i\sigma'_\ell=\frac{2}{Np}\sum_{i:i\neq\ell}\hat{J}_{i\ell}\sigma_i\sigma_\ell,
\tag{A.96}
$$

$$
\tilde{H}_N(\sigma')-\tilde{H}_N(\sigma)=\sigma_\ell\left[\frac{2}{N}\sum_{i:i\neq\ell}\sigma_i+2h\right]=\sigma_\ell\left[\frac{2}{N}(Nm-\sigma_\ell)+2h\right].
\tag{A.97}
$$

Moreover, using (A.44), (A.45), the definition of $\hat{J}_{ij}$ below (A.45), the second equality in (A.96) and the first equality in (A.97) we can write

$$
H_N(\sigma')-H_N(\sigma)=\tilde{H}_N(\sigma')-\tilde{H}_N(\sigma)+\Delta(\sigma')-\Delta(\sigma)=\sigma_\ell\left[\frac{2}{Np}\sum_{i:i\neq\ell}J_{i\ell}\sigma_i+2h\right].
\tag{A.98}
$$

Due to the presence of the indicator function $\mathbb{1}_{\sigma\sim\sigma'}$, $G(\sigma,m')$ vanishes if $m'\notin\{m\pm\frac{2}{N}\}$. Moreover, we can rewrite the sum $\sum_{\sigma'\in\mathcal{S}_N[m']}\mathbb{1}_{\sigma\sim\sigma'}$ in terms of the single vertex $\ell\in[N]$ on which σ and σ' differ. Notice that if $m(\sigma')=m+\frac{2}{N}$ then $\sigma_\ell=-1=-\sigma'_\ell$ and if $m(\sigma')=m-\frac{2}{N}$ then $\sigma_\ell=1=-\sigma'_\ell$.

Therefore, calling $i^{\pm}(\sigma):=\{j\in[N]:\sigma_j=\pm1\}$, and using (A.10), (A.97) and (A.98), we obtain

$$
G(\sigma,m+\tfrac{2}{N})=\sum_{\ell\in i^-(\sigma)}\frac{\exp\left(-\beta\left[-\frac{2}{Np}\sum_{i:i\neq\ell}J_{i\ell}\sigma_i-2h\right]_+\right)}{\exp\left(-\beta\left[-\frac{2}{N}(Nm+1)-2h\right]_+\right)}\le N\,\frac{1-m}{2}\,\mathrm{e}^{2\beta},
\tag{A.99}
$$

$$G(\sigma, m - \tfrac{2}{N}) = \sum_{\ell \in i^+(\sigma)} \frac{\exp(-\beta\left[\frac{2}{Np}\sum_{i:i\neq\ell} J_{i\ell}\sigma_i + 2h\right]_+)}{\exp(-\beta\left[\frac{2}{N}(Nm-1)+2h\right]_+)} \leq N\,\frac{1+m}{2}\,\mathrm{e}^{2\beta(1+h)}. \qquad \text{(A.100)}$$

To obtain the inequalities we used the fact that, for any σ in $\mathcal{S}_N$, the cardinalities of $i^-(\sigma)$ and $i^+(\sigma)$ are respectively $N\,\frac{1-m(\sigma)}{2}$ and $N\,\frac{1+m(\sigma)}{2}$. Moreover, for the inequality in (A.99) we used the following elementary facts holding asymptotically in N,

$$\exp\left(-\beta\left[-\frac{2}{Np}\sum_{i:i\neq\ell} J_{i\ell}\sigma_i - 2h\right]_+\right) \leq 1, \qquad \text{(A.101)}$$

$$\exp\left(\beta\left[-\frac{2}{N}(Nm+1)-2h\right]_+\right) \leq \exp\left(\beta\left[-2m-\frac{2}{N}-2h\right]_+\right) \leq \mathrm{e}^{2\beta}. \qquad \text{(A.102)}$$

Similar inequalities were used to prove (A.100).

Thus, using (A.95), (A.99), (A.100) we obtain

$$\begin{aligned} &Z_{\beta,N}\ \mathrm{cap}\left(\mathcal{S}_N[m_1], \mathcal{S}_N[m_2]\right) \\ &\quad\leq \min_{v\in\tilde{\mathcal{H}}} \tilde{Z}_{\beta,N} \sum_{m,m'\in\Gamma_N} \frac{\exp\left(-\beta N E(m)\right)}{\tilde{Z}_{\beta,N} N} \exp\left(-\beta N\left[E(m')-E(m)\right]_+\right)\left[v(m)-v(m')\right]^2 \\ &\qquad \times \mathrm{e}^{2\beta(1+h)} \sum_{\sigma\in\mathcal{S}_N[m]} \exp\left(-\beta\Delta(\sigma)\right)\left[N\tfrac{1+m}{2}\mathbb{1}_{m-\frac{2}{N}}(m') + N\tfrac{1-m}{2}\mathbb{1}_{m+\frac{2}{N}}(m')\right]. \end{aligned} \qquad \text{(A.103)}$$

Using the upper bound in Corollary A.2.5 with

$$\begin{aligned} g(m) = \sum_{m'\in\Gamma_N} &\frac{\exp\left(-\beta N E(m)\right)}{\tilde{Z}_{\beta,N} N} \exp\left(-\beta N\left[E(m')-E(m)\right]_+\right)\left[v(m)-v(m')\right]^2 \\ &\times \mathrm{e}^{2\beta(1+h)}\left[N\tfrac{1+m}{2}\mathbb{1}_{m-\frac{2}{N}}(m') + N\tfrac{1-m}{2}\mathbb{1}_{m+\frac{2}{N}}(m')\right] \end{aligned} \qquad \text{(A.104)}$$

we obtain

$$\begin{aligned} &Z_{\beta,N}\ \mathrm{cap}\left(\mathcal{S}_N[m_1], \mathcal{S}_N[m_2]\right) \\ &\overset{P(s)}{\leq}\ \mathrm{e}^{s+2\beta(1+h)+\alpha}\tilde{Z}_{\beta,N}\min_{v\in\tilde{\mathcal{H}}}\sum_{m,m'\in\Gamma_N}\frac{\exp\left(-\beta N E(m) - N I_N(m)\right)}{\tilde{Z}_{\beta,N} N}\exp(-\beta N\left[E(m')-E(m)\right]_+) \\ &\qquad \times \left[v(m)-v(m')\right]^2\left[N\frac{1+m}{2}\mathbb{1}_{m-\frac{2}{N}}(m') + N\frac{1-m}{2}\mathbb{1}_{m+\frac{2}{N}}(m')\right](1+o(1)) \\ &= \mathrm{e}^{s+2\beta(1+h)+\alpha}\tilde{Z}_{\beta,N}\min_{v\in\tilde{\mathcal{H}}}\sum_{m,m'\in\Gamma_N}\tilde{\mathcal{Q}}(m)\,\tilde{r}(m,m')\left[v(m)-v(m')\right]^2\ (1+o(1)) \\ &= \mathrm{e}^{s+2\beta(1+h)+\alpha}\,\tilde{Z}_{\beta,N}\ \mathrm{cap}^{CW}(\mathcal{S}_N[m_1], \mathcal{S}_N[m_2])\,(1+o(1))\,, \end{aligned} \qquad \text{(A.105)}$$

where we used the notation (A.38) and in the middle step we used (A.13), the first equality in (A.14) and (A.20). Furthermore, we noticed that the variational form appearing in the previous formula is the Dirichlet principle (see Bovier and den Hollander [18, (7.1.29), (7.3.1)]) applied to the random walk performed by the projection of the CW model dynamics onto the magnetisation space. See Section A.1.2 for the CW model.

We conclude that the minimum equals the capacity of the CW model using lumping techniques. More precisely, here we used Bovier and den Hollander [18, (9.3.6)], stating that the capacity for the dynamics projected onto the magnetisation space equals the capacity for the CW dynamics on the configuration space, which holds because of the CW model mean-field property. For reference on lumping see Bovier and den Hollander [18, Section 9.3]. □

A.3.2 Asymptotics on capacity: lower bound

In this section we prove Theorem A.1.6, obtaining the lower bound on the capacity of the RDCW model in terms of the capacity of the CW model. We will prove it without loss of generality, only for $m_1 < m_2 \in \Gamma_N$, because the capacity can be proven to be symmetric, using the reversibility of the dynamics.

The main idea of the proof is to find a lower bound on the capacity of the RDCW model via the Thomson principle (see e.g. Bovier and den Hollander [18, Theorem 7.37]. For $\mathrm{cap}\,(\mathcal{S}_N[m_1], \mathcal{S}_N[m_2])$ it reads

$$\mathrm{cap}\,(\mathcal{S}_N[m_1], \mathcal{S}_N[m_2]) = \sup\left\{\frac{1}{\mathcal{D}(\bar{\Psi})} : \bar{\Psi} \in \mathcal{U}_{\mathcal{S}_N[m_1],\mathcal{S}_N[m_2]}\right\}, \tag{A.106}$$

where we denote by $\mathcal{U}_{\mathcal{S}_N[m_1],\mathcal{S}_N[m_2]}$ the space of all unitary antisymmetric flows from $\mathcal{S}_N[m_1]$ to $\mathcal{S}_N[m_2]$ and $\mathcal{D}$ is defined by

$$\mathcal{D}(\psi) = \frac{1}{2}\sum_{\sigma,\sigma'\in\mathcal{S}_N} \mathbb{1}_{\sigma'\sim\sigma}\,\frac{\psi(\sigma,\sigma')^2}{\mu_{\beta,N}(\sigma)p_N(\sigma,\sigma')} \tag{A.107}$$

for any $\psi : \mathcal{S}_N^2 \to \mathbb{R}$ antisymmetric flow. Thus, in order to find a lower bound in terms of the capacity of the CW model we have to find a unitary flow from which we could reconstruct the CW capacity term.

For all $\sigma, \sigma' \in \mathcal{S}_N$, we define the candidate flow Ψ_N as follows

$$\Psi_N(\sigma,\sigma') = \phi_N(m(\sigma), m(\sigma')), \tag{A.108}$$

where, for all $m, m' \in \Gamma_N$,

$$\phi_N(m,m') = \begin{cases} \left[\frac{(1-m)N}{2}\exp\left(-NI_N(m)\right)\right]^{-1} & \text{if } m_1 \le m \le m_2 - \frac{2}{N}, m' = m + \frac{2}{N} \\ -\left[\frac{(1+m)N}{2}\exp\left(-NI_N(m)\right)\right]^{-1} & \text{if } m_1 + \frac{2}{N} \le m \le m_2, m' = m - \frac{2}{N} \\ 0 & \text{otherwise.} \end{cases} \tag{A.109}$$

The proof of Theorem A.1.6 is postponed after two technical intermediate results which are essential for it. The following lemma allows us to use Ψ_N in the Thomson principle.

Lemma A.3.1. *Let $m_1 < m_2 \in \Gamma_N$. The flow Ψ_N on $\mathcal{S}_N$, defined in* (A.108) *is a unitary antisymmetric flow from $\mathcal{S}_N[m_1]$ to $\mathcal{S}_N[m_2]$, i.e. $\Psi_N \in \mathcal{U}_{\mathcal{S}_N[m_1],\mathcal{S}_N[m_2]}$.*

Proof. Ψ_N is *antisymmetric* because for all $m \in \Gamma_N$, i.e.

$$\frac{(1+m)}{2} N \exp\left(-N I_N(m)\right) = \frac{\left(1-\left(m-\frac{2}{N}\right)\right)}{2} N \exp\left(-N I_N\left(m-\frac{2}{N}\right)\right). \tag{A.110}$$

Indeed, using (A.15), the right hand side of (A.110) writes

$$\begin{aligned}
&\frac{\left(1-\left(m-\frac{2}{N}\right)\right)}{2} N \exp\left(-N I_N\left(m-\frac{2}{N}\right)\right) = \tfrac{\left(1-\left(m-\frac{2}{N}\right)\right)N}{2} \binom{N}{\frac{1-\left(m-\frac{2}{N}\right)}{2}N} \\
&= \tfrac{\left(1-\left(m-\frac{2}{N}\right)\right)N}{2} \frac{N!}{\left[\frac{(1-m)N}{2}+1\right]!\left[\frac{(1+m)N}{2}-1\right]!} = \frac{N!}{\left[\frac{(1-m)N}{2}\right]!\left[\frac{(1+m)N}{2}-1\right]!} \\
&= \frac{(1+m)}{2} N \binom{N}{\frac{1+m}{2}N} = \frac{(1+m)}{2} N \exp\left(-N I_N(m)\right).
\end{aligned} \tag{A.111}$$

Next we prove that the *Kirchhoff law* holds, i.e., for all $\sigma \in \mathcal{S}_N \setminus (\mathcal{S}_N[m_1] \cup \mathcal{S}_N[m_2])$

$$\sum_{\sigma' \in \mathcal{S}_N : \sigma \sim \sigma'} \Psi_N(\sigma, \sigma') = 0. \tag{A.112}$$

For all $\sigma \in \mathcal{S}_N$ such that $m(\sigma) \notin (m_1, m_2)$, (A.112) holds trivially being all terms zero, by (A.109). Now, for all $\sigma \in \mathcal{S}_N$ such that $m(\sigma) \in (m_1, m_2)$,

$$\begin{aligned}
\sum_{\sigma' \in \mathcal{S}_N : \sigma \sim \sigma'} \Psi_N(\sigma, \sigma') &= \sum_{\substack{\sigma' \in \mathcal{S}_N : \sigma \sim \sigma', \\ m(\sigma') = m(\sigma) + \frac{2}{N}}} \phi_N\left(m(\sigma), m(\sigma')\right) + \sum_{\substack{\sigma' \in \mathcal{S}_N : \sigma \sim \sigma', \\ m(\sigma') = m(\sigma) - \frac{2}{N}}} \phi_N\left(m(\sigma), m(\sigma')\right) \\
&= \frac{(1-m(\sigma))N}{2} \left[\frac{(1-m(\sigma))N}{2} \exp\left(-N I_N(m(\sigma))\right)\right]^{-1} \\
&\quad - \frac{(1+m(\sigma))N}{2} \left[\frac{(1+m(\sigma))N}{2} \exp\left(-N I_N(m(\sigma))\right)\right]^{-1} \\
&= 0,
\end{aligned} \tag{A.113}$$

where $\frac{(1\mp m(\sigma))N}{2}$ in the second equality are the cardinalities of the set over which we were summing, namely the number of negative, respectively positive, spins in a configuration $\sigma \in \mathcal{S}_N$.

We are left to show that Ψ_N is *unitary* from $\mathcal{S}_N[m_1]$ to $\mathcal{S}_N[m_2]$, namely

$$\sum_{a \in \mathcal{S}_N[m_1]} \sum_{\sigma' \in \mathcal{S}_N : a \sim \sigma'} \Psi_N(a, \sigma') = 1 = \sum_{b \in \mathcal{S}_N[m_2]} \sum_{\sigma \in \mathcal{S}_N : \sigma \sim b} \Psi_N(\sigma, b). \tag{A.114}$$

The left hand side of (A.114) equals

$$\begin{aligned}
&\sum_{a \in \mathcal{S}_N[m_1]} \sum_{\sigma' \in \mathcal{S}_N : a \sim \sigma'} \phi_N\left(m(a), m(\sigma')\right) = \sum_{a \in \mathcal{S}_N[m_1]} \sum_{\substack{\sigma' \in \mathcal{S}_N : a \sim \sigma', \\ m(\sigma') = m_1 + \frac{2}{N}}} \left[\tfrac{(1-m_1)N}{2} \exp\left(-N I_N(m_1)\right)\right]^{-1} \\
&= \exp\left(-N I_N(m_1)\right) \tfrac{(1-m_1)N}{2} \left[\tfrac{(1-m_1)N}{2} \exp\left(-N I_N(m_1)\right)\right]^{-1} = 1.
\end{aligned} \tag{A.115}$$

The right hand side of (A.114) equals

$$\begin{aligned}
&\sum_{b\in\mathcal{S}_N[m_2]}\sum_{\sigma\in\mathcal{S}_N:\sigma\sim b}\phi_N\left(m(\sigma),m(b)\right)\\
&=\sum_{b\in\mathcal{S}_N[m_2]}\sum_{\substack{\sigma\in\mathcal{S}_N:\sigma\sim b,\\ m(\sigma)=m_2-\frac{2}{N}}}\left[\frac{\left(1-\left(m_2-\frac{2}{N}\right)\right)}{2}N\exp\left(-NI_N\left(m_2-\frac{2}{N}\right)\right)\right]^{-1}\\
&=\exp\left(-NI_N(m_2)\right)\frac{(1+(m_2))}{2}N\left[\frac{\left(1-\left(m_2-\frac{2}{N}\right)\right)}{2}N\exp\left(-NI_N\left(m_2-\frac{2}{N}\right)\right)\right]^{-1}.
\end{aligned}\tag{A.116}$$

We use (A.110) to conclude the proof. ☐

Lemma A.3.2. *For all $\sigma\in\mathcal{S}_N$ and $m'\in\Gamma_N$, the following holds*

$$\begin{aligned}
&\sum_{\sigma'\in\mathcal{S}_N[m']}\mathbb{1}_{\sigma'\sim\sigma}\frac{\exp\left(-\beta\left[\tilde{H}_N(\sigma')-\tilde{H}_N(\sigma)\right]_+\right)}{\exp\left(-\beta\left[H_N(\sigma')-H_N(\sigma)\right]_+\right)}\\
&\qquad\leq \mathrm{e}^{2\beta(1+h)}\left[N\frac{1+m(\sigma)}{2}\mathbb{1}_{m(\sigma)-\frac{2}{N}}(m')+N\frac{1-m(\sigma)}{2}\mathbb{1}_{m(\sigma)+\frac{2}{N}}(m')\right].
\end{aligned}\tag{A.117}$$

Proof. Let $m=m(\sigma)$. The left hand side is non-zero only if $m'\in\left\{m+\frac{2}{N},m-\frac{2}{N}\right\}$. Recalling the definition $i^{\pm}(\sigma)=\{j\in[N]:\sigma_j=\pm1\}$, if $m'=m+\frac{2}{N}$, we have

$$\begin{aligned}
&\sum_{\sigma'\in\mathcal{S}_N\left[m+\frac{2}{N}\right]}\mathbb{1}_{\sigma'\sim\sigma}\frac{\exp\left(-\beta\left[\tilde{H}_N(\sigma')-\tilde{H}_N(\sigma)\right]_+\right)}{\exp\left(-\beta\left[H_N(\sigma')-H_N(\sigma)\right]_+\right)}\\
&=\sum_{\ell\in i^-(\sigma)}\frac{\exp\left(-\beta\left[-\frac{2p}{N}(Nm+1)-2h\right]_+\right)}{\exp\left(-\beta\left[-\frac{2}{N}\sum_{i:i\neq\ell}J_{i\ell}\sigma_i-2h\right]_+\right)}\\
&\leq\sum_{\ell\in i^-(\sigma)}\exp\left(\beta\left[-\frac{2}{N}\sum_{i:i\neq\ell}J_{i\ell}\sigma_i-2h\right]_+\right)\leq\sum_{\ell\in i^-(\sigma)}\mathrm{e}^{2\beta}=N\frac{1-m}{2}\mathrm{e}^{2\beta},
\end{aligned}\tag{A.118}$$

where we have used that, since $h>0$,

$$-\frac{2}{N}\sum_{i:i\neq\ell}J_{i\ell}\sigma_i-2h\leq\frac{2}{N}\sum_{i:i\neq\ell}|J_{i\ell}\sigma_i|\leq\frac{2(N-1)}{N}\leq 2.\tag{A.119}$$

Similarly, if $m'=m-\frac{2}{N}$, we get

$$\begin{aligned}
&\sum_{\sigma'\in\mathcal{S}_N\left[m-\frac{2}{N}\right]}\mathbb{1}_{\sigma'\sim\sigma}\frac{\exp\left(-\beta\left[\tilde{H}_N(\sigma')-\tilde{H}_N(\sigma)\right]_+\right)}{\exp\left(-\beta\left[H_N(\sigma')-H_N(\sigma)\right]_+\right)}\\
&\qquad=\sum_{\ell\in i^+(\sigma)}\frac{\exp\left(-\beta\left[\frac{2p}{N}(Nm-1)+2h\right]_+\right)}{\exp\left(-\beta\left[\frac{2}{N}\sum_{i:i\neq\ell}J_{i\ell}\sigma_i+2h\right]_+\right)}\leq N\frac{1+m}{2}\mathrm{e}^{2\beta(1+h)}.
\end{aligned}\tag{A.120}$$

□

With the previous lemmas at hand, we are now ready to prove the lower bound on the capacity.

Proof of Theorem A.1.6. As we mentioned above, since the capacity is symmetric, we will prove the result only for $m_1 < m_2 \in \Gamma_N$.

Let Ψ_N be the test flow defined in (A.108), which by Lemma A.3.1 is in $\mathcal{U}_{\mathcal{S}_N[m_1],\mathcal{S}_N[m_2]}$. Thus, using the Thomson principle (A.106), we obtain the following bound

$$\mathrm{cap}\left(\mathcal{S}_N[m_1], \mathcal{S}_N[m_2]\right) \geq \frac{1}{\mathcal{D}(\Psi_N)}. \tag{A.121}$$

Therefore, we are interested in upper bounds on $\mathcal{D}(\Psi_N)$ which, using (A.3), (A.5) and (A.46), can be written as follows

$$\mathcal{D}(\Psi_N) = \frac{1}{2} N \sum_{\sigma,\sigma' \in \mathcal{S}_N} \mathbb{1}_{\sigma'\sim\sigma} \frac{\phi_N(m(\sigma), m(\sigma'))^2}{\exp\left(-\beta(\tilde{H}_N(\sigma) + \Delta(\sigma))\right)} \frac{Z_{\beta,N}}{\exp\left(-\beta[H_N(\sigma') - H_N(\sigma)]_+\right)}. \tag{A.122}$$

By multiplying and dividing by $\exp(-\beta[\tilde{H}_N(\sigma') - \tilde{H}_N(\sigma)]_+)\, \tilde{Z}_{\beta,N}$, and using (A.10), (A.11) and (A.18), we get

$$\begin{aligned}
\mathcal{D}(\Psi_N) &= N \frac{Z_{\beta,N}}{2\tilde{Z}_{\beta,N}} \sum_{m,m' \in \Gamma_N} \frac{\phi_N(m,m')^2}{\tilde{\mathcal{Q}}_{\beta,N}(m) \exp\left(N I_N(m)\right) \exp\left(-\beta N \left[E(m') - E(m)\right]_+\right)} \\
&\quad \times \sum_{\sigma \in \mathcal{S}_N[m]} \exp(\beta\Delta(\sigma)) \sum_{\sigma' \in \mathcal{S}_N[m']} \mathbb{1}_{\sigma'\sim\sigma} \frac{\exp\left(-\beta\left[\tilde{H}_N(\sigma') - \tilde{H}_N(\sigma)\right]_+\right)}{\exp\left(-\beta\left[H_N(\sigma') - H_N(\sigma)\right]_+\right)} \\
&\leq N \frac{Z_{\beta,N}}{2\tilde{Z}_{\beta,N}} \mathrm{e}^{2\beta(1+h)} \sum_{m,m' \in \Gamma_N} \frac{\phi_N(m,m')^2}{\tilde{\mathcal{Q}}_{\beta,N}(m) \exp\left(N I_N(m)\right) \exp\left(-\beta N \left[E(m') - E(m)\right]_+\right)} \\
&\quad \times \left[N \frac{1+m}{2} \mathbb{1}_{m-\frac{2}{N}}(m') + N \frac{1-m}{2} \mathbb{1}_{m+\frac{2}{N}}(m')\right] \sum_{\sigma \in \mathcal{S}_N[m]} \exp(\beta\Delta(\sigma)),
\end{aligned} \tag{A.123}$$

where we used Lemma A.3.2 to bound the sum over σ', uniformly in $\sigma \in \mathcal{S}_N[m]$. Then, to bound the remaining sum over σ, we use the upper bound in Corollary A.2.5 (in the version with $\mathrm{e}^{\beta\Delta(\sigma)}$, motivated by the Remark therein) with

$$\begin{aligned}
g(m) = &\sum_{m' \in \Gamma_N} \frac{\phi_N(m,m')^2}{\tilde{\mathcal{Q}}_{\beta,N}(m) \exp\left(N I_N(m)\right) \exp\left(-\beta N \left[E(m') - E(m)\right]_+\right)} \\
&\times \left[N \frac{1+m}{2} \mathbb{1}_{m-\frac{2}{N}}(m') + N \frac{1-m}{2} \mathbb{1}_{m+\frac{2}{N}}(m')\right],
\end{aligned} \tag{A.124}$$

obtaining, with notation (A.38),

$$\begin{aligned}\mathcal{D}(\Psi_N) &\overset{P(s)}{\le} N\,\frac{Z_{\beta,N}}{2\tilde{Z}_{\beta,N}}\mathrm{e}^{s+2\beta(1+h)+\alpha}\sum_{m,m'\in\Gamma_N}\frac{\phi_N(m,m')^2\exp\left(-NI_N(m)\right)}{\tilde{\mathcal{Q}}_{\beta,N}(m)\exp\left(NI_N(m)\right)\exp\left(-\beta N\left[E(m')-E(m)\right]_+\right)}\\ &\quad\times\left[N\frac{1+m}{2}\mathbb{1}_{m-\frac{2}{N}}(m')+N\frac{1-m}{2}\mathbb{1}_{m+\frac{2}{N}}(m')\right](1+o(1))\\ &=\frac{Z_{\beta,N}}{2\tilde{Z}_{\beta,N}}\mathrm{e}^{s+2\beta(1+h)+\alpha}\sum_{m,m'\in\Gamma_N}\frac{\phi_N(m,m')^2\exp\left(-2NI_N(m)\right)}{\tilde{\mathcal{Q}}_{\beta,N}(m)\tilde{r}_N(m,m')}\\ &\quad\times\left[N\frac{1+m}{2}\mathbb{1}_{m-\frac{2}{N}}(m')+N\frac{1-m}{2}\mathbb{1}_{m+\frac{2}{N}}(m')\right]^2(1+o(1))\,,\end{aligned}\tag{A.125}$$

where in the equality we only used (A.20).

Now we first substitute ϕ_N defined in (A.109) into (A.125) and then use reversibility to obtain

$$\begin{aligned}\mathcal{D}(\Psi_N) &\overset{P(s)}{\le}\frac{Z_{\beta,N}}{2\tilde{Z}_{\beta,N}}\,\mathrm{e}^{s+2\beta(1+h)+\alpha}\sum_{\substack{m_1\le m<m_2,\\ m\in\Gamma_N}}\frac{1}{\tilde{\mathcal{Q}}_{\beta,N}(m)\tilde{r}_N\left(m,m+\frac{2}{N}\right)}(1+o(1))\\ &\quad+\frac{Z_{\beta,N}}{2\tilde{Z}_{\beta,N}}\,\mathrm{e}^{s+2\beta(1+h)+\alpha}\sum_{\substack{m_1<m\le m_2,\\ m\in\Gamma_N}}\frac{1}{\tilde{\mathcal{Q}}_{\beta,N}(m-\frac{2}{N})\tilde{r}_N\left(m-\frac{2}{N},m\right)}(1+o(1))\\ &=\frac{Z_{\beta,N}}{\tilde{Z}_{\beta,N}}\mathrm{e}^{s+2\beta(1+h)+\alpha}\sum_{m_1\le m<m_2}\frac{1}{\tilde{\mathcal{Q}}_{\beta,N}(m)\tilde{r}_N\left(m,m+\frac{2}{N}\right)}(1+o(1))\,,\end{aligned}\tag{A.126}$$

where the last equality follows noticing that the two sums in the previous step are equal.

Therefore, by (A.121) and (A.126), we obtain

$$\begin{aligned}Z_{\beta,N}\ \mathrm{cap}(\mathcal{S}_N[m_1],\mathcal{S}_N[m_2]) &\ge\frac{Z_{\beta,N}}{\mathcal{D}(\Psi_N)}\\ &\overset{P(s)}{\ge}\tilde{Z}_{\beta,N}\,\mathrm{e}^{-s-2\beta(1+h)-\alpha}\left[\sum_{m_1\le m<m_2}\frac{1}{\tilde{\mathcal{Q}}_{\beta,N}(m)\tilde{r}_N\left(m,m+\frac{2}{N}\right)}\right]^{-1}(1+o(1))\\ &=\tilde{Z}_{\beta,N}\,\mathrm{e}^{-s-2\beta(1+h)-\alpha}\ \mathrm{cap}^{\mathrm{CW}}(\mathcal{S}_N[m_1],\mathcal{S}_N[m_2])\,(1+o(1))\,,\end{aligned}\tag{A.127}$$

where we used the notation (A.38) and we noticed that the inverse of the expression appearing in brackets in (A.127) gives exactly the capacity for the CW model. Indeed, that expression gives exactly the capacity for the one-dimensional random walk in Γ_N which is the projection of the CW dynamics onto the magnetisation space Γ_N (see the formula for the capacity in Bovier and den Hollander [18, Section 7.1.4, (7.1.60)]). Using lumping techniques exactly as at the end of the proof of Theorem A.1.5 (end of Section A.3.1), we have that the aforementioned capacity equals the CW capacity. □

A.4 Estimates on the harmonic function

As pointed out in Section A.1.4, the proof of Theorem A.1.4 relies on sharp estimates on capacities, carried out in Section A.3, and estimates on the harmonic function. We entirely devote this section to obtain asymptotic upper and lower bounds on the numerator in (A.40), which is given by the following sum

$$\sum_{\sigma\in\mathcal{S}_N} \mu_{\beta,N}(\sigma)h^N_{m_-,m_+}(\sigma), \tag{A.128}$$

that is to give the proof of Theorem A.1.7.

In order to control the sum (A.128), one generally uses a renewal argument which relies again on estimates over capacities. However, in our case this is not possible, due to the fact that capacities of single spins are too small.

We first prove the upper bound and then give some details about how to prove the lower bound, which is very similar and more straightforward. Our proof follows Bianchi, Bovier and Ioffe [7, Section 6].

A.4.1 Notation and decomposition of the space

Before starting with the proof, we introduce some notation. We refer to Figure A.1 below for a better visual understanding of the objects we are defining.

Recall that we denote by m_+ the global minimum, by m_- the local minimum, and by m^* the local maximum of $f_\beta(\cdot)$ in $[-1,1]$, where $f_\beta(\cdot) = \lim_{N\to\infty} f_{\beta,N}(\cdot)$, defined in (A.14). We want to decompose the space Γ_N (and eventually the set of spin configurations $\mathcal{S}_N$) according to the values of f_β. The notation and the decomposition are organised in 4 steps.

Step 1. First, let $\delta > 0$ be small in a way which will become clear later, and define the set

$$U_\delta = \{m \in [-1,1] : f_\beta(m) \le f_\beta(m_-) + \delta\}. \tag{A.129}$$

We write $U_\delta^c = [-1,1]\backslash U_\delta$ and we denote by $U_\delta(m)$ the connected component of U_δ containing m. Note that $\{m_-, m_+\} \in U_\delta$. In general, $U_\delta(m_-)$ and $U_\delta(m_+)$ may have non empty intersection, but we *choose* δ such that $m^* \notin U_\delta$, implying that U_δ is partitioned by the *disjoint* sets $U_\delta(m_-)$ and $U_\delta(m_+)$. For this to hold, it suffices to take $\delta < f_\beta(m^*) - f_\beta(m_-)$. Moreover, we *choose* δ also such that $-1 \notin U_\delta(m_-)$. For this to hold, it suffices to take $\delta < f_\beta(-1) - f_\beta(m_-)$. Thus, we *choose* $\delta < \min(f_\beta(-1), f_\beta(m^*)) - f_\beta(m_-)$. These conditions are needed to prove (A.138) below.

Let us denote by m_δ the unique point in (m^*, m_+) such that

$$f_\beta(m_\delta) = f_\beta(m_-) + \delta. \tag{A.130}$$

Step 2. With δ chosen as above, we define a sequence $(\delta_N)_{N\in\mathbb{N}}$, converging to δ from below, such that the left extreme of $U_{\delta_N}(m_+)$ is in Γ_N. Specifically, we define δ_N as follows:

$$\delta_N = \max\left\{\bar\delta \in (0,\delta] \,:\, \exists\, m \in U_\delta(m_+) \cap \Gamma_N \setminus [m_+, 1] \text{ s.t. } f_\beta(m) = f_\beta(m_-) + \bar\delta\right\}, \tag{A.131}$$

for N sufficiently large. Moreover, set

$$U_{\delta,N} = U_{\delta_N} \cap \Gamma_N, \qquad U^c_{\delta,N} = \Gamma_N \setminus U_{\delta,N} \qquad \text{and} \qquad U_{\delta,N}(m) = U_{\delta_N}(m) \cap \Gamma_N, \tag{A.132}$$

for all $m \in [-1,1]$. Thus, we have the partitions

$$\Gamma_N = U_{\delta,N}(m_-) \cup U_{\delta,N}(m_+) \cup U^c_{\delta,N} \tag{A.133}$$

and

$$\mathcal{S}_N = \mathcal{S}_N\left[U_{\delta,N}(m_-)\right] \cup \mathcal{S}_N\left[m_+(N)\right] \cup \mathcal{S}_N\left[U^c_{\delta,N}\right] \cup \mathcal{S}_N\left[U_{\delta,N}(m_+) \setminus \{m_+(N)\}\right]. \tag{A.134}$$

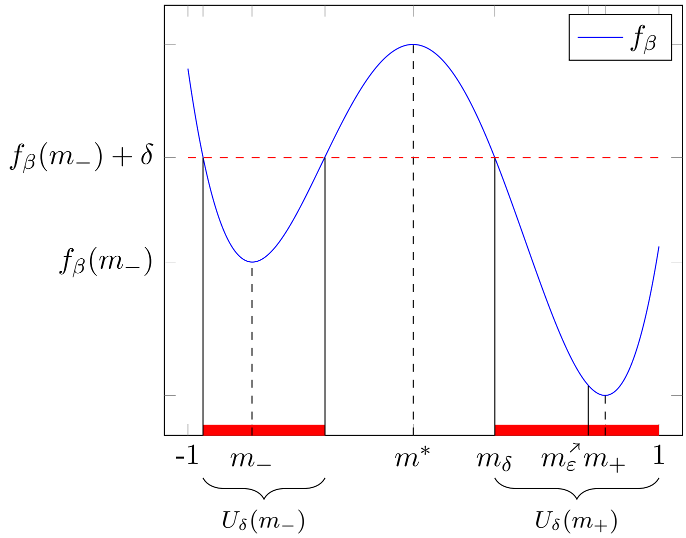


Figure A.1: Graph of f_β and decomposition of the magnetisation space $[-1,1]$: the two intervals $U_\delta(m_-)$ and $U_\delta(m_+)$ around the two minima are drawn, together with the special points m_δ, m_ε. U_δ is painted in red.

Remark A.4.1. *Notice that, for N sufficiently large, $U_{\delta,N}(m_-(N)) = U_{\delta,N}(m_-)$ and $U_{\delta,N}(m_+(N)) = U_{\delta,N}(m_+)$. Furthermore, with these definitions, $m_{\delta_N} \in U_{\delta,N}$ and it is the left extreme of $U_{\delta,N}(m_+)$.*

Step 3. Let $\varepsilon > 0$ be arbitrarily small (the choice of ε will be relevant in Section A.4.2). We denote by m_ε the only point in a small left neighbourhood of m_+, more precisely in $U_\delta(m_+) \setminus [m_+, 1]$, such that

$$f_\beta(m_\varepsilon) = f_\beta(m_+) + \varepsilon. \tag{A.135}$$

Let us define an ε-dependent parameter $\theta > 0$ by

$$\theta = m_+ - m_\varepsilon. \tag{A.136}$$

Step 4. Similarly to Step 2, fixed $\varepsilon > 0$, we want to define a sequence $(\varepsilon_N)_{N\in\mathbb{N}}$ converging to ε from below such that m_{ε_N} is in Γ_N. More precisely, we define ε_N as follows

$$\varepsilon_N = \max\left\{\bar{\varepsilon} \in (0,\varepsilon] \,:\, \exists\, m \in U_{\delta,N}(m_+) \setminus [m_+,1] \text{ s.t. } f_\beta(m) = f_\beta(m_+) + \bar{\varepsilon}\right\}. \tag{A.137}$$

We will use later that $m_{\varepsilon_N} \in U_{\delta,N}(m_+)$ and it satisfies $f_\beta(m_{\varepsilon_N}) = f_\beta(m_+) + \varepsilon_N$.

Moreover, given $\varepsilon > 0$, we define the sequence $(\theta_N)_{N\in\mathbb{N}}$, analogously to (A.136), by setting $\theta_N = m_+(N) - m_{\varepsilon_N}$. θ_N plays an important role in Lemma A.4.4 below. Notice that $\lim_{N\to\infty}\theta_N = \theta$ and, if $m_+ \neq m_+(N)$, then $f(m_{\varepsilon_N}) - f(m_+(N)) \neq \varepsilon_N$.

A.4.2 Upper bound on the harmonic sum

In this section we prove the first part of Theorem A.1.7 by giving an upper bound on the harmonic sum in (A.128).

We will estimate the contribution of each set of the partition in (A.134) to the sum in (A.128). As one expects, the only relevant contribution will be given by the terms in $\mathcal{S}_N[U_{\delta,N}(m_-)]$.

Indeed, $\mu_{\beta,N}$ is very small in $\mathcal{S}_N[U^c_{\delta,N}]$ while $h^N_{m_-,m_+}$ is very small in $\mathcal{S}_N[U_{\delta,N}(m_+)]$ and we will see the two contributions on these two sets turn out to be irrelevant.

The main ingredients in the proof of the upper bound are Corollary A.2.7 and Lemma A.4.2 below. The proof of the latter result is quite technical and it is postponed to Section A.4.3.

Proof of Theorem A.1.7. Upper bound. We are ready to start estimating the contributions of each disjoint set of the partition in (A.134) to the sum in (A.128).

Part 1. Sum on $\mathcal{S}_N[U_{\delta,N}(m_-)]$. This will be the relevant part. Using first that $h^N_{m_-,m_+}(\sigma) \leq 1$, (A.43) and (A.61) of Corollary A.2.7 with $\bar{g}(m) = \mathbb{1}_{m\in U_{\delta,N}(m_-)}(m)$ we obtain

$$\begin{aligned}\sum_{\sigma\in\mathcal{S}_N[U_{\delta,N}(m_-)]} \mu_{\beta,N}(\sigma)\, h^N_{m_-,m_+}(\sigma) &\leq \sum_{m\in U_{\delta,N}(m_-)} \mathcal{Q}_{\beta,N}(m) \\ &\overset{P(s)}{\leq} \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))}{Z_{\beta,N}} \sum_{m\in U_{\delta,N}(m_-)} \exp\left(-\beta N f_\beta(m)\right)\sqrt{\frac{2}{\pi N(1-m^2)}} \\ &= \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))\exp\left(-\beta N f_\beta(m_-)\right)}{Z_{\beta,N}\sqrt{(1-m_-^2)\,\beta f''_\beta(m_-)}}.\end{aligned} \tag{A.138}$$

In the second line we used our assumption of δ_N being small enough such that $-1 \notin U_{\delta,N}(m_-)$ (see Section A.4.1, Step 1). To obtain the last equality we first approximated, for N sufficiently large, the sum with an integral and then applied the saddle point method (see, for instance de Bruijn [29, Chp 5.7]), where m_- is the maximum point of $-\beta f_\beta$ on the considered domain. Notice that here we use the fact that $m^* \notin U_{\delta,N}(m_-)$, which holds again for δ_N small enough (see Section A.4.1, Step 1). More precisely,

$$\begin{aligned}\sum_{m\in U_{\delta,N}(m_-)} &\exp\left(-\beta N f_\beta(m)\right)\frac{1}{\sqrt{(1-m^2)}} \\ &\approx \frac{N}{2}\int_a^b \exp\Big(-\beta N f_\beta(x)\Big)\frac{1}{\sqrt{(1-x^2)}}\mathtt{d}x \\ &= \exp\left(-\beta N f_\beta(m_-)\right)\frac{1}{\sqrt{(1-m_-^2)}}\sqrt{\frac{\pi N}{2\beta f''_\beta(m_-)}}\,(1+o(1))\,,\end{aligned} \tag{A.139}$$

where $-1 < a, b \in \Gamma_N$ are the left and right extremes of $U_{\delta,N}(m_-)$, respectively.

Part 2. Sum on $\mathcal{S}_N[m_+(N)]$. Being by definition $h^N_{m_-,m_+}(\sigma) = 0$ for all $\sigma \in \mathcal{S}_N[m_+(N)]$, we trivially have

$$\sum_{\sigma\in\mathcal{S}_N[m_+(N)]} \mu_{\beta,N}(\sigma)\, h^N_{m_-,m_+}(\sigma) = 0. \tag{A.140}$$

Part 3. Sum on $\mathcal{S}_N[U^c_{\delta,N}]$.

Using $h^N_{m_-,m_+} \leq 1$ and (A.43), we have

$$\begin{aligned}\sum_{\sigma\in\mathcal{S}_N[U^c_{\delta,N}]} \mu_{\beta,N}(\sigma)\, h^N_{m_-,m_+}(\sigma) &\leq \sum_{\sigma\in\mathcal{S}_N[U^c_{\delta,N}]} \mu_{\beta,N}(\sigma) = \sum_{m\in U^c_{\delta,N}} \mathcal{Q}_{\beta,N}(m) \\ = \sum_{m\in U^c_{\delta,N}\setminus\{1,-1\}} \mathcal{Q}_{\beta,N}(m) &+ \sum_{m\in U^c_{\delta,N}\cap\{1,-1\}} \mathcal{Q}_{\beta,N}(m).\end{aligned} \tag{A.141}$$

We bound the right hand side using (A.61) of Corollary A.2.7 with $\bar{g}(m) = \mathbb{1}_{m\in U^c_{\delta,N}}(m)$ obtaining

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{S}_N[U^c_{\delta,N}]} \mu_{\beta,N}(\sigma)\, h^N_{m_-,m_+}(\sigma) \\
&\overset{P(s)}{\leq} \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))}{Z_{\beta,N}} \sum_{m\in U^c_{\delta,N}\setminus\{1,-1\}} \exp\Big(-\beta N f_\beta(m)\Big)\sqrt{\frac{2}{\pi N(1-m^2)}} \\
&\quad + \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))}{Z_{\beta,N}} \sum_{m\in U^c_{\delta,N}\cap\{1,-1\}} \exp\left(-\beta N\left(f_\beta(m)\right)\right) \\
&\leq \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))}{Z_{\beta,N}} \exp\left(-\beta N\left(f_\beta(m_-)+\delta_N\right)\right)\left(\sqrt{\frac{2}{\pi N}}\sum_{m\in U^c_{\delta,N}\setminus\{1,-1\}}\frac{1}{\sqrt{(1-m^2)}}+2\right),
\end{aligned} \tag{A.142}$$

where in the last inequality we used the bound $f_\beta(m) \geq f_\beta(m_-) + \delta_N$ given by the definition of $U^c_{\delta,N}$ (see (A.129)).

Part 4. Sum on $\mathcal{S}_N[U_{\delta,N}(m_+) \setminus \{m_+(N)\}]$. Using (A.32) and the fact that, for any $\sigma \in \mathcal{S}_N$ such that $m(\sigma) > m_+(N)$, $\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right) = 0$, we get

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{S}_N[U_{\delta,N}(m_+)\setminus\{m_+(N)\}]} \mu_{\beta,N}(\sigma)\, h^N_{m_-,m_+}(\sigma) \\
&= \sum_{\sigma\in\mathcal{S}_N\left[[m_{\delta_N},m_+(N))\right]} \mu_{\beta,N}(\sigma)\,\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right).
\end{aligned} \tag{A.143}$$

Thus, applying Lemma A.4.2 below, the following holds for any $\gamma \in (0,1)$

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{S}_N[U_{\delta,N}(m_+)\setminus\{m_+(N)\}]} \mu_{\beta,N}(\sigma)\, h^N_{m_-,m_+}(\sigma) \\
&\leq \exp\left(-\beta N(1-\gamma)f_\beta(m_-)\right) \sum_{m\in[m_{\delta_N},m_+(N))} \mathcal{Q}_{\beta,N}(m)\,\Big[\exp\left(\beta N(1-\gamma)f_\beta(m)\right) \\
&\qquad + \mathrm{e}^{\beta N(1-\gamma)3\varepsilon_N + N\ell_N(\theta_N)}\exp\left(\beta N(1-\gamma)f_\beta(m_+)\right)\Big]\,\mathrm{e}^{-\beta N(1-\gamma)\,\delta_N}(1+o(1)).
\end{aligned} \tag{A.144}$$

Using (A.61) of Corollary A.2.7 with

$$\begin{aligned}
\bar{g}(m) &= \mathbb{1}_{m\in[m_{\delta_N},m_+(N))} \\
&\quad\times (m)\left[\exp\left(\beta N(1-\gamma)f_\beta(m)\right) + \mathrm{e}^{\beta N(1-\gamma)3\varepsilon_N+N\ell_N(\theta_N)}\exp\left(\beta N(1-\gamma)f_\beta(m_+)\right)\right]
\end{aligned} \tag{A.145}$$

we obtain

$$
\begin{aligned}
&\sum_{\sigma\in\mathcal{S}_N[U_{\delta,N}(m_+)\setminus\{m_+(N)\}]} \mu_{\beta,N}(\sigma)\, h^N_{m_-,m_+}(\sigma) \\
&\overset{P(s)}{\leq} \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))}{Z_{\beta,N}} \exp\left(-\beta N(1-\gamma) f_\beta(m_-)\right) \mathrm{e}^{-\beta N(1-\gamma)\,\delta_N} \sum_{m\in[m_{\delta_N},m_+(N))} \exp\Big(-\beta N f_\beta(m)\Big) \\
&\times \sqrt{\frac{2}{\pi N(1-m^2)}}\Big[\exp\left(\beta N(1-\gamma) f_\beta(m)\right) + \mathrm{e}^{\beta N(1-\gamma)3\varepsilon_N + N\ell_N(\theta_N)} \exp\left(\beta N(1-\gamma) f_\beta(m_+)\right)\Big] \\
&\leq \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))}{Z_{\beta,N}} \exp\left(-\beta N(1-\gamma) f_\beta(m_-)\right) \mathrm{e}^{-\beta N(1-\gamma)\,\delta_N} N\sqrt{\frac{2}{\pi N(1-m_+^2)}} \\
&\times \Big[\exp\Big(-\gamma\,\beta N f_\beta(m_+)\Big) + \mathrm{e}^{\beta N(1-\gamma)3\varepsilon_N + N\ell_N(\theta_N)} \exp\Big(-\beta N f_\beta(m_+)\Big) \exp\left(\beta N(1-\gamma) f_\beta(m_+)\right)\Big] \\
&= \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))}{Z_{\beta,N}} \exp\Big(-\beta N f_\beta(m_-)\Big)\sqrt{\frac{2N}{\pi(1-m_+^2)}} \exp\Big(-\gamma\,\beta N[f_\beta(m_+)-f_\beta(m_-)]\Big) \\
&\times \mathrm{e}^{-\beta N(1-\gamma)(\delta_N-3\varepsilon_N)+N\ell_N(\theta_N)}\Big[\mathrm{e}^{-\beta N(1-\gamma)3\varepsilon_N - N\ell_N(\theta_N)} + 1\Big] \\
&\leq \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))}{Z_{\beta,N}} \exp\Big(-\beta N f_\beta(m_-)\Big)\sqrt{\frac{2}{\pi(1-m_+^2)}} \\
&\times \exp\Big[-\beta N\Big(\gamma[f_\beta(m_+)-f_\beta(m_-)] + (1-\gamma)(\delta_N-3\varepsilon_N) - \tfrac{1}{\beta}\ell_N(\theta_N) - \varepsilon_N\Big)\Big].
\end{aligned}
\tag{A.146}
$$

In the last step we embedded $[\mathrm{e}^{-\beta N(1-\gamma)3\varepsilon_N - N\ell_N(\theta_N)}+1]$ in the already present $(1+o(1))$ and bounded $\sqrt{N}$ by $\mathrm{e}^{-\beta N(-\varepsilon_N)}$, because for N large enough $\frac{\log(N)}{2\beta N} \leq \varepsilon_N$ (which converges to $\varepsilon > 0$, see Step 4 in Section A.4.1).

