# Contributions to the Potential Theoretic Approach to Metastability with Applications to the Random Field Curie-Weiss-Potts Model

vorgelegt von Diplom-Physiker

## **Martin Slowik**

aus Berlin-Pankow

Von der Fakultät II – Mathematik und Naturwissenschaften der Technischen Universität Berlin zur Erlangung des akademischen Grades

> Doktor der Naturwissenschaften Dr. rer. nat.

> > genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Reinhold Schneider Berichter: Prof. Dr. Anton Bovier Berichter: Prof. Dr. Dmitry Ioffe

Tag der wissenschaftlichen Aussprache: 24.02.2012

Berlin 2012 D 83

## Contents

Preface Motivation An outline of this thesis Acknowledgements	v v ix xi
<ul> <li>Chapter 1. Basic ingredients of the potential theoretic approach</li> <li>1.1. Introduction</li> <li>1.2. Boundary value problems</li> <li>1.3. Laplace equation and Poisson equation</li> <li>1.4. Green function and Green's identities</li> <li>1.5. Connection between equilibrium potential, capacity, and mean hitting times</li> <li>1.6. The Dirichlet and Thomson principle: a dual variational principle</li> <li>1.7. The Berman-Konsowa principle: a variational principle for the lower bound</li> <li>1.8. Renewal equation</li> <li>1.9. Almost ultrametricity of capacities</li> </ul>	1 2 4 8 9 11 14 17 18 21
<ul> <li>Chapter 2. Coupling arguments and pointwise estimates</li> <li>2.1. Introduction</li> <li>2.2. Construction of the Coupling</li> <li>2.3. Bounds on harmonic functions and local recurrence</li> <li>2.4. Cycle decomposition of <i>σ</i>-paths</li> <li>2.5. Bounds on the mean hitting time and the Laplace transform</li> </ul>	25 26 28 33 34 38
<ul> <li>Chapter 3. Metastability</li> <li>3.1. Introduction</li> <li>3.2. Ultrametricity</li> <li>3.3. Sharp estimates on mean hitting times</li> <li>3.4. Asymptotic exponential distribution</li> </ul>	41 42 44 46 50
<ul> <li>Chapter 4. Metastability and Spectral Theory</li> <li>4.1. Introduction</li> <li>4.2. Rough localization of eigenvalues</li> <li>4.3. Characterization of small eigenvalues</li> </ul>	55 55 60 64
<ul> <li>Chapter 5. Random field Curie–Weiss–Potts model</li> <li>5.1. Introduction</li> <li>5.2. Induced measure and free energy landscape</li> <li>5.3. Upper bounds on capacities</li> <li>5.4. Lower bounds on capacities</li> <li>5.5. Sharp estimates on metastable exit times and exponential distribution</li> </ul>	71 72 77 86 97 116
Bibliography	119

## Preface

Man schreibt nicht, weil man etwas zu sagen hat, sondern weil man Lust hat, etwas zu sagen. —Emile M. Cioran

This text is my PhD thesis. It is concerned with several aspects of metastability in stochastic dynamics with a particular focus on the dynamical behavior of disordered mean field spin systems at finite temperature.

### Motivation

**Phenomenology.** *Metastable behavior* of a complex system, either natural or artificial, is causal for a number of interesting phenomena as diverse as delayed freezing of super cooled liquids, folding of large bio-molecules, kinetics of chemical reactions, changes in global climate systems, apparent stability of stock markets, to name but a few. Despite such a diversity of scientific areas ranging from physics, chemistry, biology to economics in which metastable behavior can be observed, the common feature of all these situations is the existence of multiple, well separated time scales. On the corresponding short time scales, the system can only explore at limited part of the available state space. Depending on the time scale under consideration, the state space can be decomposed into several disjoint regions, in which the system is effectively trapped. Within such a region, that may be viewed as *metastable set* to which one may associate a *metastable state*, the system appears to be in a *quasi-equilibrium*. At larger time scales, rapid transitions between metastable states occur which are induced by *random fluctuations*. The main mathematical task we want to address is to analyse such a system on long time scales. In particular, we need some understanding of how the process manages to escape from a metastable state.

In order to illustrate the phenomenon of metastability in more detail, let us consider the *dynamical* behavior of a ferromagnet close to its *first order phase transition*, see [97, 94]. In the framework of equilibrium statistical mechanics, the defining characteristic of a first-order phase transition is a discontinuity in an extensive variable, such as the magnetization, as a function of an intensive variable, for instance the magnetic field.

It is well known that for a ferromagnet above the critical temperature,  $T_c$ , there exists an unique *paramagnetic* phase. Experimentally, one observes that the magnetization m(h) vanishes when switching off the external magnetic field h, i.e.  $m(h) \rightarrow 0$  as  $h \rightarrow 0$ . Below  $T_c$  the behavior of the system is markedly different. Depending on whether the external field approaches zero from below or from above there exist two different phases that can be distinguished by their different magnetization,

$$\lim_{h \downarrow 0} m(h) = m_+ > 0, \qquad \lim_{h \uparrow 0} m(h) = m_- < 0.$$

This phenomenon is called *spontaneous* magnetization. Now, consider the following experiment. Suppose the system is prepared in such a way that it reveals the magnetization  $m(0) = m_+$ . After switching on a small negative magnetic field  $|h| < h_c$ , one may observe that instead of undergoing a rapid phase transition the system remains in an apparently stationary state for a (macroscopic) long time that is characterized by a positive magnetization m(h) > 0.