Now we prove that this part is not relevant compared to the right hand side of (A.138). In particular, we show that, for a certain choice of γ,

$$
c_N = \gamma[f_\beta(m_+)-f_\beta(m_-)] + (1-\gamma)(\delta_N-3\varepsilon_N) - \tfrac{1}{\beta}\ell_N(\theta_N) - \varepsilon_N \tag{A.147}
$$

is positive and its limit,

$$
\lim_{N\to\infty} c_N = \gamma[f_\beta(m_+)-f_\beta(m_-)] + (1-\gamma)(\delta-3\varepsilon) - \tfrac{\theta}{2\beta}\left(\log(2)+3-\log(1-m_+)\right) - \varepsilon, \tag{A.148}
$$

is positive and finite. In order to achieve this, we choose $\gamma\in(0,1)$ small enough, such that c_N and its limit are positive, definitely in N. In particular, we want to impose

$$
0 < \gamma < \frac{\delta_N - 4\varepsilon_N - \frac{1}{\beta}\ell_N(\theta_N)}{f_\beta(m_-)-f_\beta(m_+)+\delta_N-3\varepsilon_N} < 1, \tag{A.149}
$$

definitely in N, and

$$
0 < \gamma < \frac{\delta - 4\varepsilon - \frac{1}{\beta}\lim_{N\to\infty}\ell_N(\theta)}{f_\beta(m_-)-f_\beta(m_+)+\delta-3\varepsilon} < 1. \tag{A.150}
$$

First, we notice that it is easy to check that the previous quantities are strictly smaller than 1. Second, we want to show that a strictly positive γ satisfying (A.149)-(A.150) exists. Note that $\ell_N(\theta_N)$, defined in (A.164), has the following trivial upper bound for every N,

$$
\ell_N(\theta_N) \leq \theta_N\left(\beta + \log 2 + O(\theta_N)\right). \tag{A.151}
$$

Thus, a sufficient condition is to choose, for N large enough, $\gamma \geq \gamma_0$, where

$$\gamma_0 = \frac{\delta - 4\varepsilon - \theta\left(1 + \frac{\log 2}{\beta} + O(\theta)\right)}{f_\beta(m_-) - f_\beta(m_+) + \delta} \tag{A.152}$$

is clearly strictly positive. Indeed, we can choose $\varepsilon > 0$ sufficiently small for the numerator on the left hand side of (A.152) to be positive, while θ is small accordingly to ε (see Section A.4.1). We conclude by obtaining, for N sufficiently large,

$$\sum_{\sigma \in \mathcal{S}_N[U_{\delta,N}(m_+)\setminus\{m_+(N)\}]} \mu_{\beta,N}(\sigma)\, h^N_{m_-,m_+}(\sigma) \overset{P(s)}{\leq} \frac{\mathrm{e}^{s+\alpha+o(1)}}{Z_{\beta,N}} \exp\left(-\beta N(f_\beta(m_-) + c_N)\right)\sqrt{\tfrac{2}{\pi(1-m_+^2)}}, \tag{A.153}$$

where $0 < c_N = O(1)$.

Conclusion.

With the previous bounds at hand, we are now ready to conclude the proof of the upper bound. Decomposing the sum over $\mathcal{S}_N$ using (A.134), and inserting the estimates we computed above into (A.128), we obtain

$$\begin{aligned}
&\sum_{\sigma \in \mathcal{S}_N} \mu_{\beta,N}(\sigma)\, h^N_{m_-,m_+}(\sigma) \\
&\overset{P(s)}{\leq} \frac{\mathrm{e}^{s+\alpha}\,(1+o(1))}{Z_{\beta,N}} \exp\Big(-\beta N f_\beta(m_-)\Big)\Bigg[\mathrm{e}^{-\beta N\delta_N}\Bigg(\sqrt{\frac{2}{\pi N}} \sum_{m \in U^c_{\delta,N}\setminus\{1,-1\}} \frac{1}{\sqrt{(1-m^2)}} + 2\Bigg) \\
&\qquad + \sqrt{\frac{2}{\pi\,(1-m_+^2)}}\mathrm{e}^{-\beta N c_N} + \frac{1}{\sqrt{(1-m_-^2)\,\beta f''_\beta(m_-)}}\Bigg] \\
&\leq \frac{\mathrm{e}^{s+\alpha}}{Z_{\beta,N}} \exp\Big(-\beta N f_\beta(m_-)\Big)\frac{1}{\sqrt{(1-m_-^2)\,\beta f''_\beta(m_-)}}\,(1+o(1))\,,
\end{aligned} \tag{A.154}$$

concluding the proof. □

A.4.3 Some technical results

In this section we prove Lemma A.4.2, which is pivotal in obtaining the upper bound in Theorem A.1.7, (see (A.144)). The proof is quite involved, therefore we split it into subsequent technical results. Before starting the proof, we give a brief outline of this section. First, we state Lemma A.4.2 and prove it via Lemmas A.4.4, A.4.5 and A.4.6, which follow later on. Second, we give the proof of Lemmas A.4.4 and A.4.5. The latter relies on Lemma A.4.6, which we subsequently prove using Lemma A.4.7. We conclude the section proving Lemma A.4.7. Throughout this section we will use the notation introduced in Section A.4.1.

Lemma A.4.2. *For all $\sigma \in \mathcal{S}_N\,[[m_{\delta_N}, m_+(N))]$, for all $\gamma \in (0,1)$ and $\varepsilon > 0$,*

$$\begin{aligned}
\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right) &\leq \exp\left(-\beta N(1-\gamma)[f_\beta(m_-) + \delta_N]\right)(1+o(1)) \\
&\quad\times \Bigg[\exp\left(\beta N(1-\gamma) f_\beta(m(\sigma))\right) + \exp\left(\beta N(1-\gamma)[f_\beta(m_+) + 3\varepsilon_N] + N\ell_N(\theta_N)\right)\Bigg],
\end{aligned} \tag{A.155}$$

where $\ell_N(\cdot)$ *is defined in* (A.164).

Proof. For all $\sigma \in \mathcal{S}_N\left[[m_{\delta_N}, m_+(N))\right]$, we have

$$\begin{aligned}
&\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right)\\
&= \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}, \tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_{\varepsilon_N}]}\right)\\
&\quad + \sum_{\eta\in\mathcal{S}_N[m_{\varepsilon_N}]} \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}, \tau_\eta < \tau_{\mathcal{S}_N[\{m_{\varepsilon_N}, m_-(N), m_+(N)\}]}\right)\\
&= \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\right)\\
&\quad + \sum_{\eta\in\mathcal{S}_N[m_{\varepsilon_N}]} \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]} \,\middle|\, \tau_\eta < \tau_{\mathcal{S}_N[\{m_{\varepsilon_N}, m_-(N), m_+(N)\}]}\right)\\
&\qquad \times \mathbb{P}_\sigma\left(\tau_\eta < \tau_{\mathcal{S}_N[\{m_{\varepsilon_N}, m_-(N), m_+(N)\}]}\right),
\end{aligned} \tag{A.156}$$

where we notice that,

$$\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]} \,\middle|\, \tau_\eta < \tau_{\mathcal{S}_N[\{m_{\varepsilon_N}, m_-(N), m_+(N)\}]}\right) = \mathbb{P}_\eta\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right). \tag{A.157}$$

Using the Markov property and taking the maximum of the first factor out of the sum, we have that, for all $\sigma \in \mathcal{S}_N\left[[m_{\delta_N}, m_+(N))\right]$,

$$\begin{aligned}
&\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right)\\
&\le \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\right)\\
&\quad + \left(\max_{\eta\in\mathcal{S}_N[m_{\varepsilon_N}]} \mathbb{P}_\eta\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right)\right) \sum_{\eta\in\mathcal{S}_N[m_{\varepsilon_N}]} \mathbb{P}_\sigma\left(\tau_\eta < \tau_{\mathcal{S}_N[\{m_{\varepsilon_N}, m_-(N), m_+(N)\}]}\right)\\
&= \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\right)\\
&\quad + \left(\max_{\eta\in\mathcal{S}_N[m_{\varepsilon_N}]} \mathbb{P}_\eta\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right)\right) \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_{\varepsilon_N}]} < \tau_{\mathcal{S}_N[\{m_-(N), m_+(N)\}]}\right)\\
&\le \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\right)\\
&\quad + \left(\max_{\eta\in\mathcal{S}_N[m_{\varepsilon_N}]} \mathbb{P}_\eta\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right)\right) \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_{\varepsilon_N}]} < \tau_{\mathcal{S}_N[m_+(N)]}\right).
\end{aligned} \tag{A.158}$$

We first consider the case $\sigma \in \mathcal{S}_N[m_{\varepsilon_N}]$. By Lemma A.4.4, we get

$$\begin{aligned}
\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right) &\le \mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\right)\\
&\quad + \left(\max_{\eta\in\mathcal{S}_N[m_{\varepsilon_N}]} \mathbb{P}_\eta\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[m_+(N)]}\right)\right)\left(1 - \mathrm{e}^{-N\ell_N(\theta_N)}(1+o(1))\right).
\end{aligned} \tag{A.159}$$

Taking the maximum over σ and noticing that the same term appears in both right and left hand side of the inequality, we obtain

$$\begin{aligned}&\max_{\sigma\in\mathcal{S}_N[m_{\varepsilon_N}]}\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]}<\tau_{\mathcal{S}_N[m_+(N)]}\right)\\&\le\max_{\sigma\in\mathcal{S}_N[m_{\varepsilon_N}]}\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]}<\tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}\right)\mathrm{e}^{N\ell_N(\theta_N)}(1+o(1))\\&\le\exp\left[-\beta N\left(1-\gamma\right)\left[f_\beta(m_-)+\delta_N-f_\beta\left(m_{\varepsilon_N}-\tfrac{2}{N}\right)-\varepsilon_N\right]-N\ell_N(\theta_N)\right](1+o(1)),\end{aligned}\tag{A.160}$$

where we used Lemma A.4.5.

By Taylor expansion of $f_\beta\left(m_{\varepsilon_N}-\frac{2}{N}\right)$ and definition of m_{ε_N}, we get

$$\begin{aligned}&\max_{\sigma\in\mathcal{S}_N[m_{\varepsilon_N}]}\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]}<\tau_{\mathcal{S}_N[m_+(N)]}\right)\\&\le\exp\left[-\beta N\left(1-\gamma\right)\left[f_\beta(m_-)+\delta_N-3\varepsilon_N-f_\beta(m_+)\right]-N\ell_N(\theta_N)\right](1+o(1)),\end{aligned}\tag{A.161}$$

where the last inequality holds for N sufficiently large. Here we bounded the Taylor expansion error $O\left(\frac{1}{N}\right)$ with ε_N, which converges to $\varepsilon>0$ (see Step 4 in Section A.4.1).

Now we consider the case where $\sigma\in\mathcal{S}_N\left[[m_{\delta_N},m_+(N))\setminus\{m_{\varepsilon_N}\}\right]$. Going back to (A.158) and using again (A.161) we obtain

$$\begin{aligned}\mathbb{P}_\sigma&\left(\tau_{\mathcal{S}_N[m_-(N)]}<\tau_{\mathcal{S}_N[m_+(N)]}\right)\le\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]}<\tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}\right)\\&\qquad+\exp\left[-\beta N(1-\gamma)\left[f_\beta(m_-)+\delta_N-3\varepsilon_N-f_\beta(m_+)\right]-N\ell_N(\theta_N)\right](1+o(1))\\&\le\exp\left(-\beta N(1-\gamma)[f_\beta(m_-)+\delta_N]\right)(1+o(1))\\&\qquad\times\left[\exp\left(\beta N(1-\gamma)f_\beta(m(\sigma))\right)+\exp\left(\beta N(1-\gamma)[f_\beta(m_+)+3\varepsilon_N]+N\ell_N(\theta_N)\right)\right].\end{aligned}\tag{A.162}$$

In the last inequality we used Lemma A.4.6, which holds for $\sigma\in\mathcal{S}_N[[m_{\delta_N},m_{\varepsilon_N})]$, and that $\mathbb{P}_\sigma(\tau_{\mathcal{S}_N[m_-(N)]}<\tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]})=0$ for all $\sigma\in\mathcal{S}_N[(m_{\varepsilon_N},m_+(N))]$.

□

Remark A.4.3. *In Lemma A.4.2 one might try to further bound the r.h.s. of* (A.155) *using that $f_\beta(m(\sigma))$ is bounded by $f_\beta(m_{\delta_N})=f_\beta(m_-)+\delta_N$. This would yield to the trivial upper bound 1 on $\mathbb{P}_\sigma(\tau_{\mathcal{S}_N[m_-(N)]}<\tau_{\mathcal{S}_N[m_+(N)]})$, which is not sufficient for our purpose of proving that the second term in* (A.154) *is negligible with respect to the last one. The way to go is, therefore, to keep the dependence on $m(\sigma)$ in order to obtain later a more suitable bound, uniform in m, by exploiting the smallness of $\mathcal{Q}_{\beta,N}(m(\sigma))$ in* (A.144) *and* (A.146).

In order for (A.159) to be true, we have to prove the following result.

Lemma A.4.4. *For all $\sigma\in\mathcal{S}_N[m_{\varepsilon_N}]$, for ε sufficiently small and for N sufficiently large,*

$$\mathbb{P}_\sigma(\tau_{\mathcal{S}_N[m_+(N)]}<\tau_{\mathcal{S}_N[m_{\varepsilon_N}]})\ge\mathrm{e}^{-N\ell_N(\theta_N)}(1+o(1)),\tag{A.163}$$

where $\ell_N:\mathbb{R}\to\mathbb{R}$ is defined by

$$\begin{aligned}\ell_N(x)=\tfrac{1}{2}\Big[&x\left(\log 2+\beta\,|2-2h|+1\right)-(1-m_+(N)+x)\log(1-m_+(N)+x)\\&+(1-m_+(N))\log(1-m_+(N))\Big].\end{aligned}\tag{A.164}$$

Proof. Recall that $\{\sigma(t)\}_{t\geq 0}$ is the Markov chain with transition probabilities (A.5) and, for $\sigma \in \mathcal{S}_N$ with $m(\sigma) < m_+(N)$, let

$$A_N(\sigma) = \Big\{(\sigma(0), \sigma(1), \sigma(2), \dots) : \sigma(0) = \sigma, \forall i \in \mathbb{N}, \sigma(i) \in \mathcal{S}_N, \sigma(i) \sim \sigma(i+1), \\ \exists\, k \in \mathbb{N} \text{ s.t. } \sigma(k) \in \mathcal{S}_N[m_+(N)], \text{ and } \forall i \leq k-1,\ m(\sigma(i+1)) = m(\sigma(i)) + \tfrac{2}{N}\Big\} \quad \text{(A.165)}$$

be the set of infinite paths starting in σ and having increasing magnetisation until the set $\mathcal{S}_N[m_+(N)]$ is reached.

Notice that, for fixed σ and N, the number k of steps of increasing magnetisation to reach $\mathcal{S}_N[m_+(N)]$ is fixed, namely $k = \frac{N}{2}(m_+(N) - m(\sigma))$.

We want to partition $A_N(\sigma)$ according to the values of the first $k+1$ elements of its paths. Given a sequence $\pi \in \mathcal{S}_N^{k+1}$, let us denote by $\{\pi\}$ the set of all paths in $A_N(\sigma)$ in which the first $k+1$ elements are exactly given by π, namely

$$\{\pi\} = \{(\sigma(0), \sigma(1), \dots, \sigma(k), \sigma(k+1), \dots) \in A_N(\sigma) : (\sigma(0), \dots, \sigma(k)) = \pi\}. \quad \text{(A.166)}$$

Notice that, by definition of $A_N(\sigma)$, $\{\pi\}$ is empty for many $\pi \in \mathcal{S}_N^{k+1}$. We denote by $B_N(\sigma)$ the set of all the sequences $\pi \in \mathcal{S}_N^{k+1}$ such that $\{\pi\}$ is not empty. Thus, we obtain the following partition of $A_N(\sigma)$

$$A_N(\sigma) = \bigsqcup_{\pi \in B_N(\sigma)} \{\pi\}. \quad \text{(A.167)}$$

Fix $\sigma \in \mathcal{S}_N[m_{\varepsilon_N}]$, then one simply notices that

$$\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_+(N)]} < \tau_{\mathcal{S}_N[m_{\varepsilon_N}]}\right) \geq \mathbb{P}_\sigma(A_N(\sigma)) = \sum_{\pi \in B_N(\sigma)} \mathbb{P}_\sigma(\{\pi\}). \quad \text{(A.168)}$$

Thus, we first find a lower bound on $\mathbb{P}_\sigma(\{\pi\})$ independent of π in $B_N(\sigma)$ and later we compute the cardinality of $B_N(\sigma)$. Fix $\pi = (\sigma(0), \sigma(1), \sigma(2), \dots, \sigma(k)) \in B_N(\sigma)$, then we have

$$\mathbb{P}_\sigma(\{\pi\}) = \prod_{i=1}^k p_N(\sigma(i-1), \sigma(i)) = \frac{1}{N^k}\prod_{i=1}^k \exp\left(-\beta\,[H(\sigma(i)) - H(\sigma(i-1))]_+\right) \\ \geq \frac{C^k}{N^k}\prod_{i=1}^k \exp\left(-\beta N[E(m_i) - E(m_{i-1})]_+\right) = \frac{C^k}{N^k}\prod_{i=1}^k \exp\left(-\beta\left[-2m_{i-1} - \frac{2}{N} - 2h\right]_+\right), \quad \text{(A.169)}$$

where $m_i = m(\sigma(i))$, $C = \exp(-\beta\,|2-2h|)$ and we used the following fact

$$\frac{\exp\big(-\beta[H(\sigma(i)) - H(\sigma(i-1))]_+\big)}{\exp\big(-\beta N[E(m_i) - E(m_{i-1})]_+\big)} = \frac{\exp\big(-\beta[H(\sigma(i)) - H(\sigma(i-1))]_+\big)}{\exp\big(-\beta\big[-2m_{i-1} - \frac{2}{N} - 2h\big]_+\big)} \\ \geq \exp\big(-\beta[H(\sigma(i)) - H(\sigma(i-1))]_+\big) = \exp\left(-\beta\left[-\frac{2}{N}\sum_{j:j\neq r} J_{jr}\sigma(i-1)_j - 2h\right]_+\right) \quad \text{(A.170)} \\ \geq \exp\left(-\beta\left[2 - 2h - \tfrac{2}{N}\right]_+\right) \geq \exp\left(-\beta\,|2-2h|\right),$$

where r is the index of the spin to be flipped to go from $\sigma(i-1)$ to $\sigma(i)$. Therefore, recalling that $m_i \in [m_{\varepsilon_N}, m_+(N)]$, we obtain the following lower bound independent of π

$$\mathbb{P}_\sigma(\{\pi\}) \geq \frac{C^k}{N^k} \prod_{i=1}^{k} \exp\left(-\beta\left[-2m_{\varepsilon_N} - \frac{2}{N} - 2h\right]_+\right) = \frac{C^k}{N^k}. \tag{A.171}$$

Indeed, for ε_N sufficiently small, m_{ε_N} is close to $m_+(N) > 0$, allowing us to assume $m_{\varepsilon_N} > 0$. Therefore, $-2m_{\varepsilon_N} - \frac{2}{N} - 2h < 0$, which implies the last equality in (A.171).

We are left to compute the cardinality of $B_N(\sigma)$, with $\sigma \in \mathcal{S}_N[m_{\varepsilon_N}]$, namely we have to count all paths from σ to $\mathcal{S}_N[m_+(N)]$ with increasing magnetisation and length $k+1$. Any of these paths is characterised by a final spin $\bar{\sigma} \in \mathcal{S}_N[m_+(N)]$ and a sequence of negative spins which are flipped. Notice that $\bar{\sigma}$ is reachable by σ through a path with increasing magnetisation if and only if the two following properties are satisfied: $\bar{\sigma}$ has k positive spins more than σ and, for all $i \in [N]$, $\sigma_i = +1$ implies $\bar{\sigma}_i = +1$. Thus, a configuration $\bar{\sigma} \in \mathcal{S}_N[m_+(N)]$ reachable by σ through a path with increasing magnetisation is characterised by the k spins which are negative in σ and positive in $\bar{\sigma}$. Therefore, the number of reachable configurations $\bar{\sigma}$ is

$$\binom{\frac{1}{2}N\,(1-m_{\varepsilon_N})}{k} = \binom{\frac{1}{2}N\,[1-m_+(N)+\theta_N]}{\frac{1}{2}N\theta_N}, \tag{A.172}$$

being $\frac{1}{2}N(1-m_{\varepsilon_N})$ the number of negative spins of $\sigma \in \mathcal{S}_N[m_{\varepsilon_N}]$ and $k = \frac{1}{2}N\theta_N$, where θ_N has been defined in Section A.4.1.

The number of paths with increasing magnetisation from $\sigma \in \mathcal{S}_N[m_{\varepsilon_N}]$ to a reachable $\bar{\sigma} \in \mathcal{S}_N[m_+(N)]$, both fixed, is $k!$, namely the number of permutations of the k negative spins which are flipped along a path. Thus, being $k = \frac{1}{2}N\theta_N$, the cardinality of $B_N(\sigma)$ is

$$\left(\tfrac{1}{2}N\theta_N\right)! \binom{\frac{1}{2}N[1-m_+(N)+\theta_N]}{\frac{1}{2}N\theta_N}. \tag{A.173}$$

Going back to (A.168), we obtain

$$\begin{aligned}\mathbb{P}_\sigma\big(\tau_{\mathcal{S}_N[m_+(N)]} < \tau_{\mathcal{S}_N[m_{\varepsilon_N}]}\big) &\geq \sum_{\pi \in B_N(\sigma)} \mathbb{P}_\sigma(\{\pi\}) \\ &\geq \left(\frac{C}{N}\right)^{\frac{1}{2}N\theta_N} \left(\tfrac{1}{2}N\theta_N\right)! \binom{\frac{1}{2}N[1-m_+(N)+\theta_N]}{\frac{1}{2}N\theta_N} \\ &= \mathrm{e}^{-\frac{1}{2}N\theta_N \log\frac{N}{C}} \frac{N(1-m_+(N)+\theta_N)}{2}! \left[\frac{N(1-m_+(N))}{2}!\right]^{-1}.\end{aligned} \tag{A.174}$$

Using Stirling's approximation $n! = \sqrt{2\pi n}\, n^n \mathrm{e}^{-n}(1+o(1)) = \sqrt{2\pi n}\, \mathrm{e}^{n(\log n - 1)}(1+o(1))$ and the notation

$$k_{\theta_N} = \frac{1-m_+(N)+\theta_N}{1-m_+(N)}, \tag{A.175}$$

we obtain

$$\begin{aligned}&\frac{N(1-m_+(N)+\theta_N)}{2}! \left[\frac{N(1-m_+(N))}{2}!\right]^{-1} \\ &\quad= \sqrt{k_{\theta_N}} \exp\left[\tfrac{N(1-m_+(N))}{2}\log(k_{\theta_N}) + \tfrac{1}{2}N\theta_N \log\left(\tfrac{N(1-m_+(N)+\theta_N)}{2}\right) - \tfrac{1}{2}N\theta_N\right](1+o(1)).\end{aligned} \tag{A.176}$$

Thus, since $k_{\theta_N} \geq 1$ and $C = \exp(-\beta\,|2-2h|)$, we conclude by

$$\begin{aligned}
&\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_+(N)]} < \tau_{\mathcal{S}_N[m_{\varepsilon_N}]}\right)\\
&\geq \sqrt{k_{\theta_N}}\,\mathrm{e}^{-\frac{N}{2}\left(\theta_N \log\left(\frac{N}{C}\right)+\theta_N-(1-m_+(N))\log(k_{\theta_N})-\theta_N\log\left(\frac{N}{2}(1-m_+(N)+\theta_N)\right)\right)}(1+o(1))\\
&\geq \mathrm{e}^{-\frac{N}{2}\left(\theta_N \log\left(\frac{N}{C}\right)+\theta_N-(1-m_+(N))\log(k_{\theta_N})-\theta_N\log\left(\frac{N}{2}\right)-\theta_N\log(1-m_+(N)+\theta_N)\right)}(1+o(1))\\
&= \mathrm{e}^{-\frac{N}{2}(\theta_N\log(2)+\theta_N\,\beta|2-2h|+\theta_N-(1-m_+(N)+\theta_N)\log(1-m_+(N)+\theta_N)+(1-m_+(N))\log(1-m_+(N)))}(1+o(1))\\
&= \mathrm{e}^{-N\ell_N(\theta_N)}(1+o(1)). \qquad \text{(A.177)}
\end{aligned}$$

□

To prove Lemma A.4.2 we used the following fact.

Lemma A.4.5. *For $\sigma \in \mathcal{S}_N[m_{\varepsilon_N}]$, for N sufficiently large and any $\gamma \in (0,1)$,*

$$\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}\right) \leq \exp\left(-\beta N(1-\gamma)\left[f_\beta(m_-)+\delta_N - f_\beta\left(m_{\varepsilon_N}-\tfrac{2}{N}\right)-\varepsilon_N\right]\right). \qquad \text{(A.178)}$$

Proof. Let us denote by $W_N(m)$ the event of making the first flip in $\mathcal{S}_N[m]$.

For $\sigma \in \mathcal{S}_N[m_{\varepsilon_N}]$, conditioning on the first step, we obtain

$$\begin{aligned}
&\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}\right)\\
&\quad= \mathbb{P}_\sigma\left(W_N\left(m_{\varepsilon_N}+\tfrac{2}{N}\right)\right)\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}\,\middle|\,W_N\left(m_{\varepsilon_N}+\tfrac{2}{N}\right)\right)\\
&\qquad+\mathbb{P}_\sigma\left(W_N\left(m_{\varepsilon_N}-\tfrac{2}{N}\right)\right)\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}\,\middle|\,W_N\left(m_{\varepsilon_N}-\tfrac{2}{N}\right)\right)\\
&\quad= \mathbb{P}_\sigma\left(W_N\left(m_{\varepsilon_N}+\tfrac{2}{N}\right)\right)\sum_{\sigma'\in\mathcal{S}_N\left[m_{\varepsilon_N}+\frac{2}{N}\right],\sigma\sim\sigma'}\mathbb{P}_{\sigma'}\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}\right)\\
&\qquad+\mathbb{P}_\sigma\left(W_N\left(m_{\varepsilon_N}-\tfrac{2}{N}\right)\right)\sum_{\sigma'\in\mathcal{S}_N\left[m_{\varepsilon_N}-\frac{2}{N}\right],\sigma\sim\sigma'}\mathbb{P}_{\sigma'}\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}\right).
\end{aligned} \qquad \text{(A.179)}$$

The first term vanishes because all the probabilities in the sum are zero. Thus, we get the upper bound

$$\mathbb{P}_\sigma(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}) \leq \sum_{\sigma'\in\mathcal{S}_N\left[m_{\varepsilon_N}-\frac{2}{N}\right],\sigma\sim\sigma'}\mathbb{P}_{\sigma'}(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}). \qquad \text{(A.180)}$$

Using first Lemma A.4.6 which gives bounds uniform in σ', we obtain

$$\begin{aligned}
&\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N),m_{\varepsilon_N}\}]}\right)\\
&\leq \sum_{\sigma'\in\mathcal{S}_N\left[m_{\varepsilon_N}-\frac{2}{N}\right],\sigma\sim\sigma'}\exp\left(-\beta N(1-\gamma)\left[f_\beta(m_-)+\delta_N-f_\beta(m(\sigma'))\right]\right)\\
&= N\frac{1+m_{\varepsilon_N}}{2}\exp\left(-\beta N(1-\gamma)\left[f_\beta(m_-)+\delta_N-f_\beta\left(m_{\varepsilon_N}-\tfrac{2}{N}\right)\right]\right)\\
&= \exp\left(-\beta N(1-\gamma)\left[f_\beta(m_-)+\delta_N-f_\beta\left(m_{\varepsilon_N}-\tfrac{2}{N}\right)-\tfrac{\log N+O(1)}{\beta N(1-\gamma)}\right]\right)\\
&\leq \exp\left(-\beta N(1-\gamma)\left[f_\beta(m_-)+\delta_N-f_\beta\left(m_{\varepsilon_N}-\tfrac{2}{N}\right)-\varepsilon_N\right]\right),
\end{aligned} \qquad \text{(A.181)}$$

where in the last inequality we used that, for N large enough, $\frac{\log N+O(1)}{\beta N(1-\gamma)} \leq \varepsilon_N$ (which converges to $\varepsilon > 0$, see Step 4 in Section A.4.1). □

In the proofs of Lemmas A.4.2 and A.4.5, we use the following fact.

Lemma A.4.6. *For $\sigma \in \mathcal{S}_N\left[[m_{\delta_N}, m_{\varepsilon_N})\right]$, for N sufficiently large and any $\gamma \in (0,1)$,*

$$\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\right) \leq \exp\left(-\beta N(1-\gamma)\left[f_\beta(m_-) + \delta_N - f_\beta(m(\sigma))\right]\right). \tag{A.182}$$

Proof. For $\sigma \in \mathcal{S}_N\left[[m_{\delta_N}, m_{\varepsilon_N})\right]$,

$$\mathbb{P}_\sigma\big(\tau_{\mathcal{S}_N[m_-(N)]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\big) \leq \mathbb{P}_\sigma\big(\tau_{\mathcal{S}_N[m_{\delta_N}]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\big), \tag{A.183}$$

being $m_-(N) < m^* < m_{\delta_N} < m_{\varepsilon_N} < m_+(N)$, for N sufficiently large. Therefore, we focus on finding an upper bound on the right hand side of (A.183). Assume that there exists a function ψ super-harmonic in $\mathcal{S}_N\left[[m_{\delta_N}, m_{\varepsilon_N})\right]$. As a consequence, $0 > L\psi(\sigma) = \frac{\partial}{\partial t}\mathbb{E}_\sigma\left[\psi(\sigma(t))\right]$. This implies $\mathbb{E}_\sigma\left[\psi(\sigma(t))\right] \leq \mathbb{E}_\sigma\left[\psi(\sigma(s))\right]$, for all $s < t$. Take $s = 0$, and $\sigma(0) = \sigma$, therefore $\mathbb{E}_\sigma\left[\psi(\sigma(t))\right] \leq \psi(\sigma)$, for all $t > 0$. Thus, $\psi(\sigma(t))$ is a super-martingale. For the integrable stopping time $T = \tau_{\mathcal{S}_N[m_{\delta_N}]} \wedge \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}$, we use Doob's optional stopping theorem for super-martingales to show that, for all σ in the domain $\mathcal{S}_N\left[[m_{\delta_N}, m_{\varepsilon_N})\right]$ of ψ, $\mathbb{E}_\sigma\left[\psi(\sigma(T))\right] \leq \psi(\sigma)$. Therefore,

$$\psi(\sigma) \geq \mathbb{E}_\sigma\left[\psi(\sigma(T))\right] \geq \min_{\sigma' \in \mathcal{S}_N[m_{\delta_N}]} \psi(\sigma')\mathbb{P}_\sigma\big(\tau_{\mathcal{S}_N[m_{\delta_N}]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\big), \tag{A.184}$$

which implies that

$$\mathbb{P}_\sigma\big(\tau_{\mathcal{S}_N[m_{\delta_N}]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\big) \leq \frac{\psi(\sigma)}{\min_{\sigma' \in \mathcal{S}_N[m_{\delta_N}]} \psi(\sigma')}. \tag{A.185}$$

For a suitably chosen ψ the latter inequality will yield the desired upper bound. Now we are left with the choice of a suitable $\psi : \mathcal{S}_N \to \mathbb{R}$ such that $L\psi(x) < 0$, for all $x \in \mathcal{S}_N\left[[m_{\delta_N}, m_{\varepsilon_N})\right]$. We define a function ψ which depends on a parameter $\gamma \in (0,1)$ and is constant on fixed magnetisation sets, i.e, for all $\sigma \in \mathcal{S}_N$,

$$\psi(\sigma) = \phi(m(\sigma)), \tag{A.186}$$

where $\phi : [-1,1] \to \mathbb{R}$ is defined by

$$\phi(m) = \exp\left(\beta N\,(1-\gamma) f_\beta(m)\right). \tag{A.187}$$

Our choice of ψ is similar to the one used by Bianchi, Bovier and Ioffe in [7, Proposition 6.4]. The choice of γ is relevant in (A.149).

We claim and prove later in Lemma A.4.7 that ψ is super-harmonic in $\mathcal{S}_N\left[[m_{\delta_N}, m_{\varepsilon_N})\right]$. Therefore, we conclude the proof by inserting ψ in (A.185) and obtaining

$$\begin{aligned}\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_{\delta_N}]} < \tau_{\mathcal{S}_N[\{m_+(N), m_{\varepsilon_N}\}]}\right) &\leq \frac{\exp\left(\beta N(1-\gamma) f_\beta(m(\sigma))\right)}{\min_{\sigma' \in \mathcal{S}_N[m_{\delta_N}]} \exp\left(\beta N(1-\gamma) f_\beta(m(\sigma'))\right)} \\ &= \exp\left(\beta N(1-\gamma)\left[f_\beta(m(\sigma)) - f_\beta(m_{\delta_N})\right]\right) \\ &= \exp\left(-\beta N(1-\gamma)\left[f_\beta(m_-) + \delta_N - f_\beta(m(\sigma))\right]\right),\end{aligned} \tag{A.188}$$

where we used the definition of m_{δ_N} (see Section A.4.1). □

We are now left with the proof of the super-harmonicity of ψ, which is used in the proof of Lemma A.4.6.

Lemma A.4.7. *ψ defined in* (A.186) *is super-harmonic in* $\mathcal{S}_N\left[[m_{\delta_N}, m_{\varepsilon_N})\right]$.

Proof. We have to prove that $L\psi(x) < 0$, for all $x \in \mathcal{S}_N\left[[m_{\delta_N}, m_{\varepsilon_N})\right]$. Fix x in $\mathcal{S}_N\left[[m_{\delta_N}, m_{\varepsilon_N})\right]$ and use the notation $\bar{m} = m(x)$. As usual, we try to rewrite the terms appearing in the expression for $L\psi(x)$ in terms of their mean-field version.

$$
\begin{aligned}
L\psi(x) &= \sum_{y\in\mathcal{S}_N} p(x,y)[\psi(y)-\psi(x)] \\
&= \frac{1}{N}\sum_{y\in\mathcal{S}_N} \mathbb{1}_{y\sim x} \exp\left(-\beta[H(y)-H(x)]_+\right) \\
&\qquad \times \Big[\exp\left(\beta(1-\gamma)Nf_\beta(m(y))\right) - \exp\left(\beta(1-\gamma)Nf_\beta(m(x))\right)\Big] \\
&= \frac{1}{N}\sum_{m\in\Gamma_N} \exp\left(-\beta N[E(m)-E(\bar{m})]_+\right) \\
&\qquad \times \Big[\exp\left(\beta N(1-\gamma)f_\beta(m)\right) - \exp\left(\beta N(1-\gamma)f_\beta(\bar{m})\right)\Big] \\
&\qquad\qquad \times \sum_{y:m(y)=m} \mathbb{1}_{x\sim y}\frac{\exp\left(-\beta[H(y)-H(x)]_+\right)}{\exp\left(-\beta N[E(m)-E(\bar{m})]_+\right)} \\
&\le \sum_{m\in\Gamma_N} \exp\left(-\beta N[E(m)-E(\bar{m})]_+\right)\phi(\bar{m})\Big[\exp\left(\beta N(1-\gamma)[f_\beta(m)-f_\beta(\bar{m})]\right)-1\Big] \\
&\qquad\qquad \times e^{2\beta}\left[\frac{1+\bar{m}}{2}\mathbb{1}_{\bar{m}-\frac{2}{N}}(m) + \frac{1-\bar{m}}{2}\mathbb{1}_{\bar{m}+\frac{2}{N}}(m)\right],
\end{aligned}
\tag{A.189}
$$

where ϕ is defined in (A.187) and we used the upper bound $\exp(2\beta)$ on $G(\sigma, m')$ as in the proof of the upper bound on capacity (see (A.99), (A.100)).

Now, recalling definition (A.20), we use the following notation

$$
r_+ = \tilde{r}_N\left(\bar{m}, \bar{m}+\tfrac{2}{N}\right) = \exp\left(-2\beta\left[-\frac{1}{N}-(\bar{m}+h)\right]_+\right)\frac{1-\bar{m}}{2}, \tag{A.190}
$$

$$
r_- = \tilde{r}_N\left(\bar{m}, \bar{m}-\tfrac{2}{N}\right) = \exp\left(-2\beta\left[-\frac{1}{N}+\bar{m}+h\right]_+\right)\frac{1+\bar{m}}{2}, \tag{A.191}
$$

and, for all $m \in \Gamma_N \setminus \{1\}$,

$$
g(m) = \frac{N}{2}\left[f_\beta\left(m+\tfrac{2}{N}\right) - f_\beta(m)\right]. \tag{A.192}
$$

Therefore, we can rewrite (A.189) as

$$
\begin{aligned}
L\psi(x) &\le e^{2\beta}\,\phi(\bar{m})\,r_+\Big[\exp\left(2\beta(1-\gamma)g(\bar{m})\right)-1\Big] \\
&\qquad + e^{2\beta}\,\phi(\bar{m})\,r_-\Big[\exp\left(-2\beta(1-\gamma)g\left(\bar{m}-\tfrac{2}{N}\right)\right)-1\Big] \\
&= e^{2\beta}\,\phi(\bar{m})\,r_+\,G_+,
\end{aligned}
\tag{A.193}
$$

where

$$G_+ = \left(\mathrm{e}^{2\beta(1-\gamma)g(\bar{m})} - 1\right) + \frac{r_-}{r_+}\left(\mathrm{e}^{-2\beta(1-\gamma)g\left(\bar{m}-\frac{2}{N}\right)} - 1\right). \tag{A.194}$$

Being $\mathrm{e}^{2\beta}$, $\phi(\bar{m})$ and r_+ positive, we have only to show that $G_+ < 0$. First we notice that

$$g(m) = -m - h + \frac{1}{\beta}I'(m) + O\left(\tfrac{1}{N}\right) \tag{A.195}$$

and similarly

$$g\left(m - \tfrac{2}{N}\right) = -m - h + \frac{1}{\beta}I'(m) + O\left(\tfrac{1}{N}\right). \tag{A.196}$$

Therefore,

$$g(m) - g\left(m - \tfrac{2}{N}\right) = O\left(\tfrac{1}{N}\right). \tag{A.197}$$

Then, since $I'(m) = \frac{1}{2}\log\left(\frac{1+m}{1-m}\right)$ (see (A.16)), and using (A.196) we have

$$\begin{aligned}
\frac{r_-}{r_+} &= \frac{1+\bar{m}}{1-\bar{m}}\frac{\exp\left(2\beta\left[-\frac{1}{N}-(\bar{m}+h)\right]_+\right)}{\exp\left(2\beta\left[-\frac{1}{N}+\bar{m}+h\right]_+\right)} \\
&= \frac{1+\bar{m}}{1-\bar{m}}\exp\left(-2\beta(\bar{m}+h)\right)\left(1+O\left(\tfrac{1}{N}\right)\right) \\
&= \exp\left(2I'(\bar{m}) - 2\beta(\bar{m}+h)\right)\left(1+O\left(\tfrac{1}{N}\right)\right) \\
&= \exp\left(2\beta\left[g\left(\bar{m}-\tfrac{2}{N}\right)+\bar{m}+h+O\left(\tfrac{1}{N}\right)\right] - 2\beta(\bar{m}+h)\right)\left(1+O\left(\tfrac{1}{N}\right)\right) \\
&= \exp\left(2\beta g(\bar{m}-\tfrac{2}{N})\right)\left(1+O\left(\tfrac{1}{N}\right)\right).
\end{aligned} \tag{A.198}$$

Therefore, rearranging (A.194) and then using (A.197) and (A.198), we obtain

$$\begin{aligned}
G_+ &= \left[\exp\left(2\beta(1-\gamma)g(\bar{m})\right)-1\right]\left[1-\frac{r_-}{r_+}\exp\left(-2\beta(1-\gamma)g\left(\bar{m}-\tfrac{2}{N}\right)\right)\right] \\
&\quad + \frac{r_-}{r_+}\left[\exp\left(2\beta(1-\gamma)\left[g(\bar{m})-g\left(\bar{m}-\tfrac{2}{N}\right)\right]\right)-1\right] \\
&= \left[\exp\left(2\beta(1-\gamma)g(\bar{m})\right)-1\right]\left[1-\exp\left(2\beta\gamma g\left(\bar{m}-\tfrac{2}{N}\right)\right)\left(1+O\left(\tfrac{1}{N}\right)\right)\right] \\
&\quad + \frac{r_-}{r_+}\left[\exp\left(2\beta(1-\gamma)\,O\left(\tfrac{1}{N}\right)\right)-1\right].
\end{aligned} \tag{A.199}$$

Notice that, for every $m \in [m_{\delta_N}, m_{\varepsilon_N}) \subset [m^*, m_+)$, $g(m)$ is negative, being f_β strictly decreasing in $[m^*, m_+)$. As a consequence, $\mathrm{e}^{2\beta(1-\gamma)g(\bar{m})} - 1 < 0$. Furthermore, for N sufficiently large, $1 - \mathrm{e}^{2\beta\gamma g(\bar{m}-\frac{2}{N})}\left(1+O\left(\frac{1}{N}\right)\right) > 0$, implying that the first term in (A.199) is negative.

Moreover, $\frac{r_-}{r_+} \geq 0$ is uniformly bounded from above, for N sufficiently large. Therefore, since β is finite, $\gamma \in (0,1)$ and the term $\left[\exp\left(2\beta(1-\gamma)\,O\left(\frac{1}{N}\right)\right)-1\right]$ is positive but converging to zero as N grows to infinity, the second term in (A.199) is negligible.

Therefore, for N sufficiently large, G_+ is negative, concluding the proof.

□

A.4.4 Lower bound on the harmonic sum

In this section we provide the main ideas to prove the second part of Theorem A.1.7, namely the lower bound on the harmonic sum in (A.128).

Proof of Theorem A.1.7. Lower bound. The proof is very similar to the proof of the upper bound we gave in Section A.4.2, therefore we omit the details. The main contribution is given once again by the sum on $\mathcal{S}_N\left[U_{\delta,N}(m_-)\right]$.

We have,

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{S}_N}\mu_{\beta,N}(\sigma)\,h^N_{m_-,m_+}(\sigma)\geq\sum_{\sigma\in\mathcal{S}_N[U_{\delta,N}(m_-)]}\mu_{\beta,N}(\sigma)\,h^N_{m_-,m_+}(\sigma)\\
&=\sum_{\sigma\in\mathcal{S}_N[U_{\delta,N}(m_-)]}\mu_{\beta,N}(\sigma)-\sum_{\sigma\in\mathcal{S}_N[U_{\delta,N}(m_-)]}\mu_{\beta,N}(\sigma)(1-h^N_{m_-,m_+}(\sigma))\\
&\geq\sum_{m\in U_{\delta,N}(m_-)\setminus[-1,m_-(N))}\mathcal{Q}_{\beta,N}(m)\\
&\quad-\sum_{\sigma\in\mathcal{S}_N[U_{\delta,N}(m_-)\setminus[-1,m_-(N))]}\mu_{\beta,N}(\sigma)\,\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_+(N)]}<\tau_{\mathcal{S}_N[m_-(N)]}\right).
\end{aligned}\tag{A.200}$$

The first term, i.e. the sum on the mesoscopic measure $\mathcal{Q}_{\beta,N}$, gives the main contribution. This sum can be estimated from below using the lower bound in Corollary A.2.5, obtaining a lower bound similar to the second upper bound in Corollary A.2.7 and applying the saddle point method as in (A.138). More precisely, using the notation (A.38), we have the following lower bound for $s>0$:

$$\sum_{m\in U_{\delta,N}(m_-)\setminus[-1,m_-(N))}\mathcal{Q}_{\beta,N}(m)\overset{P(s)}{\geq}\mathrm{e}^{\kappa-s}\,\frac{\exp\left(-\beta Nf_\beta(m_-)\right)}{Z_{\beta,N}\sqrt{(1-m_-^2)\,\beta f''_\beta(m_-)}}\,(1+o(1)).\tag{A.201}$$

The second term in (A.200), appearing with a negative sign in front, is estimated via an upper bound, obtaining

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{S}_N[U_{\delta,N}(m_-)\setminus[-1,m_-(N))]}\mu_{\beta,N}(\sigma)\,\mathbb{P}_\sigma\left(\tau_{\mathcal{S}_N[m_+(N)]}<\tau_{\mathcal{S}_N[m_-(N)]}\right)\\
&\qquad\leq\frac{\mathrm{e}^{s+\alpha}\exp\left(-\beta Nf_\beta(m_-)\right)}{Z_{\beta,N}}\sqrt{\frac{2}{\pi(1-m_+^2)}}\,\mathrm{e}^{-\beta Nc}(1+o(1)),
\end{aligned}\tag{A.202}$$

which is negligible compared to the right hand side of (A.201), concluding the proof.

We omit the proof of (A.202) being it again technical and very similar to the proof of the upper bound (A.153) in Part 4 of Section A.4.2. An analogue construction to the one given in Section A.4.1 and similar proofs to those in Section A.4.3 are needed. The main difference consists in restricting the analysis on a right neighbourhood of $m_-(N)$ instead of a left neighbourhood of $m_+(N)$.

□

Appendix B

Publication: Metastability for Glauber Dynamics on the Complete Graph with Coupling Disorder

This appendix reproduces exactly the content of the paper [19] with title "Metastability for Glauber Dynamics on the Complete Graph with Coupling Disorder", authored by Anton Bovier, Frank den Hollander and Saeda Marello, and published in *Communications in Mathematical Physics, 2022, 392, 307-345*, `https://doi.org/10.1007/s00220-022-04351-8`. This paper was summarised in Chapter 4.

B.1 Introduction and main results

B.1.1 Background

Interacting particle systems evolving according to a Metropolis dynamics associated with an energy functional called the Hamiltonian, may be trapped for a long time near a state that is a local minimum of the free energy, but not a global minimum. The deepest local minima are called *metastable states*, the global minimum is called the *stable state*. The transition from a metastable state to the stable state marks the relaxation of the system to equilibrium. To describe this relaxation, one needs to identify the set of critical configurations the system must attain in order to achieve this transition and to compute the crossover time. These critical configurations correspond to saddle points in the free energy landscape.

Metastability for interacting particle systems on *lattices* has been studied intensively in the past. For a summary, we refer the reader to the monographs by Olivieri and Vares [64], and Bovier and den Hollander [18]. Successful attempts towards understanding metastable behaviour in random environments were made for the random field Curie-Weiss model, by Mathieu and Picco [58], Bovier, Eckhoff, Gayrard and Klein [12] and Bianchi, Bovier and Ioffe [7, 8]. Recently, there has been interest in metastability for interacting particle systems on *random graphs*. This is challenging, because the crossover times typically depend on the realisation of the graph. In den Hollander and Jovanovksi [50] and Bovier, Marello and Pulvirenti [22], Glauber dynamics on dense *Erdős-Rényi* random graphs was analysed. Earlier work on metastability for Glauber dynamics on sparse random graphs can be found in Dommers [35] (random regular graph) and Dommers, den Hollander, Jovanovski and Nardi [38] (configuration model). The present paper is a first step towards the study of metastability for Glauber dynamics on *Chung-Lu*-like random graphs.

To the best of our knowledge, Tindemans and Capel [76] and Dommers, Giardinà, Giberti, van der Hofstad and Prioriello [36] are the only references where the model with the interaction Hamiltonian in (B.2) below has been studied in detail. Both focus on equilibrium properties only.

B.1.2 Glauber dynamics on the complete graph with coupling disorder

Let $\mathcal{K}_n$ be the complete graph on n vertices. Each vertex carries an Ising spin that can take the values -1 or $+1$. Let $S_n = \{-1,+1\}^{[n]}$ denote the set of spin configurations on $\mathcal{K}_n$, where $[n] = \{1,2,\dots,n\}$. Let $(\Omega,\mathcal{F},\mathcal{P})$ be an abstract probability space, and let $J = (J(i))_{i\in[n]}$ be a sequence of i.i.d. random variables on this probability space taking values in a *finite* set $\{a_1,\dots,a_k\} \subset [0,\infty)$ of cardinality $k \in \mathbb{N}$. The distribution of these random variables is given by

$$\mathcal{P}\left(J(i) = a_\ell\right) = \omega_\ell \in (0,1), \quad i \in [n], \ell \in [k], \tag{B.1}$$

with $\sum_{\ell\in[k]} \omega_\ell = 1$.

Let $H_n\colon S_n \to \mathbb{R}$ be the interaction Hamiltonian defined by

$$H_n(\sigma) \equiv -\frac{1}{n} \sum_{\substack{i,j\in[n]\\ i<j}} J(i)J(j)\,\sigma(i)\sigma(j) - h \sum_{i\in[n]} \sigma(i), \qquad \sigma \in S_n, \tag{B.2}$$

where $h \in [0,\infty)$ is the magnetic field. We consider *Glauber dynamics* on S_n, defined as the continuous-time Markov process with transition rates

$$r_n(\sigma,\sigma') = \begin{cases} \mathrm{e}^{-\beta[H_n(\sigma')-H_n(\sigma)]_+}, & \text{if } \sigma' \sim \sigma, \\ 0, & \text{otherwise,} \end{cases} \qquad \sigma,\sigma' \in S_n, \tag{B.3}$$

where $\beta \in (0,\infty)$ is the inverse temperature, $\sigma' \sim \sigma$ means that σ' differs from σ by a single spin-flip and $[\cdot]_+$ is the positive part. This dynamics is *reversible* with respect to the *Gibbs measure*

$$\mu_n(\sigma) \equiv \frac{1}{Z_n}\,\mathrm{e}^{-\beta H_n(\sigma)}, \qquad \sigma \in S_n, \tag{B.4}$$

where the normalising constant Z_n is called the partition sum. Note that the reference measure for (B.4) is the *counting measure* on S_n. We write

$$(\sigma_t)_{t\ge 0}, \qquad \sigma_t \in S_n, \tag{B.5}$$

to denote a path of the Glauber dynamics on S_n, and $\mathbb{P}_\sigma$ and $\mathbb{E}_\sigma$ to denote probability and expectation on path space given $\sigma_0 = \sigma$ (we suppress J,h,β and n from the notation).

For fixed n, if $h = 0$ the Hamiltonian in (B.2) has two global minima, at $\sigma \equiv +1$ and $\sigma \equiv -1$, while if $h > 0$ it achieves a global minimum at $\sigma \equiv +1$ and a local minimum at $\sigma \equiv -1$. The latter is the deepest local minimum not equal to the global minimum (at least for h small enough). However, in the limit as $n \to \infty$, these do *not* form a metastable pair of configurations because *entropy* comes into play.

B.1.3 Metastability on the complete graph with coupling disorder

In this section we state our main results.

Empirical magnetisations

The relevant quantity to monitor in order to characterise the metastable behaviour is the *disorder weighted magnetisation*

$$K_n(\sigma) = \frac{1}{n} \sum_{i \in [n]} J(i)\sigma(i), \qquad \sigma \in S_n. \tag{B.6}$$

The following quantities will be essential for *coarse-graining.* Define the *level sets*

$$A_{\ell,n} \equiv \{i \in [n] \colon \, J(i) = a_\ell\}, \qquad \ell \in [k], \tag{B.7}$$

and the *level magnetisations*

$$m_{\ell,n}(\sigma) \equiv \frac{1}{|A_{\ell,n}|} \sum_{i \in A_{\ell,n}} \sigma(i), \qquad \ell \in [k],\, \sigma \in S_n. \tag{B.8}$$

Put

$$m_n(\sigma) = \big(m_{\ell,n}(\sigma)\big)_{\ell \in [k]} \in [-1,1]^k, \qquad \sigma \in S_n, \tag{B.9}$$

and note that $K_n(\sigma) = \frac{1}{n}\sum_{\ell \in [k]} a_\ell \; |A_{\ell,n}| \; m_{\ell,n}(\sigma)$ depends on σ only through $m_n(\sigma)$. Thus, with abuse of notation, we may define

$$K_n(m) \equiv \frac{1}{n} \sum_{\ell \in [k]} a_\ell \, |A_{\ell,n}| \, m_\ell, \qquad m = (m_\ell)_{\ell \in [k]} \in [-1,1]^k, \tag{B.10}$$

so that $K_n(\sigma) = K_n(m_n(\sigma))$.

Thermodynamic limit

As $n \to \infty$, by the law of large numbers the random function K_n converges uniformly in probability to a deterministic function K given by

$$K(m) = \sum_{\ell \in [k]} a_\ell \, \omega_\ell \, m_\ell, \qquad m = (m_\ell)_{\ell \in [k]} \in [-1,1]^k. \tag{B.11}$$

Similarly, the random free energy function F_n converges uniformly in probability to a deterministic function $F_{\beta,h}$ (see (B.36) and (B.47) below for explicit formulas). In Section B.3, we show that the stationary points of $F_{\beta,h}$ are given by $\mathbf{m} = (\mathbf{m}_\ell)_{\ell \in [k]}$, where

$$\mathbf{m}_\ell = \tanh(\beta[a_\ell K(\mathbf{m}) + h]), \qquad \ell \in [k]. \tag{B.12}$$

Note that, via (B.12), the k-dimensional vector $\mathbf{m}$ is fully determined by the real number $K(\mathbf{m})$. Therefore, finding the stationary points of $F_{\beta,h}$ reduces to finding the solutions of the equation

$$K = T_{\beta,h}(K), \qquad T_{\beta,h}(K) = \sum_{\ell \in [k]} a_\ell \, \omega_\ell \, \tanh(\beta[a_\ell K + h]). \tag{B.13}$$

Metastable regime

It turns out that the critical inverse temperature β_c is given by

$$\beta_c = \left[\sum_{\ell\in[k]} a_\ell^2\omega_\ell\right]^{-1}. \tag{B.14}$$

Namely, if $\beta \in (0, \beta_c]$, then the system is not in the metastable regime for any $h \in [0,\infty)$, while if $\beta \in (\beta_c,\infty)$, then, for $h \in [0,\infty)$ small enough, it is in the metastable regime (i.e., (B.13) has more than one solution at which $T_{\beta,h}$ is not tangent to the diagonal). Given $\beta \in (\beta_c,\infty)$, the critical magnetic field $h_c(\beta)$ is the minimal value of h for which the system is not metastable. The *metastable regime* is thus

$$\beta \in (\beta_c,\infty), \qquad h \in [0, h_c(\beta)). \tag{B.15}$$

In Section B.3, we show that $\beta \mapsto h_c(\beta)$ is continuous on (β_c,∞), with

$$\lim_{\beta\downarrow\beta_c} h_c(\beta) = 0, \qquad \lim_{\beta\to\infty} h_c(\beta) = C \in (0,\infty), \tag{B.16}$$

where the explicit value of C is given in (B.65) below. Interestingly, $\beta \mapsto h_c(\beta)$ is not necessarily monotone, i.e., the metastable crossover may be *re-entrant*.

It turns out that there exists an $\ell \in [k]$ (depending on β, h and on the law of the components of J), such that $F_{\beta,h}$ has $2\ell + 1$ stationary points.

Metastable crossover

Let $\mathcal{M}_n$ be the set of minima of F_n. Given $\mathbf{m} \in \mathcal{M}_n$, define

$$\mathcal{M}_n(\mathbf{m}) \equiv \{m \in \mathcal{M}_n\backslash\mathbf{m}\colon\ F_n(m) \leq F_n(\mathbf{m})\}. \tag{B.17}$$

Let $\mathcal{G}(A,B)$ be the gate between two disjoint subsets A and B of $\mathcal{M}_n$. We refer to [18, Section 10.1] for a precise definition of the gate.

Fix $\mathbf{m}_n \in \mathcal{M}_n$ as the initial magnetisation. Throughout the paper we assume that the following hypotheses hold for $\mathbf{m}_n$.

Hypothesis B.1.1.

1. *$\mathcal{M}_n(\mathbf{m}_n)$ is non-empty.*
2. *The Hessian of F_n has only non-zero eigenvalues at $\mathbf{m}_n$ and at all the points in $\mathcal{G}(\mathbf{m}_n, \mathcal{M}_n(\mathbf{m}_n))$.*
3. *There is a* unique *point $\mathbf{t}_n$ in $\mathcal{G}(\mathbf{m}_n, \mathcal{M}_n(\mathbf{m}_n))$, which will often be called simply saddle point.*
4. *The saddle point $\mathbf{t}_n$ is such that $r_\ell\big[\,|A_{\ell,n}|\,(1-\mathbf{t}_{\ell,n}^2)\big]^{-1}$ takes distinct values for different $\ell \in [k]$, where r_ℓ is defined in* (B.89) *below.*

Hypothesis B.1.1(2) and (3) are made to avoid complications. Hypothesis B.1.1(4) is needed in the proof of Lemma B.4.2 below (as in [18, Lemma 14.9]). Neither is very restrictive: if for some parameter choice they fail, then after an infinitesimal parameter change they hold. Moreover, if

Hypothesis B.1.1(3) fails, it is sufficient to compute separately the contribution to the crossover time of the various saddle points in the gate.

Let $S_n[\mathbf{m}_n]$ and $S_n[\mathcal{M}_n(\mathbf{m}_n)]$ denote the sets of configurations in S_n for which the level magnetisations are $\mathbf{m}_n$ and are contained in $\mathcal{M}_n(\mathbf{m}_n)$, respectively. For $A \subset S_n$, write

$$\tau_A = \{t \geq 0\colon\, \sigma_t \in A,\, \sigma_{t-} \notin A\} \tag{B.18}$$

to denote the first hitting time or return time of A.

We next state our main results for the crossover time. Theorem B.1.2 provides a sharp asymptotics for the average crossover time from any metastable state to the set of states with lower free energy. Theorem B.1.3 shows that asymptotically the crossover time is exponential on the scale of its mean, a property that is standard for metastable behaviour.

Theorem B.1.2 (**Average crossover time with coupling disorder).**
Let $\mathbb{A}_n(\cdot)$ be the $k \times k$ Hessian matrix defined in (B.82) *below, and γ_n the unique negative solution of the equation in* (B.100) *below. For every $\mathbf{m_n} \in \mathcal{M}_n$ satisfying Hypothesis B.1.1 and within the metastable regime* (B.15)*, uniformly in $\sigma \in S_n[\mathbf{m_n}]$, and with $\mathcal{P}$-probability tending to* 1*,*

$$\mathbb{E}_\sigma\left[\tau_{S_n[\mathcal{M}_n(\mathbf{m_n})]}\right] = [1 + o_n(1)]\sqrt{\frac{[-\det(\mathbb{A}_n(\mathbf{t}_n))]}{\det(\mathbb{A}_n(\mathbf{m}_n))}}\left(\frac{\pi}{2\beta(-\gamma_n)}\right)\mathrm{e}^{\beta n[F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)]}. \tag{B.19}$$

Theorem B.1.3 (**Exponential law with coupling disorder).**
For every $\mathbf{m_n} \in \mathcal{M}_n$ satisfying Hypothesis B.1.1 and within the metastable regime (B.15)*, uniformly in $\sigma \in S_n[\mathbf{m}_n]$ and with $\mathcal{P}$-probability tending to* 1*,*

$$\mathbb{P}_\sigma\left(\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]} > t\,\mathbb{E}_\sigma\left[\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]}\right]\right) = [1 + o_n(1)]\,\mathrm{e}^{-t}, \qquad t \geq 0. \tag{B.20}$$

As the average crossover time estimated in Theorem B.1.2 is a random variable, we next provide more information on the randomness of the quantity in the right-hand side of (B.19), which depends on the realisation of the random variable J. The prefactor in (B.19) converges with $\mathcal{P}$-probability tending to 1 to a deterministic limit, which depends on the law of J but not on the realisation of J. However, the exponent does not converge to a deterministic limit. In Theorem B.1.4 we compute the exponent up to order $O(1)$. Recall that $F_n \to F_{\beta,h}$, $\mathbf{m}_n \to \mathbf{m}$ and $\mathbf{t}_n \to \mathbf{t}$ as $n \to \infty$.

Theorem B.1.4 (**Randomness of the exponent).**
For every $\mathbf{m_n} \in \mathcal{M}_n$ satisfying Hypothesis B.1.1 and within the metastable regime (B.15)*, in distribution,*

$$n[F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)] = n[F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})] + Z\sqrt{n} + O(1), \tag{B.21}$$

where Z is a normal random variable with mean zero and variance in $(0,\infty)$*, defined on* $(\Omega, \mathcal{F}, \mathcal{P})$ *and independent of J.*

The variance of Z turns out to be a complicated function of β, h and the distribution of J. We refer to Section B.6.3 for further details. Computing the exponent up to order 1 is in principle possible, but the formulas become rather complicated. Without this precision the prefactor in (B.19) is asymptotically negligible. Still, knowing this prefactor allows us to determine what the leading order behaviour of the randomness is.

B.1.4 Discussion on the continuous case

Bianchi, Bovier and Ioffe [7, 8] study the Curie-Weiss model with a *random magnetic field* whose distribution is continuous. Lumping techniques work for discrete distributions but not for continuous distributions. The latter require coarse-graining techniques to approximate the continuous distribution by a sequence of discrete distributions. In the present paper we consider pair interaction random variables with a discrete distribution only. It seems hard to obtain results with a similar precision for continuous distributions. The techniques employed in [7, 8] do not carry over, because the error introduced by the coarse-graining turns out to be quadratic rather than linear.

B.1.5 Techniques and outline

In order to prove Theorems B.1.2–B.1.4 we use the potential-theoretic approach to metastability developed in Bovier, Eckhoff, Gayrard and Klein [13, 14]. More specifically, we first find a sharp approximation of the Dirichlet form associated with the coarse-grained dynamics. We use these results, together with lumpability properties and well-known variational principles, to obtain sharp capacity estimates that are key quantities in the proof. For a more detailed overview of the methods, we refer the reader to the monograph by Bovier and den Hollander [18].

The remainder of the paper is organised as follows. Section B.2 provides quantities and notations that are needed throughout the paper. Section B.3 identifies the metastable regime. Section B.4 provides a sharp approximation of the Dirichlet form associated with the Glauber dynamics in the presence of the disorder. Section B.5 provides estimates on capacity and on the metastable valley measure. Section B.6 proves Theorems B.1.2–B.1.4. Appendix B.7 contains a brief overview on known results for the standard CW model, which corresponds to the setting without disorder. Appendix B.8 gives numerical evidence for the presence of multiple metastable states for suitable choices of β, h and of the law of the components of J. Appendix B.9 contains an example in which $\beta \mapsto h_c(\beta)$ is not increasing, implying the possibility of a re-entrant metastable crossover. Appendix B.10 provides the limit as $n \to \infty$ of the prefactor in (B.19).

B.2 Preparations

Section B.2.1 introduces further notation and writes the Hamiltonian in terms of the level magnetisations. Section B.2.2 introduces the Dirichlet form associated with the Glauber dynamics and rewrites this in terms of the level magnetisations. Section B.2.3 computes gradients and Hessians of the free energy as a function of the level magnetisations. Section B.2.4 closes with an approximation of the free energy that will be needed later on.

B.2.1 Hamiltonian

Recall (B.7). Abbreviate

$$\omega_{\ell,n} = \frac{|A_{\ell,n}|}{n}. \tag{B.22}$$

Since, by the law of large numbers, $(\omega_{\ell,n})_{\ell\in[k]} \to (\omega_\ell)_{\ell\in[k]} \in (0,\infty)^k$ as $n \to \infty$ with $\mathcal{P}$-probability tending to 1, we may and will assume that $A_{\ell,n} \neq \emptyset$ for all $\ell \in [k]$ and all n large enough. Recall (B.8)–(B.9). Note that $m_{\ell,n}(\sigma)$ takes values in the set

$$\Gamma_{\ell,n} = \left\{-1, -1 + \tfrac{2}{|A_{\ell,n}|}, \dots, 1 - \tfrac{2}{|A_{\ell,n}|}, 1\right\}. \tag{B.23}$$

Hence $m_n(\sigma)$ takes values in the set

$$\Gamma_n = \bigtimes_{\ell\in[k]} \Gamma_{\ell,n}. \tag{B.24}$$

The configurations corresponding to $M \subseteq \Gamma_n$ are denoted by

$$S_n[M] = \{\sigma \in S_n \colon\ m_n(\sigma) \in M\}. \tag{B.25}$$

For singletons $M = \{m\}$ we write $S_n[m]$ instead of $S_n[\{m\}]$.

Let

$$H_n(\sigma) = -\frac{1}{2n}\sum_{i,j\in[n]} J(i)J(j)\,\sigma(i)\sigma(j) - h\sum_{i\in[n]}\sigma(i), \qquad \sigma\in S_n, \tag{B.26}$$

which is the Hamiltonian in (B.2), except for the diagonal term $-\frac{1}{2n}\sum_{i\in[n]} J^2(i)$, which is a constant shift. Using the notation above, we can write the Hamiltonian in (B.26) as

$$H_n(\sigma) = -n\left[\frac{1}{2}\left(\sum_{\ell\in[k]} a_\ell\,\omega_{\ell,n}\,m_{\ell,n}(\sigma)\right)^2 + h\sum_{\ell\in[k]}\omega_{\ell,n}\,m_{\ell,n}(\sigma)\right] = nE_n(m_n(\sigma)), \tag{B.27}$$

where we abbreviate

$$E_n(m) = -\frac{1}{2}\left(\sum_{\ell\in[k]} a_\ell\,\omega_{\ell,n}\,m_\ell\right)^2 - h\sum_{\ell\in[k]}\omega_{\ell,n}\,m_\ell, \qquad m = (m_\ell)_{\ell\in[k]}\in\Gamma_n. \tag{B.28}$$

B.2.2 Dirichlet form and mesoscopic dynamics

By (B.3)–(B.4), the *Dirichlet form* associated with the Glauber dynamics equals

$$\begin{aligned}\mathcal{E}_{S_n}(h,h) &= \frac{1}{2}\sum_{\sigma,\sigma'\in S_n}\mu_n(\sigma)r_n(\sigma,\sigma')\,[h(\sigma)-h(\sigma')]^2\\ &= \frac{1}{2Z_n}\sum_{\sigma\in S_n}\sum_{\substack{\sigma'\in S_n,\\ \sigma'\sim\sigma}} \mathrm{e}^{-\beta H_n(\sigma)}\mathrm{e}^{-\beta[H_n(\sigma')-H_n(\sigma)]_+}[h(\sigma)-h(\sigma')]^2,\end{aligned} \tag{B.29}$$

where h is a test function on S_n taking values in $[0,1]$. Because of (B.27), for any h such that $h(\sigma) = \bar h(m_n(\sigma))$, with $\bar h$ a test function on Γ_n, we have

$$\mathcal{E}_{S_n}(h,h) = \frac{1}{2Z_n}\sum_{m\in\Gamma_n}\sum_{m'\in\Gamma_n}\mathrm{e}^{-\beta\,nE_n(m)}\mathrm{e}^{-\beta\,n[E_n(m')-E_n(m)]_+}\left[\,\bar h(m)-\bar h(m')\right]^2\sum_{\substack{\sigma\in S_n,\\ m_n(\sigma)=m}}\ \sum_{\substack{\sigma'\in S_n,\,\sigma'\sim\sigma,\\ m_n(\sigma')=m'}} 1, \tag{B.30}$$

where $m = (m_\ell)_{\ell\in[k]}$. If $\sigma'\sim\sigma$, then $\sigma' = \sigma^i$ for some $i\in[n]$, with σ^i obtained from σ by flipping the spin with label i. Let $\ell'\in[k]$ be such that $i\in A_{\ell',n}$. If $\sigma(i) = \pm1 = -\sigma^i(i)$, then

$$m_{\ell,n}(\sigma^i) = \begin{cases} m_{\ell',n}(\sigma) \mp \frac{2}{|A_{\ell',n}|}, & \ell = \ell',\\ m_{\ell,n}(\sigma), & \ell\neq\ell'.\end{cases} \tag{B.31}$$

For $m, m' \in \Gamma_n$, we write $m \sim m'$ when there exists an $\ell' \in [k]$ such that $m' = m^{\ell',+}$ or $m' = m^{\ell',-}$, where

$$m_\ell^{\ell',\pm} = \begin{cases} m_{\ell'} \pm \frac{2}{|A_{\ell',n}|}, & \ell = \ell', \\ m_\ell, & \ell \neq \ell'. \end{cases} \tag{B.32}$$

Moreover, for $\ell \in [k]$ and $\sigma \in S_n$ with $m_n(\sigma) = m$, the cardinality of the set $\{\sigma' \in S_n \colon \sigma' \sim \sigma,\ m_n(\sigma') = m^{\ell,\pm}\}$ equals $\frac{1 \mp m_\ell}{2}|A_{\ell,n}|$, namely, the number of (∓ 1)-spins in σ with index in $A_{\ell,n}$. Furthermore,

$$|\{\sigma \in S_n \colon m_n(\sigma) = m\}| = \prod_{\ell \in [k]} \binom{|A_{\ell,n}|}{\frac{1+m_\ell}{2}|A_{\ell,n}|}, \qquad m \in \Gamma_n, \tag{B.33}$$

as is seen by counting the number of (-1)-spins with label in $A_{\ell,n}$ of a configuration with ℓ-th level magnetisation m_ℓ. Using these observations, we can rewrite (B.30) as

$$\begin{aligned} \mathcal{E}_{S_n}(h,h) = \frac{1}{2Z_n} \sum_{m \in \Gamma_n} \mathrm{e}^{-\beta n E_n(m)} \sum_{m' \in \Gamma_n} \mathrm{e}^{-\beta n [E_n(m') - E_n(m)]_+} \left[\bar{h}(m) - \bar{h}(m')\right]^2 \\ \times \prod_{\ell \in [k]} \binom{|A_{\ell,n}|}{\frac{1+m_\ell}{2}|A_{\ell,n}|} \sum_{\ell \in [k]} |A_{\ell,n}| \left[\frac{1-m_\ell}{2} \mathbb{1}(m' = m^{\ell,+}) + \frac{1+m_\ell}{2} \mathbb{1}(m' = m^{\ell,-})\right]. \end{aligned} \tag{B.34}$$

Next, abbreviate

$$I_n(m) = -\frac{1}{n} \log \left[\prod_{\ell \in [k]} \binom{|A_{\ell,n}|}{\frac{1+m_\ell}{2}|A_{\ell,n}|}\right], \qquad m \in \Gamma_n, \tag{B.35}$$

and put

$$F_n(m) = E_n(m) + \frac{1}{\beta} I_n(m) = -\frac{1}{2} \left(\sum_{\ell \in [k]} a_\ell\, \omega_{\ell,n}\, m_\ell\right)^2 - h \sum_{\ell \in [k]} \omega_{\ell,n}\, m_\ell + \frac{1}{\beta} I_n(m), \quad m \in \Gamma_n, \tag{B.36}$$

where $E_n(m)$ is defined in (B.28). Moreover, define

$$\bar{r}_n(m,m') = \mathrm{e}^{-\beta n [E_n(m') - E_n(m)]_+} \sum_{\ell \in [k]} |A_{\ell,n}| \left[\frac{1-m_\ell}{2} \mathbb{1}(m' = m^{\ell,+}) + \frac{1+m_\ell}{2} \mathbb{1}(m' = m^{\ell,-})\right]. \tag{B.37}$$

With this notation, we can write the *mesoscopic measure* $Q_n(\cdot) = \mu_n \circ m_n^{-1}(\cdot)$ on Γ_n, with μ_n defined in (B.4), as

$$Q_n(m) = \mu_n(S_n[m]) = \frac{1}{Z_n} \mathrm{e}^{-\beta n F_n(m)}, \qquad m \in \Gamma_n, \tag{B.38}$$

and so the Dirichlet form in (B.34) becomes

$$\mathcal{E}_{S_n}(h,h) = \frac{1}{2} \sum_{m \in \Gamma_n} Q_n(m) \sum_{m' \in \Gamma_n} \bar{r}_n(m,m') \left[\bar{h}(m) - \bar{h}(m')\right]^2. \tag{B.39}$$

B.2.3 Gradients and Hessians

Denote the Cramér entropy by

$$I_{\mathbf{C}}(x) = \frac{1-x}{2}\log\left(\frac{1-x}{2}\right) + \frac{1+x}{2}\log\left(\frac{1+x}{2}\right). \tag{B.40}$$

Define

$$\bar{I}_n(m) = \sum_{\ell\in[k]} \omega_{\ell,n} I_{\mathbf{C}}(m_\ell). \tag{B.41}$$

Since $|A_{\ell,n}| = [1+o_n(1)]\,\omega_\ell n$, we can use Stirling's formula $N! = [1+o_N(1)]\,N^N \mathrm{e}^{-N}\sqrt{2\pi N}$ to obtain

$$I_n(m) = \bar{I}_n(m) + \sum_{\ell\in[k]} \frac{1}{2n}\log\left(\frac{\pi(1-m_\ell^2)|A_{\ell,n}|}{2}\right) + o(n^{-1}) = \bar{I}_n(m) + O(n^{-1}\log n), \tag{B.42}$$

where the error term is *uniform* in $m\in\Gamma_n$. For $\ell,\bar{\ell}\in[k]$, we compute

$$\frac{\partial \bar{I}_n(m)}{\partial m_\ell} = \frac{\omega_{\ell,n}}{2}\log\left(\frac{1+m_\ell}{1-m_\ell}\right) \tag{B.43}$$

and

$$\begin{aligned}\frac{\partial^2 \bar{I}_n(m)}{\partial m_\ell \partial m_{\bar{\ell}}} &= 0, \qquad \ell\neq\bar{\ell},\\ \frac{\partial^2 \bar{I}_n(m)}{\partial {m_\ell}^2} &= \frac{\omega_{\ell,n}}{1-m_\ell^2}.\end{aligned} \tag{B.44}$$

Recalling (B.28), we compute

$$\frac{\partial E_n(m)}{\partial m_\ell} = -a_\ell\,\omega_{\ell,n}\left(\sum_{\ell'\in[k]} a_{\ell'}\omega_{\ell',n}\,m_{\ell'}\right) - \omega_{\ell,n}h. \tag{B.45}$$

Define

$$\bar{F}_n(m) = E_n(m) + \frac{1}{\beta}\bar{I}_n(m) = -\frac{1}{2}\left(\sum_{\ell\in[k]} a_\ell\,\omega_{\ell,n}\,m_\ell\right)^2 - h\sum_{\ell\in[k]}\omega_{\ell,n}m_\ell + \frac{1}{\beta}\bar{I}_n(m). \tag{B.46}$$

Remark B.2.1. By (B.42), $F_n(m) = \bar{F}_n(m) + O(n^{-1}\log n)$, where F_n is defined in (B.36). ♠

For $m\in[-1,1]^k$, define

$$F_{\beta,h}(m) = -\frac{1}{2}\left(\sum_{\ell\in[k]} a_\ell\,\omega_\ell\,m_\ell\right)^2 - h\sum_{\ell\in[k]}\omega_\ell\,m_\ell + \frac{1}{\beta}\sum_{\ell\in[k]}\omega_\ell I_{\mathbf{C}}(m_\ell), \tag{B.47}$$

which corresponds to the uniform limit in probability of F_n as $n\to\infty$. Compute

$$\frac{\partial \bar{F}_n(m)}{\partial m_\ell} = \omega_{\ell,n}\left[\frac{1}{2\beta}\log\left(\frac{1+m_\ell}{1-m_\ell}\right) - a_\ell\left(\sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell',n}\,m_{\ell'}\right) - h\right] \tag{B.48}$$

and

$$\begin{aligned}\frac{\partial^2 \bar{F}_n(m)}{\partial m_\ell\,\partial m_{\ell'}} &= -a_\ell\,\omega_{\ell,n}\,a_{\ell'}\,\omega_{\ell',n}, \qquad \ell\neq\ell',\\ \frac{\partial^2 \bar{F}_n(m)}{\partial {m_\ell}^2} &= \frac{\omega_{\ell,n}}{\beta}\frac{1}{1-m_\ell^2} - a_\ell^2\,\omega_{\ell,n}^2.\end{aligned} \tag{B.49}$$

The same formulas apply for I_n, F_n, with an error term $O(n^{-1})$.

B.2.4 Additional computation

We conclude with a computation that will be useful later on. Recalling (B.32), we write

$$\begin{aligned} & n[\bar{I}_n(m^{\ell,\pm}) - \bar{I}_n(m)] \\ &= n\,\omega_{\ell,n}\left[\frac{1+m_\ell}{2}\log\left(1 \pm \frac{2}{|A_{\ell,n}|\,(1+m_\ell)}\right) + \frac{1-m_\ell}{2}\log\left(1 \mp \frac{2}{|A_{\ell,n}|\,(1-m_\ell)}\right) \pm \frac{1}{|A_{\ell,n}|}A^{\pm}_{\ell,n}\right] \\ &= n\,\omega_{\ell,n}\left[\pm\frac{1}{|A_{\ell,n}|} \mp \frac{1}{|A_{\ell,n}|} + O(n^{-2}) \pm \frac{1}{|A_{\ell,n}|}\Delta^{\pm}_{\ell,n}\right] = \Delta^{\pm}_{\ell,n} + O(n^{-1}), \end{aligned} \tag{B.50}$$

where

$$\Delta^{\pm}_{\ell,n} = \log\left(1 + \frac{2m_\ell \pm \frac{4}{|A_{\ell,n}|}}{1 - m_\ell \mp \frac{2}{|A_{\ell,n}|}}\right). \tag{B.51}$$

The same formula applies for I_n with an error term of order $O(n^{-1})$, and hence

$$n[I_n(m^{\ell,\pm}) - I_n(m)] = \Delta^{\pm}_{\ell,n} + O(n^{-1}). \tag{B.52}$$

Note that $\Delta^{\pm}_{\ell,n} = O(1)$. Therefore, using (B.36), we get

$$\begin{aligned} n\left[E_n(m^{\ell,\pm}) - E_n(m)\right] &= n\left[F_n(m^{\ell,\pm}) - F_n(m)\right] - \frac{1}{\beta}n\left[I_n(m^{\ell,\pm}) - I_n(m)\right] \\ &= n\left[F_n(m^{\ell,\pm}) - F_n(m)\right] - \frac{1}{\beta}\Delta^{\pm}_{\ell,n} + O(n^{-1}). \end{aligned} \tag{B.53}$$

B.3 Metastable regime

Section B.3.1 identifies the stationary points of $\bar{F}_n$. Section B.3.2 identifies the metastable regime. Section B.3.3 provides details on the 1-dimensional metastable landscape.

B.3.1 Stationary points of $\bar{F}_n$ and $F_{\beta,h}$

By (B.48), the critical points $m = (m_\ell)_{\ell\in[k]}$ of $\bar{F}_n$ solve the system of equations (with $\omega_{\ell,n} \neq 0$)

$$0 = \frac{\partial \bar{F}_n(m)}{\partial m_\ell} = \omega_{\ell,n}\left[\frac{1}{2\beta}\log\left(\frac{1+m_\ell}{1-m_\ell}\right) - a_\ell\left(\sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell',n}\,m_{\ell'}\right) - h\right], \qquad \ell \in [k]. \tag{B.54}$$

Hence

$$\frac{1}{2}\log\left(\frac{1+m_\ell}{1-m_\ell}\right) = \beta\left[a_\ell\left(\sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell',n}\,m_{\ell'}\right) + h\right]. \tag{B.55}$$

Since $\operatorname{arctanh} x = \frac{1}{2}\log\frac{1+x}{1-x}$, $x \in (-1,+1)$, (B.55) can be rewritten as

$$m_\ell = \tanh\left(\beta\left[a_\ell\left(\sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell',n}\,m_{\ell'}\right) + h\right]\right), \qquad \ell \in [k]. \tag{B.56}$$

Similarly, the critical points $m = (m_\ell)_{\ell\in[k]}$ of $F_{\beta,h}$ solve the deterministic equation

$$m_\ell = \tanh\left(\beta\left[a_\ell\left(\sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell'}\,m_{\ell'}\right)+h\right]\right), \qquad \ell\in[k]. \tag{B.57}$$

Note that this can also be obtained directly from (B.56) after replacing $\omega_{\ell,n}$ by its mean value ω_ℓ.

B.3.2 Metastable regime

We are interested in identifying the metastable regime, i.e., the set of pairs (β, h) for which $F_{\beta,h}$ has more than one minimum. Put

$$K = K(m) = \sum_{\ell\in[k]} a_\ell\,\omega_\ell\,m_\ell. \tag{B.58}$$

From the characterisation of the critical points of $F_{\beta,h}$ in (B.57) it follows that

$$K = T_{\beta,h}(K), \qquad T_{\beta,h}(K) = \sum_{\ell\in[k]} a_\ell\,\omega_\ell\,\tanh(\beta[a_\ell K + h]). \tag{B.59}$$

Note that any critical point $m = (m_\ell)_{\ell\in[k]} \in [-1,1]^k$ of $F_{\beta,h}$ is uniquely determined by $K(m) \in \mathbb{R}$. Consequently, the problem of solving the k-dimensional system in (B.57) can be reduced to solving the 1-dimensional equation (B.59). Recalling Hypothesis B.1.1(2), the system is in the metastable regime if and only if (B.59) has more than one solution that is not tangent to the diagonal.

Compute

$$\begin{aligned} T'_{\beta,h}(K) &= \beta \sum_{\ell\in[k]} a_\ell^2\,\omega_\ell\,(1-\tanh^2(\beta[a_\ell K + h])),\\ T''_{\beta,h}(K) &= -2\beta^2 \sum_{\ell\in[k]} a_\ell^3\,\omega_\ell\,\tanh(\beta[a_\ell K + h])\,(1-\tanh^2(\beta[a_\ell K + h])). \end{aligned} \tag{B.60}$$

For $h = 0$, the system is metastable when

$$\beta > \frac{1}{\sum_{\ell\in[k]} a_\ell^2\,\omega_\ell}, \tag{B.61}$$

in which case $T_{\beta,h}$ has a unique inflection point at $K = 0$, implying that (B.59) has three solutions $K \in \{-K^*, 0, +K^*\}$ with $K^* > 0$. Otherwise (B.59) has only one solution $K = 0$.

We proceed with the more interesting case $h > 0$.

Number of solutions

Lemma B.3.1 (Number of solutions). *For $h > 0$, the number of critical points of $F_{\beta,h}$, i.e., solutions of* (B.59)*, varies in $\{1, 3, \ldots, 2\ell+1\}$, where $\ell \in [k]$ and $2\ell - 1$ is the number of inflection points of $T_{\beta,h}$.*

Proof. For $h > 0$ and K positive and large enough, $T''_{\beta,h}(K) < 0$. Moreover, for $h > 0$ and K negative with $|K|$ large enough, $T''_{\beta,h}(K) > 0$. Therefore, $T_{\beta,h}$ has at least one inflection point and that the number of inflection points of $T_{\beta,h}$ cannot be even: it takes values in $\{1, 3, \ldots, 2k-1\}$ depending on β, h and the law of the components of J. Consequently, if $2\ell - 1$ ($\ell \in [k]$) is the number of inflection points, then the cardinality of the solutions of (B.59) takes values in $\{1, 3, \ldots, 2\ell + 1\}$ depending on β, h and on the law of the components of J. □ □

We conjecture that for any finite k there exist β, h and a law of the components of J such that (B.59) has any number of solutions in the set $\{1, 3, \ldots, 2k + 1\}$. We found numerical evidence for this fact for $k \in \{2, 3, 4\}$. See Appendix B.8.

Lemma B.3.2 (Unique strictly positive solution). *For every $\beta > 0$ and $h > 0$,* (B.59) *has exactly one strictly positive solution.*

Proof. Put $W(K) = T_{\beta,h}(K) - K$. The solutions of (B.59) are the roots of W. Clearly, $W(0) > 0$. Moreover, $\lim_{K\to\infty} W(K) = -\infty$ because $\lim_{K\to\infty} T_{\beta,h}(K) = \sum_{\ell\in[k]} a_\ell\,\omega_\ell > 0$ is finite. Therefore, by continuity, a root of $W(K)$ exists in $(0, \infty)$.

Let $\tilde{K}$ be the smallest positive root of W. Next we will prove that this root is unique. Indeed, $W(K)'' < 0$ when $K \in [0, \infty)$, meaning that $K \mapsto W(K)'$ is strictly decreasing. By continuity, since $W(K) > 0$ for all $K \in [0, \tilde{K})$, we have $W(\tilde{K})' \le 0$ and $\lim_{K\to\infty} W(K)' = -1$. Therefore, $W(K)' < 0$ for all $K \in (\tilde{K}, \infty)$, and so W is strictly decreasing. Moreover, $W(K) < W(\tilde{K}) = 0$ for all $K \in (\tilde{K}, \infty)$. Thus, $\tilde{K}$ is the only positive root of W. □ □

Metastable regime

Lemma B.3.3 (Characterisation of the metastable regime).
(B.59) *has at least three solutions not tangent to the diagonal if and only if there exists $\bar{K} < 0$ such that $\bar{K} > T_{\beta,h}(\bar{K})$, i.e.,*

$$\bar{K} > \sum_{\ell\in[k]} a_\ell\,\omega_\ell \tanh(\beta[a_\ell \bar{K} + h]). \tag{B.62}$$

Proof. Using Lemma B.3.2, we see that (B.59) has at least three solutions if and only if it has at least two strictly negative solutions. As above, we define $W(K) = T_{\beta,h}(K) - K$. The solutions of (B.59) are the roots of W. Now, *assume* that there exists a $\bar{K} < 0$ such that $\bar{K} > T_{\beta,h}(\bar{K})$. Since $W(\bar{K}) < 0$ and $W(0) > 0$, $W(K)$ has a root in $(\bar{K}, 0)$, implying that (B.59) has at least one solution in $(\bar{K}, 0)$. Moreover, since $\lim_{K\to-\infty} T_{\beta,h}(K) = -\sum_{\ell\in[k]} a_\ell\,\omega_\ell$ is finite, we have $\lim_{K\to-\infty} W(K) = \infty$. Because $W(\bar{K}) < 0$, it follows that W has at least one root in $(-\infty, \bar{K})$. With the same argument it can be shown that the negative roots of W are always even. The opposite implication is trivial. □ □

Remark B.3.4. Applying the intermediate value theorem to the derivative of $W(K) = T_{\beta,h}(K) - K$, we get that if the condition in Lemma B.3.3 is satisfied, then there exists a $\bar{K} < 0$ such that $T'_{\beta,h}(\bar{K}) = 1$ and $\bar{K} > T_{\beta,h}(\bar{K})$. ♠

Theorem B.3.5 (Metastable regime). *Define, as in* (B.14)*,*

$$\beta_c = \frac{1}{\sum_{\ell\in[k]} a_\ell^2\,\omega_\ell}. \tag{B.63}$$

The metastable regime is

$$\beta \in (\beta_c, \infty), \qquad h \in [0, h_c(\beta)), \tag{B.64}$$

with $\beta \mapsto \beta h_c(\beta)$ non-decreasing on $[\beta_c, \infty)$. Furthermore, if the support of the law of the components of J is put into increasing order, i.e., $a_1 < a_2 < \cdots < a_k$, then

$$\lim_{\beta\to\infty} h_c(\beta) = \min_{\ell\in[k]^*} \left(\sum_{\ell'=\ell}^{k} a_\ell \, a_{\ell'} \, \omega_{\ell'} - \sum_{\ell'=1}^{\ell-1} a_\ell \, a_{\ell'} \, \omega_{\ell'} \right), \tag{B.65}$$

where the minimum is over all $\ell \in [k]$ such that the quantity between brackets is positive.

Proof. Recalling Lemma B.3.3, we look for conditions for the existence of a $K < 0$ satisfying (B.62). If such a K exists, then by Remark B.3.4 there exists a $\bar{K} < 0$ satisfying (B.62) such that $T'_{\beta,h}(\bar{K}) = 1$, which reads

$$\sum_{\ell\in[k]} a_\ell^2 \, \omega_\ell \, \tanh^2(\beta[a_\ell \bar{K} + h]) = \sum_{\ell\in[k]} a_\ell^2 \, \omega_\ell - \frac{1}{\beta}. \tag{B.66}$$

Since the left-hand side of (B.66) is positive, it admits solutions only if

$$\frac{1}{\beta} < \sum_{\ell\in[k]} a_\ell^2 \, \omega_\ell = \frac{1}{\beta_c}. \tag{B.67}$$

Therefore, (B.67) is a necessary condition for the metastable regime.

Now assume (B.67). Since $\tanh x \sim x$, $x \to 0$, for $|K| \ll \beta(\max_{\ell\in[k]} a_\ell)^{-1}$ and $h \downarrow 0$, we have

$$K = T_{\beta,h}(K) = \sum_{\ell\in[k]} a_\ell \, \omega_\ell \, \tanh(\beta[a_\ell K + h]) \sim \sum_{\ell\in[k]} a_\ell \, \omega_\ell \, \beta[a_\ell K + h], \tag{B.68}$$

which reads

$$K \sim - \left(\sum_{\ell\in[k]} a_\ell \, \omega_\ell \right) \left(\frac{1}{\beta_c} - \frac{1}{\beta} \right)^{-1} h \tag{B.69}$$

and proves the existence of a negative solution. A positive solution is guaranteed by Lemma B.3.2. The existence of a third (strictly negative) solution of (B.57), for every $\beta > \beta_c$ and for $h \downarrow 0$, follows as in the proof of Lemma B.3.3. Therefore, the lower bound on β_c is sharp.

Since $h \mapsto T_{\beta,h}(K)$ is strictly increasing for every fixed $\beta > 0$ and $K \in \mathbb{R}$, there exists a unique critical curve $\beta \mapsto h_c(\beta)$ such that the system is metastable for $0 \le h < h_c(\beta)$ and not metastable for $h \ge h_c(\beta)$. We know that $h_c(\beta) > 0$ for $\beta > \beta_c$. By passing to the parametrisation $g = h\beta$, we get that $\beta \mapsto T_{\beta,g}(K)$ is strictly decreasing for every g and for every $K < 0$, from which it follows that $\beta \mapsto g_c(\beta) = \beta h_c(\beta)$ is non-decreasing.

We next focus on the limit of $h_c(\beta)$ as $\beta \to \infty$. By Lemma B.3.3, we may focus on the existence of $\bar{K}$ satisfying (B.62). In the limit as $\beta \to \infty$, $\tanh(\beta[a_\ell \bar{K} + h]) \to 2\Theta_{-h/a_\ell}(\bar{K}) - 1$, where $\Theta_x(\cdot)$ is the Heaviside function centred in x. Thus, for all $\ell \in [k+1]$,

$$\lim_{\beta\to\infty} \sum_{\ell'\in[k]} a_{\ell'} \, \omega_{\ell'} \, \tanh(\beta[a_{\ell'} K + h]) = - \sum_{\ell'=\ell}^{k} a_{\ell'} \, \omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'} \, \omega_{\ell'}, \qquad K \in \left(-\frac{h}{a_{\ell-1}}, -\frac{h}{a_\ell} \right), \tag{B.70}$$

and, for all $\ell \in [k]$,

$$\lim_{\beta\to\infty} \sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell'}\ \tanh(\beta[a_{\ell'}K+h]) = -\sum_{\ell'=\ell+1}^{k} a_{\ell'}\,\omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'}\,\omega_{\ell'}, \qquad K = -\frac{h}{a_\ell}, \tag{B.71}$$

where we set $-\frac{h}{a_0} = -\infty$ and $-\frac{h}{a_{k+1}} = \infty$. Thus, for $\bar{K} \in \left(-\frac{h}{a_{\ell-1}}, -\frac{h}{a_\ell}\right)$, (B.62) can be written as

$$\bar{K} > -\sum_{\ell'=\ell}^{k} a_{\ell'}\,\omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'}\,\omega_{\ell'}. \tag{B.72}$$

Therefore, (B.62) has a solution if and only if there exists an $\ell \in [k]$ such that

$$-\sum_{\ell'=\ell}^{k} a_{\ell'}\,\omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'}\,\omega_{\ell'} < -\frac{h}{a_\ell}, \tag{B.73}$$

in which case a solution $\bar{K}$ of (B.62) exists in

$$\left(-\sum_{\ell'=\ell}^{k} a_{\ell'}\,\omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'}\,\omega_{\ell'}, -\frac{h}{a_\ell}\right). \tag{B.74}$$

Note that the quantity between brackets in (B.65) is always positive for $\ell = 1$. Thus, the minimum is always finite.

The proof is complete after we show why we may drop the case where $\bar{K} = -\frac{h}{a_\ell}$ for some $\ell \in [k]$. In this case the condition for $\bar{K}$ to satisfy (B.62) is

$$-\sum_{\ell'=\ell+1}^{k} a_{\ell'}\,\omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'}\,\omega_{\ell'} < -\frac{h}{a_\ell}, \tag{B.75}$$

which implies (B.73). Thus, if $\bar{K} = \frac{-h}{a_\ell}$ satisfies (B.62), then also some other K in (B.74) satisfies (B.62). Therefore, the condition in (B.73) is equivalent to having metastability. □ □

Lemma B.3.6 (Re-entrant crossover). *The function $\beta \mapsto h_c(\beta)$ is not necessarily non-decreasing.*

Proof. In Appendix B.9 we provide an example of $\beta \mapsto h_c(\beta)$ that is not increasing. □ □

Bounds on the inflection points and on the critical curve

Lemma B.3.7 (Bounds on inflection points). *All solutions of $T''_{\beta,h}(K) = 0$ are contained in the interval*

$$\left[-\frac{h}{\min_{\ell\in[k]} a_\ell}, -\frac{h}{\max_{\ell\in[k]} a_\ell}\right]. \tag{B.76}$$

In particular, they are all strictly negative.

Proof. If $K > -\frac{h}{\max_{\ell\in[k]} a_\ell}$, then $\tanh(\beta[a_\ell K + h]) > 0$ for all $\ell \in [k]$, which implies $T''_{\beta,h}(K) < 0$. If $K < -\frac{h}{\min_{\ell\in[k]} a_\ell}$, then $\tanh(\beta[a_\ell K + h]) < 0$ for all $\ell \in [k]$, which implies $T''_{\beta,h}(K) > 0$. □ □

Lemma B.3.8 (Upper bound on h_c). $\sup_{\beta\in(\beta_c,\infty)} h_c(\beta) < \left(\max_{\ell\in[k]} a_\ell\right) \sum_{\ell\in[k]} a_\ell\,\omega_\ell$.

Proof. Use Lemma B.3.3 to characterise the metastable regime and Remark B.3.4. We *claim* that if a solution $\bar{K}$ of (B.62) with $T'_{\beta,h}(\bar{K}) = 1$ exists, then it must be negative and strictly less than an inflection point. Using this fact, together with Lemma B.3.7 and the inequality in (B.62), we obtain a necessary upper bound on h:

$$\sum_{\ell\in[k]} a_\ell\,\omega_\ell\,\tanh(\beta[a_\ell\bar{K}+h]) < -\frac{h}{\max_{\ell\in[k]} a_\ell}. \tag{B.77}$$

Using that $\tanh(\beta[a_\ell\bar{K}+h]) > -1$, we conclude the proof.

We are left to prove the claim. By Lemma B.3.7, all inflection points are negative, and $T''_{\beta,h}(K) < 0$ for $K \geq 0$. Assume, by contradiction, that $T''_{\beta,h}(K) < 0$ for all $K \in (\bar{K}, \infty)$. Then $T'_{\beta,h}$ is strictly decreasing. Therefore, $T'_{\beta,h}(K) < 1$ for all $K \in (\bar{K}, \infty)$, which implies $T_{\beta,h}(K) - T_{\beta,h}(0) < K$. Since $T_{\beta,h}(0) > 0$, there exists a $\tilde{K} \in (\bar{K}, 0)$ such that $T_{\beta,h}(\tilde{K}) > 0 > \tilde{K}$. Thus, $T_{\beta,h}(\tilde{K}) - T_{\beta,h}(0) > \tilde{K}$, which contradicts what we have proved for all $K \in (\bar{K}, \infty)$. □ □

B.3.3 Quasi 1-dimensional landscape

Given $K \in \mathbb{R}$, by standard saddle point approximation, the leading order of

$$-\frac{1}{\beta n}\log \mu_n(\{\sigma\colon\ K_n(m_n(\sigma)) = K\}) \tag{B.78}$$

turns out to be the function $G_n\colon \mathbb{R} \to \mathbb{R}$ defined by

$$G_n(K) = \inf_{m\colon\ K_n(m)=K} \bar{F}_n(m). \tag{B.79}$$

Recalling definitions (B.46) and (B.58), using Lagrange multipliers and integrating the condition $K_n(m) = K$, we obtain

$$G_n(K) = -\frac{1}{2}K^2 - \frac{\log 2}{\beta} - \inf_{t\in\mathbb{R}}\left(Kt + \sum_{\ell\in[k]} \frac{\omega_{\ell,n}}{\beta}\log\cosh\left[\beta(h - ta_\ell)\right]\right). \tag{B.80}$$

Lemma B.3.9 (**Alternative characterisation for the critical points).**

1. *If m^* is a (not maximal) critical point for F_n, then $K_n(m^*)$ is a critical point for G_n.*
2. *If K is a critical point for G_n, then $m^* = (m^*_\ell)_{\ell\in[k]}$ with $m^*_\ell = \tanh\left(\beta\left[a_\ell K + h\right]\right)$ (recall* (B.56)*) is a critical point for F_n.*
3. *$F_n(m^*) = G_n(K_n(m^*))$ for any (not maximal) critical point m^*.*

Proof. Similar to [12, Lemma 7.4]. □ □

We have already seen that $K_n(m)$ fully determines any critical value m of F_n, and is useful to order them. Lemma B.3.9 exhibits the one-dimensional structure underlying the metastable landscape and provides a tool to describe the nature of the critical points of F_n.

Remark B.3.10. The above results extend to the limit $n \to \infty$: replace F_n by $F_{\beta,h}$ and G_n by $G_{\beta,\ell}$, obtained after replacing $\omega_{\ell,n}$ by ω_ℓ in (B.80), and $K_n(\cdot)$ by $K(\cdot)$. ♠

B.4 Approximation of the Dirichlet form near the saddle point

In this section we approximate the Dirichlet form associated with the coarse-grained dynamics near the saddle point. This is a key step to obtain capacity estimates in the following section. Further details and examples on the techniques we use here can be found in [18, Chapters 9, 10 and 14].

Section B.4.1 introduces some key quantities that are needed to express the mesoscopic measure. Section B.4.2 introduces an approximate mesoscopic measure that leads to an approximate dynamics. Section B.4.3 approximates the harmonic functions associated with this dynamics. Section B.4.4 computes an approximate Dirichlet form. Section B.4.5 uses the latter to approximate the full Dirichlet form.

B.4.1 Key quantities

Let $\mathbf{m}_n = (\mathbf{m}_{\ell,n})_{\ell\in[k]}$ and $\mathbf{t}_n = (\mathbf{t}_{\ell,n})_{\ell\in[k]}$ in Γ_n be a local minimum of F_n and the correspondent saddle point, respectively, as defined in Section B.1.3. Note that both $\mathbf{m}_n$ and $\mathbf{t}_n$ satisfy (B.56). Consider the neighbourhood of $\mathbf{t}_n$ defined by

$$\mathcal{D}_n = \left\{ m \in \Gamma_n \colon\ d(m, \mathbf{t}_n) \leq C' n^{-1/2} \log^{1/2} n \right\}, \tag{B.81}$$

where d is the Euclidean distance and $C' \in (0, \infty)$ is a constant. Abbreviate the Hessian of F_n

$$\mathbb{A}_n(m) = (\nabla^2 F_n)(m), \qquad m \in \Gamma_n, \tag{B.82}$$

and put

$$\mathbb{A}_n = \mathbb{A}_n(\mathbf{t}_n). \tag{B.83}$$

By (B.49),

$$\begin{aligned}
(\mathbb{A}_n(m))_{\ell,\ell'} &= -a_\ell\, \omega_{\ell,n}\, a_{\ell'}\, \omega_{\ell',n} + O(n^{-1}), \qquad \ell \neq \ell', \\
(\mathbb{A}_n(m))_{\ell,\ell} &= \frac{\omega_{\ell,n}}{\beta} \frac{1}{1-m_\ell^2} - a_\ell^2\, \omega_{\ell,n}^2 + O(n^{-1}) = \frac{1}{\beta} \frac{\partial^2 \bar{I}_n(m)}{\partial {m_\ell}^2} - a_\ell^2\, \omega_{\ell,n}^2 + O(n^{-1}).
\end{aligned} \tag{B.84}$$

Note that $\mathbb{A}_n(m)$ is a diagonal matrix minus a rank one matrix. Compute

$$\det \mathbb{A}_n(m) = \left(1 - \sum_{\ell\in[k]} \beta\, a_\ell^2\, \omega_{\ell,n} [1 - m_\ell^2] \right) \prod_{\ell'\in[k]} \frac{1}{\beta} \frac{\omega_{\ell',n}}{1 - m_{\ell'}^2} [1 + O(n^{-1})]. \tag{B.85}$$

B.4.2 Approximate dynamics and Dirichlet form

For any two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^k$, let $\langle \mathbf{v}, \mathbf{w} \rangle$ denote their scalar product. For any $k \times k$ matrix $\mathbf{M}$ and any $\mathbf{v} \in \mathbb{R}^k$, let $\mathbf{M} \cdot \mathbf{v}$ denote their matrix product, as $\mathbf{v}$ was in $\mathbb{R}^k \times 1$.

For $m \in \mathcal{D}_n$, define

$$\tilde{Q}_n(m) = \frac{1}{Z_n} \exp\left[-\tfrac{\beta n}{2} \langle [m - \mathbf{t}_n], \mathbb{A}_n \cdot [m - \mathbf{t}_n] \rangle \right] \exp\left[-\beta n F_n(\mathbf{t}_n) \right], \tag{B.86}$$

and

$$\tilde{r}_n(m, m') = \begin{cases} \bar{r}_n(\mathbf{t}_n, \mathbf{t}_n^{\ell,+}), & m' = m^{\ell,+}, \\ \bar{r}_n(\mathbf{t}_n^{\ell,-}, \mathbf{t}_n) \frac{\tilde{Q}_n(m^{\ell,-})}{\tilde{Q}_n(m)}, & m' = m^{\ell,-}, \\ 0, & \text{else}, \end{cases} \tag{B.87}$$

where $\bar{r}_n$ is defined in (B.37). The transition rates $\tilde{r}_n$ define a random dynamics on $\mathcal{D}_n$ that is reversible with respect to the mesoscopic measure $\tilde{Q}_n$. The corresponding *Dirichlet form* is

$$\tilde{\mathcal{E}}_{\mathcal{D}_n}(u,u) = \sum_{m\in\mathcal{D}_n} \tilde{Q}_n(m) \sum_{\ell\in[k]} \tilde{r}_n(m, m^{\ell,+}) \left[u(m) - u(m^{\ell,+})\right]^2, \tag{B.88}$$

where u is a test function on $\mathcal{D}_n$. Put

$$r_\ell = \tilde{r}_n(m, m^{\ell,+}) = \bar{r}_n(\mathbf{t}_n, \mathbf{t}_n^{\ell,+}). \tag{B.89}$$

Using (B.28) and (B.37), we get

$$r_\ell = |A_{\ell,n}| \, \frac{1-\mathbf{t}_{\ell,n}}{2} \, \exp\left[-2\beta \left(-h - a_\ell \left(\frac{a_\ell}{n} + \sum_{\ell'\in[k]} a_{\ell'}\omega_{\ell',n}\mathbf{t}_{\ell',n}\right)\right)_+\right]. \tag{B.90}$$

Approximation estimates

Next we estimate how close the pairs $(\bar{r}_n, \tilde{r}_n)$ and $(Q_n, \tilde{Q}_n)$ are. By Taylor expansion around $\mathbf{t}_n$, we have

$$F_n(m) - F_n(\mathbf{t}_n) = \tfrac{1}{2}\Big\langle [m-\mathbf{t}_n], \mathbb{A}_n \cdot [m-\mathbf{t}_n]\Big\rangle + O(d(m,\mathbf{t}_n)^3). \tag{B.91}$$

In particular,

$$\begin{aligned} F_n(\mathbf{t}_n^{\ell,\pm}) - F_n(\mathbf{t}_n) &= \frac{1}{2}\frac{4}{|A_{\ell,n}|^2}(\mathbb{A}_n)_{\ell,\ell} + O(\,|A_{\ell,n}|^{-3}\,) \\ &= \frac{2}{n^2\omega_{\ell,n}^2}\left[\frac{\omega_{\ell,n}}{\beta}\frac{1}{1-\mathbf{t}_{\ell,n}^2} - a_\ell^2\,\omega_{\ell,n}^2 + o\left((n\,\omega_{\ell,n})^{-1}\right)\right] + O((n\,\omega_{\ell,n})^{-3}) \\ &= \frac{2}{n^2}\left(\frac{1}{\beta\,\omega_{\ell,n}(1-\mathbf{t}_{\ell,n}^2)} - a_\ell^2\right) + O((n\,\omega_{\ell,n})^{-3}), \end{aligned} \tag{B.92}$$

where the second equality uses (B.84). Moreover, for $m \in \mathcal{D}_n$ (e_ℓ is the unitary vector in $\mathbb{R}^k$ whose ℓ-th component is non-zero),

$$\begin{aligned} &F_n(m^{\ell,\pm}) - F_n(m) \\ &= \left\langle \left[\pm\tfrac{2}{|A_{\ell,n}|}\,\mathrm{e}_\ell\right], \mathbb{A}_n\cdot[m-\mathbf{t}_n]\right\rangle + \frac{1}{2}\left\langle \left[\pm\tfrac{2}{|A_{\ell,n}|}\,\mathrm{e}_\ell\right], \mathbb{A}_n\cdot\left[\pm\tfrac{2}{|A_{\ell,n}|}\,\mathrm{e}_\ell\right]\right\rangle + O(d(m,\mathbf{t}_n)^3) \\ &= \pm\frac{2}{|A_{\ell,n}|}\sum_{\ell'\in[k]}(\mathbb{A}_n)_{\ell,\ell'}(m_{\ell'} - \mathbf{t}_{\ell',n}) + \frac{2}{|A_{\ell,n}|^2}(\mathbb{A}_n)_{\ell,\ell} + O(d(m,\mathbf{t}_n)^3) \\ &= \left(\pm\frac{2}{n\omega_{\ell,n}}(m_\ell - \mathbf{t}_{\ell,n}) + \frac{2}{n^2\omega_{\ell,n}^2}\right)\left(\frac{\omega_{\ell,n}}{\beta}\frac{1}{1-\mathbf{t}_{\ell,n}^2} - a_\ell^2\,\omega_{\ell,n}^2 + o(n^{-1})\right) \\ &\quad \pm \frac{2}{n\omega_{\ell,n}}\sum_{\ell'\in[k],\,\ell'\neq\ell}(-a_\ell\,\omega_{\ell,n}\,a_{\ell'}\,\omega_{\ell',n})(m_{\ell'} - \mathbf{t}_{\ell',n}) + O(n^{-3/2}\log^{3/2} n) \\ &= \mp\frac{2}{n}\sum_{\ell'\in[k]} a_\ell\,a_{\ell'}\,\omega_{\ell',n}(m_{\ell'} - \mathbf{t}_{\ell',n}) \pm \frac{2(m_\ell - \mathbf{t}_{\ell,n})}{\beta\,n(1-\mathbf{t}_{\ell,n}^2)} + O(n^{-3/2}\log^{3/2} n), \end{aligned} \tag{B.93}$$

where the third equality uses (B.84). For $m \in \mathcal{D}_n$, we have $d(m, \mathbf{t}_n)^3 = O(n^{-3/2} \log^{3/2} n)$. Therefore, combining (B.38), (B.86) and (B.91), we have

$$\left| \frac{Q_n(m)}{\tilde{Q}_n(m)} - 1 \right| \leq C'' n^{-1/2} \log^{3/2} n, \qquad m \in \mathcal{D}_n, \tag{B.94}$$

for some $C'' \in (0, \infty)$ constant. Using (B.37) and (B.53), we can write

$$\bar{r}_n(m, m^{\ell,\pm}) = \exp\left[-\beta \left[n \left[F_n(m^{\ell,\pm}) - F_n(m) \right] - \frac{1}{\beta} \Delta^{\pm}_{\ell,n} + O\left(n^{-1}\right) \right]_+ \right] \frac{1 \mp m_\ell}{2}, \tag{B.95}$$

where $\Delta^{\pm}_{\ell,n}$ is defined in (B.51).

Using (B.87), (B.92), (B.93) and (B.95), we find that, for all $m \in \mathcal{D}_n$,

$$\begin{aligned}
\left| \frac{\bar{r}_n\left(m, m^{\ell,+}\right)}{\tilde{r}_n\left(m, m^{\ell,+}\right)} - 1 \right| &= \left| \frac{\bar{r}_n\left(m, m^{\ell,+}\right)}{\bar{r}_n(\mathbf{t}_n, \mathbf{t}_n^{\ell,+})} - 1 \right| \\
&= \left| \frac{(1 - m_\ell) \exp\left\{ - \left[I_1 + O(n^{-1/2} \log^{3/2} n) - \Delta^{\pm}_{\ell,n} + o_n(1) \right]_+ \right\}}{(1 - \mathbf{t}_{\ell,n}) \exp\left\{ - \left[I_2 + O(n^{-2}\, \omega^{-3}_{\ell,n}) - \Delta^{\pm}_{\ell,n} + o_n(1) \right]_+ \right\}} - 1 \right| \\
&= \left| \frac{(1 - m_\ell) \exp\left\{ - \left[I_1 - \Delta^{\pm}_{\ell,n} + o_n(1) \right]_+ \right\}}{(1 - \mathbf{t}_{\ell,n}) \exp\left\{ - \left[- \Delta^{\pm}_{\ell,n} + o_n(1) \right]_+ \right\}} - 1 \right| \\
&\leq C''' n^{-1/2} \log^{1/2} n,
\end{aligned} \tag{B.96}$$

where $C''' \in (0, \infty)$ is a constant and we abbreviate

$$\begin{aligned}
I_1 &= -2\beta \sum_{\ell' \in [k]} a_\ell\, a_{\ell'}\, \omega_{\ell',n} (m_{\ell'} - \mathbf{t}_{\ell',n}) + \frac{2(m_\ell - \mathbf{t}_{\ell,n})}{1 - \mathbf{t}^2_{\ell,n}}, \\
I_2 &= \frac{2}{n} \left(\frac{1}{\omega_{\ell,n}(1 - \mathbf{t}^2_{\ell,n})} - \beta\, a^2_\ell \right).
\end{aligned} \tag{B.97}$$

Equations (B.94) and (B.96) are relevant for the following approximation.

B.4.3 Approximate harmonic function

Let $\mathbb{B}_n$ be the $k \times k$ matrix defined by

$$(\mathbb{B}_n)_{\ell\ell'} = \frac{\sqrt{r_\ell r_{\ell'}}}{n\, \omega_{\ell,n} \omega_{\ell',n}} (\mathbb{A}_n)_{\ell\ell'}, \tag{B.98}$$

where $\mathbb{A}_n$ is defined in (B.83). Note that

$$\det \mathbb{B}_n = (\det \mathbb{A}_n) \prod_{\ell \in [k]} \frac{r_\ell}{n\, \omega^2_{\ell,n}}. \tag{B.99}$$

Let $\gamma_n^{(\ell)}$, $\ell \in [k]$, be the eigenvalues of $\mathbb{B}_n$, ordered in increasing order. Let $\gamma_n = \gamma_n^{(1)}$ denote the unique negative eigenvalue of $\mathbb{B}_n$, and $\hat{v}$ the corresponding unitary eigenvector. Define $v = (v_\ell)_{\ell \in [k]}$ by $v_\ell = \hat{v}_\ell \frac{\omega_{\ell,n} \sqrt{n}}{\sqrt{r_\ell}}$.

Remark B.4.1. As in [18, Remark 10.4], it follows by Hypothesis B.1.1 that $\mathbb{A}_n$ has all strictly positive eigenvalues but one strictly negative. It can be seen that the same property holds for the eigenvalues of $\mathbb{B}_n$. ♠

Lemma B.4.2 (Eigenvalue). *The eigenvalue γ_n is the unique solution of the equation*

$$\frac{1}{n}\sum_{\ell\in[k]}\frac{a_\ell^2}{\frac{1}{n\beta\,\omega_{\ell,n}(1-\mathbf{t}_{\ell,n}^2)}-\frac{\gamma_n}{r_\ell}} = 1+O(n^{-1}). \tag{B.100}$$

Proof. We follow the line of proof of [18, Lemma 14.9], using the last point in Hypothesis B.1.1. In our case, [18, Eq. (14.7.12)] reads

$$-\frac{1}{n}a_\ell\sqrt{r_\ell}\sum_{\ell'\in[k]}a_{\ell'}\sqrt{r_{\ell'}}u_{\ell'}+\left(r_\ell\frac{1}{n\beta\omega_{\ell,n}(1-\mathbf{t}_{\ell,n})^2}-\gamma_n\right)u_\ell+O(n^{-1})=0, \qquad \ell\in[k]. \tag{B.101}$$

□ □

Remark B.4.3. As in [18, Lemma 14.9], since the left-hand side of (B.100) is increasing in γ_n for $\gamma_n\geq 0$, a negative solution of (B.100) exists if and only if

$$\beta\sum_{\ell\in[k]}a_\ell^2\,\omega_{\ell,n}(1-\mathbf{t}_{\ell,n}^2)>1. \tag{B.102}$$

Using (B.85), (B.102) holds if and only if $\det\mathbb{A}_n<0$. By Remark B.4.1 the latter holds true. ♠

Define $f\colon\,\mathbb{R}\to[0,1]$ as

$$f(x)=\sqrt{\frac{(-\gamma_n)\beta n}{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{1}{2}(-\gamma_n)\beta n u^2}\,\mathrm{d}u \tag{B.103}$$

and $g\colon\,\mathbb{R}^k\to[0,1]$ as

$$g(m)=f(\langle v,m-\mathbf{t}_n\rangle). \tag{B.104}$$

Recall the definition of $\mathcal{M}_n(\mathbf{m}_n)$ given in (B.17).

Let W_0 be a strip in Γ_n of width $Cn^{-1/2}\log^{1/2}n$ such that $\mathbf{t}_n\in W_0$, $\mathcal{M}_n(\mathbf{m}_n)\cap W_0$ is empty and W_0^c consists in two non-neighbouring parts: W_1 containing $\mathbf{m}_n$ and W_2 containing $\mathcal{M}_n(\mathbf{m}_n)$. Moreover, we require that, for some fixed constant $c>1$, $W_0\cap\mathcal{D}_n^c\subseteq\{m\in\Gamma_n\colon F_n(m)>F_n(\mathbf{t}_n)+cn^{-1}\log n\}$. Define

$$\tilde g(m)=\begin{cases}0, & m\in W_1,\\ 1, & m\in W_2,\\ g(x), & m\in W_0\cap\mathcal{D}_n,\\ 0, & m\in W_0\cap\mathcal{D}_n^c.\end{cases} \tag{B.105}$$

By choosing W_0 and $\mathcal{D}_n$ suitably we have, for $m\sim m'$ (i.e., $\bar r_n(m,m')>0$) and $c\in(0,\infty)$ large enough (coming from the definition of W_0),

$$Q_n(m)\leq Q_n(\mathbf{t}_n)n^{-c\beta}, \qquad m\in W_0\cap\mathcal{D}_n^c, \tag{B.106}$$

$$(\tilde g(m)-\tilde g(m'))^2\,\bar r_n(m,m')Q_n(m)\leq Q_n(\mathbf{t}_n)n^{-c\beta}, \qquad m\in W_0\cap\mathcal{D}_n, m'\in W_0^c. \tag{B.107}$$

B.4.4 Computation of the approximate Dirichlet form

In this section we follow [18, Sections 10.2.2–10.2.3] to approximate $\tilde{\mathcal{E}}_{\mathcal{D}_n}(g,g)$ defined in (B.88). As in [18, Eq. (10.2.24)], for $m \in \mathcal{D}_n$ and $\ell \in [k]$ such that $m^{\ell,+} \in D_n$, compute

$$\begin{aligned} g(m^{\ell,+}) - g(m) &= \frac{2}{|A_{\ell,n}|} v_\ell f'(\langle v, m - \mathbf{t}_n\rangle) + \frac{2}{|A_{\ell,n}|^2} v_\ell^2 f''(\langle v, m - \mathbf{t}_n\rangle) + \frac{4}{3\,|A_{\ell,n}|^3} v_\ell^3 f'''(\langle v, \tilde{m} - \mathbf{t}_n\rangle) \\ &= v_\ell \sqrt{\frac{2(-\gamma_n)\beta}{\pi n\, \omega_{\ell,n}^2}} \exp\left(-\frac{\beta n}{2}(-\gamma_n)\,\langle v, m - \mathbf{t}_n\rangle^2\right) \\ &\quad \times \left(1 - \frac{1}{\omega_{\ell,n}} v_\ell(-\gamma_n)\beta\,\langle v, m-\mathbf{t}_n\rangle + O(\omega_{\ell,n}^{-2}\, n^{-1}\log n)\right). \end{aligned} \tag{B.108}$$

Recalling (B.88)–(B.89), we have

$$\begin{aligned} \tilde{\mathcal{E}}_{\mathcal{D}_n}(g,g) &= \sum_{m\in\mathcal{D}_n} \tilde{Q}_n(m) \sum_{\ell\in[k]} r_\ell \left[g(m^{\ell,+}) - g(m)\right]^2 \\ &= \frac{1}{Z_n} \sum_{m\in\mathcal{D}_n} \exp\left[-\tfrac{\beta n}{2}\langle [m-\mathbf{t}_n], \mathbb{A}_n \cdot [m - \mathbf{t}_n]\rangle\right] \mathrm{e}^{-\beta n F_n(\mathbf{t}_n)} \\ &\quad \times \sum_{\ell\in[k]} r_\ell v_\ell^2 \frac{2(-\gamma_n)\beta}{\pi n\,\omega_{\ell,n}^2} \exp\left(-\beta n(-\gamma_n)\,\langle v, m-\mathbf{t}_n\rangle^2\right) \\ &\quad \times \left(1 - \frac{v_\ell}{\omega_{\ell,n}}(-\gamma_n)\beta\,\langle v, m - \mathbf{t}_n\rangle + O(\omega_{\ell,n}^{-2}\, n^{-1}\log n)\right)^2 \\ &= \frac{1}{Z_n}\frac{2(-\gamma_n)\beta}{\pi} \sum_{m\in\mathcal{D}_n} \exp\left[-\tfrac{\beta n}{2}\langle [m-\mathbf{t}_n], \mathbb{A}_n\cdot[m-\mathbf{t}_n]\rangle\right] \mathrm{e}^{-\beta n F_n(\mathbf{t}_n)} \\ &\quad \times \exp\left(-\beta n(-\gamma_n)\,\langle v, m-\mathbf{t}_n\rangle^2\right)\left[1 + O\left(\omega_{\ell,n}^{-1}\, n^{-1/2}\log^{1/2} n\right)\right] \\ &= \frac{1}{Z_n}\frac{2(-\gamma_n)\beta}{\pi}\left[1 + O\left(\omega_{\ell,n}^{-1}\; n^{-1/2}\log^{1/2} n\right)\right] \mathrm{e}^{-\beta n F_n(\mathbf{t}_n)} \left(\prod_{\ell\in[k]} \frac{|A_{\ell,n}|}{2}\right) \\ &\quad \times \int_{\mathcal{D}_n} \mathrm{d}m \exp\left[-\tfrac{\beta n}{2}\langle [m-\mathbf{t}_n], \mathbb{A}_n\cdot[m-\mathbf{t}_n])\rangle\right] \exp\left(-\beta n(-\gamma_n)\,\langle v, m-\mathbf{t}_n\rangle^2\right) \\ &= \frac{1}{Z_n}\mathrm{e}^{-\beta n F_n(\mathbf{t}_n)} \frac{(-\gamma_n)n}{\sqrt{[-\det \mathbb{A}_n]}} \left(\frac{\pi n}{2\beta}\right)^{\frac{k}{2}-1} \left(\prod_{\ell\in[k]} \omega_{\ell,n}\right)\left[1 + O\left(\omega_{\ell,n}^{-1}\, n^{-1/2}\log^{1/2} n\right)\right], \end{aligned} \tag{B.109}$$

where we use [18, Eq. (10.2.33)] with $\varepsilon = \frac{1}{\beta n}$ and $d = k$. Here $\frac{1}{2}\,|A_{\ell,n}|$ is the inverse of the step in the ℓ–direction, while in [18, Eq. (10.2.33)] the step is ε.

Remark B.4.4. Note that

$$\tilde{\mathcal{E}}_{\mathcal{D}_n}(g,g) = \tilde{\mathcal{E}}_{\mathcal{D}_n}(\tilde{g},\tilde{g})\,[1 + o(1)] \tag{B.110}$$

because $\tilde{g}(m) = g(m)\,[1 + o(1)]$ for all $m \in W_0^c \cap \mathcal{D}_n$. The latter can be proved by approximating the Gaussian integral by 0 or 1 when $\langle v, m - \mathbf{t}_n\rangle$ is proportional to $-n^{-1/2}\log^{1/2} n$ or $n^{-1/2}\log^{1/2} n$, respectively. ♠

B.4.5 Final Dirichlet form approximation

We are now ready to compare $\mathcal{E}_{S_n}$ with $\tilde{\mathcal{E}}_{\mathcal{D}_n}$. Let $h\colon S_n \to [0,1]$ be such that $h(\sigma) = \tilde{g}(m_n(\sigma))$, $\sigma \in S_n$. We split the sum in (B.39) into four subsets of $\Gamma_n \times \Gamma_n$: $m \in W_0 \cap \mathcal{D}_n^c$, $m' \in \Gamma_n$; $m \in W_0 \cap \mathcal{D}_n$, $m' \in W_1$; $m \in W_0 \cap \mathcal{D}_n$, $m' \in W_2$; $m \in W_0 \cap \mathcal{D}_n$, $m' \in W_0 \cap \mathcal{D}_n$. Then, using (B.105)–(B.107), we obtain

$$\mathcal{E}_{S_n}(h,h) = O(n^{-c\beta}) + \frac{1}{2} \sum_{m \in W_0 \cap \mathcal{D}_n} \sum_{m' \in W_0 \cap \mathcal{D}_n} Q_n(m)\, \bar{r}_n(m,m')\, \big[\tilde{g}(m) - \tilde{g}(m')\big]^2. \tag{B.111}$$

Using (B.94) and (B.96), we obtain

$$\begin{aligned} \mathcal{E}_{S_n}(h,h) &= O(n^{-c\beta}) + \frac{1}{2} \sum_{m \in W_0 \cap \mathcal{D}_n} \Big[1 + O\big(n^{-1/2} \log^{3/2} n\big)\Big]\, \tilde{Q}_n(m) \\ &\quad \times \sum_{m' \in W_0 \cap \mathcal{D}_n} \Big(1 + O\big(n^{-1/2} \log^{1/2} n\big)\Big)\, \tilde{r}_n(m,m')\, \big[\tilde{g}(m) - \tilde{g}(m')\big]^2 \\ &= \Big[1 + O\big(n^{-1/2} \log^{1/2} n\big)\Big] \frac{1}{2} \sum_{m,m' \in W_0 \cap \mathcal{D}_n} \tilde{Q}_n(m) \tilde{r}_n(m,m')\, \big[\tilde{g}(m) - \tilde{g}(m')\big]^2 \\ &= \Big[1 + O\big(n^{-1/2} \log^{1/2} n\big)\Big] \frac{1}{2} \sum_{m,m' \in \mathcal{D}_n} \tilde{Q}_n(m) \tilde{r}_n(m,m')\, \big[\tilde{g}(m) - \tilde{g}(m')\big]^2 \\ &= \tilde{\mathcal{E}}_{\mathcal{D}_n}(\tilde{g},\tilde{g}) \Big[1 + O\big(n^{-1/2} \log^{1/2} n\big)\Big] \\ &= [1 + o_n(1)]\, \frac{1}{Z_n} \exp\left[-\beta n F_n(\mathbf{t}_n)\right] \frac{(-\gamma_n) n}{\sqrt{[-\det \mathbb{A}_n]}} \left(\frac{\pi n}{2\beta}\right)^{\frac{k}{2}-1} \left(\prod_{\ell \in [k]} \omega_{\ell,n}\right), \end{aligned} \tag{B.112}$$

where the third equality follows from (B.105)–(B.107) together with (B.94), and the last equality follows from (B.109)–(B.110).

B.5 Capacity and valley estimates

Section B.5.1 provides sharp asymptotic upper bounds and lower bounds on the capacity of the metastable pair between which the crossover is being considered. These estimates use the results of the Section B.4 together with the Dirichlet principle and the Berman-Konsowa principle, which are variational representations of capacity. Section B.5.2 provides a sharp asymptotic estimate for the mesoscopic measure of the valleys of the minima of F_n, which leads to a sharp asymptotic estimate for F_n inside this valley.

B.5.1 Capacity estimates

Given a Markov process $(x_t)_{t \geq 0}$ with state space S, a key quantity in the potential-theoretic approach to metastability is the *capacity* $\mathrm{cap}(A,B)$ of two disjoint subsets A, B of S. This is defined by (see [18, Eq. (7.1.39)])

$$\mathrm{cap}(A,B) = \sum_{x \in A} \mu(x) \mathbb{P}_x(\tau_B < \tau_A), \tag{B.113}$$

where μ is the invariant measure and $\mathbb{P}_x$ is the probability distribution of the Markov process starting in x.

Recall that $\mathcal{M}_n$ is the set of local minima of F_n.

Proposition B.5.1 (**Asymptotics of the capacity**). *Let $\mathbf{m}_n = (\mathbf{m}_{\ell,n})_{\ell\in[k]} \in \mathcal{M}_n$ and $M_n \subset \mathcal{M}_n \backslash \mathbf{m}_n$, such that the gate $\mathcal{G}(\mathbf{m}_n, M_n)$ consists of a unique point $\mathbf{t}_n = (\mathbf{t}_{\ell,n})_{\ell\in[k]}$. Suppose that $\beta \in (\beta_c, \infty)$ and $h \in [0, h_c(\beta))$. Then, as $n \to \infty$,*

$$\mathrm{cap}(S_n[\mathbf{m}_n], S_n[M_n]) = [1 + o_n(1)] \, \frac{1}{Z_n} \mathrm{e}^{-\beta n F_n(\mathbf{t}_n)} \frac{(-\gamma_n)\, n}{\sqrt{[-\det(\mathbb{A}_n(\mathbf{t}_n))]}} \left(\frac{\pi n}{2\beta}\right)^{\frac{k}{2}-1} \left(\prod_{\ell\in[k]} \omega_{\ell,n}\right). \tag{B.114}$$

Remark B.5.2. Proposition B.5.1 holds for any subset $M_n \subseteq \mathcal{M}_n \backslash \mathbf{m}_n$, separated from $\mathbf{m}_n$ by $\mathbf{t}_n$, independently on the values of F_n on M_n. ♠

Upper bound: Dirichlet principle

An important characterisation of the capacity between two disjoint sets is given by the *Dirichlet principle.* For our quantity of interest this states that

$$\mathrm{cap}(S_n[\mathbf{m}_n], S_n[M_n]) = \inf_{u\in\tilde{\mathcal{H}}} \mathcal{E}_{S_n}(u, u), \tag{B.115}$$

where $\tilde{\mathcal{H}}$ is the set of functions from S_n to $[0, 1]$ that are equal to 1 on $S_n[\mathbf{m}_n]$ and 0 on $S_n[M_n]$.

Given that, by assumption, $\mathcal{G}(\mathbf{m}_n, M_n) = \{\mathbf{t}_n\}$, we use the Dirichlet principle in (B.115) to obtain an upper bound on the capacity. We take as test function $h \in \tilde{\mathcal{H}}$ defined in Section B.4.5 and, using (B.112), we obtain

$$\begin{aligned}\mathrm{cap}(S_n[\mathbf{m}_n], S_n[M_n]) &\le \mathcal{E}_{S_n}(h, h) \\ &= [1 + o_n(1)] \, \frac{1}{Z_n} \mathrm{e}^{-\beta n F_n(\mathbf{t}_n)} \frac{(-\gamma_n) n}{\sqrt{[-\det(\mathbb{A}_n(\mathbf{t}_n))]}} \left(\frac{\pi n}{2\beta}\right)^{\frac{k}{2}-1} \left(\prod_{\ell\in[k]} \omega_{\ell,n}\right).\end{aligned} \tag{B.116}$$

Lower bound: Berman-Konsowa principle

We first note that the process $(\sigma_t)_{t\ge0}$ is lumpable. Indeed, the process $(m_n(\sigma_t))_{t\ge0}$ is Markovian because the Hamiltonian $H_n(\sigma)$ depends on $m_n(\sigma)$ only (see (B.27)). Therefore, for $\mathbf{A} = S_n[A]$ and $\mathbf{B} = S_n[B]$ with A and B disjoint subsets of Γ_n,

$$\mathrm{cap}(\mathbf{A}, \mathbf{B}) = \mathrm{cap}_\Gamma(A, B), \tag{B.117}$$

where cap_Γ denotes the capacity for the process $(m_n(\sigma_t))_{t\ge0}$, i.e., the projection of the process $(\sigma_t)_{t\ge0}$ on the magnetisation space Γ_n. We write $\mathbb{P}^\Gamma$ and $\mathbb{E}^\Gamma$ to denote the law of $(m_n(\sigma_t))_{t\ge0}$ induced by the law $\mathbb{P}$ of $(\sigma_t)_{t\ge0}$, and its expectation, respectively. By the lumpability, we can focus on the dynamics on Γ_n.

Following the line of argument in [18, Section 10.3] (with $\varepsilon = \frac{2}{n}$ and $d = k$), we obtain the lower bound

$$\begin{aligned}\mathrm{cap}(S_n[\mathbf{m}_n], S_n[M_n]) &= \mathrm{cap}_\Gamma(\mathbf{m}_n, M_n) \ge \tilde{\mathcal{E}}_{\mathcal{D}_n}(\tilde{g}, \tilde{g}) \left[1 + O(n^{-1/2} \log^{1/2} n)\right] \\ &= \frac{1}{Z_n} \mathrm{e}^{-\beta n F_n(\mathbf{t}_n)} \frac{(-\gamma_n) n}{\sqrt{[-\det(\mathbb{A}_n(\mathbf{t}_n))]}} \left(\frac{\pi n}{2\beta}\right)^{\frac{k}{2}-1} \left(\prod_{\ell\in[k]} \omega_{\ell,n}\right) [1 + o_n(1)],\end{aligned} \tag{B.118}$$

where we use (B.109) and (B.110).

We sketch the proof. The main idea is to use the Berman-Konsowa principle for a suitable defective flow. More precisely, given disjoint subsets A, B of the state space, for any *defective loop-free unit flow* $f_{A,B}$ from A to B with defect function δ (as defined in [18, Definition 9.2]), we can estimate (see [18, Lemma 9.4], and notation therein)

$$\mathrm{cap}(A,B) \geq \prod_{i=1}^{M} \left(1 + \left[\max_{y \in A_i} \frac{\delta(y)}{\mathcal{F}(y)}\right]_+\right)^{-1} \sum_{\gamma} \mathbb{P}^{f_{A,B}}(\gamma) \left[\left(\sum_{(x,y)\in\gamma} \frac{f_{A,B}((x,y))}{\mu(x)p(x,y)}\right)^{-1}\right], \tag{B.119}$$

where $[\cdot]_+$ denotes the positive part and γ is a self-avoiding path from A to B. It turns out that, with a suitable choice of the flow f, the product in the right-hand side of (B.119) is bounded from below by $1 + O(n^{-1/2}\log^{1/2} n)$, and the sum over γ from below by $\tilde{\mathcal{E}}_{\mathcal{D}_n}(\tilde{g},\tilde{g})[1+o_n(1)]$. This proves (B.118).

We give a sketch of the test flow definition in our setting. Here $A = \{\mathbf{m}_n\}$ and $B = M_n$. Let v^* be the eigenvector corresponding to the unique negative eigenvalue of the Hessian of F_n at the saddle point $\mathbf{t}_n$ (unique gate point in $\mathcal{G}(\{\mathbf{m}_n\}, M_n)$). Let G_n be the cylinder in $\mathbb{R}^k$ intersected with Γ_n, centred at $\mathbf{t}_n$, with axis v^*, radius $\rho = C\, n^{-1/2}\log^{1/2} n$ and length $\rho' = C'\, n^{-1/2}\log^{1/2} n$. We will denote by $\partial_B G_n$ the base facing B and by $\partial_A G_n$ the central part of radius $C''\, n^{-1/2}\log^{1/2} n$ of the base facing A, with $C'' < C$. Choose the constants so that G_n is contained in $\mathcal{D}_n$ defined in (B.81).

We define a defective flow $f_{A,B}$ from A to B consisting of three parts: f_A, a unitary flow from A to $\partial_A G_n$; f, a defective loop-free unit flow from $\partial_A G_n$ to $\partial_B G_n$ inside G_n; f_B, a unitary flow from $\partial_B G_n$ to B. This choice implies that the sum over γ in (B.119) is relevant only on the paths entering G_n in $\partial_A G_n$, exiting G_n in $\partial_B G_n$, and afterwards reaching B without going back to G_n. For this purpose we choose f_A and f_B such that $f_A((x,y))$ and $f_B((x,y))$ are proportional to $Q_n(x)$. For $m \in G_n$ such that $m^{\ell,+} \in G_n$, define

$$f((m, m^{\ell,+})) = \frac{\tilde{Q}_n(m) r_\ell \left[g(m^{\ell,+}) - g(m)\right]_+}{N(g)}, \tag{B.120}$$

where g is defined in (B.104), $\tilde{Q}_n$ in (B.86), r_ℓ in (B.89) and

$$N(g) = \sum_{m \in \partial_A G_n} \sum_{\substack{\ell \in [k]: \\ m^{\ell,+} \in G_n}} \tilde{Q}_n(m) r_\ell \left[g(m^{\ell,+}) - g(m)\right]_+ . \tag{B.121}$$

The contribution to the sum in brackets in (B.119) turns out to be negligible outside G_n. Therefore, no further conditions on the flows f_A and f_B are necessary, provided the total flow out of A is 1 and the total flow $f_{A,B}$ is defective and loop-free.

B.5.2 Measure of the valley

In order to prove Theorem B.1.2, we need the following estimate on the measure of the valley of the minima of F_n. For $\mathbf{m}_n \in \mathcal{M}_n$, let $A(\mathbf{m}_n) \subset \Gamma_n$ be the valley of $\mathbf{m}_n$ as defined in [18, Eq. (8.2.10)].

Lemma B.5.3 (Gibbs weight of the valley). *Given* $\mathbf{m}_n \in \mathcal{M}_n$*,*

$$Q_n(A(\mathbf{m}_n)) = \frac{1}{Z_n} \frac{\exp\left(-\beta n F_n(\mathbf{m}_n)\right)}{\sqrt{\det(\mathbb{A}_n(\mathbf{m}_n))}} \left(\frac{n\pi}{2\beta}\right)^{\frac{k}{2}} \left(\prod_{\ell\in[k]} \omega_{\ell,n}\right) \left[1 + O(n^{-1/2}\log^{3/2} n)\right], \tag{B.122}$$

where Q_n is the mesoscopic measure defined in (B.38)*, and $\mathbb{A}_n(\mathbf{m}_n)$ is the $k \times k$ Hessian matrix defined in* (B.82)*.*

Proof. The proof follows that of [18, Lemma 10.12 and (10.2.33)]. The relevant contribution to $Q_n(A(\mathbf{m}_n))$ is given by the measure of a ball B_ρ of radius $\rho = C\, n^{-1/2} \log^{1/2} n$ centred in $\mathbf{m}_n$, with C constant, contained in $A(\mathbf{m}_n)$. Indeed, if $y \in A(\mathbf{m}_n)$ and $d(\mathbf{m}_n, y) > \rho$, then by Taylor expansion of F_n around $\mathbf{m}_n$ we have

$$\begin{aligned} Q_n(y) &= \frac{1}{Z_n} \exp[-\beta n F_n(y)] = \frac{1}{Z_n} \exp\left[-\beta n [F_n(\mathbf{m}_n) + c\, d(\mathbf{m}_n, y)^2]\right] \\ &\leq \frac{1}{Z_n} \exp\left[-\beta n [F_n(\mathbf{m}_n) + c\rho^2]\right] = \frac{n^{-\beta c C^2}}{Z_n} \exp\left[-\beta n F_n(\mathbf{m}_n)\right], \end{aligned} \tag{B.123}$$

where c is a constant. The condition $y \in A(\mathbf{m}_n)$ is needed to ensure that $F_n(y) > F_n(\mathbf{m}_n)$, implying that c is positive. Therefore, we obtain the rough estimate

$$Q_n(A(\mathbf{m}_n) \backslash B_\rho) \leq n^k \frac{n^{-\beta c C^2}}{Z_n} \exp\left[-\beta n F_n(\mathbf{m}_n)\right], \tag{B.124}$$

where we use that $|\Gamma_n| \leq n^k$. The bound in (B.124) is sufficient to show that $Q_n(A(\mathbf{m}_n) \backslash B_\rho)$ is negligible in $Q_n(A(\mathbf{m}_n))$.

Compute

$$\begin{aligned} Z_n Q_n(A(\mathbf{m}_n) \cap B_\rho) &= Z_n Q_n(B_\rho) = Z_n \sum_{y \in B_\rho} Q_n(y) = \sum_{y \in B_\rho} \mathrm{e}^{-\beta n F_n(y)} \\ &= \mathrm{e}^{-\beta n F_n(\mathbf{m}_n)} \sum_{y \in B_\rho} \exp\left[-\frac{\beta n}{2} \langle y - \mathbf{m}_n, (\mathbb{A}_n(\mathbf{m}_n)) \cdot (y - \mathbf{m}_n) \rangle + O(n\rho^3)\right] \\ &= \mathrm{e}^{-\beta n F_n(\mathbf{m}_n)} [1 + O(n\rho^3)] \sum_{y \in B_\rho} \exp\left[-\frac{\beta n}{2} \langle y - \mathbf{m}_n, (\mathbb{A}_n(\mathbf{m}_n)) \cdot (y - \mathbf{m}_n) \rangle\right] \\ &= \mathrm{e}^{-\beta n F_n(\mathbf{m}_n)} \left(\prod_{\ell \in [k]} \frac{|A_{\ell,n}|}{2}\right) [1 + O(n\rho^3)] \\ &\times \int_{B_\rho} \mathtt{d} y \exp\left[-\frac{\beta n}{2} \langle y - \mathbf{m}_n, (\mathbb{A}_n(\mathbf{m}_n)) \cdot (y - \mathbf{m}_n) \rangle\right] \\ &= \mathrm{e}^{-\beta n F_n(\mathbf{m}_n)} \left(\frac{n}{2}\right)^k \left(\prod_{\ell \in [k]} \omega_{\ell,n}\right) [1 + O(n\rho^3)] \left(\frac{2\pi}{n\beta}\right)^{\frac{k}{2}} \sqrt{\frac{1}{\det(\mathbb{A}_n(\mathbf{m}_n))}} \\ &= \frac{\mathrm{e}^{-\beta n F_n(\mathbf{m}_n)}}{\sqrt{\det(\mathbb{A}_n(\mathbf{m}_n))}} \left(\frac{n\pi}{2\beta}\right)^{\frac{k}{2}} \left(\prod_{\ell \in [k]} \omega_{\ell,n}\right) [1 + O(n\rho^3)], \end{aligned} \tag{B.125}$$

where we use the Taylor expansion

$$F_n(y) = F_n(\mathbf{m}_n) + \frac{1}{2} \left\langle y - \mathbf{m}_n, (\nabla^2 F_n) \cdot (\mathbf{m}_n)(y - \mathbf{m}_n) \right\rangle + O(\rho^3), \qquad y \in B_\rho, \tag{B.126}$$

and the approximation of the sum by an integral is correct up to an error $1 + O(\rho)$. In the last lines we approximated the Gaussian integral on intervals $[-\rho, \rho]$ by the Gaussian integral on $\mathbb{R}$, with an error $1 + O(n^{-c})$. We conclude by looking at (B.124) and (B.125), and noting that for C large enough $Q_n(A(\mathbf{m}_n) \backslash B_\rho)$ is negligible compared to $Q_n(A(\mathbf{m}_n) \cap B_\rho)$. □ □

B.6 Proof of the theorems

In this section we prove Theorems B.1.2–B.1.4. Section B.6.1 uses the asymptotics for the capacity of the metastable pair from Section B.5.1 and the asymptotics for the mesoscopic measure from Section B.5.2 to prove Theorem B.1.2. Section B.6.2 proves Theorem B.1.3. Section B.6.3 proves Theorem B.1.4.

B.6.1 Average crossover time

Let us return to the notation of Theorem B.1.2, where $\mathbf{m}_n \in \mathcal{M}_n$ and $\mathcal{M}_n(\mathbf{m}_n) = \{m \in \mathcal{M}_n \backslash \mathbf{m}_n \colon F_n(m) \leq F_n(\mathbf{m}_n)\}$. To prove Theorem B.1.2 we use the relation

$$\mathbb{E}^{\Gamma}_{\mathbf{m}_n}(\tau_{\mathcal{M}_n(\mathbf{m}_n)}) = [1+o_n(1)]\,\frac{\mu(A(\mathbf{m}_n))}{\mathrm{cap}_\Gamma(\mathbf{m}_n, \mathcal{M}_n(\mathbf{m}_n))}, \tag{B.127}$$

Recall notation introduced in Section B.5.1. Because $F_n(m) \leq F_n(\mathbf{m}_n)$ for all $m \in \mathcal{M}_n(\mathbf{m}_n)$, (B.127) follows from [18, Theorem 8.15] after proving that $\mathcal{M}_n$ is a set of metastable points in the sense of [18, Definition 8.2]. The latter follows along the lines of the proof of [18, Theorem 10.6], where similar values of capacities and invariant measures occur.

Using (B.127) in combination with Proposition B.5.1 and Lemma B.5.3, we obtain that, for all $\sigma \in S_n[\mathbf{m}_n]$,

$$\begin{aligned}
\mathbb{E}_\sigma(\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]}) = \mathbb{E}^{\Gamma}_{\mathbf{m}_n}(\tau_{\mathcal{M}_n(\mathbf{m}_n)}) &= [1+o_n(1)]\,\frac{Q_n(A(\mathbf{m}_n))}{\mathrm{cap}_\Gamma(\mathbf{m}_n, \mathcal{M}_n(\mathbf{m}_n))}\\
&= [1+o_n(1)]\,\frac{Q_n(A(\mathbf{m}_n))}{\mathrm{cap}(S_n[\mathbf{m}_n], S_n[\mathcal{M}_n(\mathbf{m}_n)])}\\
&= [1+o_n(1)]\,\frac{\frac{1}{Z_n}\frac{\exp(-\beta n F_n(\mathbf{m}_n))}{\sqrt{\det(\mathbb{A}_n(\mathbf{m}_n))}}\left(\frac{n\pi}{2\beta}\right)^{\frac{k}{2}}\left(\prod_{\ell\in[k]}\omega_{\ell,n}\right)}{\frac{1}{Z_n}\exp\left[-\beta n F_n(\mathbf{t}_n)\right]\frac{(-\gamma_n)n}{\sqrt{[-\det(\mathbb{A}_n(\mathbf{t}_n))]}}\left(\frac{\pi n}{2\beta}\right)^{\frac{k}{2}-1}\left(\prod_{\ell\in[k]}\omega_{\ell,n}\right)}\\
&= [1+o_n(1)]\sqrt{\frac{[-\det(\mathbb{A}_n(\mathbf{t}_n))]}{\det(\mathbb{A}_n(\mathbf{m}_n))}}\left(\frac{\pi}{2\beta(-\gamma_n)}\right)\exp\left[\beta n(F_n(\mathbf{t}_n)-F_n(\mathbf{m}_n))\right],
\end{aligned} \tag{B.128}$$

where we use that the dynamics depends on the starting configuration $\sigma \in S_n[\mathbf{m}_n]$ only, through its level magnetisations $m_n(\sigma) = \mathbf{m}_n$ (see (B.27)), and also use the lumpability.

B.6.2 Exponential law

In this section we prove Theorem B.1.3. Since the dynamics depends on the starting configuration $\sigma \in S_n[\mathbf{m}_n]$ through its level magnetisation $m_n(\sigma) = \mathbf{m}_n$ only (see (B.27)), we have

$$\lim_{n\to\infty}\mathbb{P}_\sigma\left(\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]} > t\,\mathbb{E}_\sigma\left[\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]}\right]\right) = \lim_{n\to\infty}\mathbb{P}^{\Gamma}_{\mathbf{m}_n}\left(\bar\tau_{\mathcal{M}_n(\mathbf{m}_n)} > t\,\mathbb{E}^{\Gamma}_{\mathbf{m}_n}\left[\bar\tau_{\mathcal{M}_n(\mathbf{m}_n)}\right]\right), \tag{B.129}$$

where $\bar\tau$ is the hitting time of the process projected on Γ_n. Given the non-degeneracy hypothesis (Hypothesis B.1.1 in Section B.1.3) and the one-dimensional landscape analysis (in Section B.3.3), we can apply [18, Theorem 8.45] to the right-hand side of (B.129) and conclude the proof.

B.6.3 Randomness of the exponent

In this section we prove Theorem B.1.4. In particular, we compute $F_n(\mathbf{t}_n)-F_n(\mathbf{m}_n)-[F_{\beta,h}(\mathbf{t})-F_{\beta,h}(\mathbf{m})]$ to leading order.

Recalling definitions (B.47) and (B.58), we have

$$F_{\beta,h}(m) = -\frac{1}{2}K(m)^2 - h\sum_{\ell\in[k]}\omega_\ell\, m_\ell + \frac{1}{\beta}\sum_{\ell\in[k]}\omega_\ell I_{\mathbf{C}}(m_\ell). \tag{B.130}$$

Let $\mathbf{m}=(\mathbf{m}_\ell)_{\ell\in[k]}, \mathbf{t}=(\mathbf{t}_\ell)_{\ell\in[k]}\in[-1,1]^k$ be the critical points of $F_{\beta,h}$ closest to $\mathbf{m}_n, \mathbf{t}_n$ (i.e., the critical points of F_n defined above), respectively. Note that $\mathbf{m}$ and $\mathbf{t}$ satisfy (B.57), while $\mathbf{m}_n$ and $\mathbf{t}_n$ satisfy (B.56). Using (B.42), we get

$$\begin{aligned} F_n(\mathbf{t}_n)-F_{\beta,h}(\mathbf{t}_n) = &-\frac{1}{2}[K_n(\mathbf{t}_n)^2-K(\mathbf{t}_n)^2] - h\sum_{\ell\in[k]}[\omega_{\ell,n}-\omega_\ell]\,\mathbf{t}_{\ell,n} \\ &+\frac{1}{\beta}\left[\sum_{\ell\in[k]}[\omega_{\ell,n}-\omega_\ell]I_{\mathbf{C}}(\mathbf{t}_{\ell,n}) + \sum_{\ell\in[k]}\frac{1}{2n}\log\left(\frac{\pi(1-\mathbf{t}_{\ell,n}^2)}{2}\right)\omega_{\ell,n} - \frac{k}{2n} + o\left(n^{-1}\right)\right] \end{aligned} \tag{B.131}$$

and

$$F_{\beta,h}(\mathbf{t}_n)-F_{\beta,h}(\mathbf{t}) = -\frac{1}{2}[K(\mathbf{t}_n)^2-K(\mathbf{t})^2] + \frac{1}{\beta}\sum_{\ell\in[k]}\omega_\ell[I_{\mathbf{C}}(\mathbf{t}_{\ell,n})-I_{\mathbf{C}}(\mathbf{t}_\ell)]. \tag{B.132}$$

By (B.55), we have

$$\begin{aligned} \frac{1}{2}\log\left(\frac{1+\mathbf{t}_{\ell,n}}{1-\mathbf{t}_{\ell,n}}\right) &= \beta\left[a_\ell K_n(\mathbf{t}_n)+h\right], \\ \frac{1}{2}\log\left(\frac{1+\mathbf{t}_\ell}{1-\mathbf{t}_\ell}\right) &= \beta\left[a_\ell K(\mathbf{t})+h\right]. \end{aligned} \tag{B.133}$$

Thus,

$$\begin{aligned} I_{\mathbf{C}}(\mathbf{t}_{\ell,n})-I_{\mathbf{C}}(\mathbf{t}_\ell) &= (\mathbf{t}_{\ell,n}-\mathbf{t}_\ell)I'_{\mathbf{C}}(\mathbf{t}_\ell)+O((\mathbf{t}_{\ell,n}-\mathbf{t}_\ell)^2) \\ &= (\mathbf{t}_{\ell,n}-\mathbf{t}_\ell)\frac{1}{2}\log\left(\frac{1+\mathbf{t}_\ell}{1-\mathbf{t}_\ell}\right)+O((\mathbf{t}_{\ell,n}-\mathbf{t}_\ell)^2) \\ &= (\mathbf{t}_{\ell,n}-\mathbf{t}_\ell)\beta\left[a_\ell K(\mathbf{t})+h\right]+O((\mathbf{t}_{\ell,n}-\mathbf{t}_\ell)^2). \end{aligned} \tag{B.134}$$

Moreover,

$$\begin{aligned} K(\mathbf{t}_n)^2-K(\mathbf{t})^2 &= \sum_{\ell,\ell'\in[k]} a_\ell\, a_{\ell'}\,\omega_\ell\,\omega_{\ell'}[\mathbf{t}_{\ell,n}\mathbf{t}_{\ell',n}-\mathbf{t}_\ell\mathbf{t}_{\ell'}] \\ &= \sum_{\ell,\ell'\in[k]} a_\ell\, a_{\ell'}\,\omega_\ell\,\omega_{\ell'}\left(\mathbf{t}_\ell[\mathbf{t}_{\ell',n}-\mathbf{t}_{\ell'}]+\mathbf{t}_{\ell'}[\mathbf{t}_{\ell,n}-\mathbf{t}_\ell]+[\mathbf{t}_{\ell,n}-\mathbf{t}_\ell][\mathbf{t}_{\ell',n}-\mathbf{t}_{\ell'}]\right) \end{aligned} \tag{B.135}$$

and

$$\begin{aligned} K_n(\mathbf{t}_n)^2-K(\mathbf{t}_n)^2 &= \sum_{\ell,\ell'\in[k]} a_\ell\, a_{\ell'}\left[\omega_{\ell,n}\,\omega_{\ell',n}-\omega_\ell\,\omega_{\ell'}\right]\mathbf{t}_{\ell,n}\mathbf{t}_{\ell',n} \\ &= \sum_{\ell,\ell'\in[k]} a_\ell\, a_{\ell'}\,\mathbf{t}_{\ell,n}\mathbf{t}_{\ell',n}\left(\omega_\ell[\omega_{\ell',n}-\omega_{\ell'}]+\omega_{\ell'}[\omega_{\ell,n}-\omega_\ell]+[\omega_{\ell,n}-\omega_\ell][\omega_{\ell',n}-\omega_{\ell'}]\right). \end{aligned} \tag{B.136}$$

Similar equalities hold after we replace $\mathbf{t}$ by $\mathbf{m}$ and $\mathbf{t}_n$ by $\mathbf{m}_n$. Using the previous computations, we obtain

$$
\begin{aligned}
&F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n) - [F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})] \\
&= F_n(\mathbf{t}_n) - F_{\beta,h}(\mathbf{t}_n) + F_{\beta,h}(\mathbf{t}_n) - F_{\beta,h}(\mathbf{t}) - [F_n(\mathbf{m}_n) - F_{\beta,h}(\mathbf{m}_n) + F_{\beta,h}(\mathbf{m}_n) - F_{\beta,h}(\mathbf{m})] \\
&= -\frac{1}{2} \sum_{\ell,\ell' \in [k]} a_\ell\, a_{\ell'} \left[\mathbf{t}_{\ell,n}\mathbf{t}_{\ell',n} - \mathbf{m}_{\ell,n}\mathbf{m}_{\ell',n}\right] \left(\omega_\ell[\omega_{\ell',n} - \omega_{\ell'}] + \omega_{\ell'}[\omega_{\ell,n} - \omega_\ell] + [\omega_{\ell,n} - \omega_\ell][\omega_{\ell',n} - \omega_{\ell'}]\right) \\
&\quad - \frac{1}{2} \sum_{\ell,\ell' \in [k]} a_\ell\, a_{\ell'}\, \omega_\ell\, \omega_{\ell'} \left[\mathbf{t}_{\ell,n}\mathbf{t}_{\ell',n} - \mathbf{t}_\ell \mathbf{t}_{\ell'} + \mathbf{m}_\ell \mathbf{m}_{\ell'} - \mathbf{m}_{\ell,n}\mathbf{m}_{\ell',n}\right] \\
&\quad - h \sum_{\ell \in [k]} [\omega_{\ell,n} - \omega_\ell]\, [\mathbf{t}_{\ell,n} - \mathbf{m}_{\ell,n}] \\
&\quad + \frac{1}{\beta} \sum_{\ell \in [k]} [\omega_{\ell,n} - \omega_\ell]\, [I_{\mathbf{C}}(\mathbf{t}_{\ell,n}) - I_{\mathbf{C}}(\mathbf{m}_{\ell,n})] + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{1}{2n} \log\left(\frac{1 - \mathbf{t}_{\ell,n}^2}{1 - \mathbf{m}_{\ell,n}^2}\right) \\
&\quad + \frac{1}{\beta} \sum_{\ell \in [k]} \omega_\ell\, [I_{\mathbf{C}}(\mathbf{t}_{\ell,n}) - I_{\mathbf{C}}(\mathbf{t}_\ell) + I_{\mathbf{C}}(\mathbf{m}_\ell) - I_{\mathbf{C}}(\mathbf{m}_{\ell,n})] + o\left(n^{-1}\right).
\end{aligned}
\tag{B.137}
$$

Using(B.134), we find

$$
\begin{aligned}
&[F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)] - [F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})] \\
&= -\frac{1}{2} \sum_{\ell,\ell' \in [k]} a_\ell\, a_{\ell'} \left[\mathbf{t}_{\ell,n}\mathbf{t}_{\ell',n} - \mathbf{m}_{\ell,n}\mathbf{m}_{\ell',n}\right] \left(\omega_\ell[\omega_{\ell',n} - \omega_{\ell'}] + \omega_{\ell'}[\omega_{\ell,n} - \omega_\ell] + [\omega_{\ell,n} - \omega_\ell][\omega_{\ell',n} - \omega_{\ell'}]\right) \\
&\quad - \frac{1}{2} \sum_{\ell,\ell' \in [k]} a_\ell\, a_{\ell'}\, \omega_\ell\, \omega_{\ell'} \left[\mathbf{t}_{\ell,n}\mathbf{t}_{\ell',n} - \mathbf{t}_\ell \mathbf{t}_{\ell'} + \mathbf{m}_\ell \mathbf{m}_{\ell'} - \mathbf{m}_{\ell,n}\mathbf{m}_{\ell',n}\right] \\
&\quad - h \sum_{\ell \in [k]} [\omega_{\ell,n} - \omega_\ell]\, [\mathbf{t}_{\ell,n} - \mathbf{m}_{\ell,n}] \\
&\quad + \frac{1}{\beta} \sum_{\ell \in [k]} [\omega_{\ell,n} - \omega_\ell]\, [I_{\mathbf{C}}(\mathbf{t}_{\ell,n}) - I_{\mathbf{C}}(\mathbf{m}_{\ell,n})] + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{1}{2n} \log\left(\frac{1 - \mathbf{t}_{\ell,n}^2}{1 - \mathbf{m}_{\ell,n}^2}\right) \\
&\quad + \frac{1}{\beta} \sum_{\ell \in [k]} \omega_\ell \Big[(\mathbf{t}_{\ell,n} - \mathbf{t}_\ell)\beta\, [a_\ell K(\mathbf{t}) + h] + O((\mathbf{t}_{\ell,n} - \mathbf{t}_\ell)^2) - (\mathbf{m}_{\ell,n} - \mathbf{m}_\ell)\beta\, [a_\ell K(\mathbf{m}) + h] \\
&\qquad\qquad + O((\mathbf{m}_{\ell,n} - \mathbf{m}_\ell)^2)\Big] \\
&\quad + o\left(n^{-1}\right).
\end{aligned}
\tag{B.138}
$$

Since

$$
\mathbf{t}_{\ell,n}\mathbf{t}_{\ell',n} - \mathbf{t}_\ell \mathbf{t}_{\ell'} = \left(\mathbf{t}_\ell[\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'}] + \mathbf{t}_{\ell'}[\mathbf{t}_{\ell,n} - \mathbf{t}_\ell] + [\mathbf{t}_{\ell,n} - \mathbf{t}_\ell][\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'}]\right), \tag{B.139}
$$

we focus on estimating $\mathbf{t}_{\ell,n} - \mathbf{t}_\ell$.

From Taylor expansion, we get

$$\begin{aligned}\mathbf{t}_{\ell,n} - \mathbf{t}_\ell &= \tanh\left(\beta\left[a_\ell \sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell',n}\,\mathbf{t}_{\ell',n} + h\right]\right) - \tanh\left(\beta\left[a_\ell \sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell'}\,\mathbf{t}_{\ell'} + h\right]\right)\\ &= \beta\, a_\ell \sum_{\ell'\in[k]} a_{\ell'}[\omega_{\ell',n}\,\mathbf{t}_{\ell',n} - \omega_{\ell'}\,\mathbf{t}_{\ell'}]\left[1 - \tanh\left(\beta\left[a_\ell \sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell'}\,\mathbf{t}_{\ell'} + h\right]\right)^2\right]\\ &\quad - \beta^2\, a_\ell^2 \left(\sum_{\ell'\in[k]} a_{\ell'}[\omega_{\ell',n}\,\mathbf{t}_{\ell',n} - \omega_{\ell'}\,\mathbf{t}_{\ell'}]\right)^2 \tanh\left(\beta\left[a_\ell \sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell'}\,\mathbf{t}_{\ell'} + h\right]\right)\\ &\quad \times \left[1 - \tanh\left(\beta\left[a_\ell \sum_{\ell'\in[k]} a_{\ell'}\,\omega_{\ell'}\,\mathbf{t}_{\ell'} + h\right]\right)^2\right]\\ &\quad + O\left(a_\ell^3\left(\sum_{\ell'\in[k]} a_{\ell'}[\omega_{\ell',n}\,\mathbf{t}_{\ell',n} - \omega_{\ell'}\,\mathbf{t}_{\ell'}]\right)^3\right).\end{aligned} \tag{B.140}$$

Since

$$\omega_{\ell',n}\,\mathbf{t}_{\ell',n} - \omega_{\ell'}\,\mathbf{t}_{\ell'} = (\omega_{\ell',n} - \omega_{\ell'})\,\mathbf{t}_{\ell'} + \omega_{\ell',n}(\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'}), \tag{B.141}$$

we have

$$\begin{aligned}\mathbf{t}_{\ell,n} - \mathbf{t}_\ell = \beta\, a_\ell \left[1 - \mathbf{t}_\ell^2\right] &\sum_{\ell'\in[k]} a_{\ell'}[(\omega_{\ell',n} - \omega_{\ell'})\,\mathbf{t}_{\ell'} + \omega_{\ell',n}(\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'})]\\ &- \beta^2\, a_\ell^2\,\mathbf{t}_\ell\left[1 - \mathbf{t}_\ell^2\right]\left(\sum_{\ell'\in[k]} a_{\ell'}[(\omega_{\ell',n} - \omega_{\ell'})\,\mathbf{t}_{\ell'} + \omega_{\ell',n}(\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'})]\right)^2\\ &+ O\left(a_\ell^3\left(\sum_{\ell'\in[k]} a_{\ell'}[(\omega_{\ell',n} - \omega_{\ell'})\,\mathbf{t}_{\ell'} + \omega_{\ell',n}(\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'})]\right)^3\right).\end{aligned} \tag{B.142}$$

Suppose that $\mathbf{t}_{\ell,n} - \mathbf{t}_\ell \sim \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}}$. By the Central Limit Theorem, $\omega_{\ell,n} - \omega_\ell \sim \frac{Z_\ell}{\sqrt{n}}$, where Z_ℓ is the normal random variable $N(0,\omega_\ell(1-\omega_\ell))$. Hence

$$
\begin{aligned}
\frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} &= \beta\, a_\ell \left[1-\mathbf{t}_\ell^2\right] \sum_{\ell'\in[k]} a_{\ell'} \left[\frac{Z_{\ell'}}{\sqrt{n}}\mathbf{t}_{\ell'} + \left(\frac{Z_{\ell'}}{\sqrt{n}}+\omega_{\ell'}\right)\frac{Y_{\ell'}^{\mathbf{t}}}{\sqrt{n}}\right] \\
&\quad - \beta^2\, a_\ell^2\, \mathbf{t}_\ell \left[1-\mathbf{t}_\ell^2\right] \left(\sum_{\ell'\in[k]} a_{\ell'} \left[\frac{Z_{\ell'}}{\sqrt{n}}\,\mathbf{t}_{\ell'} + \left(\frac{Z_{\ell'}}{\sqrt{n}}+\omega_{\ell'}\right)\frac{Y_{\ell'}^{\mathbf{t}}}{\sqrt{n}}\right]\right)^2 \\
&\quad + O\left(a_\ell^3 \left(\sum_{\ell'\in[k]} a_{\ell'} \left[\frac{Z_{\ell'}}{\sqrt{n}}\,\mathbf{t}_{\ell'} + \left(\frac{Z_{\ell'}}{\sqrt{n}}+\omega_{\ell'}\right)\frac{Y_{\ell'}^{\mathbf{t}}}{\sqrt{n}}\right]\right)^3\right) \\
&= \frac{1}{\sqrt{n}}\beta\, a_\ell \left[1-\mathbf{t}_\ell^2\right] \sum_{\ell'\in[k]} a_{\ell'} \left(\mathbf{t}_{\ell'} Z_{\ell'} + \omega_{\ell'} Y_{\ell'}^{\mathbf{t}}\right) \\
&\quad + \frac{1}{n}\beta\, a_\ell \left[1-\mathbf{t}_\ell^2\right] \sum_{\ell'\in[k]} a_{\ell'} Z_{\ell'} \left(Y_{\ell'}^{\mathbf{t}} - \beta\, a_\ell \mathbf{t}_\ell \mathbf{t}_{\ell'} \sum_{\ell''\in[k]} a_{\ell''}\omega_{\ell''} Y_{\ell''}^{\mathbf{t}}\right) + o(n^{-1})
\end{aligned}
\tag{B.143}
$$

and so

$$
Y_\ell^{\mathbf{t}} = \beta\, a_\ell \left[1-\mathbf{t}_\ell^2\right] \frac{\sum_{\ell'\in[k]} a_{\ell'} \mathbf{t}_{\ell'} Z_{\ell'}}{1-\beta \sum_{\ell'\in[k]} a_{\ell'}^2 \omega_{\ell'} \left[1-\mathbf{t}_{\ell'}^2\right]} + O(n^{-\frac{1}{2}}), \tag{B.144}
$$

where the denominator does not vanish because of Remark B.4.3. Thus, up to a factor $O(n^{-\frac{1}{2}})$, $Y_\ell^{\mathbf{t}}$ is a normal random variable with mean 0 and variance

$$
\beta^2\, a_\ell^2 \left[1-\mathbf{t}_\ell^2\right]^2 \frac{\sum_{\ell'\in[k]} a_{\ell'}^2 \mathbf{t}_{\ell'}^2 \omega_{\ell'}(1-\omega_{\ell'})}{\left(1-\beta \sum_{\ell'\in[k]} a_{\ell'}^2 \omega_{\ell'} \left[1-\mathbf{t}_{\ell'}^2\right]\right)^2}. \tag{B.145}
$$

Similar results hold after we replace $\mathbf{t}$ by $\mathbf{m}$.

Going back to (B.138), using (B.139) and (B.144), and inserting $\mathbf{t}_{\ell,n} - \mathbf{t}_\ell \sim \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}}$ and $\mathbf{m}_{\ell,n} -$

$\mathbf{m}_\ell \sim \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}}$ and $\omega_{\ell,n} - \omega_\ell \sim \frac{Z_\ell}{\sqrt{n}}$, we obtain

$$\begin{aligned}
&[F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)] - [F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})] \\
&\sim -\frac{1}{2} \sum_{\ell,\ell' \in [k]} a_\ell \, a_{\ell'} \left[\left(\mathbf{t}_\ell + \frac{Y^{\mathbf{t}}_\ell}{\sqrt{n}} \right) \left(\mathbf{t}_{\ell'} + \frac{Y^{\mathbf{t}}_{\ell'}}{\sqrt{n}} \right) - \left(\mathbf{m}_\ell + \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}} \right) \left(\mathbf{m}_{\ell'} + \frac{Y^{\mathbf{m}}_{\ell'}}{\sqrt{n}} \right) \right] \\
&\quad \times \left(\omega_\ell \frac{Z_{\ell'}}{\sqrt{n}} + \omega_{\ell'} \frac{Z_\ell}{\sqrt{n}} + \frac{Z_\ell Z_{\ell'}}{n} \right) \\
&\quad - \frac{1}{2} \sum_{\ell,\ell' \in [k]} a_\ell \, a_{\ell'} \, \omega_\ell \, \omega_{\ell'} \left(\mathbf{t}_\ell \frac{Y^{\mathbf{t}}_{\ell'}}{\sqrt{n}} + \mathbf{t}_{\ell'} \frac{Y^{\mathbf{t}}_\ell}{\sqrt{n}} + \frac{Y^{\mathbf{t}}_\ell Y^{\mathbf{t}}_{\ell'}}{n} - \mathbf{m}_\ell \frac{Y^{\mathbf{m}}_{\ell'}}{\sqrt{n}} - \mathbf{m}_{\ell'} \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}} - \frac{Y^{\mathbf{m}}_\ell Y^{\mathbf{m}}_{\ell'}}{n} \right) \\
&\quad - h \sum_{\ell \in [k]} \frac{Z_\ell}{\sqrt{n}} \left(\mathbf{t}_\ell + \frac{Y^{\mathbf{t}}_\ell}{\sqrt{n}} - \mathbf{m}_\ell - \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}} \right) \\
&\quad + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{Z_\ell}{\sqrt{n}} \left[I_{\mathbf{C}} \left(\mathbf{t}_\ell + \frac{Y^{\mathbf{t}}_\ell}{\sqrt{n}} \right) - I_{\mathbf{C}} \left(\mathbf{m}_\ell + \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}} \right) \right] + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{1}{2n} \log \left(\frac{1 - \left(\mathbf{t}_\ell + \frac{Y^{\mathbf{t}}_\ell}{\sqrt{n}} \right)^2}{1 - \left(\mathbf{m}_\ell + \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}} \right)^2} \right) \\
&\quad + \frac{1}{\beta} \sum_{\ell \in [k]} \omega_\ell \left[\frac{Y^{\mathbf{t}}_\ell}{\sqrt{n}} \beta \left[a_\ell K(\mathbf{t}) + h \right] + O\left(\frac{(Y^{\mathbf{t}}_\ell)^2}{n} \right) - \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}} \left[a_\ell K(\mathbf{m}) + h \right] + O\left(\frac{(Y^{\mathbf{m}}_\ell)^2}{n} \right) \right] \\
&\quad + o\left(n^{-1} \right).
\end{aligned} \tag{B.146}$$

Thus,

$$\begin{aligned}
&[F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)] - [F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})] \\
&= -\frac{1}{2} \sum_{\ell,\ell' \in [k]} a_\ell \, a_{\ell'} \left[\mathbf{t}_\ell \mathbf{t}_{\ell'} - \mathbf{m}_\ell \mathbf{m}_{\ell'} \right] \left(\omega_\ell \frac{Z_{\ell'}}{\sqrt{n}} + \omega_{\ell'} \frac{Z_\ell}{\sqrt{n}} \right) \\
&\quad - \frac{1}{2} \sum_{\ell,\ell' \in [k]} a_\ell \, a_{\ell'} \, \omega_\ell \, \omega_{\ell'} \left(\mathbf{t}_\ell \frac{Y^{\mathbf{t}}_{\ell'}}{\sqrt{n}} + \mathbf{t}_{\ell'} \frac{Y^{\mathbf{t}}_\ell}{\sqrt{n}} - \mathbf{m}_\ell \frac{Y^{\mathbf{m}}_{\ell'}}{\sqrt{n}} - \mathbf{m}_{\ell'} \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}} \right) \\
&\quad - h \sum_{\ell \in [k]} \left[\mathbf{t}_\ell - \mathbf{m}_\ell \right] \frac{Z_\ell}{\sqrt{n}} + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{Z_\ell}{\sqrt{n}} \left[I_{\mathbf{C}} \left(\mathbf{t}_\ell + \frac{Y^{\mathbf{t}}_\ell}{\sqrt{n}} \right) - I_{\mathbf{C}} \left(\mathbf{m}_\ell + \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}} \right) \right] \\
&\quad + \sum_{\ell \in [k]} \omega_\ell \left[\frac{Y^{\mathbf{t}}_\ell}{\sqrt{n}} \left[a_\ell K(\mathbf{t}) + h \right] - \frac{Y^{\mathbf{m}}_\ell}{\sqrt{n}} \left[a_\ell K(\mathbf{m}) + h \right] \right] + O\left(n^{-1} \right).
\end{aligned} \tag{B.147}$$

Since the random variables $Y^{\mathbf{t}}_\ell$, $Y^{\mathbf{m}}_\ell$, Z_ℓ are centred normal, this concludes the proof of Theorem B.1.4.

From (B.147) it is possible to compute explicitly the variance of Z defined in Theorem B.1.4, because the variances of all the random variables involved are known (at least to leading order).

B.7 Appendix A: Metastability on the complete graph without disorder

We give a brief overview of well-known results for the standard Curie-Weiss model. We refer to [18, Chapter 13] for more details.

The Glauber dynamics is defined as in Section B.1.2, but with $J \equiv 1$. For convenience we write the Curie-Weiss Hamiltonian as

$$H_n(\sigma) = -\frac{1}{2n} \sum_{i,j\in[n]} \sigma(i)\sigma(j) - h \sum_{i\in[n]} \sigma(i), \qquad \sigma \in S_n, \tag{B.148}$$

which is as (B.26) when $J \equiv 1$. What makes this case easier than the one with disorder is that the interaction is *mean-field*. Indeed, we may write

$$H_n(\sigma) = n\big[-\tfrac{1}{2}m_n(\sigma)^2 - hm_n(\sigma)\big], \tag{B.149}$$

with

$$m_n(\sigma) = \frac{1}{n} \sum_{i\in[n]} \sigma(i) \in [-1, 1] \tag{B.150}$$

the magnetisation. In this case the magnetisation process $(m_n(t))_{t\geq 0}$, defined by

$$m_n(t) = m_n(\sigma_t), \tag{B.151}$$

is Markovian. More specifically, it is a nearest-neighbour random walk on the grid

$$\Gamma_n = \Big\{-1, -1 + \tfrac{2}{n}, \dots, +1 - \tfrac{2}{n}, +1\Big\}. \tag{B.152}$$

In the limit as $n \to \infty$, (B.151) converges to a Brownian motion on $[-1, +1]$ in the potential $F_{\beta,h}$ given by

$$F_{\beta,h}(m) = -\frac{1}{2}m^2 - hm + \frac{1}{\beta}I(m), \tag{B.153}$$

with

$$I(m) = \frac{1-m}{2} \log\left(\frac{1-m}{2}\right) + \frac{1+m}{2} \log\left(\frac{1+m}{2}\right) \tag{B.154}$$

the relative entropy of the Bernoulli measure on $\{-1, +1\}$ with parameter m with respect to the counting measure on $\{-1, +1\}$. $F_{\beta,h}(m)$ is the *free energy* at magnetisation m, consisting of an *energy term* $-\frac{1}{2}m^2 - hm$ and an *entropy term* $\frac{1}{\beta}I(m)$. See [18, Chapter 13] for more details.

Since

$$F'_{\beta,h}(m) = -m - h + \frac{1}{2\beta} \log\left(\frac{1+m}{1-m}\right), \qquad F''_{\beta,h}(m) = -1 - \frac{1}{\beta}\frac{m}{1-m^2}, \tag{B.155}$$

the stationary points of $F_{\beta,h}$ are the solutions to the equation

$$m = T_{\beta,h}(m), \qquad T_{\beta,h}(m) = \tanh[\beta(m + h)]. \tag{B.156}$$

Since

$$T'_{\beta,h}(m) = \beta\big[1 - T^2_{\beta,h}(m)\big], \tag{B.157}$$

$T_{\beta,h}$ is strictly increasing and has a unique inflection point at $m = -h$. Consequently, (B.156) has either one or three solutions. The latter occurs if and only if

$$\beta \in (\bar\beta_c, \infty) \qquad \text{and} \qquad h \in (0, h_c(\beta)), \tag{B.158}$$

where $\bar\beta_c = 1$ is the *critical inverse temperature* and $\bar h_c(\beta)$ is the *critical magnetic field*, i.e., the unique value of h for which $T_{\beta,h}$ touches the diagonal at a unique value of the magnetisation, say $-m(\beta)$. Clearly, $1 = \beta(1 - m^2(\beta))$, i.e.,

$$m(\beta) = \sqrt{1 - \beta^{-1}}, \tag{B.159}$$

and so $\bar{h}_c(\beta)$ solves the equation $T_{\beta,\bar{h}_c(\beta)}(-m(\beta)) = -m(\beta)$. Hence (see Fig. B.1)

$$\bar{h}_c(\beta) = m(\beta) - \frac{1}{2\beta}\log\left(\frac{1+m(\beta)}{1-m(\beta)}\right), \qquad \beta \geq 1. \tag{B.160}$$

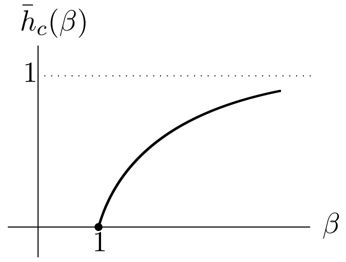


Figure B.1: Plot of $\beta \mapsto \bar{h}_c(\beta)$.

The range of parameters in (B.158) represents the *metastable regime* in which $F_{\beta,h}$ has a *double-well* shape and, in the limit as $n \to \infty$, the Gibbs measure μ_n in (B.4) has two phases given by the two minima of $F_{\beta,h}$: the *metastable phase* with magnetisation $\mathbf{m} < 0$ and the *stable phase* with magnetisation $\mathbf{s} > 0$. The unique *saddle point* in the gate $\mathcal{G}(\mathbf{m}, \mathbf{s})$ has magnetisation $\mathbf{t} < 0$ (see Fig. B.2).

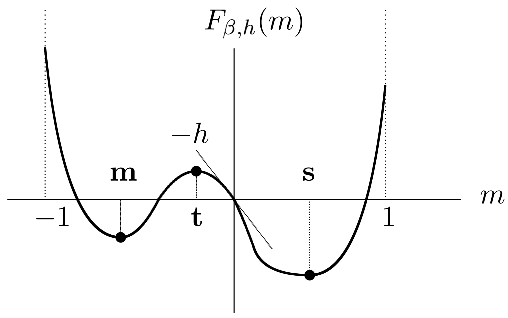


Figure B.2: Plot of $m \mapsto F_{\beta,h}(m)$ for β, h in the metastable regime.

Theorems B.7.1–B.7.2 can be found in Bovier and den Hollander [18, Chapter 13]. Here the notation is the same as the one in SectionB.1. Let $S_n[\mathbf{m}]$, $S_n[\mathbf{s}]$ denote the sets of configurations in S_n for which the magnetisation is closest to $\mathbf{m}$, $\mathbf{s}$, respectively.

Theorem B.7.1 **(Average crossover time).**
Subject to (B.158)*, uniformly in* $\sigma \in S_n[\mathbf{m}]$,

$$\mathbb{E}_\sigma\left[\tau_{S_n[\mathbf{s}]}\right] = [1 + o_n(1)]\,\frac{\pi}{1-\mathbf{t}}\sqrt{\frac{1-\mathbf{t}^2}{1-\mathbf{m}^2}}\,\frac{1}{\beta\sqrt{F''_{\beta,h}(\mathbf{m})[-F''_{\beta,h}(\mathbf{t})]}}\,\mathrm{e}^{\beta n[F_{\beta,h}(\mathbf{t})-F_{\beta,h}(\mathbf{m})]}. \tag{B.161}$$

Theorem B.7.2 **(Exponential law).**
Subject to (B.158)*, uniformly in* $\sigma \in S_n[\mathbf{m}]$,

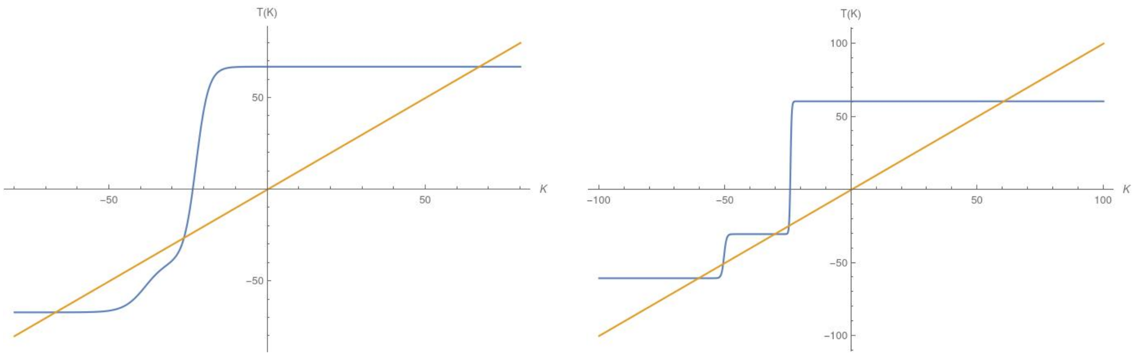
$$\mathbb{P}_\sigma\left(\tau_{S_n[\mathbf{s}]} > t\,\mathbb{E}_\sigma\left[\tau_{S_n[\mathbf{s}]}\right]\right) = [1 + o_n(1)]\,\mathrm{e}^{-t}, \qquad t \geq 0. \tag{B.162}$$

Fig. B.2 illustrates the setting: the average crossover time from $S_n[\mathbf{m}]$ to $S_n[\mathbf{s}]$ depends on the energy barrier $F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})$ and on the curvature of $F_{\beta,h}$ at $\mathbf{m}$ and $\mathbf{t}$. The crossover time is exponential on the scale of its average.

B.8 Appendix B: Examples with multiple metastable states

We provide examples of distributions and parameter choices (in the metastable regime) for which the model with disorder has multiple critical points. More specifically, we provide numerical evidence that, for $k \in \{2, 3, 4\}$, (B.59) can have any number of solutions in the set $\{3, 5 \ldots, 2k+1\}$. The cases with strictly more than 3 solutions present multiple minimal critical points, i.e. multiple metastable states.

B.8.1 Case k=2



(a) *3 critical points.* (b) *5 critical points.*

Figure B.3: $T_{\beta,h}$, $k = 2$.

- Figure B.3a: 3 critical points, parameters $a_1 = 77$, $a_2 = 45$, $\omega_1 = 0.688$, $h = 1740$, $\beta = 113\,\beta_c$.

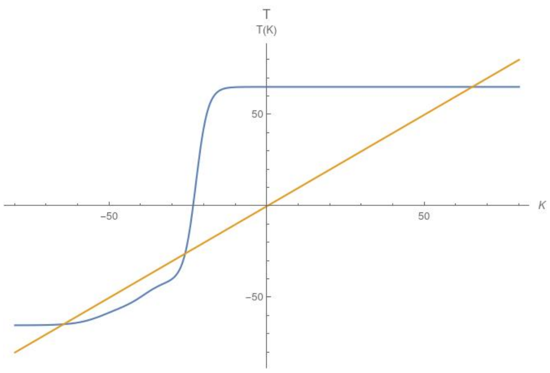
- Figure B.3b: 5 critical points, parameters $a_1 = 774$, $a_2 = 36.84$, $\omega_1 = 0.59$, $h = 1740$, $\beta = 131\,\beta_c$.

B.8.2 Case k=3

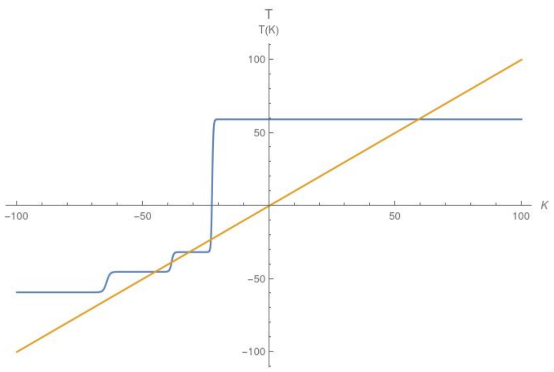
- Figure B.4a: 3 critical points, parameters $a_1 = 77$, $a_2 = 45$, $a_3 = 33.5$, $\omega_1 = 0.688$, $\omega_2 = 0.15$, $h = 1740$, $\beta = 113\,\beta_c$.

- Figure B.4b: 5 critical points, parameters $a_1 = 77$, $a_2 = 45$, $a_3 = 27$, $\omega_1 = 0.59$, $\omega_2 = 0.15$, $h = 1740$, $\beta = 113\,\beta_c$.

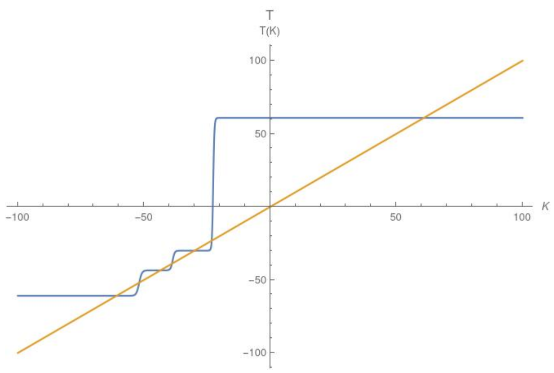
- Figure B.4c: 7 critical points, parameters $a_1 = 77$, $a_2 = 45$, $a_3 = 33.5$, $\omega_1 = 0.59$, $\omega_2 = 0.15$, $h = 1740$, $\beta = 113\,\beta_c$.



(a) *3 critical points.*

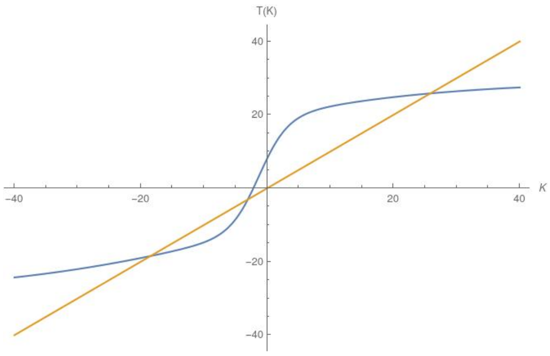


(b) *5 critical points.*

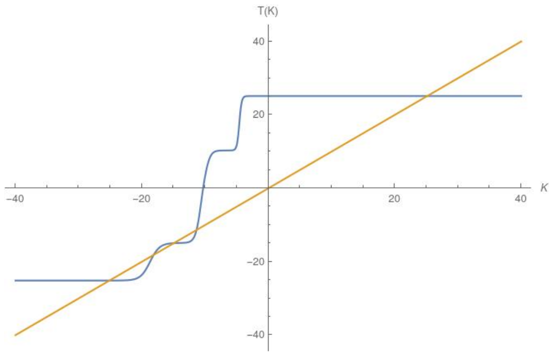


(c) *7 critical points.*

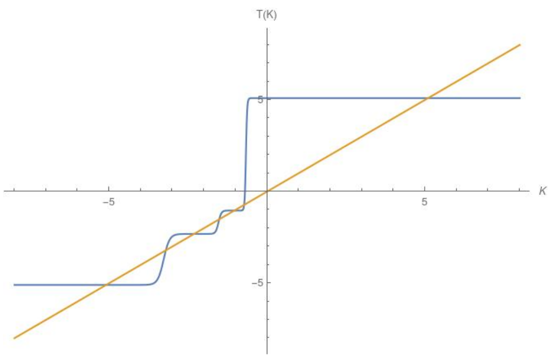
Figure B.4: $T_{\beta,h}$, $k = 3$.



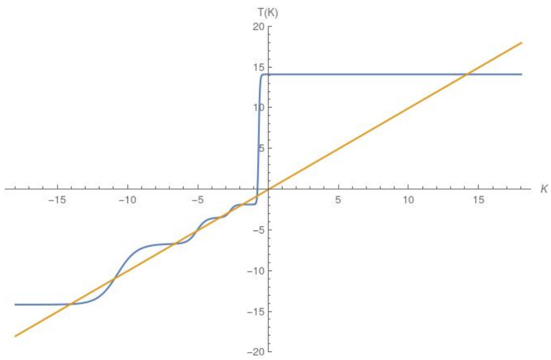
(a) *3 critical points.*



(b) *5 critical points.*



(c) *7 critical points.*



(d) *9 critical points.*

Figure B.5: $T_{\beta,h}$, $k = 4$.

B.8.3 Case k=4

- Figure B.5a: 3 critical points, parameters $a_1 = 12$, $a_2 = 16$, $a_3 = 139.5$, $a_4 = 24.5$, $\omega_1 = 0.474$, $\omega_2 = 0.22$, $\omega_3 = 0.111$, $h = 178$, $\beta = 3.8\,\beta_c$.

- Figure B.5b: 5 critical points, parameters $a_1 = 14$, $a_2 = 27$, $a_3 = 57$, $a_4 = 24.5$, $\omega_1 = 0.366$, $\omega_2 = 0.1$, $\omega_3 = 0.13$, $h = 262$, $\beta = 38.4\,\beta_c$.

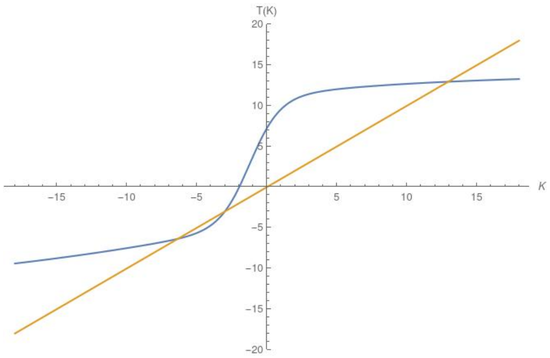
- Figure B.5c: 7 critical points, parameters $a_1 = 2.32$, $a_2 = 4.92$, $a_3 = 5$, $a_4 = 11.32$, $\omega_1 = 0.6$, $\omega_2 = 0.096$, $\omega_3 = 0.033$, $h = 7.6$, $\beta = 95.2\,\beta_c$.

- Figure B.5d: 9 critical points, parameters $a_1 = 12$, $a_2 = 16$, $a_3 = 50.5$, $a_4 = 24.5$, $\omega_1 = 0.474$, $\omega_2 = 0.22$, $\omega_3 = 0.111$, $h = 178$, $\beta = 63.2\,\beta_c$.

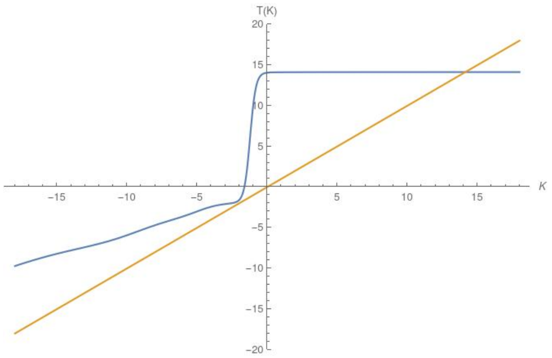
B.9 Appendix C: Example of $h_c(\beta)$ not increasing

We provide here an example of choice of the law of J for which the critical threshold $\beta \mapsto h_c(\beta)$ is not monotone increasing. This implies the possibility of a re-entrant metastable crossover.

For $k = 4$, pick $a_1 = 12$, $a_2 = 16$, $a_3 = 50.5$, $a_4 = 24.5$ and $\omega_1 = 0.474$, $\omega_2 = 0.22$, $\omega_3 = 0.111$. Take $h = 100$, and plot the function $K \mapsto T_{\beta,h}(K)$ varying β. For $\beta_1 = 4\,\beta_c = 0.00762336$ the system is metastable: $T_{\beta,h}$ intersects the diagonal three times (see Figure B.6a), which implies that $h < h_c(\beta_1)$. For $\beta_2 = 21\,\beta_c = 0.04002264 > \beta_1$ the system is not metastable: $T_{\beta,h}$ intersects the diagonal only once (see Figure B.6b), which implies that $h > h_c(\beta_2)$. This shows that $h_c(\beta)$ is not necessarily an increasing function of β.



(a) $h = 100$, $\beta = 0.00762336$.



(b) $h = 100$, $\beta = 0.04002264$.

Figure B.6: $T_{\beta,h}$, fixed h and law of the components of J, varying β.

B.10 Appendix D: Limit of the prefactor

Below Theorem B.1.3 we stated that the prefactor in (B.19) converges. For completeness, in this Appendix we compute its limit, although, as we mentioned after Theorem B.1.4, it is negligible because of the order of approximation of the exponent.

We focus first on γ_n. Recall notation in (B.10), (B.11) and (B.22). Then (B.100) can be written as

$$\begin{aligned}1+O(n^{-1}) &= \sum_{\ell\in[k]} \frac{a_\ell^2\omega_{\ell,n}(1-\mathbf{t}_{\ell,n})\exp\left[-2\beta\left(-a_\ell\left(\frac{a_\ell}{n}+K_n(\mathbf{t}_n)\right)-h\right)_+\right]}{\frac{\exp\left[-2\beta\left(-a_\ell\left(\frac{a_\ell}{n}+K_n(\mathbf{t}_n)\right)-h\right)_+\right]}{\beta(1+\mathbf{t}_{\ell,n})}-2\gamma_n} \\ &= \sum_{\ell\in[k]} \frac{a_\ell^2\omega_{\ell,n}\left(1-\tanh\left(\beta\left[a_\ell K_n(\mathbf{t}_n)+h\right]\right)\right)\exp\left[-2\beta\left(-a_\ell\left(\frac{a_\ell}{n}+K_n(\mathbf{t}_n)\right)-h\right)_+\right]}{\frac{\exp\left[-2\beta\left(-a_\ell\left(\frac{a_\ell}{n}+K_n(\mathbf{t}_n)\right)-h\right)_+\right]}{\beta(1+\tanh(\beta[a_\ell K_n(\mathbf{t}_n)+h]))}-2\gamma_n}.\end{aligned} \tag{B.163}$$

In the first equality we use (B.56) for $\mathbf{t}_n$, i.e., the approximation of the stationary points of F_n by the stationary points of $\bar{F}_n$. This makes $\mathbf{t}_{\ell,n}$ independent of ℓ, so that we can use the law of large numbers in the limit as $n\to\infty$. Thus, we obtain that γ_n converges to γ, the solution of the equation

$$\mathfrak{E}\left(\frac{J(1)^2(1+\tanh U)\,\mathrm{e}^{-2U_+}}{\frac{1}{\beta(1-\tanh U)}\,\mathrm{e}^{-2U_+}-2\gamma}\right)=1, \tag{B.164}$$

where $\mathfrak{E}$ denotes expectation with respect to $\mathcal{P}$ and $U=-\beta[J(1)\,K(\mathbf{t})+h]$, with $\mathbf{t}$ solving (B.57). Note that (B.164) is similar to [18, Eq. (14.4.14)].

We are left to find the limit of the determinants ratio. By (B.85),

$$\det\mathbb{A}_n(m)=\left(1-\sum_{\ell\in[k]}\beta\,a_\ell^2\,\omega_{\ell,n}[1-(m_\ell)^2]\right)\prod_{\ell'\in[k]}\frac{1}{\beta}\frac{\omega_{\ell',n}}{1-(m_{\ell'})^2}\left[1+O(n^{-1})\right]. \tag{B.165}$$

Using (B.56) for $m\in\{\mathbf{t}_n,\mathbf{m}_n\}$, we have

$$\sum_{\ell\in[k]}\beta\,a_\ell^2\,\omega_{\ell,n}[1-(m_{\ell,n})^2]=\sum_{\ell\in[k]}\beta\,a_\ell^2\,\omega_{\ell,n}\left[1-\tanh^2\left(\beta\left[a_\ell\sum_{\ell'\in[k]}a_{\ell'}\,\omega_{\ell',n}\,m_{\ell',n}+h\right]\right)\right]. \tag{B.166}$$

Using the law of large numbers as above and with the same notation, we find

$$\lim_{n\to\infty}\frac{[-\det(\mathbb{A}_n(\mathbf{t}_n))]}{\det(\mathbb{A}_n(\mathbf{m}_n))}=\frac{-1+\mathfrak{E}\left(\beta\,J(1)^2\left[1-\tanh^2\left[U(\mathbf{t})\right]\right]\right)}{1-\mathfrak{E}\left(\beta\,J(1)^2\left[1-\tanh^2\left[U(\mathbf{m})\right]\right]\right)}\prod_{\ell'\in[k]}\frac{1-(\mathbf{m}_{\ell'})^2}{1-(\mathbf{t}_{\ell'})^2}, \tag{B.167}$$

where $U(\mathbf{x})=-\beta(J(1)K(\mathbf{x})+h)$.

Appendix C

Preprint: Metastability of Glauber dynamics with inhomogeneous coupling disorder

This appendix reproduces exactly the content of the paper [20] with title "Metastability of Glauber dynamics with inhomogeneous coupling disorder", authored by Anton Bovier, Frank den Hollander, Saeda Marello, Elena Pulvirenti and Martin Slowik, and available as a preprint on *https://arxiv.org/abs/2209.09827, [math.PR], 2022* and not peer-reviewed yet. This work was summarised in Chapter 5.

C.1 Introduction

Over the last decade there has been increasing interest in metastability under Glauber dynamics of the Ising model with *random interactions*, in particular, of the Ising model on random graphs. Dommers [35] considered the case of random regular graphs, Dommers, den Hollander, Jovanovski, and Nardi [38] the configuration model, in both cases in finite volume and at low temperature. Mossel and Sly [59, 60] computed mixing times on sparse Erdős-Rényi random graphs and on random regular graphs, in both cases in finite volume and at high temperature. Recently, Can, van der Hofstad, and Kumagai [25] analysed mixing times on random regular graphs, in large volumes and at fixed temperature.

Metastability under Glauber dynamics of the Ising model on dense random graphs has so far only been studied for the Erdős-Rényi random graph with fixed edge retention probability, by den Hollander and Jovanovski [50] and by Bovier, Marello, and Pulvirenti [22]. In both papers, mean metastable exit times of the random model are compared to those of the standard Curie-Weiss model, in large volumes and at fixed temperature. In [50] the pathwise approach to metastability (see Olivieri and Vares [65]) was used to prove that mean metastable exit times are asymptotically equal to those of the Curie-Weiss model, multiplied by a random prefactor of polynomial order in the size of the system. The prefactor estimate was improved in [22] by using the potential-theoretic approach to metastability (see Bovier and den Hollander [18]), at the expense of losing generality in the initial distribution. Recently, Bovier, den Hollander and Marello [19] studied metastability under Glauber dynamics of the Ising model on the complete graph with random independent couplings in large volumes and at fixed temperature. In that model, the product structure of the couplings allows for lumping of states, and for combining the potential-theoretic approach with coarse-graining techniques, to obtain sharp estimates on mean metastable exit times.

The present paper extends the results for the Erdős–Rényi random graph to inhomogeneous dense random graphs and to more general random interactions. We compare the metastable behaviour of a class of spin systems whose Hamiltonian has random and conditionally independent coupling coefficients, called *quenched model*, with the corresponding *annealed model* in which the coupling coefficients are replaced by their conditional mean. More precisely, we prove that metastability of the annealed model implies, in large volumes and at fixed temperature, almost sure metastability of the quenched model with respect to the metastable sets of the annealed model. Moreover, assuming metastability of the annealed model, we consider the ratio between the mean hitting times of the quenched model and the annealed model, and estimate both its tail behaviour and its moments, again in large volumes and at fixed temperature.

As in [22], we follow the potential-theoretic approach to metastability, which allows us to estimate mean metastable exit times by estimating *capacities* and weighted sums of the equilibrium potential called *harmonic sums*. Estimates on the former can be obtained with the help of well-known variational principles, while estimates on the latter are more involved. See, for instance, Bianchi, Bovier and Ioffe [7] and [22], where long and model-dependent computations were needed to prove that the relevant contribution of the harmonic sum is localised around the starting metastable set. Schlichting and Slowik in [68], using an alternative definition of metastable sets, prove that localisation of the harmonic sum around the starting metastable set holds in large generality. Their work allows us to derive results for a large class of models. A second novelty of the present paper compared to [22] concerns the techniques that are used to prove concentration results. In [22], Talagrand's concentration inequality was used, while here we use McDiarmid's concentration inequality.

C.2 Model, results and methods

This section is structured as follows. In Section C.2.1, we introduce the model. In Section C.2.2, we define metastability, introduce relevant quantities, and state our main results. In Section C.2.3, we summarise our strategy and methods, and give an outline of the rest of the paper.

C.2.1 The model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an abstract probability space. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$ and $J = (J_{ij})_{1 \le i < j < \infty}$ be a triangular array of real random variables that are *conditionally independent* given $\mathcal{G}$ and *uniformly bounded*, i.e., there exists a $k_J \in (0, \infty)$ such that $|J_{ij}| \le k_J$ $\mathbb{P}$-a.s. for all $1 \le i < j < \infty$. We write $\mathbb{P}_{\mathcal{G}}[\,\cdot\,]$ to denote a regular conditional distribution for J given $\mathcal{G}$ (which exists because J is a sequence of real random variables; see Chow and Teicher [27, p. 218]). Write $\mathbb{E}$ to denote the expectation with respect to $\mathbb{P}$, and $\mathbb{E}_{\mathcal{G}}$ and $\mathbb{V}\mathrm{ar}_{\mathcal{G}}$ to denote expectation and variance with respect to $\mathbb{P}_{\mathcal{G}}$.

Given $2 \le N \in \mathbb{N}$, consider the Ising Hamiltonian with random couplings of the form

$$H_N(\sigma) \coloneqq -\frac{1}{N} \sum_{\substack{i,j=1 \\ i<j}}^{N} J_{ij} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i, \qquad \sigma \in \mathcal{S}_N, \tag{C.1}$$

with $h \in \mathbb{R}$ the magnetic field and $\mathcal{S}_N := \{-1, 1\}^N$ the set of spin configurations. The corresponding Gibbs measure on $\mathcal{S}_N$ is denoted by

$$\mu_N(\sigma) := Z_N^{-1}\, \mathrm{e}^{-\beta H_N(\sigma)}, \qquad \sigma \in \mathcal{S}_N, \tag{C.2}$$

with $\beta \in (0, \infty)$ the inverse temperature and Z_N the normalizing partition function. The spin configurations evolve in time as a continuous-time Markov chain $(\Sigma_N(t))_{t\geq 0}$ with state space $\mathcal{S}_N$ and Glauber-Metropolis transition rates given by

$$p_N(\sigma, \sigma') := \begin{cases} \exp\big(-\beta[H_N(\sigma') - H_N(\sigma)]_+\big), & \text{if } \sigma \sim \sigma', \\ 0, & \text{otherwise}, \end{cases} \qquad \sigma, \sigma' \in \mathcal{S}_N, \tag{C.3}$$

where $\sigma \sim \sigma'$ means that σ' is obtained from σ by a flip of a single spin. The associated generator L_N acts on bounded functions $f : \mathcal{S}_N \to \mathbb{R}$ as

$$(L_N f)(\sigma) := \sum_{\sigma' \in \mathcal{S}_N} p_N(\sigma, \sigma')\big(f(\sigma') - f(\sigma)\big), \qquad \sigma \in \mathcal{S}_N. \tag{C.4}$$

Note that the stochastic process $(\Sigma_N(t))_{t\geq 0}$ is irreducible and reversible with respect to the Gibbs measure μ_N. We denote by P_ν^N the law of the Markov chain $(\Sigma_N(t))_{t\geq 0}$ on $\mathcal{D}([0,\infty), \mathcal{S}_N)$ (the space of $\mathcal{S}_N$-valued càdlàg functions on $[0, \infty)$) with initial distribution ν. The corresponding expectation is denoted by E_ν^N. If the initial distribution is concentrated on a single configuration $\sigma \in \mathcal{S}_N$, then we write P_σ^N and E_σ^N, respectively. For a non-empty subset $\mathcal{A} \subset \mathcal{S}_N$, let $\tau_\mathcal{A}^N$ be the first return time to $\mathcal{A}$, i.e.,

$$\tau_\mathcal{A}^N \equiv \tau_\mathcal{A}^N\big((\Sigma_N(t))_{t\geq 0}\big) := \inf\{t > 0 : \Sigma_N(t) \in \mathcal{A}, \Sigma_N(t-) \notin \mathcal{A}\}. \tag{C.5}$$

Our main objective is to compare the evolution of the model with Hamiltonian H_N and the model with Hamiltonian $\tilde{H}_N$ defined by

$$\widetilde{H}_N(\sigma) := \mathbb{E}_\mathcal{G}[H_N(\sigma)] = -\frac{1}{N} \sum_{\substack{i,j=1\\ i<j}}^N \mathbb{E}_\mathcal{G}[J_{ij}]\sigma_i\sigma_j - h\sum_{i=1}^N \sigma_i, \quad \sigma \in \mathcal{S}_N, \quad \mathbb{P}\text{-a.s.} \tag{C.6}$$

Throughout the paper, we use the superscript $\sim$ to denote quantities that refer to the model defined in terms of $\widetilde{H}_N$. For instance, $((\widetilde{\Sigma}_N(t))_{t\geq 0}, \widetilde{\mathrm{P}}_\sigma^N : \sigma \in \mathcal{S}_N)$ denotes the continuous-time Markov chain with transitions rates $(\tilde{p}_N(\sigma, \sigma'))_{\sigma,\sigma' \in \mathcal{S}_N}$, which is reversible with respect to $\tilde{\mu}_N$, where both the transition rates and the Gibbs measure are defined as in (C.1) and (C.2), but with H_N replaced by $\widetilde{H}_N$. For lack of better names and with abuse of terminology, we refer to the models defined in terms of H_N and $\widetilde{H}_N$ as the *quenched model* and the *annealed model*, respectively.

In the sequel, we provide a list of motivating examples for which the results stated later hold. For this purpose, consider two sequences

$$(A_{ij})_{1\leq i<j<\infty}, \qquad (P_{ij})_{1\leq i<j<\infty}, \tag{C.7}$$

of triangular arrays with $|A_{ij}| \leq k_J$ and $P_{ij} \in (0, 1)$ for $i, j \in \mathbb{N}$ with $i < j$, and let $\mathcal{G} := \sigma(A_{ij}, P_{ij} : 1 \leq i < j \leq \infty)$ be the σ-algebra generated by these sequences. Moreover, let

$$(U_{ij})_{1\leq i<j<\infty} \tag{C.8}$$

be a triangular array of i.i.d. random variables distributed uniformly in $(0,1)$. Define

$$J_{ij} \coloneqq A_{ij}B_{ij}, \qquad B_{ij} \coloneqq \mathbb{1}_{\{U_{ij}\leq P_{ij}\}}, \qquad 1 \leq i < j < \infty. \tag{C.9}$$

Note that $(J_{ij})_{1\leq i<j<\infty}$ and $(B_{ij})_{1\leq i<j<\infty}$ are triangular arrays of *conditionally independent* random variables given $\mathcal{G}$. In particular, B_{ij} are Bernoulli random variables with mean P_{ij}.

Example C.2.1 (Ising model on the Erdős–Rényi random graph)**.** By choosing $A_{ij} \coloneqq 1$ and $P_{ij} \coloneqq p \in (0,1]$ for $1 \leq i < j < \infty$, $\mathcal{G}$ becomes the set $\{\emptyset, \Omega\}$ and H_N in (C.1) becomes the Hamiltonian of the Ising model on the Erdős–Rényi random graph with edge retention probability p, known as the *randomly diluted Curie-Weiss model.* Its metastable behaviour was studied in [50] and [22]. In this case the annealed model is the Curie-Weiss model.

Example C.2.2 (Ising model on inhomogeneous random graphs)**.** By taking $A_{ij} \coloneqq 1$ for $1 \leq i < j < \infty$, H_N in (C.1) becomes the Hamiltonian of the Ising model on an inhomogeneous random graph, in which an edge (ij) is present with probability P_{ij}. Of particular interest is the case $P_{ij} = V_iV_j$ with $(V_i)_{i\in\mathbb{N}}$ a sequence of i.i.d. random variables with support in $(0,1)$, known as the Ising model on the *Chung-Lu random graph* [28]. The metastable behaviour of the corresponding annealed model was studied in [19] for the case where the random variables V_i have finite support.

Example C.2.3 (Randomly diluted Hopfield model)**.** Given $n \in \mathbb{N}$ random patterns $\xi^1, \dots, \xi^n$, with $\xi^k = (\xi^k_i)_{i\in\mathbb{N}}$ and $\xi^k_i \in [-1,1]$ for $1 \leq k \leq n$, set $A_{ij} \coloneqq \sum_{k=1}^n \xi^k_i \xi^k_j$. By taking $P_{ij} \equiv p \in (0,1)$ for $1 \leq i < j < \infty$, H_N in (C.1) becomes the Hamiltonian of a *Hopfield model* in which the interaction coefficients are *randomly diluted* by i.i.d. Bernoulli random variables with mean p. See Bovier and Gayrard [16] for a review on the Hopfield model. The metastable behaviour of the annealed model, i.e., the undiluted Hopfield model, was studied by an der Heiden in [2] in a restricted (β,h)-regime. We plan to address the metastable behaviour in a more general (β,h)-regime in a future paper.

C.2.2 Metastability and main results

Before stating our main results, we recall the definition of metastable Markov chains and metastable sets put forward in Schlichting and Slowik [68, Definition 1.1].

Definition C.2.4 (ρ_N-metastability and metastable sets)**.** *For $\rho_N > 0$ and $2 \leq K \in \mathbb{N}$, let $\{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}$ be a set of subsets of $\mathcal{S}_N$ such that $\mathcal{M}_{i,N} \cap \mathcal{M}_{j,N} = \emptyset$ for all $1 \leq i \neq j \leq K$. The Markov process $(\Sigma_N(t))_{t\geq 0}$ is called ρ_N-*metastable *with respect to $\{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}$ when*

$$K \frac{\max_{j\in\{1,\dots,K\}} \mathrm{P}^N_{\mu_N|\mathcal{M}_{j,N}}\left[\tau^N_{\mathcal{M}_N\setminus\mathcal{M}_{j,N}} < \tau^N_{\mathcal{M}_{j,N}}\right]}{\min_{\mathcal{X}\subset\mathcal{S}_N\setminus\mathcal{M}_N} \mathrm{P}^N_{\mu_N|\mathcal{X}}\left[\tau^N_{\mathcal{M}_N} < \tau^N_{\mathcal{X}}\right]} \leq \rho_N \ll 1, \tag{C.10}$$

where $\mathcal{M}_N \coloneqq \bigcup_{i=1}^K \mathcal{M}_{i,N}$ and, for a non-empty set $\mathcal{X} \subseteq \mathcal{S}_N$, $\mu_N|\mathcal{X}$ denotes the Gibbs measure μ_N conditioned on the set $\mathcal{X}$.

Remark C.2.5. The advantage of this definition compared to the one given in Bovier and den Hollander [18, Definition 8.2] is twofold: it allows for direct control of $\ell^p(\mu_N)$-norms of functions, and does not depend on the cardinality of the state space. For a more detailed comparison of the two definitions of metastability we refer to [68, Remark 1.2].

For fixed $2 \leq K \in \mathbb{N}$ and $k_1 > 0$, define

$$\widetilde{\Omega}_{\mathrm{meta}}(N) := \Big\{\omega \colon \exists \{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}(\omega) \text{ non-empty disjoint subsets of } \mathcal{S}_N \\ \text{s.t. } (\widetilde{\Sigma}_N(t))_{t\geq 0}(\omega) \text{ is } \mathrm{e}^{-k_1 N}\text{-metastable w.r.t. } \{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}(\omega)\Big\}, \tag{C.11}$$

i.e., the event that the Markov chain $(\widetilde{\Sigma}_N(t))_{t\geq 0}$ is $\tilde{\rho}_N$-metastable with respect to some $\{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}$, where we abbreviate

$$\tilde{\rho}_N := \mathrm{e}^{-k_1 N}. \tag{C.12}$$

Remark C.2.6. Note that both $\widetilde{\Omega}_{\mathrm{meta}}(N)$ and $\{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}$ – playing the role of *metastable sets* – are $\mathcal{G}$-measurable, because they are defined in terms of the annealed Hamiltonian $\widetilde{H}_N$.

In our main results we impose the following assumption on the annealed model.

Assumption C.2.7 (Metastability of the annealed model)**.** *For some* $(\beta, h) \in (0,\infty) \times \mathbb{R}$*, the following holds for the Markov chain* $(\widetilde{\Sigma}_N(t))_{t\geq 0}$ *of the annealed model. There exist* $2 \leq K \in \mathbb{N}$ *and* $k_1 > 0$ *such that,*

$$\mathbb{P}\Big[\liminf_{N\to\infty} \widetilde{\Omega}_{\mathrm{meta}}(N)\Big] = 1, \tag{C.13}$$

where $\widetilde{\Omega}_{\mathrm{meta}}(N)$ *is defined in* (C.11) *and depends on* K *and* k_1*.*

Remark C.2.8. Assertion of Assumption C.2.7 can be rephrased as follows: $\mathbb{P}$-a.s., there exists a finite random variable $N_0 \in \mathbb{N}$ and a sequence $(\{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\})_{N\geq N_0}$ of K non-empty mutually disjoint subsets of $\mathcal{S}_N$ (possibly depending on ω) such that for any $N \geq N_0$ the process $(\widetilde{\Sigma}_N(t))_{t\geq 0}$ is $\tilde{\rho}_N := \mathrm{e}^{-k_1 N}$-metastable with respect to $\{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}$.

Remark C.2.9. Let us illustrate in the case of Example C.2.1 how to identify candidates of metastable sets. It is well know (cf. [11, Section 3.5]) that, for any $\beta \in (0,\infty)$ and $h \in \mathbb{R}$,

$$\lim_{N\to\infty} \frac{1}{\beta N} \log \widetilde{Z}_N = -\inf_{x\in[-1,1]} \widetilde{F}_{\beta,h}(x), \tag{C.14}$$

where $\widetilde{F}_{\beta,h} : [-1,1] \to \mathbb{R}$ denotes the *free energy* per vertex of the annealed model given by

$$\widetilde{F}_{\beta,h}(x) := -\frac{1}{2}x^2 - hx + \frac{1}{\beta}\Big(\frac{1-x}{2}\log\frac{1-x}{2} + \frac{1+x}{2}\log\frac{1+x}{2}\Big) + \log 2. \tag{C.15}$$

In particular, for any $\beta > \beta_c := 1$ and $h \in (-h_c(\beta), h_c(\beta))$, where the critical strength of the magnetic field is given by

$$h_c(\beta) := \sqrt{1-\beta^2} - \frac{1}{2\beta}\log(\beta(1+\sqrt{1-1/\beta})^2),$$

the free energy $\widetilde{F}_{\beta,h}$ admits two local minima $m_1, m_2 \in (-1,1)$. For $N \in \mathbb{N}$, let $m_{1,N}$ and $m_{2,N}$ be the closest point in $\{-1, -1+2/N, \dots, 1-2/N, 1\}$ to m_1 and m_2, respectively. Define the sets $\mathcal{M}_{1,N}, \mathcal{M}_{2,N} \subset \mathcal{S}_N$ as the (set-valued) pre-image of the empirical magnetization $\mathcal{S}_N \ni \sigma \mapsto m_N(\sigma) := \frac{1}{N}\sum_{i=1}^N \sigma_i$, i.e.,

$$\mathcal{M}_{1,N} := m_N^{-1}(m_{1,N}) \qquad \text{and} \qquad \mathcal{M}_{2,N} := m_N^{-1}(m_{2,N}). \tag{C.16}$$

By using arguments similar to those given in [68, Lemma 4.1], it follows that $\{\mathcal{M}_{1,N}, \mathcal{M}_{2,N}\}$ forms a pair of metastable sets in the sense of Definition C.2.4.

For fixed $N \in \mathbb{N}$, given the metastable sets $\{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}$, we can decompose the state space $\mathcal{S}_N$ into the domains of attraction with to respect the dynamics of the annealed model. More precisely, by following [68, Definition 1.4], within the event $\widetilde{\Omega}_{\mathrm{meta}}(N)$ the metastable sets $\{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}$ give rise to a *metastable partition* $\{\mathcal{S}_{1,N}, \dots, \mathcal{S}_{K,N}\}$ of the state space $\mathcal{S}_N$ such that

$$\mathcal{M}_{i,N} \subseteq \mathcal{S}_{i,N} \subset \mathcal{V}_{i,N}, \qquad i \in \{1, \dots, K\}. \tag{C.17}$$

The *local valley* $\mathcal{V}_{i,N}$ around the metastable set $\mathcal{M}_{i,N}$ is defined as

$$\mathcal{V}_{i,N} := \mathcal{M}_{i,N} \cup \Big\{ \sigma \in \mathcal{M}_N^c : \widetilde{\mathrm{P}}_\sigma^N \Big[\tau^N_{\mathcal{M}_{i,N}} < \tau^N_{\mathcal{M}_N \setminus \mathcal{M}_{i,N}} \Big] \geq \max_{j \neq i} \widetilde{\mathrm{P}}_\sigma^N \Big[\tau^N_{\mathcal{M}_{j,N}} < \tau^N_{\mathcal{M}_N \setminus \mathcal{M}_{j,N}} \Big] \Big\}, \tag{C.18}$$

where we recall that $\mathcal{M}_N := \bigcup_{i=1}^K \mathcal{M}_{i,N}$.

Our first theorem says that, subject to Assumption C.2.7, $(\Sigma_N(t))_{t \geq 0}$ also exhibits metastable behaviour in the sense of Definition C.2.4.

Theorem C.2.10 (Metastability)**.** *Suppose that $(\beta, h) \in (0, \infty) \times \mathbb{R}$ satisfies Assumption C.2.7. Then, for any $c_0 \in (0, k_1)$, the event*

$$\Omega_{\mathrm{meta}}(N) := \Big\{ \omega \in \widetilde{\Omega}_{\mathrm{meta}}(N) \colon (\Sigma_N(t))_{t \geq 0}(\omega) \text{ is } \mathrm{e}^{-c_0 N}\text{-metastable w.r.t. } \{\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}\}(\omega) \Big\} \tag{C.19}$$

satisfies

$$\mathbb{P}\Big[\liminf_{N \to \infty} \Omega_{\mathrm{meta}}(N) \Big] = 1. \tag{C.20}$$

Let the following assumption hold for all $N \in \mathbb{N}$, in $\widetilde{\Omega}_{\mathrm{meta}}(N)$.

Assumption C.2.11 (Non-degeneracy)**.** *Label the metastable sets $\mathcal{M}_{1,N}, \dots, \mathcal{M}_{K,N}$ in such a way that they are ordered decreasingly according to their weights under the Gibbs measure of the annealed model, i.e.,*

$$\tilde{\mu}_N[\mathcal{M}_{1,N}] \geq \tilde{\mu}_N[\mathcal{M}_{2,N}] \geq \dots \geq \tilde{\mu}_N[\mathcal{M}_{K,N}], \qquad \mathbb{P}\text{-a.s.}, \forall N \in \mathbb{N}, \tag{C.21}$$

and assume that, for some $i \in \{2, \dots, K\}$, there exists a $k_2 \in (0, \infty)$ such that

$$\tilde{\mu}_N[\mathcal{S}_{j,N}] \leq \mathrm{e}^{-k_2 N} \tilde{\mu}_N[\mathcal{S}_{i,N}], \qquad \mathbb{P}\text{-a.s.}, \forall j \in \{i+1, \dots, K\},\ N \in \mathbb{N}. \tag{C.22}$$

Remark C.2.12. The non-degeneracy condition in (C.22) can be relaxed by replacing $\mathrm{e}^{-k_2 N}$ with some $\tilde{\delta}_N$ satisfying $\mathrm{e}^{-k_2 N} \leq \tilde{\delta}_N < \mathrm{e}^{-c\sqrt{N}}$ for some sufficiently large $c \in (0, \infty)$. The technical reasons can be found in the proofs in Section C.5.

In view of Assumption C.2.11, fix $i \in \{2, \dots, K\}$ such that (C.22) holds and put, for $N \in \mathbb{N}$, $\mathbb{P}$-a.s. in $\widetilde{\Omega}_{\mathrm{meta}}(N)$,

$$\mathcal{A}_N := \mathcal{M}_{i,N}, \qquad \mathcal{B}_N := \bigcup_{j=1}^{i-1} \mathcal{M}_{j,N}. \tag{C.23}$$

Note that $\mathcal{B}_N$ is the union of all metastable sets with weight not smaller than the weight of $\mathcal{A}_N$.

Before proceeding, we define, for $N \in \mathbb{N}$ and non-empty disjoint sets $\mathcal{A}, \mathcal{B} \subset \mathcal{S}_N$, the so-called *last-exit biased distribution* on $\mathcal{A}$ for the transition from $\mathcal{A}$ to $\mathcal{B}$ by

$$\nu_{\mathcal{A},\mathcal{B}}(\sigma) \equiv \nu^N_{\mathcal{A},\mathcal{B}}(\sigma) = \frac{\mu_N(\sigma)\, \mathrm{P}^N_\sigma\big[\tau^N_{\mathcal{B}} < \tau^N_{\mathcal{A}}\big]}{\sum_{\sigma\in\mathcal{A}} \mu_N(\sigma)\, \mathrm{P}^N_\sigma\big[\tau^N_{\mathcal{B}} < \tau^N_{\mathcal{A}}\big]}, \qquad \sigma \in \mathcal{A}. \tag{C.24}$$

This distribution plays an essential role in the potential-theoretic approach to metastability, as can be seen in (C.29) below.

We are now ready to state our second theorem, in which we compare the mean hitting time of $\mathcal{B}_N$ for the Markov chain $(\Sigma_N(t))_{t\geq 0}$ starting from the set $\mathcal{A}_N$ according to the distribution $\nu_{\mathcal{A}_N,\mathcal{B}_N}$ with the corresponding quantity for the Markov chain $(\widetilde{\Sigma}_N(t))_{t\geq 0}$. Under the regular conditional distribution $\mathbb{P}_{\mathcal{G}}$, we obtain for the ratio of these metastable hitting times estimates both on its tail behaviour and on its moments.

Theorem C.2.13. *Suppose that $(\beta, h) \in (0,\infty) \times \mathbb{R}$ and $i \in \{2,\dots,K\}$ satisfy Assumptions C.2.7 and C.2.11. Set*

$$\alpha_N := \frac{\beta^2}{2N^2} \sum_{\substack{i,j=1\\ i<j}}^{N} \mathbb{V}\mathrm{ar}_{\mathcal{G}}[J_{ij}]. \tag{C.25}$$

(i) For $t \geq 0$, $\mathbb{P}$-a.s.,

$$\lim_{N\to\infty} \mathbb{P}_{\mathcal{G}}\left[\mathrm{e}^{-t-\alpha_N} \leq \frac{\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]}{\tilde{\mathrm{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tilde{\tau}^N_{\mathcal{B}_N}\big]} \leq \mathrm{e}^{+t+2\alpha_N}\right] \geq 1 - 4\,\mathrm{e}^{-t^2/(2\beta k_J)^2}. \tag{C.26}$$

(ii) For any $q \geq 1$ and $c \in (0,\infty)$, let

$$\Omega_{q,c}(N) :=$$
$$\left\{\omega\colon \mathrm{e}^{-\alpha_N}\Big(1 - \tfrac{c}{\sqrt{N}}\Big) \leq \frac{\mathbb{E}_{\mathcal{G}}\Big[\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]^q\Big]^{1/q}(\omega)}{\tilde{\mathrm{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tilde{\tau}^N_{\mathcal{B}_N}\big](\omega)} \leq \mathrm{e}^{4q\alpha_N}\Big(1 + \tfrac{c}{\sqrt{N}}\Big)\right\}. \tag{C.27}$$

Then, for any $q \geq 1$ there exists $c_1 \in (0,\infty)$ such that

$$\mathbb{P}\left[\liminf_{N\to\infty} \Omega_{q,c_1}(N)\right] = 1. \tag{C.28}$$

Remark C.2.14. (a) Since the random variables $(J_{ij})_{1\leq i<j<\infty}$ are assumed to be uniformly bounded, it follows that $\alpha_N = O(1)$.

(b) Assumption C.2.11 ensures that the metastable sets $\mathcal{M}_{j,N}$, $j \in \{i+1,\dots,K\}$, are not relevant for the analysis of the crossing times from $\mathcal{A}_N$ to $\mathcal{B}_N$.

C.2.3 Methods and outline

Key notions from the potential-theoretic approach to metastability

To prove Theorems C.2.10 and C.2.13, we crucially rely on potential theory, which allows us to express probabilistic objects of interest in terms of solutions of certain boundary value problems. It is well-known (see e.g. [18, Corollary 7.11]) that, for any $N \in \mathbb{N}$ and any non-empty disjoint $\mathcal{A}, \mathcal{B} \subset \mathcal{S}_N$, the mean hitting time of $\mathcal{B}$ starting from the last-exit biased distribution $\nu^N_{\mathcal{A},\mathcal{B}}$ on $\mathcal{A}$ is given by

$$\mathrm{E}^N_{\nu_{\mathcal{A},\mathcal{B}}}[\tau^N_{\mathcal{B}}] = \frac{\|h^N_{\mathcal{A},\mathcal{B}}\|_{\mu_N}}{\mathrm{cap}_N(\mathcal{A},\mathcal{B})}, \tag{C.29}$$

where $\|h^N_{\mathcal{A},\mathcal{B}}\|_{\mu_N}$ denotes the $\ell^1(\mu_N)$-norm of the *equilibrium potential* $h^N_{\mathcal{A},\mathcal{B}}$ of the pair $(\mathcal{A},\mathcal{B})$, i.e., the function $h^N_{\mathcal{A},\mathcal{B}} : \mathcal{S}_N \to [0,1]$ that is the unique solution of the boundary value problem

$$\begin{cases} (L_N f)(\sigma) = 0, & \sigma \in \mathcal{S}_N \setminus (\mathcal{A} \cup \mathcal{B}), \\ f(\sigma) = \mathbb{1}_{\mathcal{A}}(\sigma), & \sigma \in \mathcal{A} \cup \mathcal{B}. \end{cases} \tag{C.30}$$

Note that the equilibrium potential has a natural interpretation in terms of *hitting probabilities*, namely, $h^N_{\mathcal{A},\mathcal{B}}(\sigma) = \mathrm{P}^N_\sigma[\tau^N_{\mathcal{A}} < \tau^N_{\mathcal{B}}]$, for all $\sigma \in \mathcal{S}_N \setminus (\mathcal{A} \cup \mathcal{B})$. The *capacity* $\mathrm{cap}_N(\mathcal{A},\mathcal{B})$ of the pair $(\mathcal{A},\mathcal{B})$ is defined by

$$\mathrm{cap}_N(\mathcal{A},\mathcal{B}) \coloneqq \sum_{\sigma \in \mathcal{A}} \mu_N(\sigma)\, \mathrm{P}^N_\sigma[\tau^N_{\mathcal{B}} < \tau^N_{\mathcal{A}}]. \tag{C.31}$$

From this definition it is clear that

$$\mathrm{P}^N_{\mu_N|\mathcal{A}}[\tau^N_{\mathcal{B}} < \tau^N_{\mathcal{A}}] = \frac{\mathrm{cap}_N(\mathcal{A},\mathcal{B})}{\mu_N[\mathcal{A}]}, \tag{C.32}$$

where $\mu_N|\mathcal{A}$ denotes the Gibbs measure μ_N conditioned on the set $\mathcal{A}$.

Capacities can be expressed in terms of variational principles that are very useful to obtain upper und lower bounds (see [18, Section 7.3] for more details). Upper bounds are obtained by using the *Dirichlet principle*, which states that

$$\mathrm{cap}_N(\mathcal{A},\mathcal{B}) = \inf\{\mathcal{E}_N(f) \colon f \in \mathcal{H}^N_{\mathcal{A},\mathcal{B}}\}. \tag{C.33}$$

Here, $\mathcal{H}^N_{\mathcal{A},\mathcal{B}}$ denotes the set of all functions from $\mathcal{S}_N$ to $\mathbb{R}$ that are equal to 1 on $\mathcal{A}$ and 0 on $\mathcal{B}$, and

$$\mathcal{E}_N(f) \coloneqq \langle f, -L_N f\rangle_{\mu_N} = \frac{1}{2} \sum_{\sigma,\sigma' \in \mathcal{S}_N} \mu_N(\sigma) p_N(\sigma,\sigma') \big(f(\sigma) - f(\sigma')\big)^2 \tag{C.34}$$

is the *Dirichlet form*. We recall that the transition rates p_N are defined in (C.3).

Lower bounds are obtained via the *Thomson principle*, which states that

$$\mathrm{cap}_N(\mathcal{A},\mathcal{B}) = \sup\Big\{\frac{1}{\mathcal{D}_N(\varphi)} \colon \varphi \in \mathcal{U}^N_{\mathcal{A},\mathcal{B}}\Big\} = \Big(\inf\{\mathcal{D}_N(\varphi) \colon \varphi \in \mathcal{U}^N_{\mathcal{A},\mathcal{B}}\}\Big)^{-1}, \tag{C.35}$$

where $\mathcal{U}^N_{\mathcal{A},\mathcal{B}}$ is the space of all unit antisymmetric $\mathcal{AB}$-flows $\varphi : \mathcal{S}_N \times \mathcal{S}_N \to \mathbb{R}$, and

$$\mathcal{D}_N(\varphi) \coloneqq \frac{1}{2} \sum_{\substack{\sigma,\sigma' \in \mathcal{S}_N \\ \sigma \sim \sigma'}} \frac{\varphi(\sigma,\sigma')^2}{\mu_N(\sigma)\, p_N(\sigma,\sigma')}. \tag{C.36}$$

Strategy of proofs

The proof of Theorem C.2.10 relies on Definition C.2.4 and on (C.32), together with an application of the Dirichlet principle in combination with a comparison of the quenched Hamiltonian H_N and the annealed Hamiltonian $\widetilde{H}_N$ on a particular event of high probability.

We prove Theorem C.2.13(i) by combining concentration inequalities for the logarithm of the mean hitting time $\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}[\tau^N_{\mathcal{B}_N}]$ of the quenched model with bounds on the distance between the (conditional on $\mathcal{G}$) mean of that logarithm and the logarithm of the mean hitting time $\tilde{\mathrm{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}[\tilde{\tau}^N_{\mathcal{B}_N}]$ of the annealed model. Estimates of the latter type, comparing conditional means with average means, will be called *annealed estimates*. The results in Theorem C.2.13(ii) are annealed estimates as well. In view of (C.29), estimates on the mean hitting time $\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}[\tau^N_{\mathcal{B}_N}]$, or on its logarithm, will follow once we have separately proven corresponding estimates for both $Z_N \operatorname{cap}_N(\mathcal{A}_N, \mathcal{B}_N)$ and $Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}$.

To prove concentration inequalities for the quantities $\log[Z_N \operatorname{cap}_N(\mathcal{A}_N, \mathcal{B}_N)]$ and $\log[Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}]$, we use a conditional version of McDiarmid's bounded differences inequality (see Proposition C.7.1 below). This strategy for proving concentration is different from the one used in [22], where Talagrand's concentration inequality was used. The advantage of McDiarmid over Talagrand is twofold. First, McDiarmid's inequality provides exact constants. Second, it does not require convexity of the map $J \mapsto \log(Z_N \operatorname{cap}_N(\mathcal{A}_N, \mathcal{B}_N))$, which is crucial because we do not know how to prove convexity.

Estimates on capacities for Theorem C.2.13 are proven by using the Dirichlet principle and the Thomson principle, and do not require any assumption on metastability. Finding estimates on the equilibrium potential, however, is more involved. We use a result that is similar to [68, Theorem 1.7] (Proposition C.5.2 below), for which Assumption C.2.11 is required, together with the same comparison of the Hamiltonians H_N and $\widetilde{H}_N$ that is used in the proof of Theorem C.2.10, both holding with high probability. We emphasise that the constants appearing in our statements may depend on the parameters of the model.

Outline

The remainder of the paper is organised as follows. In Section C.3 we provide the proof of Theorem C.2.10 on metastability of the quenched model. In Section C.4 we provide estimates on capacities. Section C.5 is devoted to stating and proving estimates on weighted sums of the equilibrium potential, called harmonic sums. In Section C.6 we prove Theorem C.2.13 by using the results of the previous sections. Appendix C.7 contains the conditional version of the McDiarmid's inequality that is used in the paper.

C.3 Metastability

Before proving Theorem C.2.10, we provide in Lemma C.3.1 a comparison of the quenched Hamiltonian H_N and the annealed Hamiltonian $\widetilde{H}_N$. This lemma will be used both in the proof of Theorem C.2.10 below and in Section C.5, where we deal with estimates on the equilibrium potential.

Given a positive real sequence $(a_N)_{N\in\mathbb{N}}$, let

$$\Xi(a_N) := \Big\{ \max_{\sigma\in\mathcal{S}_N} |H_N(\sigma) - \widetilde{H}_N(\sigma)| < a_N \Big\} \subset \Omega, \qquad 2 \le N \in \mathbb{N}, \tag{C.37}$$

denote the event that, uniformly in $\sigma \in \mathcal{S}_N$, H_N differs from $\widetilde{H}_N$ by at most a_N. On the event $\Xi(a_N)$ we have control on the difference between the quantities determining $(\Sigma_N(t))_{t\geq 0}$ and $(\widetilde{\Sigma}_N(t))_{t\geq 0}$. Moreover, for suitably chosen sequences $(a_N)_{N\in\mathbb{N}}$, the event $\Xi(a_N)^c$ turns out to be negligible in the limit as $N \to \infty$.

Lemma C.3.1. *For a positive real sequence $(a_N)_{N\in\mathbb{N}}$, set $b_N := a_N^2/(2k_J) - N\log 2$. Then, $\mathbb{P}$-a.s.,*

$$\mathbb{P}_{\mathcal{G}}\big[\Xi(a_N)^c\big] \leq \mathrm{e}^{-b_N} \wedge 1, \qquad \forall N \in \mathbb{N}, N \geq 2. \tag{C.38}$$

Proof. Fix $2 \leq N \in \mathbb{N}$. Clearly, it suffices to prove (C.38) in case $b_N > 0$. Using that the triangular array $J = (J_{ij})_{1\leq i<j<\infty}$ is assumed to be conditionally independent given $\mathcal{G}$ and that the map $J \longmapsto H_N(\sigma)$ depends linearly on the random coupling $(J_{ij})_{1\leq i<j\leq N}$, we can apply McDiarmid's concentration inequality (Theorem C.7.1), together with a union bound, to get that, $\mathbb{P}$-a.s.,

$$\begin{aligned}\mathbb{P}_{\mathcal{G}}\Big[\max_{\sigma\in\mathcal{S}_N}\big|H_N(\sigma) - \widetilde{H}_N(\sigma)\big| \geq a_N\Big] &\leq \sum_{\sigma\in\mathcal{S}_N}\mathbb{P}_{\mathcal{G}}\Big[\big|H_N(\sigma) - \widetilde{H}_N(\sigma)\big| \geq a_N\Big]\\ &\leq 2^N\exp\Big(-\frac{a_N^2N^2}{2k_JN(N-1)}\Big).\end{aligned} \tag{C.39}$$

Since $2k_J(N-1)/N \leq 2k_J$, (C.38) follows. □

Proof of Theorem C.2.10. We will prove that

$$\mathbb{P}\Big[\limsup_{N\to\infty}\Omega_{\mathrm{meta}}(N)^c\Big] = 0, \tag{C.40}$$

which is equivalent to (C.20). First note that, by the choice of μ_N and p_N in (C.2) and (C.3),

$$Z_N\mu_N(\sigma)\,p_N(\sigma,\sigma') = \mathrm{e}^{-\beta(H_N(\sigma)\vee H_N(\sigma'))}, \qquad \sigma\sim\sigma' \in \mathcal{S}_N. \tag{C.41}$$

An elementary computation yields that, on the event $\Xi(a_N)$,

$$\widetilde{H}_N(\sigma)\vee\widetilde{H}_N(\sigma') - a_N \leq H_N(\sigma)\vee H_N(\sigma') \leq \widetilde{H}_N(\sigma)\vee\widetilde{H}_N(\sigma') + a_N, \quad \sigma,\sigma' \in \mathcal{S}_N. \tag{C.42}$$

Thus, by a comparison of Dirichlet forms it follows that

$$\widetilde{Z}_N\widetilde{\mathcal{E}}_N(f)\,\mathrm{e}^{-\beta a_N} \leq Z_N\mathcal{E}_N(f) \leq \widetilde{Z}_N\widetilde{\mathcal{E}}_N(f)\,\mathrm{e}^{\beta a_N}, \tag{C.43}$$

for any $f : \mathcal{S}_N \to \mathbb{R}$. In view of the Dirichlet principle (C.33), we deduce that, on the event $\Xi(a_N)$,

$$\mathrm{e}^{-\beta a_N} \leq \frac{Z_N\,\mathrm{cap}_N(\mathcal{X},\mathcal{Y})}{\widetilde{Z}_N\widetilde{\mathrm{cap}}_N(\mathcal{X},\mathcal{Y})} \leq \mathrm{e}^{\beta a_N}, \qquad \emptyset \neq \mathcal{X},\mathcal{Y} \subset \mathcal{S}_N \text{ disjoint.} \tag{C.44}$$

Moreover, for any $2 \leq N \in \mathbb{N}$, on the event $\Xi(a_N)$,

$$\mathrm{e}^{-\beta a_N} \leq \frac{Z_N\mu_N[\mathcal{X}]}{\widetilde{Z}_N\widetilde{\mu}_N[\mathcal{X}]} \leq \mathrm{e}^{\beta a_N}, \qquad \emptyset \neq \mathcal{X} \subset \mathcal{S}_N. \tag{C.45}$$

It follows from (C.32), (C.44) and (C.45) that, on the event $\Xi(a_N)$,

$$e^{-2\beta a_N} \le \frac{P^N_{\mu_N|\mathcal{X}}[\tau^N_{\mathcal{Y}} < \tau^N_{\mathcal{X}}]}{\widetilde{P}^N_{\mu_N|\mathcal{X}}[\tilde{\tau}^N_{\mathcal{Y}} < \tilde{\tau}^N_{\mathcal{X}}]} \le e^{2\beta a_N}, \qquad \emptyset \neq \mathcal{X}, \mathcal{Y} \subset \mathcal{S}_N \text{ disjoint.} \tag{C.46}$$

Thus, on the event $\Xi(a_N) \cap \widetilde{\Omega}_{\text{meta}}(N)$,

$$\begin{aligned} &\max_{j\in\{1,\ldots,K\}} P^N_{\mu_N|\mathcal{M}_{j,N}}\Big[\tau^N_{\mathcal{M}_N\setminus\mathcal{M}_{j,N}} < \tau^N_{\mathcal{M}_{j,N}}\Big] \\ &\qquad \le \frac{\tilde{\rho}_N}{K}\, e^{4\beta a_N} \min_{\mathcal{X}\subset\mathcal{S}_N\setminus\mathcal{M}_N} P^N_{\mu_N|\mathcal{X}}\Big[\tau^N_{\mathcal{M}_N} < \tau^N_{\mathcal{X}}\Big]. \end{aligned} \tag{C.47}$$

Now set $a_N = \sqrt{2k_J(k_1 + \log 2)N}$ for $2 \le N \in \mathbb{N}$, and note that, with this choice of a_N, Lemma C.3.1 implies that

$$\mathbb{P}[\Xi(a_N)^c] = \mathbb{E}[\mathbb{P}_{\mathcal{G}}[\Xi(a_N)^c]] \le e^{-b_N} = e^{-k_1 N}, \qquad N \in \mathbb{N}, N \ge 2. \tag{C.48}$$

By choosing $N(k_1, c_0, \beta, k_J) \in \mathbb{N}$ in such a way that $c_0 < k_1 - 4\beta a_N/N$ for all $N \ge N(k_1, c_0, \beta, k_J)$, it follows from (C.47) that $\Xi(a_N) \cap \widetilde{\Omega}_{\text{meta}}(N) \subseteq \Omega_{\text{meta}}(N)$ for all $N \ge N(k_1, c_0, \beta, k_J)$. In particular,

$$\Omega_{\text{meta}}(N)^c \subseteq \Xi(a_N)^c \cup \widetilde{\Omega}_{\text{meta}}(N)^c, \qquad N \ge N(k_1, c_0, \beta, k_J). \tag{C.49}$$

Therefore, using continuity of the probability measure, we get

$$\begin{aligned} \mathbb{P}\Big[\limsup_{N\to\infty} \Omega_{\text{meta}}(N)^c\Big] &\overset{(C.49)}{\le} \mathbb{P}\Big[\limsup_{N\to\infty}\Big(\Xi(a_N)^c \cup \widetilde{\Omega}_{\text{meta}}(N)^c\Big)\Big] \\ &\le \lim_{N\to\infty}\Big(\mathbb{P}\Big[\bigcup\nolimits_{m\ge N} \Xi(a_m)^c\Big] + \mathbb{P}\Big[\bigcup\nolimits_{m\ge N} \widetilde{\Omega}_{\text{meta}}(m)^c\Big]\Big) \\ &= \mathbb{P}\Big[\limsup_{N\to\infty} \Xi(a_N)^c\Big] + \mathbb{P}\Big[\limsup_{N\to\infty} \widetilde{\Omega}_{\text{meta}}(N)^c\Big]. \end{aligned} \tag{C.50}$$

Since, by (C.48), $\sum_{N=1}^{\infty} \mathbb{P}[\Xi(a_N)^c] < \infty$, an application of the Borel Cantelli Lemma and Assumption C.2.7 yields that

$$\mathbb{P}\Big[\limsup_{N\to\infty} \Omega_{\text{meta}}(N)^c\Big] \le \mathbb{P}\Big[\limsup_{N\to\infty} \Xi(a_N)^c\Big] + \mathbb{P}\Big[\limsup_{N\to\infty} \widetilde{\Omega}_{\text{meta}}(N)^c\Big] = 0. \tag{C.51}$$

□

C.4 Capacity estimates

In this section we provide general estimates on the capacity of the quenched model compared to the annealed model. These estimates are general because we do not require any assumption on metastability, and the sets involved in the estimates are general disjoint subsets of the configuration space.

In Section C.4.1 we prove concentration for the logarithm of the capacities by using the Dirichlet principle and McDiarmid's concentration inequality. In Section C.4.2 we first estimate the conditional mean of $Z_N\mu_N$ and p_N in terms of the corresponding quantities of the annealed model $\widetilde{Z}_N\tilde{\mu}_N$ and $\tilde{p}_N$ in Lemma C.4.2, and afterwards prove annealed capacity estimates by using both the Dirichlet and the Thomson principle, together with Lemma C.4.2. The latter is crucial also in the proof of annealed estimates of $\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}$ in Section C.5.3.

All formulas in this Section C.4 are intended to hold $\mathbb{P}$-a.s. In order to lighten notation, we refrain from repeating that.

C.4.1 Concentration of quenched capacities

Proposition C.4.1. *Let $2 \leq N \in \mathbb{N}$, and consider two non-empty disjoint subsets $\mathcal{X}, \mathcal{Y} \subset \mathcal{S}_N$. Then, for any $t \geq 0$,*

$$\mathbb{P}_{\mathcal{G}}\Big[\big|\log(Z_N \operatorname{cap}_N(\mathcal{X},\mathcal{Y})) - \mathbb{E}_{\mathcal{G}}\big[\log(Z_N \operatorname{cap}_N(\mathcal{X},\mathcal{Y}))\big]\big| > t\Big] \leq 2\,\mathrm{e}^{-t^2/(\beta k_J)^2}. \tag{C.52}$$

Proof. First, recall that the triangular array $(J_{ij})_{1\leq i<j<\infty}$ is assumed to be conditionally independent given $\mathcal{G}$. Hence, in view of McDiarmid's concentration inequality (Theorem C.7.1), the assertion in (C.52) is immediate once we show that, for any $2 \leq N \in \mathbb{N}$, the mapping

$$(J_{ij})_{1\leq i<j\leq N} \longmapsto F_N\big((J_{ij})_{1\leq i<j\leq N}\big) \coloneqq \log(Z_N \operatorname{cap}_N(\mathcal{X},\mathcal{Y})) \tag{C.53}$$

satisfies a bounded difference estimate. More precisely, it is sufficient to show that, for any $1 \leq k < l \leq N$,

$$\big|F_N\big((J_{ij})_{1\leq i<j\leq N}\big) - F_N\big((J'_{ij})_{1\leq i<j\leq N}\big)\big| \leq \frac{2\beta k_J}{N}, \tag{C.54}$$

where $J'_{ij} \coloneqq J_{ij}$ for all $1 \leq i < j \leq N$ such that $(i,j) \neq (k,l)$, and J'_{kl} is a conditionally independent copy of $(J_{ij})_{1\leq i<j\leq N}$ given $\mathcal{G}$. In the sequel, we write H_N^J, Z_N^J, $\mathcal{E}_N^J$ and $\operatorname{cap}_N^J(\mathcal{X},\mathcal{Y})$ to emphasise the dependence on the random coupling $J = (J_{ij})_{1\leq i<j\leq N}$.

We proceed by following the same line of argument that led to (C.44) in the proof of Theorem C.2.10. Since

$$\big|H_N^J(\sigma) - H_N^{J'}(\sigma)\big| = \frac{|J_{kl} - J'_{kl}|}{N} \leq \frac{2k_J}{N}, \qquad \sigma \in \mathcal{S}_N, \tag{C.55}$$

an elementary computation yields that, for any $\sigma, \sigma' \in \mathcal{S}_N$,

$$H_N^{J'}(\sigma) \vee H_N^{J'}(\sigma') - \frac{2k_J}{N} \leq H_N^J(\sigma) \vee H_N^J(\sigma') \leq H_N^{J'}(\sigma) \vee H_N^{J'}(\sigma') + \frac{2k_J}{N}. \tag{C.56}$$

Thus, by a comparison of Dirichlet forms, we obtain, for any $f : \mathcal{S}_N \to \mathbb{R}$,

$$Z_N^{J'} \mathcal{E}_N^{J'}(f)\,\mathrm{e}^{-2\beta k_J/N} \leq Z_N^J \mathcal{E}_N^J(f) \leq Z_N^{J'} \mathcal{E}_N^{J'}(f)\,\mathrm{e}^{2\beta k_J/N}. \tag{C.57}$$

In view of the Dirichlet principle (C.33), we deduce that

$$Z_N^{J'} \operatorname{cap}_N^{J'}(\mathcal{X},\mathcal{Y})\,\mathrm{e}^{-2\beta k_J/N} \leq Z_N^J \operatorname{cap}_N^J(\mathcal{X},\mathcal{Y}) \leq Z_N^{J'} \operatorname{cap}_N^{J'}(\mathcal{X},\mathcal{Y})\,\mathrm{e}^{2\beta k_J/N}, \tag{C.58}$$

which yields (C.54). □

C.4.2 Annealed capacity estimates

Notation 1. *For any three sequences $(a_N)_{N\geq 0}, (b_N)_{N\geq 0}$, $(c_N)_{N\geq 0}$ and $N \in \mathbb{N}$, the notation $a_N = b_N + O(c_N)$ means that there exists a $C \in (0,\infty)$ independent of ω and N such that*

$$-Cc_N \leq a_N - b_N \leq Cc_N. \tag{C.59}$$

Before proving annealed capacity estimates, we prove the following lemma which is used both in the current section and for proving further annealed estimates in Section C.5.3.

Lemma C.4.2. *For $2 \leq N \in \mathbb{N}$ the following hold:*

(i) For any $\sigma \in \mathcal{S}_N$,

$$\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\pm\beta\Delta_N(\sigma)}\Big] = \mathrm{e}^{\alpha_N}\big(1 + O(N^{-1})\big), \tag{C.60}$$

(ii) For any $\sigma, \sigma' \in \mathcal{S}_N$ *with* $\sigma \sim \sigma'$,

$$\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\pm\beta(H_N(\sigma)\vee H_N(\sigma'))}\Big] = \mathrm{e}^{\pm\beta(\widetilde{H}_N(\sigma)\vee\widetilde{H}_N(\sigma'))}\,\mathrm{e}^{\alpha_N}\big(1 + O(N^{-1/2})\big), \tag{C.61}$$

where α_N *is defined in* (C.25) *and* $\Delta_N(\sigma) \coloneqq H_N(\sigma) - \widetilde{H}_N(\sigma)$, $\sigma \in \mathcal{S}_N$.

Proof. (i) Denote by $\mathbb{R} \ni t \mapsto \Lambda_{ij}(t) \coloneqq \log \mathbb{E}_{\mathcal{G}}\big[\exp\big(t(J_{ij} - \mathbb{E}_{\mathcal{G}}[J_{ij}])\big)\big]$ the conditional log-moment generating function given $\mathcal{G}$. By a Taylor expansion up to the third order, we get that, for any $t \in \mathbb{R}$,

$$\Lambda_{ij}(t) = \frac{t^2}{2}\,\mathbb{V}\mathrm{ar}_{\mathcal{G}}\big[J_{ij}\big] + \frac{t^3}{2}\int_0^1 (1-\theta)^2 \Lambda_{ij}'''(\theta t)\mathrm{d}\theta. \tag{C.62}$$

Since the random variables are assumed to be uniformly bounded, i.e., $|J_{ij}| \le k_J$, an elementary computation exploiting Cramér's measure yields that, $|\Lambda_{ij}'''(t)| \le 6k_J^3$. Hence

$$\big|\Lambda_{ij}(t) - \frac{t^2}{2}\,\mathbb{V}\mathrm{ar}_{\mathcal{G}}[J_{ij}]\big| \le k_J^3\,|t|^3\,. \tag{C.63}$$

Since the triangular array $(J_{ij})_{1\le i<j<\infty}$ is conditionally independent given $\mathcal{G}$, we have

$$\Big|\log \mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\pm\beta\Delta_N(\sigma)}\Big] - \alpha_N\Big| \le \sum_{\substack{i,j=1\\ i<j}}^{N} \Big|\Lambda_{ij}\big(\tfrac{\mp\beta}{N}\sigma_i\sigma_j\big) - \frac{(\mp\beta\sigma_i\sigma_j)^2}{2N^2}\,\mathbb{V}\mathrm{ar}_{\mathcal{G}}\big[J_{ij}\big]\Big| \le \frac{(\beta k_J)^3}{2N}, \tag{C.64}$$

which concludes the proof of (C.60). In particular, for any $\sigma \in \mathcal{S}_N$,

$$\mathrm{e}^{\pm\beta\widetilde{H}_N(\sigma)+\alpha_N}\,\mathrm{e}^{-(\beta k_J)^3/2N} \le \mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\pm\beta H_N(\sigma)}\Big] \le \mathrm{e}^{\pm\beta\widetilde{H}_N(\sigma)+\alpha_N}\,\mathrm{e}^{(\beta k_J)^3/2N}. \tag{C.65}$$

(ii) Because the proofs of (C.61) for $\pm\beta$ are similar, we give a detailed proof for $+\beta$ only. Since the conditional expectation of the maximum of two random variables is bounded from below by the maximum of their conditional expectations, it is immediate from (C.65) that, for any $\sigma, \sigma' \in \mathcal{S}_N$,

$$\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta(H_N(\sigma)\vee H_N(\sigma'))}\Big] \ge \mathrm{e}^{\beta(\widetilde{H}_N(\sigma)\vee\widetilde{H}_N(\sigma'))}\,\mathrm{e}^{\alpha_N}\Big(1 - \frac{(\beta k_J)^3}{2N}\Big). \tag{C.66}$$

Thus, we are left with proving the desired upper bound. For this purpose, we define, for $\sigma \in \mathcal{S}_N$ and $k \in \{1, \dots, N\}$,

$$H_N^k(\sigma) \coloneqq -\sigma_k\bigg(\frac{1}{N}\sum_{\substack{j=1\\ j>k}}^{N} J_{kj}\sigma_j + \frac{1}{N}\sum_{\substack{i=1\\ i<k}}^{N} J_{ik}\sigma_i + h\bigg), \tag{C.67}$$

and set $H_N^{\neq k}(\sigma) \coloneqq H_N(\sigma) - H_N^k(\sigma)$. Denoting by $\sigma^k \in \mathcal{S}_N$ the configuration that is obtained from σ by flipping the spin at site $k \in \{1, \dots, N\}$, we get $H_N^{\neq k}(\sigma) = H_N^{\neq k}(\sigma^k)$ and $H_N^k(\sigma) = -H_N^k(\sigma^k)$.

Since for any $\sigma \in \mathcal{S}_N$ the random variables $H_N^{\neq k}(\sigma)$ and $H_N^k(\sigma)$ are conditionally independent given $\mathcal{G}$, it follows that

$$\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta(H_N(\sigma)\vee H_N(\sigma^k))}\Big] = \mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^{\neq k}(\sigma)}\Big]\,\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^k(\sigma)} \vee \mathrm{e}^{\beta H_N^k(\sigma^k)}\Big]. \tag{C.68}$$

In order to estimate the second term, note that for any two non-negative random variables X, Y with finite second moment, we have

$$\begin{aligned}
\mathbb{E}[X \vee Y] &= \mathbb{E}[X] \vee \mathbb{E}[Y] + \frac{1}{2}\Big(\mathbb{E}[|X-Y|] - |\mathbb{E}[X] - \mathbb{E}[Y]|\Big)\\
&\le \mathbb{E}[X] \vee \mathbb{E}[Y] + \frac{1}{2}\Big(\mathbb{E}[|X-\mathbb{E}[X]|] + \mathbb{E}[|Y-\mathbb{E}[Y]|]\Big)\\
&\le \mathbb{E}[X] \vee \mathbb{E}[Y] + \frac{1}{2}\Big(\sqrt{\mathbb{V}\mathrm{ar}[X]} + \sqrt{\mathbb{V}\mathrm{ar}[Y]}\Big),
\end{aligned} \tag{C.69}$$

where we use the triangular inequality and Jensen's inequality. Hence,

$$\begin{aligned}
&\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^k(\sigma)} \vee \mathrm{e}^{\beta H_N^k(\sigma^k)}\Big]\\
&\quad\le \mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^k(\sigma)}\Big] \vee \mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^k(\sigma^k)}\Big] + \frac{1}{2}\bigg(\sqrt{\mathbb{V}\mathrm{ar}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^k(\sigma)}\Big]} + \sqrt{\mathbb{V}\mathrm{ar}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^k(\sigma^k)}\Big]}\bigg).
\end{aligned} \tag{C.70}$$

Since,

$$\begin{aligned}
1 + \frac{\mathbb{V}\mathrm{ar}_{\mathcal{G}}\big[\mathrm{e}^{\pm\beta H_N^k(\sigma)}\big]}{\mathbb{E}_{\mathcal{G}}\big[\mathrm{e}^{\pm\beta H_N^k(\sigma)}\big]^2} &\le \frac{\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\pm 2\beta(H_N^k(\sigma) - \mathbb{E}_{\mathcal{G}}[H_N^k(\sigma)])}\Big]}{\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\pm\beta(H_N^k(\sigma) - \mathbb{E}_{\mathcal{G}}[H_N^k(\sigma)])}\Big]^2}\\
&= \frac{\exp\Big(\sum_{j=k+1}^N \Lambda_{kj}\big(\frac{\mp 2\beta}{N}\sigma_k\sigma_j\big) + \sum_{i=1}^{k-1}\Lambda_{ik}\big(\frac{\mp 2\beta}{N}\sigma_i\sigma_k\big)\Big)}{\exp\Big(2\sum_{j=k+1}^N \Lambda_{kj}\big(\frac{\mp\beta}{N}\sigma_k\sigma_j\big) + 2\sum_{i=1}^{k-1}\Lambda_{ik}\big(\frac{\mp\beta}{N}\sigma_i\sigma_k\big)\Big)},
\end{aligned} \tag{C.71}$$

and $|\Lambda_{ij}(t)| \le k_J^2 t^2$, which follows from a Taylor expansion of $\Lambda_{ij}(t)$ up to second order together with the estimate $|\Lambda''_{ij}(t)| \le 2k_J^2$, we obtain that

$$\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^k(\sigma)} \vee \mathrm{e}^{\beta H_N^k(\sigma^k)}\Big] \le \bigg(\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^k(\sigma)}\Big] \vee \mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta H_N^k(\sigma^k)}\Big]\bigg)\Big(1 + \sqrt{\mathrm{e}^{\frac{6(\beta k_J)^2}{N}} - 1}\Big). \tag{C.72}$$

Combining (C.72), (C.68) and (C.65), we see that there exists a $c \equiv c(\beta, k_J)$ such that for all $2 \le N \in \mathbb{N}$,

$$\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{\beta(H_N(\sigma)\vee H_N(\sigma^k))}\Big] \le \mathrm{e}^{\beta(\widetilde{H}_N(\sigma)\vee \widetilde{H}_N(\sigma^k))}\,\mathrm{e}^{\alpha_N}\,\Big(1 + \frac{c}{\sqrt{N}}\Big). \tag{C.73}$$

This concludes the proof of (C.61). □

We are ready to prove the annealed capacity estimates.

Proposition C.4.3. *Let $2 \le N \in \mathbb{N}$, and let $\mathcal{X}, \mathcal{Y} \subset \mathcal{S}_N$ be two non-empty and disjoint.*

(i) Then,

$$\Big|\mathbb{E}_{\mathcal{G}}\big[\log(Z_N \operatorname{cap}_N(\mathcal{X},\mathcal{Y}))\big] - \log(\widetilde{Z}_N \widetilde{\operatorname{cap}}_N(\mathcal{X},\mathcal{Y}))\Big| = \alpha_N + O\Big(\frac{1}{\sqrt{N}}\Big). \tag{C.74}$$

(ii) For any $q \in [1, \infty)$ there exists a $c_3 \in (0, \infty)$ such that

$$e^{-\alpha_N}\Big(1 - \frac{c_3}{\sqrt{N}}\Big) \leq \frac{\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\big)^q\Big]^{1/q}}{\widetilde{Z}_N \widetilde{\operatorname{cap}}_N(\mathcal{X}, \mathcal{Y})} \leq e^{q\alpha_N}\Big(1 + \frac{c_3}{\sqrt{N}}\Big), \tag{C.75}$$

$$e^{-\alpha_N}\Big(1 - \frac{c_3}{\sqrt{N}}\Big) \leq \frac{\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\big)^{-q}\Big]^{1/q}}{\big(\widetilde{Z}_N \widetilde{\operatorname{cap}}_N(\mathcal{X}, \mathcal{Y})\big)^{-1}} \leq e^{q\alpha_N}\Big(1 + \frac{c_3}{\sqrt{N}}\Big), \tag{C.76}$$

where α_N is defined in (C.25).

Proof. Fix $\mathbb{N} \ni N \geq 2$ and consider two non-empty disjoint subsets $\mathcal{X}, \mathcal{Y} \subset \mathcal{S}_N$. Recall from Section C.2.3 the definition of the Dirichlet form $\mathcal{E}_N(f)$ for functions $f \in \mathcal{H}_{\mathcal{X},\mathcal{Y}}$ and the Dirichlet form $\mathcal{D}_N(\varphi)$ for unit flows $\varphi \in \mathcal{U}_{\mathcal{X},\mathcal{Y}}$. In view of Lemma C.4.2(ii) we have that

$$\begin{aligned} \mathbb{E}_{\mathcal{G}}\big[Z_N \mathcal{E}_N(f)\big] &= \widetilde{Z}_N \widetilde{\mathcal{E}}_N(f)\, e^{\alpha_N}\big(1 + O(N^{-1/2})\big) && \forall\, f \in \mathcal{H}_{\mathcal{X},\mathcal{Y}}, \\ \mathbb{E}_{\mathcal{G}}\big[Z_N^{-1} \mathcal{D}_N(\varphi)\big] &= \widetilde{Z}_N^{-1} \widetilde{\mathcal{D}}_N(\varphi)\, e^{\alpha_N}\big(1 + O(N^{-1/2})\big) && \forall\, \varphi \in \mathcal{U}_{\mathcal{X},\mathcal{Y}}. \end{aligned} \tag{C.77}$$

(i) The claim in (C.74) is an immediate consequence of the Dirichlet principle and the Thomson principle combined with Jensen's inequality. Indeed, in view of (C.77) there exists a $c \equiv c(\beta, k_J)$ such that

$$\begin{aligned} \mathbb{E}_{\mathcal{G}}\big[\log\big(Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\big)\big] &\leq \inf_{f \in \mathcal{H}_{\mathcal{X},\mathcal{Y}}} \log \mathbb{E}_{\mathcal{G}}\big[Z_N \mathcal{E}_N(f)\big] \\ &\leq \inf_{f \in \mathcal{H}_{\mathcal{X},\mathcal{Y}}} \log\big(\widetilde{Z}_N \widetilde{\mathcal{E}}_N(f)\big) + \alpha_N + cN^{-1/2} \\ &= \log\big(\widetilde{Z}_N\, \widetilde{\operatorname{cap}}_N(\mathcal{X}, \mathcal{Y})\big) + \alpha_N + cN^{-1/2}. \end{aligned} \tag{C.78}$$

Likewise, we obtain that

$$\begin{aligned} \mathbb{E}_{\mathcal{G}}\big[\log\big(Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\big)\big] &\geq -\inf_{\varphi \in \mathcal{U}_{\mathcal{X},\mathcal{Y}}} \log \mathbb{E}_{\mathcal{G}}\big[Z_N^{-1} \mathcal{D}_N(\varphi)\big] \\ &\geq -\inf_{\varphi \in \mathcal{U}_{\mathcal{X},\mathcal{Y}}} \log\big(\widetilde{Z}_N^{-1} \widetilde{\mathcal{D}}_N(\varphi)\big) - \alpha_N - c\, N^{-1/2} \\ &= \log\big(\widetilde{Z}_N\, \widetilde{\operatorname{cap}}_N(\mathcal{X}, \mathcal{Y})\big) - \alpha_N - cN^{-1/2}. \end{aligned} \tag{C.79}$$

(ii) Since the proofs of (C.75) and (C.76) are similar, we present the proof for (C.76) only. To get the lower bound, note that by Jensen's inequality it is immediate that

$$\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\big)^{-q}\Big]^{1/q} \geq \mathbb{E}_{\mathcal{G}}\Big[\big(Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\big)^{-1}\Big] \geq \frac{1}{\mathbb{E}_{\mathcal{G}}\big[Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\big]}. \tag{C.80}$$

Hence, analogously to (C.78), by applying the Dirichlet principle and (C.77) we obtain that there exists a $c \equiv c(\beta, k_J)$ such that

$$\frac{1}{\mathbb{E}_{\mathcal{G}}\big[Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\big]} \geq \frac{e^{-\alpha_N}}{\widetilde{Z}_N\, \widetilde{\operatorname{cap}}_N(\mathcal{X}, \mathcal{Y})}\big(1 - cN^{-1/2}\big). \tag{C.81}$$

To get the upper bound, note that, analogously to (C.79), by the Thomson principle we have that

$$\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N \operatorname{cap}_N(\mathcal{X}, \mathcal{Y})\big)^{-q}\Big]^{1/q} \leq \inf\nolimits_{\varphi \in \mathcal{U}_{\mathcal{X},\mathcal{Y}}} \mathbb{E}_{\mathcal{G}}\Big[\big(Z_N^{-1} \mathcal{D}_N(\varphi)\big)^q\Big]^{1/q}. \tag{C.82}$$

By Minkowski's inequality and an application of (C.61) with β replaced by βq, we find that for any $q \in [1,\infty)$ there exists a $c' \equiv c'(q,\beta,k_J)$ such that, for all $\varphi \in \mathcal{U}_{\mathcal{X},\mathcal{Y}}$,

$$\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N^{-1}\mathcal{D}_N(\varphi)\big)^q\Big]^{1/q} \le \widetilde{Z}_N^{-1}\widetilde{\mathcal{D}}_N(\varphi)\,\mathrm{e}^{q\alpha_N}\,(1+c'N^{-1/2}). \tag{C.83}$$

Therefore, again applying the Thomson principle, we obtain

$$\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\,\mathrm{cap}_N(\mathcal{X},\mathcal{Y})\big)^{-q}\Big]^{1/q} \le \frac{\mathrm{e}^{q\alpha_N}}{\widetilde{Z}_N\,\widetilde{\mathrm{cap}}_N(\mathcal{X},\mathcal{Y})}(1+c'N^{-1/2}), \tag{C.84}$$

and by setting $c_3 := c \vee c'$ we conclude the proof. □

C.5 Equilibrium potential estimates

This section contains all our results concerning the $\ell^1(\mu_N)$-norm of the equilibrium potential $h^N_{\mathcal{A}_N,\mathcal{B}_N}$, which we call the *harmonic sum*. Before proving concentration estimates in Section C.5.2 and annealed estimates in Section C.5.3, we provide some preliminary estimate in Section C.5.1. We emphasise that throughout this section, contrary to Section C.4, metastability plays an essential role.

C.5.1 Preliminary estimates

As mentioned above, for estimates on the harmonic sum we restrict to the event $\Xi(a_N)$, which is defined in (C.37) and used in Section C.3 to prove Theorem C.2.10. This event has high probability for suitably chosen sequences $(a_N)_{N\in\mathbb{N}}$ (recall Lemma C.3.1). We use two facts: on $\Xi(a_N)$ we can control the quenched Gibbs measure μ_N in terms of the annealed Gibbs measure $\tilde{\mu}_N$ (recall (C.45)), and the harmonic sum localises on the metastable valley of $\mathcal{A}_N$. We state and prove the last result in Proposition C.5.2.

Notation 2. *For any two random variables X, Y depending on $N \in \mathbb{N}$, writing "$X = Y$ holds $\mathbb{P}$-a.s. on the event $\widetilde{\Omega}_{\mathrm{meta}}(N)$" means that*

$$\mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)}X = \mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)}Y, \qquad \mathbb{P}\text{-}a.s.. \tag{C.85}$$

We stress that all formulas in Section C.5 hold $\mathbb{P}$-a.s. Moreover, formulas involving the quantities $\mathcal{S}_{j,N}, \mathcal{M}_{j,N}, \mathcal{A}_N, \mathcal{B}_N$, for fixed $j \in \{1,\dots,K\}$ and $N \in \mathbb{N}$, hold $\mathbb{P}$-a.s. on the event $\widetilde{\Omega}_{\mathrm{meta}}(N)$ unless differently specified, because those quantities are not defined in $\widetilde{\Omega}_{\mathrm{meta}}(N)^c$.

Remark C.5.1. By $\mathcal{G}$-measurability of $\widetilde{\Omega}_{\mathrm{meta}}$ (see Remark C.2.6) we are allowed to compute expectations and probabilities conditioned to $\mathcal{G}$ on the event $\widetilde{\Omega}_{\mathrm{meta}}$.

Proposition C.5.2. *Suppose that $i \in \{2,\dots,K\}$ satisfies Assumption C.2.11, and that $(a_N)_{N\in\mathbb{N}}$ is a non-negative sequence that is sublinear in N. Then there exists a $C \in (0, k_1 \wedge k_2)$ such that $\mathbb{P}$-a.s. on the event $\Xi(a_N) \cap \widetilde{\Omega}_{\mathrm{meta}}(N)$, for N sufficiently large depending on $(a_N)_{N\in\mathbb{N}}, \beta, k_1, k_2$,*

$$\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N} = \mu_N[\mathcal{S}_{i,N}]\big[1+O(\mathrm{e}^{-CN})\big]. \tag{C.86}$$

Remark C.5.3. Proposition C.5.2 holds true also for $a_N = cN$, with $c > 0$, in case $c\beta$ is sufficiently small compared to k_1 and k_2. We do not use this result.

Remark C.5.4. Although Proposition C.5.2 is similar to [68, Theorem 1.7], it is not an immediate consequence of the latter. Indeed, in (C.86) both $\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}$ and μ_N refer to the *quenched* Markov chain $(\Sigma_N(t))_{t\geq 0}$, but $\mathcal{S}_{i,N}$ is a set of the metastable partition of the *annealed* Markov chain $(\widetilde{\Sigma}_N(t))_{t\geq 0}$, while in [68, Theorem 1.7] all quantities refer to the same process. We made this modification for two reasons. First of all, we are not able to prove [68, Theorem 1.7] for the *quenched* Markov chain because we cannot prove the non-degeneracy assumption needed therein. Second, even if we were able to prove it, it would not be useful later on as we do not have estimates on the measure of the metastable partition of $(\Sigma_N(t))_{t\geq 0}$. However, as we shall see later, Assumption C.2.11 and (C.45) allow us both to prove (C.86) and to use it later having estimates on its right hand side.

Remark C.5.5. Note that in the following proof of Proposition C.5.2 we use the metastability of $(\widetilde{\Sigma}_N(t))_{t\geq 0}$ and do not use the metastability of $(\Sigma_N(t))_{t\geq 0}$.

Proof of Proposition C.5.2. The proof is inspired by that of [68, Lemma 3.3] and consists of two steps.

Step 1. Let $(a_N)_{N\in\mathbb{N}}$ be any non-negative real sequence and fix $N \in \mathbb{N}$. We start by showing that the following is true $\mathbb{P}$-a.s. on the event $\Xi(a_N) \cap \widetilde{\Omega}_{\mathrm{meta}}(N)$, for any two $j \neq k \in \{1, \dots, K\}$ and any $\varepsilon \in (0, 1]$,

$$\sum_{\sigma\in\mathcal{S}_{k,N}} \frac{\mu_N(\sigma)}{\mu_N[\mathcal{S}_{k,N}]}\, h^N_{\mathcal{M}_{j,N},\mathcal{M}_{k,N}}(\sigma) \leq \varepsilon + \tilde{\rho}_N\, \mathrm{e}^{4\beta a_N} \log(1/\varepsilon) \min\Big\{1, \frac{\mu_N[\mathcal{S}_{j,N}]}{\mu_N[\mathcal{S}_{k,N}]}\Big\}. \tag{C.87}$$

Recall that, for $\ell \in \{1, \dots, K\}$, $\mathcal{V}_{\ell,N}$ and $\mathcal{S}_{\ell,N}$ are, respectively, the local valley around the metastable set $\mathcal{M}_{\ell,N}$ and a set of the metastable partition of the *annealed model*. Moreover, $\mathcal{M}_N = \bigcup_{\ell=1}^K \mathcal{M}_{\ell,N}$. By applying (C.32) and (C.46), we have that, for any $\mathcal{X} \subset \mathcal{V}_{k,N} \setminus \mathcal{M}_{k,N}$, $\mathbb{P}$-a.s. on the event $\Xi(a_N) \cap \widetilde{\Omega}_{\mathrm{meta}}(N)$,

$$\begin{aligned}
\mu_N[\mathcal{X}] &\overset{(C.46)}{\leq} \mathrm{e}^{2\beta a_N} \frac{\mathrm{cap}_N(\mathcal{X}, \mathcal{M}_{k,N})}{\widetilde{\mathrm{P}}^N_{\mu_N|\mathcal{X}}[\tilde{\tau}^N_{\mathcal{M}_{k,N}} < \tilde{\tau}^N_{\mathcal{X}}]} \\
&\leq \mathrm{e}^{2\beta a_N} \tilde{\rho}_N \Big(\max_{\ell\in\{1,\dots,K\}} \widetilde{\mathrm{P}}^N_{\mu_N|\mathcal{M}_{\ell,N}} \Big[\tilde{\tau}^N_{\mathcal{M}_N\setminus\mathcal{M}_{\ell,N}} < \tilde{\tau}^N_{\mathcal{M}_{\ell,N}}\Big]\Big)^{-1} \mathrm{cap}_N(\mathcal{X}, \mathcal{M}_{k,N}) \\
&\overset{(C.46)}{\leq} \mathrm{e}^{4\beta a_N} \tilde{\rho}_N \Big(\max_{\ell\in\{1,\dots,K\}} \mathrm{P}^N_{\mu_N|\mathcal{M}_{\ell,N}} \Big[\tau^N_{\mathcal{M}_N\setminus\mathcal{M}_{\ell,N}} < \tau^N_{\mathcal{M}_{\ell,N}}\Big]\Big)^{-1} \mathrm{cap}_N(\mathcal{X}, \mathcal{M}_{k,N}),
\end{aligned} \tag{C.88}$$

where in the second inequality we used that we are in $\widetilde{\Omega}_{\mathrm{meta}}(N)$ and $\mathcal{X} \subset \mathcal{V}_{k,N} \setminus \mathcal{M}_{k,N}$ to apply [68, Lemma 3.1]. Moreover, using (C.32) and monotonicity of capacities, we get

$$\begin{aligned}
&\max_{\ell\in\{1,\dots,K\}} \mathrm{P}^N_{\mu_N|\mathcal{M}_{\ell,N}} \Big[\tau^N_{\mathcal{M}_N\setminus\mathcal{M}_{\ell,N}} < \tau^N_{\mathcal{M}_{\ell,N}}\Big] \\
&\qquad \geq \max\Big\{ \frac{\mathrm{cap}_N(\mathcal{M}_{j,N}, \mathcal{M}_{k,N})}{\mu_N[\mathcal{M}_{k,N}]}, \frac{\mathrm{cap}_N(\mathcal{M}_{k,N}, \mathcal{M}_{j,N})}{\mu_N[\mathcal{M}_{j,N}]} \Big\}.
\end{aligned} \tag{C.89}$$

Next, for $t \in (0, 1]$ we write $\mathcal{X}_N(t) \coloneqq \{\sigma \in \mathcal{S}_N : h^N_{\mathcal{M}_{j,N},\mathcal{M}_{k,N}}(\sigma) \geq t\}$ to denote the super level-sets of $h^N_{\mathcal{M}_{j,N},\mathcal{M}_{k,N}}$. Note that, for any $t \in (0, 1]$, $\mathcal{X}_N(t) \cap \mathcal{M}_{k,N} = \emptyset$ and $\mathcal{M}_{j,N} \subseteq \mathcal{X}_N(t)$. Using reversibility, we have

$$\begin{aligned}
t\, \mathrm{cap}(\mathcal{X}_N(t), \mathcal{M}_{k,N}) &\leq \langle -L_N h^N_{\mathcal{X}_N(t),\mathcal{M}_{k,N}}, h^N_{\mathcal{M}_{j,N},\mathcal{M}_{k,N}} \rangle_{\mu_N} \\
&= \langle h^N_{\mathcal{X}_N(t),\mathcal{M}_{k,N}}, -L_N h^N_{\mathcal{M}_{j,N},\mathcal{M}_{k,N}} \rangle_{\mu_N} = \mathrm{cap}(\mathcal{M}_{j,N}, \mathcal{M}_{k,N}).
\end{aligned} \tag{C.90}$$

By expressing the expected value of a non-negative random variable in terms of the integral of the tail of its distribution, we obtain, for any $\varepsilon \in (0,1]$,

$$\sum_{\sigma \in \mathcal{S}_{k,N}} \frac{\mu_N(\sigma)}{\mu_N[\mathcal{S}_{k,N}]} h^N_{\mathcal{M}_{j,N},\mathcal{M}_{k,N}}(\sigma) \le \varepsilon + \int_\varepsilon^1 \frac{\mu_N[\mathcal{X}_N(t) \cap \mathcal{S}_{k,N}]}{\mu_N[\mathcal{S}_{k,N}]} \mathrm{d}t. \tag{C.91}$$

Using (C.88) with $\mathcal{X} = \mathcal{X}_N(t) \cap \mathcal{S}_{k,N}$, together with (C.89), the symmetry and monotonicity of capacities and (C.90), we obtain, for $t \in [\varepsilon, 1]$,

$$\begin{aligned} \mu_N[\mathcal{X}_N(t) \cap \mathcal{S}_{k,N}] &\le \tilde{\rho}_N \, \mathrm{e}^{4\beta a_N} \frac{\min\{\mu_N[\mathcal{M}_{k,N}], \mu_N[\mathcal{M}_{j,N}]\}}{\mathrm{cap}_N(\mathcal{M}_{j,N}, \mathcal{M}_{k,N})} \mathrm{cap}_N(\mathcal{X}_N(t), \mathcal{M}_{k,N}) \\ &\le \tilde{\rho}_N \, \mathrm{e}^{4\beta a_N} \min\{\mu_N[\mathcal{M}_{k,N}], \mu_N[\mathcal{M}_{j,N}]\} \frac{1}{t}. \end{aligned} \tag{C.92}$$

Therefore, recalling that $\mathcal{M}_{\ell,N} \subseteq \mathcal{S}_{\ell,N}$, $\ell \in \{1, \dots, K\}$, we obtain

$$\sum_{\sigma \in \mathcal{S}_{k,N}} \frac{\mu_N(\sigma)}{\mu_N[\mathcal{S}_{k,N}]} h^N_{\mathcal{M}_{j,N},\mathcal{M}_{k,N}}(\sigma) \le \varepsilon + \tilde{\rho}_N \, \mathrm{e}^{4\beta a_N} \min\left\{ \frac{\mu_N[\mathcal{S}_{j,N}]}{\mu_N[\mathcal{S}_{k,N}]}, 1 \right\} \int_\varepsilon^1 \frac{1}{t} \mathrm{d}t, \tag{C.93}$$

which completes the proof of (C.87).

Step 2. In view of (C.87), the proof of (C.86) runs along the same lines as the proof of [68, Theorem 1.7]. For the reader's convenience we provide the details here. Let $\mathcal{A}_N, \mathcal{B}_N$ be defined as in (C.23). In particular, recall that $\mathcal{A}_N = \mathcal{M}_{i,N}$. Then

$$\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N} = \mu_N[\mathcal{S}_{i,N}] \left(\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N|\mathcal{S}_{i,N}} + \sum_{j \ne i} \frac{\mu_N[\mathcal{S}_{j,N}]}{\mu_N[\mathcal{S}_{i,N}]} \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N|\mathcal{S}_{j,N}} \right). \tag{C.94}$$

In order to prove a lower bound, we neglect the last term in the bracket in (C.94), while the first term is bounded from below by

$$\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N|\mathcal{S}_{i,N}} = 1 - \|h^N_{\mathcal{B}_N,\mathcal{A}_N}\|_{\mu_N|\mathcal{S}_{i,N}} \ge 1 - \sum_{j=1}^{i-1} \|h^N_{\mathcal{M}_{j,N},\mathcal{M}_{i,N}}\|_{\mu_N|\mathcal{S}_{i,N}}, \tag{C.95}$$

where we used that, for all $\sigma \in \mathcal{S}_N \setminus (\mathcal{A}_N \cup \mathcal{B}_N)$,

$$\begin{aligned} h^N_{\mathcal{B}_N,\mathcal{A}_N}(\sigma) &= \mathrm{P}^N_\sigma\left[\tau^N_{\bigcup_{j=1}^{i-1} \mathcal{M}_{j,N}} < \tau^N_{\mathcal{M}_{i,N}}\right] \\ &\le \sum_{j=1}^{i-1} \mathrm{P}^N_\sigma\left[\tau^N_{\mathcal{M}_{j,N}} < \tau^N_{\mathcal{M}_{i,N}}\right] = \sum_{j=1}^{i-1} h^N_{\mathcal{M}_{j,N},\mathcal{M}_{i,N}}(\sigma). \end{aligned} \tag{C.96}$$

By applying (C.87) with $\varepsilon = \mathrm{e}^{-k_1 N}$, recalling $\tilde{\rho}_N = \mathrm{e}^{-k_1 N}$ (see (C.12)), we obtain that $\mathbb{P}$-a.s. on the event $\Xi(a_N) \cap \widetilde{\Omega}_{\mathrm{meta}}(N)$,

$$\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N|\mathcal{S}_{i,N}} \ge 1 - K\mathrm{e}^{-k_1 N}(1 + \mathrm{e}^{4\beta a_N} \log(1/\mathrm{e}^{-k_1 N})). \tag{C.97}$$

Hence, we get

$$\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N} \ge \mu_N[\mathcal{S}_{i,N}] \Big(1 - KN\mathrm{e}^{-k_1 N + \beta a_N}\big(\mathrm{e}^{-\log N - \beta a_N} + k_1\big)\Big). \tag{C.98}$$

To get the upper bound, we exploit the fact that Assumption C.2.11, together with (C.45), implies that $\mu_N[\mathcal{S}_{j,N}]/\mu_N[\mathcal{S}_{i,N}] \le \mathrm{e}^{-k_2 N}\,\mathrm{e}^{2\beta a_N}$ for all $j \in \{i+1,\dots,K\}$. Hence

$$\begin{aligned}\sum_{j\neq i} \frac{\mu_N[\mathcal{S}_{j,N}]}{\mu_N[\mathcal{S}_{i,N}]}\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N|\mathcal{S}_{j,N}} & \\ \le K\mathrm{e}^{-k_2 N}\,\mathrm{e}^{2\beta a_N} + \sum_{j=1}^{i-1}\frac{\mu_N[\mathcal{S}_{j,N}]}{\mu_N[\mathcal{S}_{i,N}]}\|h^N_{\mathcal{M}_{i,N},\mathcal{M}_{j,N}}\|_{\mu_N|\mathcal{S}_{j,N}},&\end{aligned} \tag{C.99}$$

where we used that, for $j \in \{1,\dots,i-1\}$ and $\sigma \in \mathcal{S}_N \setminus \bigcup_{\ell=1}^{i}\mathcal{M}_{\ell,N}$,

$$\begin{aligned} h^N_{\mathcal{A}_N,\mathcal{B}_N}(\sigma) &= \mathrm{P}^N_\sigma\Big[\tau^N_{\mathcal{M}_{i,N}} < \tau^N_{\bigcup_{\ell=1}^{i-1}\mathcal{M}_{\ell,N}}\Big] \\ &\le \mathrm{P}^N_\sigma\Big[\tau^N_{\mathcal{M}_{i,N}} < \tau^N_{\mathcal{M}_{j,N}}\Big] = h^N_{\mathcal{M}_{i,N},\mathcal{M}_{j,N}}(\sigma). \end{aligned} \tag{C.100}$$

Thus, applying (C.87) with $\varepsilon = \mathrm{e}^{-k_1 N}\min_{\ell\in\{1,\dots,i-1\}}\mu_N[\mathcal{S}_{i,N}]/\mu_N[\mathcal{S}_{\ell,N}]$, we get that, $\mathbb{P}$-a.s. on the event $\Xi(a_N)\cap\widetilde{\Omega}_{\mathrm{meta}}(N)$,

$$\begin{aligned}&\frac{\mu_N[\mathcal{S}_{j,N}]}{\mu_N[\mathcal{S}_{i,N}]}\|h^N_{\mathcal{M}_{i,N},\mathcal{M}_{j,N}}\|_{\mu_N|\mathcal{S}_{j,N}} \\ &\qquad\le \mathrm{e}^{-k_1 N} - \tilde{\rho}_N\,\mathrm{e}^{4\beta a_N}\log\Big(\mathrm{e}^{-k_1 N}\min_{\ell\in\{1,\dots,i-1\}}\frac{\mu_N[\mathcal{S}_{i,N}]}{\mu_N[\mathcal{S}_{\ell,N}]}\Big)\end{aligned} \tag{C.101}$$

for $j \in \{1,\dots,i-1\}$. Since $\mu_N[\mathcal{S}_{i,N}]/\mu_N[\mathcal{S}_{j,N}] \ge \mathrm{e}^{-\beta(k_J+h)N}$ for $j \in \{1,\dots,K\}$ and $\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N|\mathcal{S}_{i,N}} \le 1$, we can use (C.94) to conclude that

$$\begin{aligned}&\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N} \\ &\le \mu_N[\mathcal{S}_{i,N}]\Big(1 + K\mathrm{e}^{-k_2 N + 2\beta a_N} + K\big(\mathrm{e}^{-k_1 N} + (k_1+\beta(k_J+h))N\mathrm{e}^{-k_1 N + 4\beta a_N}\big)\Big).\end{aligned} \tag{C.102}$$

Let

$$\bar{N} := \min\{N\in\mathbb{N}\colon -k_2 N + 2\beta a_N < 0 \text{ and } -k_1 N + 4\beta a_N + \log N < 0\}. \tag{C.103}$$

Note that $\bar{N}$ depends on $(a_N)_{N\in\mathbb{N}}, \beta, k_1, k_2$ and is deterministic. The minimum exists because β, k_1, k_2 are fixed and a_N is taken sublinear in N. By combining (C.98) and (C.102), the assertion follows for all $N \ge \bar{N}$. ☐

Corollary C.5.6. *Suppose that $i \in \{2,\dots,K\}$ satisfies Assumption C.2.11. Then there exists a $C \in (0, k_1\wedge k_2)$ such that, for N sufficiently large depending on β, k_1, k_2, k_J, $\mathbb{P}$-a.s. on the event $\widetilde{\Omega}_{\mathrm{meta}}(N)$,*

$$\mathbb{E}_{\mathcal{G}}\Big[\log\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\Big] = \mathbb{E}_{\mathcal{G}}\Big[\log\big(Z_N\mu_N[\mathcal{S}_{i,N}]\big)\Big] + O(\mathrm{e}^{-CN}). \tag{C.104}$$

Proof. First observe that $\mathbb{P}$-a.s. $\mathrm{e}^{-\beta(k_J+h)N} \le Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N} \le 2^N\mathrm{e}^{\beta(k_J+h)N}$ for $2 \le N \in \mathbb{N}$. Moreover, let $a_N = \sqrt{2k_J(k_1+\log 2)N}$. In view of Proposition C.5.2, we know that there exist a $C \in (0, k_1\wedge k_2)$ and a $c' \in (0,\infty)$ such that, for all N sufficiently large depending on β, k_1, k_2, k_J,

$$\begin{aligned}&\mathbb{E}_{\mathcal{G}}\Big[\log\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\Big] \\ &\le \mathbb{E}_{\mathcal{G}}\Big[\log\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\mathbb{1}_{\Xi(a_N)}\Big] + (\beta(k_J+h)+\log 2)N\,\mathbb{P}_{\mathcal{G}}[\Xi(a_N)^c] \\ &\le \mathbb{E}_{\mathcal{G}}\Big[\log\big(Z_N\mu_N[\mathcal{S}_{i,N}]\big)\Big] + \log(1+c'(\mathrm{e}^{-CN})) + (\beta(k_J+h)+\log 2)N\,\mathrm{e}^{-k_1 N},\end{aligned} \tag{C.105}$$

where we used (C.48), which is implied by Lemma C.3.1 and our choice of a_N. Likewise, by using additionally that $Z_N \mu_N[\mathcal{S}_{i,N}] \leq 2^N \mathrm{e}^{\beta(k_J+h)N}$, we obtain that

$$\begin{aligned}
&\mathbb{E}_{\mathcal{G}}\Big[\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\Big] \\
&\geq \mathbb{E}_{\mathcal{G}}\Big[\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\mathbb{1}_{\Xi(a_N)}\Big] - \beta(k_J+h)N\,\mathbb{P}_{\mathcal{G}}\big[\Xi(a_N)^c\big] \\
&\geq \mathbb{E}_{\mathcal{G}}\Big[\log\big(Z_N \mu_N[\mathcal{S}_{i,N}]\big)\Big] + \log(1-c'(\mathrm{e}^{-CN})) - (2\beta(k_J+h)+\log 2)N\mathrm{e}^{-k_1 N}.
\end{aligned} \tag{C.106}$$

Since $C < k_1$, this concludes the proof. ☐

C.5.2 Concentration inequality

Proposition C.5.7. *Suppose that $i \in \{2,\dots,K\}$ satisfies Assumption C.2.11. Then there exist $C \in (0, k_1 \wedge k_2)$ and $c_4 \in (0,\infty)$ such that, for all N sufficiently large depending on β, k_1, k_2, k_J, and all $t \geq 0$, $\mathbb{P}$-a.s. on the event $\widetilde{\Omega}_{\mathrm{meta}}(N)$,*

$$\begin{aligned}
\mathbb{P}_{\mathcal{G}}\Big[\big|\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big) - \mathbb{E}_{\mathcal{G}}\big[\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\big]\big| > t\Big] \\
\leq 2\mathrm{e}^{-\left(\frac{t-c_N}{\beta k_J}\right)^2} + \mathrm{e}^{-k_1 N},
\end{aligned} \tag{C.107}$$

where $c_N \coloneqq c_4\,\mathrm{e}^{-CN}$.

Proof. Let $a_N = \sqrt{2k_J(k_1+\log 2)N}$. In view of Proposition C.5.2 and Corollary C.5.6, there exist a $C \in (0, k_1 \wedge k_2)$ and $c_4 \in (0,\infty)$ such that, for N sufficiently large depending on β, k_1, k_2, k_J, $\mathbb{P}$-a.s on the event $\Xi(a_N) \cap \widetilde{\Omega}_{\mathrm{meta}}(N)$,

$$\big|\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big) - \log\big(Z_N \mu_N[\mathcal{S}_{i,N}]\big)\big| \leq \frac{c_4}{2}\,\mathrm{e}^{-CN} \tag{C.108}$$

and

$$\big|\mathbb{E}_{\mathcal{G}}\big[\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\big] - \mathbb{E}_{\mathcal{G}}\big[\log\big(Z_N \mu_N[\mathcal{S}_{i,N}]\big)\big]\big| \leq \frac{c_4}{2}\,\mathrm{e}^{-CN}. \tag{C.109}$$

Hence, by setting $c_N \coloneqq c_4\,\mathrm{e}^{-CN}$, we obtain that

$$\begin{aligned}
&\mathbb{P}_{\mathcal{G}}\Big[\big|\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big) - \mathbb{E}_{\mathcal{G}}\big[\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\big]\big| > t\Big] \\
&\leq \mathbb{P}_{\mathcal{G}}\Big[\big|\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big) - \mathbb{E}_{\mathcal{G}}\big[\log\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\big]\big| > t, \Xi(a_N)\Big] \\
&\quad + \mathbb{P}_{\mathcal{G}}\big[\Xi(a_N)^c\big] \\
&\leq \mathbb{P}_{\mathcal{G}}\Big[\big|\log\big(Z_N \mu_N[\mathcal{S}_{i,N}]\big) - \mathbb{E}_{\mathcal{G}}\big[\log\big(Z_N \mu_N[\mathcal{S}_{i,N}]\big)\big]\big| > t - c_N\Big] + \mathrm{e}^{-k_1 N},
\end{aligned} \tag{C.110}$$

where, as above, we used (C.48), which is implied by Lemma C.3.1 and our choice of a_N. In order to bound the first term on the right-hand side of (C.110), recall that the triangular array $(J_{ij})_{1\leq i<j<\infty}$ is assumed to be conditionally independent given $\mathcal{G}$. Moreover, in view of (C.55), for any $2 \leq N \in \mathbb{N}$ it is immediate that the mapping

$$(J_{ij})_{1\leq i<j\leq N} \longmapsto \bar{F}_N\big((J_{ij})_{1\leq i<j\leq N}\big) \coloneqq \log\big(Z_N \mu_N[\mathcal{S}_{i,N}]\big) \tag{C.111}$$

satisfies the estimate

$$\left|\bar{F}_N\big((J_{ij})_{1\le i<j\le N}\big) - \bar{F}_N\big((J'_{ij})_{1\le i<j\le N}\big)\right| \le \frac{2\beta k_J}{N}, \tag{C.112}$$

where $J'_{ij} := J_{ij}$ for all $1 \le i < j \le N$ such that $(i,j) \neq (k,l)$ and J'_{kl} is a conditionally independent copy of $(J_{ij})_{1\le i<j\le N}$ given $\mathcal{G}$, for any $1 \le k < l \le N$. Hence, by applying McDiarmid concentration inequality in the version of Proposition C.7.1, we get that

$$\mathbb{P}_{\mathcal{G}}\Big[\big|\log(Z_N\mu_N[\mathcal{S}_{i,N}]) - \mathbb{E}_{\mathcal{G}}[\log(Z_N\mu_N[\mathcal{S}_{i,N}])]\big| > t\Big] \le 2\,\mathrm{e}^{-t^2/(\beta k_J)^2}. \tag{C.113}$$

Combining (C.110) and (C.113), we get the assertion in (C.107). □

C.5.3 Annealed estimate

Proposition C.5.8. *Suppose that* $i \in \{2,\dots,K\}$ *satisfies Assumption C.2.11. Let* α_N *be as defined in* (C.25)*. Then the following hold:*

(i) There exists a $c_5 \in (0,\infty)$ *such that, for* N *sufficiently large depending on* β, k_1, k_2, k_J*,* $\mathbb{P}$*-a.s. on the event* $\widetilde{\Omega}_{\mathrm{meta}}(N)$*,*

$$-\frac{c_5}{N} \le \mathbb{E}_{\mathcal{G}}\Big[\log\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)\Big] - \log\big(\widetilde{Z}_N\|\tilde{h}^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\tilde{\mu}_N}\big) \le \alpha_N + \frac{c_5}{N}. \tag{C.114}$$

(ii) For any $q \in [1,\infty)$ *there exists a* $c_6 \in (0,\infty)$ *such that, for* N *sufficiently large depending on* β, k_1, k_2, k_J, q*,* $\mathbb{P}$*-a.s. on the event* $\widetilde{\Omega}_{\mathrm{meta}}(N)$*,*

$$\mathrm{e}^{\alpha_N}(1 - c_6N^{-1}) \le \frac{\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)^q\Big]^{1/q}}{\widetilde{Z}_N\|\tilde{h}^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\tilde{\mu}_N}} \le \mathrm{e}^{q\alpha_N}(1 + c_6N^{-1}). \tag{C.115}$$

Proof. (i) By using Jensen's inequality, $\mathcal{G}$-measurability of $\mathcal{S}_{i,N}$ and Lemma C.4.2(i), we find that for $2 \le N \in \mathbb{N}$

$$\begin{aligned}
&\mathbb{E}_{\mathcal{G}}\Big[\log(Z_N\mu_N[\mathcal{S}_{i,N}])\Big] \le \log \mathbb{E}_{\mathcal{G}}\Big[Z_N\mu_N[\mathcal{S}_{i,N}]\Big] \\
&= \log(\widetilde{Z}_N\tilde{\mu}_N[\mathcal{S}_{i,N}]) + \log\bigg(\sum_{\sigma\in\mathcal{S}_{i,N}} \frac{\tilde{\mu}_N(\sigma)}{\tilde{\mu}_N[\mathcal{S}_{i,N}]}\,\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{-\beta\Delta_N(\sigma)}\Big]\bigg) \\
&= \log(\widetilde{Z}_N\tilde{\mu}_N[\mathcal{S}_{i,N}]) + \alpha_N + \log(1 + O(N^{-1})).
\end{aligned} \tag{C.116}$$

Likewise,

$$\begin{aligned}
&\mathbb{E}_{\mathcal{G}}\Big[\log(Z_N\mu_N[\mathcal{S}_{i,N}])\Big] \\
&= \log(\widetilde{Z}_N\tilde{\mu}_N[\mathcal{S}_{i,N}]) + \mathbb{E}_{\mathcal{G}}\bigg[\log\bigg(\sum_{\sigma\in\mathcal{S}_{i,N}} \frac{\tilde{\mu}_N(\sigma)}{\tilde{\mu}_N[\mathcal{S}_{i,N}]}\,\mathrm{e}^{-\beta\Delta_N(\sigma)}\bigg)\bigg] \\
&\ge \log(\widetilde{Z}_N\tilde{\mu}_N[\mathcal{S}_{i,N}]) + \sum_{\sigma\in\mathcal{S}_{i,N}} \frac{\tilde{\mu}_N(\sigma)}{\tilde{\mu}_N[\mathcal{S}_{i,N}]}\,\mathbb{E}_{\mathcal{G}}[-\beta\Delta_N(\sigma)] \\
&= \log(\widetilde{Z}_N\tilde{\mu}_N[\mathcal{S}_{i,N}]).
\end{aligned} \tag{C.117}$$

Moreover, since $\tilde{\mu}_N[\mathcal{S}_{j,N}]/\tilde{\mu}_N[\mathcal{S}_{i,N}] \leq \mathrm{e}^{\beta(k_J+h)N}$ for all $j \in \{1,\dots,K\}$, we deduce from [68, Theorem 1.7] that

$$\widetilde{Z}_N \|\tilde{h}^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\tilde{\mu}_N} = \widetilde{Z}_N \tilde{\mu}_N[\mathcal{S}_{i,N}]\big(1 + O(\mathrm{e}^{-k_2 N} + N\tilde{\rho}_N)\big). \tag{C.118}$$

Thus, recalling that $\tilde{\rho}_N = \mathrm{e}^{-k_1 N}$, the assertion in (C.114) follows from Corollary C.5.6 combined with (C.116), (C.117) and (C.118).
(ii) For given $q \in [1,\infty)$, take $c' \in (0,\infty)$ and

$$a_N \coloneqq \sqrt{2k_J(q(c' + 2\beta(k_J + h) + \log 2) + \log 2)N}.$$

It follows from Lemma C.3.1 that, $\mathbb{P}$-a.s.,

$$\mathbb{P}_{\mathcal{G}}\big[\Xi(a_N)^c\big]^{\frac{1}{q}} \leq \mathrm{e}^{-b_N/q} = 2^{-N}\mathrm{e}^{-2\beta(k_J+h)N}\mathrm{e}^{-c'N}. \tag{C.119}$$

To get an upper bound for the q-th conditional moment given $\mathcal{G}$ of the harmonic sum, we use Minkowski's inequality, (C.119) and the facts that $Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N} \leq 2^N \mathrm{e}^{\beta(k_J+h)N}$ and $\widetilde{Z}_N\|\tilde{h}^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\tilde{\mu}_N} \geq \mathrm{e}^{-\beta(k_J+h)N}$. This gives

$$\begin{aligned}
&\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)^q\Big]^{\frac{1}{q}} \\
&\quad\leq \mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)^q \mathbb{1}_{\Xi(a_N)}\Big]^{\frac{1}{q}} + 2^N\mathrm{e}^{\beta(k_J+h)N}\,\mathbb{P}_{\mathcal{G}}\big[\Xi(a_N)^c\big]^{\frac{1}{q}} \\
&\quad\leq \mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)^q \mathbb{1}_{\Xi(a_N)}\Big]^{\frac{1}{q}} + \widetilde{Z}_N\|\tilde{h}^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\tilde{\mu}_N}\,\mathrm{e}^{-c'N}.
\end{aligned} \tag{C.120}$$

To analyse the first term of the right-hand side of (C.120), we apply Proposition C.5.2 to obtain that there exists a $C \in (0, k_1 \wedge k_2)$ such that, for N sufficiently large depending on β, k_1, k_2, k_J, q,

$$\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)^q \mathbb{1}_{\Xi(a_N)}\Big]^{\frac{1}{q}} = \mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\mu_N[\mathcal{S}_{i,N}]\big)^q \mathbb{1}_{\Xi(a_N)}\Big]^{\frac{1}{q}}\big(1 + O(\mathrm{e}^{-CN})\big). \tag{C.121}$$

Moreover, a further application of Minkowski's inequality yields that

$$\begin{aligned}
\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\mu_N[\mathcal{S}_{i,N}]\big)^q \mathbb{1}_{\Xi(a_N)}\Big]^{\frac{1}{q}} &\leq \widetilde{Z}_N \sum_{\sigma\in\mathcal{S}_{i,N}} \tilde{\mu}_N(\sigma)\,\mathbb{E}_{\mathcal{G}}\Big[\mathrm{e}^{-\beta q\Delta_N(\sigma)}\Big]^{\frac{1}{q}} \\
&= \widetilde{Z}_N\tilde{\mu}_N[\mathcal{S}_{i,N}]\,\mathrm{e}^{q\alpha_N}\big(1 + O(N^{-1})\big),
\end{aligned} \tag{C.122}$$

where in the last step we used Lemma C.4.2(i) with β replaced by βq. Thus, combining (C.120) with (C.121), (C.122) and (C.118), we see that there exists a $c'' \in (0,\infty)$ such that for N sufficiently large depending on β, k_1, k_2, k_J, q,

$$\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)^q\Big]^{\frac{1}{q}} \leq \widetilde{Z}_N\|\tilde{h}^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\tilde{\mu}_N}\,\mathrm{e}^{q\alpha_N}\big(1 + c''N^{-1}\big). \tag{C.123}$$

We close by proving a lower bound for the q-th conditional moment given $\mathcal{G}$ of the harmonic sum. By Jensen's inequality we get that

$$\mathbb{E}_{\mathcal{G}}\Big[\big(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)^q\Big]^{\frac{1}{q}} \geq \mathbb{E}_{\mathcal{G}}\Big[Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\Big] \geq \mathbb{E}_{\mathcal{G}}\Big[Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\mathbb{1}_{\Xi(a_N)}\Big]. \tag{C.124}$$

In view of Proposition C.5.2, together with (C.119) and the facts that $Z_N\mu_N[\mathcal{S}_{i,N}] \le 2^N e^{\beta(k_J+h)N}$ and $\widetilde{Z}_N\|\tilde{h}^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\tilde{\mu}_N} \ge e^{-\beta(k_J+h)N}$, we get that there exists a $C \in (0, k_1 \wedge k_2)$ such that, for N sufficiently large depending on β, k_1, k_2, k_J, q

$$\begin{aligned}
&\mathbb{E}_{\mathcal{G}}\Big[Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\mathbb{1}_{\Xi(a_N)}\Big] \\
&\quad= \mathbb{E}_{\mathcal{G}}\Big[Z_N\mu_N[\mathcal{S}_{i,N}]\mathbb{1}_{\Xi(a_N)}\Big](1+O(e^{-CN})) \qquad (C.125)\\
&\quad\ge \Big(\mathbb{E}_{\mathcal{G}}\Big[Z_N\mu_N[\mathcal{S}_{i,N}]\Big] - \widetilde{Z}_N\|\tilde{h}^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\tilde{\mu}_N}\, e^{-c'N}\Big)(1+O(e^{-CN})).
\end{aligned}$$

Since, by Lemma C.4.2(i),

$$\mathbb{E}_{\mathcal{G}}\Big[Z_N\mu_N[\mathcal{S}_{i,N}]\Big] = \widetilde{Z}_N\tilde{\mu}_N[\mathcal{S}_{i,N}]\, e^{\alpha_N}(1+O(N^{-1})), \qquad (C.126)$$

we conclude from (C.124) combined with (C.125), (C.126) and (C.118) that there exists a $c''' \in (0,\infty)$ such that, for N sufficiently large depending on β, k_1, k_2, k_J, q

$$\mathbb{E}_{\mathcal{G}}\Big[(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N})^q\Big]^{\frac{1}{q}} \ge \widetilde{Z}_N\|\tilde{h}^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\tilde{\mu}_N}\, e^{\alpha_N}(1+c'''N^{-1}). \qquad (C.127)$$

By setting $c_6 := c'' \vee c'''$, we get the assertion. □

C.6 Estimates on mean hitting times of metastable sets

Before proving Theorem C.2.13, we state two immediate corollaries of the propositions proved in Sections C.4 and C.5.

Corollary C.6.1. *Suppose that $i \in \{2,\dots,K\}$ satisfies Assumption C.2.11. Then there exist $C \in (0, k_1 \wedge k_2)$ and $c_4 \in (0,\infty)$ such that, for all N sufficiently large depending on β, k_1, k_2, k_J, and all $t \ge 0$, $\mathbb{P}$-a.s.,*

$$\begin{aligned}
&\mathbb{P}_{\mathcal{G}}\Big[\Big|\log(\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}[\tau^N_{\mathcal{B}_N}]) - \mathbb{E}_{\mathcal{G}}\Big[\log(\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}[\tau^N_{\mathcal{B}_N}])\Big]\Big| > t, \widetilde{\Omega}_{\mathrm{meta}}(N)\Big] \\
&\qquad\le \mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)}\left[2\big(e^{-\left(\frac{t-c_N}{2\beta k_J}\right)^2} + e^{-\left(\frac{t}{2\beta k_J}\right)^2}\big) + e^{-k_1N}\right],
\end{aligned} \qquad (C.128)$$

where $c_N := c_4\, e^{-CN}$.

Proof. In view of (C.29), we have that

$$\begin{aligned}
&\mathbb{P}_{\mathcal{G}}\Big[\Big|\log(\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}[\tau^N_{\mathcal{B}_N}]) - \mathbb{E}_{\mathcal{G}}\Big[\log(\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}[\tau^N_{\mathcal{B}_N}])\Big]\Big| > t, \widetilde{\Omega}_{\mathrm{meta}}(N)\Big] \\
&\le \mathbb{P}_{\mathcal{G}}\Big[\Big|\log(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}) - \mathbb{E}_{\mathcal{G}}\Big[\log(Z_N\|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N})\Big]\Big| > \frac{t}{2}, \widetilde{\Omega}_{\mathrm{meta}}(N)\Big] \qquad (C.129)\\
&+ \mathbb{P}_{\mathcal{G}}\Big[\Big|\log(Z_N\,\mathrm{cap}_N(\mathcal{A}_N,\mathcal{B}_N)) - \mathbb{E}_{\mathcal{G}}\Big[\log(Z_N\,\mathrm{cap}_N(\mathcal{A}_N,\mathcal{B}_N))\Big]\Big| > \frac{t}{2}, \widetilde{\Omega}_{\mathrm{meta}}(N)\Big].
\end{aligned}$$

Thus, the assertion follows immediately from Proposition C.4.1 and Proposition C.5.7. □

Corollary C.6.2. *Suppose that $i \in \{2,\dots,K\}$ satisfies Assumption C.2.11. Then there exists a $c_7 \in (0,\infty)$ such that, for N sufficiently large depending on β, k_1, k_2, k_J, $\mathbb{P}$-a.s. on the event $\widetilde{\Omega}_{\mathrm{meta}}(N)$,*

$$-\alpha_N - \frac{c_7}{\sqrt{N}} \le \mathbb{E}_{\mathcal{G}}\Big[\log \mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]\Big] - \log \tilde{\mathrm{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tilde{\tau}^N_{\mathcal{B}_N}\big] \le 2\alpha_N + \frac{c_7}{\sqrt{N}}, \tag{C.130}$$

where α_N is defined in (C.25).

Proof. In view of (C.29), the assertion in (C.130) follows immediately from Proposition C.5.8(i) and Proposition C.4.3(i). ☐

Proof of Theorem C.2.13. *(i)* Recall once again the $\mathcal{G}$-measurability of $\widetilde{\Omega}_{\mathrm{meta}}(N)$ (see Remark C.2.6), and note that an application of Corollary C.6.2 yields that there exists a $c_7 \in (0,\infty)$ such that, for N sufficiently large depending on β, k_1, k_2, k_J, $\mathbb{P}$-a.s.,

$$\begin{aligned}
&\mathbb{P}_{\mathcal{G}}\left[\mathrm{e}^{-t-\alpha_N} \le \frac{\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]}{\tilde{\mathrm{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tilde{\tau}^N_{\mathcal{B}_N}\big]} \le \mathrm{e}^{+t+2\alpha_N}\right]\mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)} \\
&\overset{(\mathrm{C}.130)}{\ge} \mathbb{P}_{\mathcal{G}}\left[\mathrm{e}^{-(t-\frac{c_7}{\sqrt{N}})} \le \frac{\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]}{\exp\Big(\mathbb{E}_{\mathcal{G}}\Big[\log \mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]\Big]\Big)} \le \mathrm{e}^{+(t-\frac{c_7}{\sqrt{N}})}, \widetilde{\Omega}_{\mathrm{meta}}(N)\right] \\
&\ge \mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)} \\
&\quad - \mathbb{P}_{\mathcal{G}}\left[\Big|\log\big(\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]\big) - \mathbb{E}_{\mathcal{G}}\Big[\log\big(\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]\big)\Big]\Big| > t - \frac{c_7}{\sqrt{N}}\right]\mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)}.
\end{aligned} \tag{C.131}$$

Thus, from Corollary C.6.1 it follows that there exists a $C \in (0, k_1 \wedge k_2)$ such that, for N sufficiently large depending on β, k_1, k_2, k_J, $\mathbb{P}$-a.s.,

$$\begin{aligned}
&\mathbb{P}_{\mathcal{G}}\left[\mathrm{e}^{-t-\alpha_N} \le \frac{\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]}{\tilde{\mathrm{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tilde{\tau}^N_{\mathcal{B}_N}\big]} \le \mathrm{e}^{+t+2\alpha_N}\right]\mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)} \\
&\ge \bigg(1 - 2\Big(\mathrm{e}^{-\left(\frac{t-c_7N^{-1/2}-c_N}{2\beta k_J}\right)^2} + \mathrm{e}^{-\left(\frac{t-c_7N^{-1/2}}{2\beta k_J}\right)^2}\Big) - \mathrm{e}^{-k_1 N}\bigg)\mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)}.
\end{aligned} \tag{C.132}$$

Assumption C.2.7 implies that $\lim_{N\to\infty} \mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)}$ exists and is $\mathbb{P}$-a.s. equal to 1. Since trivially $1 \ge \mathbb{1}_{\widetilde{\Omega}_{\mathrm{meta}}(N)}$, taking the limit $N \to \infty$ of (C.132) yields (C.26).
(ii) Fix $q \in [1,\infty)$. In view of (C.29), an application of the Cauchy-Schwarz inequality yields that, $\mathbb{P}$-a.s.,

$$\begin{aligned}
&\mathbb{E}_{\mathcal{G}}\Big[\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]^q\Big]^{\frac{1}{q}} \\
&\quad\le \mathbb{E}_{\mathcal{G}}\Big[\big(Z_N \|h^N_{\mathcal{A}_N,\mathcal{B}_N}\|_{\mu_N}\big)^{2q}\Big]^{\frac{1}{2q}} \mathbb{E}_{\mathcal{G}}\Big[\big(Z_N \operatorname{cap}_N(\mathcal{A}_N,\mathcal{B}_N)\big)^{-2q}\Big]^{\frac{1}{2q}}.
\end{aligned} \tag{C.133}$$

Hence, by Proposition C.4.3(ii) and C.5.8(ii), there exists a $c \in (0,\infty)$ such that for N sufficiently large depending on β, k_1, k_2, k_J, q, $\mathbb{P}$-a.s. on the event $\widetilde{\Omega}_{\mathrm{meta}}(N)$,

$$\mathbb{E}_{\mathcal{G}}\Big[\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]^q\Big]^{\frac{1}{q}} \le \tilde{\mathrm{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tilde{\tau}^N_{\mathcal{B}_N}\big]\,\mathrm{e}^{4q\alpha_N}\Big(1 + \frac{c}{\sqrt{N}}\Big). \tag{C.134}$$

On the other hand, by using Jensen's inequality and Corollary C.6.2, we find that there exists a $c_7 \in (0,\infty)$ such that, for any $q \in [1,\infty)$ and N sufficiently large depending on β, k_1, k_2, k_J, q, $\mathbb{P}$-a.s. on the event $\widetilde{\Omega}_{\text{meta}}(N)$,

$$\mathbb{E}_{\mathcal{G}}\Big[\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]^q\Big]^{\frac{1}{q}} \geq \exp\Big(\mathbb{E}_{\mathcal{G}}\Big[\log \mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]\Big]\Big) \\ \geq \tilde{\mathrm{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tilde{\tau}^N_{\mathcal{B}_N}\big]\, \mathrm{e}^{-\alpha_N}\Big(1 - \frac{c_7}{\sqrt{N}}\Big). \tag{C.135}$$

Therefore, by letting $c_1 = c \vee c_7$, for N sufficiently large depending on β, k_1, k_2, k_J, q, $\mathbb{P}$-a.s. on the event $\widetilde{\Omega}_{\text{meta}}(N)$,

$$\mathrm{e}^{-\alpha_N}\big(1 - \tfrac{c_1}{\sqrt{N}}\big) \leq \frac{\mathbb{E}_{\mathcal{G}}\Big[\mathrm{E}^N_{\nu_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tau^N_{\mathcal{B}_N}\big]^q\Big]^{1/q}}{\tilde{\mathrm{E}}^N_{\tilde{\nu}_{\mathcal{A}_N,\mathcal{B}_N}}\big[\tilde{\tau}^N_{\mathcal{B}_N}\big]} \leq \mathrm{e}^{4q\alpha_N}\big(1 + \tfrac{c_1}{\sqrt{N}}\big). \tag{C.136}$$

Thus, the set $\Omega_{q,c_1}(N)$ defined in (C.27) contains $\widetilde{\Omega}_{\text{meta}}(N)$. Therefore, using Assumption C.2.7, and monotonicity of probability

$$1 = \mathbb{P}\Big[\liminf_{N\to\infty} \widetilde{\Omega}_{\text{meta}}(N)\Big] \leq \mathbb{P}\Big[\liminf_{N\to\infty} \Omega_{q,c_1}(N)\Big] \tag{C.137}$$

which concludes the proof of (C.28). □

C.7 Appendix A: Concentration inequality

We present a concentration inequality for functionals of conditionally independent random variables that is a slight extension of the classical McDiarmid concentration inequality for functionals of independent random variables satisfying a bounded difference estimate (cf. [10, Theorem 6.2], [33, Section 2.4.1]).

Proposition C.7.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra of $\mathcal{F}$, $1 \leq n \in \mathbb{N}$ and $\mathcal{X}$ a Polish space. Consider a vector $X = (X_1, \ldots, X_n)$ of $\mathcal{X}$-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ that are conditionally independent given $\mathcal{G}$, and let $f_n : \mathcal{X}^n \to \mathbb{R}$ be a measurable function. Suppose that, for any $i \in \{1, \ldots, n\}$,*

$$\big|f_n(X_1, \ldots, X_n) - f_n(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)\big| \leq c_i \in [0,\infty) \qquad \mathbb{P}\text{-a.s.}, \tag{C.138}$$

where $(X_1', \ldots, X_n')$ is a conditionally independent copy of $(X_1, \ldots, X_n)$ given $\mathcal{G}$. Then, $\mathbb{P}$-a.s. for all $t \geq 0$,

$$\left.\begin{array}{l} \mathbb{P}\big[f_n(X) - \mathbb{E}[f_n(X) \mid \mathcal{G}] > +t \mid \mathcal{G}\big] \\ \mathbb{P}\big[f_n(X) - \mathbb{E}[f_n(X) \mid \mathcal{G}] < -t \mid \mathcal{G}\big] \end{array}\right\} \leq \mathrm{e}^{-t^2/(2v)}, \tag{C.139}$$

where $v := \frac{1}{4}\sum_{n=1}^n c_i^2$.

Proof. Since there exists a regular conditional probability for X (see e.g. [27, p. 217]), the proof follows the line of proof of the non-conditional McDiarmid concentration inequality. □

Bibliography

[1] R. Acharyya and D. Štefankovič. Glauber dynamics for Ising model on convergent dense graph sequences. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 81 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 23, 22. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017. doi:10.4230/LIPIcs.APPROX-RANDOM.2017.23.

[2] M. an der Heiden. *Metastability of Markov chains and in the Hopfield model.* Doctoral thesis, Technische Universität Berlin, Fakultät III - Prozesswissenschaften, Berlin, 2007. URL: http://dx.doi.org/10.14279/depositonce-1513, doi:10.14279/depositonce-1513.

[3] R. J. Baxter. *Exactly solved models in statistical mechanics.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, 1989. Reprint of the 1982 original.

[4] N. Berglund. Kramers' law: validity, derivations and generalisations. *Markov Process. Related Fields*, 19(3):459–490, 2013.

[5] K. A. Berman and M. H. Konsowa. Random paths and cuts, electrical networks, and reversible Markov chains. *SIAM J. Discrete Math.*, 3(3):311–319, 1990. doi:10.1137/0403026.

[6] A. Bianchi. *Mixing time for Glauber dynamics beyond* $\mathbb{Z}^d$. PhD thesis, Università di Roma Tre, 2006.

[7] A. Bianchi, A. Bovier, and D. Ioffe. Sharp asymptotics for metastability in the random field Curie-Weiss model. *Electron. J. Probab.*, 14:no. 53, 1541–1603, 2009. doi:10.1214/EJP.v14-673.

[8] A. Bianchi, A. Bovier, and D. Ioffe. Pointwise estimates and exponential laws in metastable systems via coupling methods. *Ann. Probab.*, 40(1):339–371, 2012. doi:10.1214/10-AOP622.

[9] T. Bodineau, B. Graham, and M. Wouts. Metastability in the dilute Ising model. *Probab. Theory Related Fields*, 157(3-4):955–1009, 2013. doi:10.1007/s00440-012-0474-8.

[10] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities.* Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux. doi:10.1093/acprof:oso/9780199535255.001.0001.

[11] A. Bovier. *Statistical mechanics of disordered systems*, volume 18 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, 2006. A mathematical perspective. doi:10.1017/CBO9780511616808.

[12] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in stochastic dynamics of disordered mean-field models. *Probab. Theory Related Fields*, 119(1):99–161, 2001. doi:10.1007/PL00012740.

[13] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability and low lying spectra in reversible Markov chains. *Comm. Math. Phys.*, 228(2):219–255, 2002. doi:10.1007/s002200200609.

[14] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)*, 6(4):399–424, 2004. doi:10.4171/JEMS/14.

[15] A. Bovier and V. Gayrard. The thermodynamics of the Curie-Weiss model with random couplings. *J. Statist. Phys.*, 72(3-4):643–664, 1993. doi:10.1007/BF01048027.

[16] A. Bovier and V. Gayrard. Hopfield models as generalized random mean field models. In *Mathematical aspects of spin glasses and neural networks*, volume 41 of *Progr. Probab.*, pages 3–89. Birkhäuser Boston, Boston, MA, 1998. doi:10.1007/978-1-4612-4102-7_1.

[17] A. Bovier, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc. (JEMS)*, 7(1):69–99, 2005. doi:10.4171/JEMS/22.

[18] A. Bovier and F. den Hollander. *Metastability - A potential-theoretic approach*, volume 351 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Cham, 2015. doi:10.1007/978-3-319-24777-9.

[19] A. Bovier, F. den Hollander, and S. Marello. Metastability for Glauber Dynamics on the Complete Graph with Coupling Disorder. *Comm. Math. Phys.*, 392(1):307–345, 2022. doi:10.1007/s00220-022-04351-8.

[20] A. Bovier, F. den Hollander, S. Marello, E. Pulvirenti, and M. Slowik. Metastability of Glauber dynamics with inhomogeneous coupling disorder. *Preprint arXiv:2209.09827 [math.PR]*, 2022.

[21] A. Bovier and F. Manzo. Metastability in Glauber dynamics in the low-temperature limit: beyond exponential asymptotics. *J. Statist. Phys.*, 107(3-4):757–779, 2002. doi:10.1023/A:1014586130046.

[22] A. Bovier, S. Marello, and E. Pulvirenti. Metastability for the dilute Curie-Weiss model with Glauber dynamics. *Electronic Journal of Probability*, 26:1–38, 2021. doi:10.1214/21-EJP610.

[23] M. Brokate and J. Sprekels. *Hysteresis and phase transitions*, volume 121 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996. doi:10.1007/978-1-4612-4048-8.

[24] S. G. Brush. History of the Lenz-Ising model. *Rev. Mod. Phys.*, 39:883–893, Oct 1967. URL: https://link.aps.org/doi/10.1103/RevModPhys.39.883, doi:10.1103/RevModPhys.39.883.

[25] V. H. Can, R. van der Hofstad, and T. Kumagai. Glauber dynamics for Ising models on random regular graphs: cut-off and metastability. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 18(2):1441–1482, 2021. doi:10.30757/alea.v18-52.

[26] M. Cassandro, A. Galves, E. Olivieri, and M. E. Vares. Metastable behavior of stochastic dynamics: a pathwise approach. *J. Statist. Phys.*, 35(5-6):603–634, 1984. doi:10.1007/BF01010826.

[27] Y. S. Chow and H. Teicher. *Probability theory*. Springer Texts in Statistics. Springer-Verlag, New York, third edition, 1997. Independence, interchangeability, martingales. doi:10.1007/978-1-4612-1950-7.

[28] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, 99(25):15879–15882, 2002. doi:10.1073/pnas.252631999.

[29] N. G. de Bruijn. *Asymptotic methods in analysis*. Second edition. Bibliotheca Mathematica, Vol. IV. North-Holland Publishing Co., Amsterdam; P. Noordhoff Ltd., Groningen, 1961.

[30] L. De Sanctis. Fluctuations in the Ising model on a sparse random graph. *Stochastic Process. Appl.*, 119(10):3383–3394, 2009. doi:10.1016/j.spa.2009.06.001.

[31] L. De Sanctis and F. Guerra. Mean field dilute ferromagnet: high temperature and zero temperature behavior. *J. Stat. Phys.*, 132(5):759–785, 2008. doi:10.1007/s10955-008-9575-2.

[32] A. Dembo and A. Montanari. Ising models on locally tree-like graphs. *Ann. Appl. Probab.*, 20(2):565–592, 2010. doi:10.1214/09-AAP627.

[33] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition. doi:10.1007/978-3-642-03311-7.

[34] S. Dommers. *Spin models on random graphs*. PhD thesis, Ruhr-Universität Bochum, 2013.

[35] S. Dommers. Metastability of the Ising model on random regular graphs at zero temperature. *Probab. Theory Related Fields*, 167(1-2):305–324, 2017. doi:10.1007/s00440-015-0682-0.

[36] S. Dommers, C. Giardinà, C. Giberti, R. van der Hofstad, and M. L. Prioriello. Ising critical behavior of inhomogeneous Curie-Weiss models and annealed random graphs. *Comm. Math. Phys.*, 348(1):221–263, 2016. doi:10.1007/s00220-016-2752-2.

[37] S. Dommers, C. Giardinà, and R. van der Hofstad. Ising models on power-law random graphs. *J. Stat. Phys.*, 141(4):638–660, 2010. doi:10.1007/s10955-010-0067-9.

[38] S. Dommers, F. den Hollander, O. Jovanovski, and F. R. Nardi. Metastability for Glauber dynamics on random graphs. *Ann. Appl. Probab.*, 27(4):2130–2158, 2017. doi:10.1214/16-AAP1251.

[39] P. G. Doyle and J. L. Snell. *Random walks and electric networks*, volume 22 of *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 1984. doi:10.5948/UPO9781614440222.

[40] H. Duminil-Copin. 100 years of the (critical) Ising model on the hypercubic lattice. *to appear in Proc. Int. Cong. Math.*, 1, 2022. URL: https://www.mathunion.org/fileadmin/IMU/Prizes/Fields/2022/hdc.pdf, doi:10.4171/ICM2022/204.

[41] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[42] L. R. Fontes, P. Mathieu, and P. Picco. On the averaged dynamics of the random field Curie-Weiss model. *Ann. Appl. Probab.*, 10(4):1212–1245, 2000. doi:10.1214/aoap/1019487614.

[43] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1984. Translated from the Russian by Joseph Szücs. doi:10.1007/978-1-4684-0176-9.

[44] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems: A concrete mathematical introduction.* Cambridge University Press, Cambridge, 2018. doi:10.1017/9781316882603.

[45] A. Gaudillière. Condenser physics applied to Markov chains - a brief introduction to potential theory. *Preprint arXiv:0901.3053 [math.PR]*, 2009.

[46] A. Gaudillière and C. Landim. A Dirichlet principle for non reversible Markov chains and some recurrence theorems. *Probab. Theory Related Fields*, 158(1-2):55–89, 2014. doi:10.1007/s00440-012-0477-5.

[47] R. van der Hofstad. *Random graphs and complex networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017. doi:10.1017/9781316779422.

[48] R. van der Hofstad. *Stochastic processes on random graphs.* https://www.win.tue.nl/~rhofstad/, Book in preparation.

[49] F. den Hollander and S. Jansen. Berman-Konsowa principle for reversible Markov jump processes. *Markov Process. Related Fields*, 22(3):409–442, 2016.

[50] F. den Hollander and O. Jovanovski. Glauber dynamics on the Erdős-Rényi random graph. In *In and out of equilibrium. 3. Celebrating Vladas Sidoravicius*, volume 77 of *Progress in Probability*, pages 519–589. Birkhäuser/Springer, Cham, 2021. doi:10.1007/978-3-030-60754-8_24.

[51] E. Ising. *Beitrag zur Theorie des Ferro- und Paramagnetismus.* Doctoral thesis, Univ. Hamburg, 1924.

[52] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift. f. Physik*, 31:253–258, 1925. doi:https://doi.org/10.1007/BF02980577.

[53] Z. Kabluchko, M. Löwe, and K. Schubert. Fluctuations of the magnetization for Ising models on dense Erdős–Rényi random graphs. *J. Stat. Phys.*, 177(1):78–94, 2019. doi:10.1007/s10955-019-02358-5.

[54] Z. Kabluchko, M. Löwe, and K. Schubert. Fluctuations of the magnetization for Ising models on Erdős-Rényi random graphs—the regimes of small p and the critical temperature. *J. Phys. A*, 53(35):355004, 37, 2020. doi:10.1088/1751-8121/aba05f.

[55] Z. Kabluchko, M. Löwe, and K. Schubert. Fluctuations of the magnetization for Ising models on Erdős-Rényi random graphs— the regimes of low temperature and external magnetic field. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 19(1):537–563, 2022. doi:10.30757/alea.v19-21.

[56] J. G. Kemeny and J. L. Snell. *Finite Markov chains*. Undergraduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1976. Reprinting of the 1960 original.

[57] W. Lenz. Beitrag zum Verständnis der magnetischen Eigenschaften in festen Körpern. *Phys. Zeitschr.*, 21:613–615, 1920.

[58] P. Mathieu and P. Picco. Metastability and convergence to equilibrium for the random field Curie-Weiss model. *J. Statist. Phys.*, 91(3-4):679–732, 1998. doi:10.1023/A:1023085829152.

[59] E. Mossel and A. Sly. Rapid mixing of Gibbs sampling on graphs that are sparse on average. *Random Structures Algorithms*, 35(2):250–270, 2009. doi:10.1002/rsa.20276.

[60] E. Mossel and A. Sly. Exact thresholds for Ising-Gibbs samplers on general graphs. *Ann. Probab.*, 41(1):294–328, 2013. doi:10.1214/11-AOP737.

[61] M. Niss. History of the Lenz-Ising model 1920–1950: from ferromagnetic to cooperative phenomena. *Arch. Hist. Exact Sci.*, 59(3):267–318, 2005. doi:10.1007/s00407-004-0088-3.

[62] M. Niss. History of the Lenz-Ising model 1950–1965: from irrelevance to relevance. *Arch. Hist. Exact Sci.*, 63(3):243–287, 2009. doi:10.1007/s00407-008-0039-5.

[63] M. Niss. History of the Lenz-Ising model 1965–1971: the role of a simple model in understanding critical phenomena. *Arch. Hist. Exact Sci.*, 65(6):625–658, 2011. doi:10.1007/s00407-011-0086-1.

[64] E. Olivieri and M. E. Vares. *Large deviations and metastability*, volume 100 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005. doi:10.1017/CBO9780511543272.

[65] E. Olivieri and M. E. Vares. *Large deviations and metastability*, volume 100 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005. doi:10.1017/CBO9780511543272.

[66] R. Peierls. On Ising's model of ferromagnetism. *Mathematical Proceedings of the Cambridge Philosophical Society*, 32(3):477–481, 1936. doi:10.1017/S0305004100019174.

[67] F. N. H. Robinson, S. B. McGrayne, B. Bleaney, E. Kashy, and E. E. Suckling. Magnetism — Encyclopedia Britannica. https://www.britannica.com/science/magnetism. Online; Accessed: 2022-11-04.

[68] A. Schlichting and M. Slowik. Poincaré and logarithmic Sobolev constants for metastable Markov chains via capacitary inequalities. *Ann. Appl. Probab.*, 29(6):3438–3488, 2019. doi:10.1214/19-AAP1484.

[69] J. V. Selinger. *Ising Model for Ferromagnetism*, pages 7–24. Springer International Publishing, Cham, 2016. doi:10.1007/978-3-319-21054-4_2.

[70] I. Seo. Condensation of non-reversible zero-range processes. *Comm. Math. Phys.*, 366(2):781–839, 2019. doi:10.1007/s00220-019-03346-2.

[71] M. Slowik. *Contributions to the potential theoretic approach to metastability with applications to the random field Curie-Weiss-Potts model.* PhD thesis, Technische Universität Berlin, 2012. doi:10.14279/DEPOSITONCE-3202.

[72] D. W. Stroock. *An introduction to Markov processes*, volume 230 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 2005. doi:10.1007/978-3-642-40523-5.

[73] M. Talagrand. *Spin glasses: a challenge for mathematicians*, volume 46 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, 2003. Cavity and mean field models.

[74] T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012. doi:10.1090/gsm/132.

[75] The Editors of Encyclopaedia Britannica. Spin — Encyclopedia Britannica. https://www.britannica.com/science/spin-atomic-physics. Online; Accessed: 2022-11-04.

[76] P. Tindemans and H. Capel. An exact calculation of the free energy in systems with separable interactions. *Physica*, 72(3):433–464, 1974. URL: https://www.sciencedirect.com/science/article/pii/0031891474902092, doi:https://doi.org/10.1016/0031-8914(74)90209-2.

[77] P. Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *J. Phys. Theor. Appl.*, 6(1):661–690, 1907. doi:10.1051/jphystap:019070060066100.

[78] M. Wouts. Slow dynamics for the dilute Ising model in the phase coexistence region. *Comm. Math. Phys.*, 317(2):381–423, 2013. doi:10.1007/s00220-012-1590-0.

[79] Y. Xu and S. Mukherjee. Ising models on dense regular graphs. *Preprint arXiv:2210.13178 [math.ST]*, 2022.