This situation corresponds to a metastable state. In contrast to that, the magnetization of the equilibrium state has the same sign as the external field h < 0.

In order to avoid the impression that a metastable behavior can be described only by considering static aspects, e.g. an analytic continuation of the critical curves in the phase diagram, let us stress the fact that metastability is a *dynamical* phenomenon.

For this reason, let us now describe the dynamical behavior of a magnetic system in a metastable state. On a microscopic level, the magnetization of an individual (classical) spin can change its sign from +1 to -1 and vice versa induced by internal thermal perturbations. For this reason, at any time small droplets of downward pointing spins appear in a "sea of upward-pointing spins". While the spins in the bulk of such a droplet become aligned with the external field, spins at the surface feel both the negative external field and an positive field coming from spins outside the droplet. In other words, there is a competition between the gain of energy in the bulk of a droplet and the loss of energy at its surface. This causes the dissolution of small droplets whereas a large droplet, once created, can be seen as a gate through which the system escapes rapidly from the metastable to the stable equilibrium state. However, the spontaneous formation of a sufficiently large droplet is a rather rare event which may be seen as an explanation that an escape from a metastable state can only be observed on macroscopic time scales.

**Metastability and diffusions.** From a mathematical point of view, "Kramers' equation has become the paradigm of metastability" <sup>1</sup>. For this reason, we describe in the sequel briefly some aspects of its derivation and some of the rigorous mathematical tools that have been invented in order to study it. Due to the interdependence between experimental observations, rigorous/non-rigorous theoretical analysis and computer simulations of metastable systems, it is rather challenging to give proper credit to the people involved. In particular, the review presented below is far from being complete. For more a detailed review and an outlook of future challenges, we refer to the recent papers [**12**, **99**].

The origin of this equation can be traced back to the study of kinetic of chemical reactions. In most real-world applications, metastable systems are described by many-particle systems. Since the dynamics of a many-particle system is very difficult to analyse either analytically or numerically, model reductions play an important role. For instance, the full phase space of a system describing the dynamics of a chemical reaction has  $\sim 10^{23}$  degrees of freedom. Therefore, instead of studying the dynamics on the high-dimensional surface in the full phase space  $\Sigma$ , one of the first model reduction steps is to consider the dynamics only along the so-called *reaction path*  $\chi = (X_t, \dot{X}_t) \in \Sigma$  in an effective potential that takes into account the interaction with the (thermal) reservoir. As a consequence of such a projection from  $\Sigma$  to  $\chi$ , the resulting dynamics is in general described by a non-Markovian process [57, 58]. Assuming additionally that the noise correlation times in the reservoir are extremely short, one can further use a Markovian approximation for the reduced dynamics [116].

These reduction steps lead to the *first mathematical model* for metastability proposed in 1940 by Kramers [**78**] in order to describe a chemical reaction. It consists of a classical particle of mass one moving in an one-dimensional asymmetric double-well potential U under the influence of Gaussian white noise and friction with coefficient  $\gamma$  which models effectively the thermal reservoir at temperature T. Its equation of motion is given by

$$\gamma^{-1}\ddot{X}_t = -\dot{X}_t - U'(X_t) + \sqrt{2\varepsilon}\,\dot{B}_t,$$

where the parameter  $\varepsilon \equiv \varepsilon(T)$  is temperature dependent and tends to zero when T approaches the absolute zero. In the limit when the friction becomes infinitely strong, this equation results

<sup>&</sup>lt;sup>1</sup>Bovier [**12**, p.2]

#### MOTIVATION

in the simple, one-dimensional diffusion equation

$$dX_t = b(X_t) dt + \sqrt{2\varepsilon} dB_t$$
(0.1)

to which we refer to as *Kramers' equation*, where b(x) = -U'(x) and  $B_t$  is assumed to be a Brownian motion. Notice that local minima of U correspond to metastable states/points in this model. Kramers studied various interesting questions in the context of this model. In particular, he identified the mechanism of noise-assisted reactions and derived an explicit formula, also called *Kramers formula*, for the averaged transition time from the local minimum a to the global minimum b passing through the maximum z that lies in between them,

$$\mathbb{E}_{a}\left[\tau_{b}\right] = \frac{2\pi}{\sqrt{U''(a)|U''(z)|}} \exp\left(\varepsilon^{-1}\left(U(z) - U(a)\right)\right) \left(1 + \mathcal{O}_{\varepsilon}(1)\right). \tag{0.2}$$

Based on experimental data, Arrhenius had already suggested in 1889 that the reaction-rates are proportional to the inverse of the temperature on logarithmic scales [2]. The exponential term in (0.2) reflects *Arrhenius' law*. Let us now have a look at the prefactor. In comparison with a sharply-peaked maximum, it is more likely that the particle, when slightly passing by a flat maximum, returns to its starting well. On the other hand, in a flat minimum the particle gets less often close to the barrier than in a peaked one. This explains why a smaller curvature of the local minimum as well as the maximum leads to a larger averaged transition time.

Let us remark that, in the context of *reaction rate theory*, Eyring had given previously a heuristic derivation of the prefactor based on quantum and classical statistical mechanics computations expressing the prefactor as the ratio of the partition function of the reactants and the "activated complex", respectively, [47]. For a discussion of past and possible future developments in the reaction rate theory we refer the reader to the review by Hänngi, Talkner and Borkovec [61] and the recent paper by Pollak and Talkner [99].

In the context of randomly perturbed dynamical systems, generalizing Kramers' equation to higher dimensions, various dynamical aspects concerning the exit problem of a domain were first analysed rigorously in the seminal work by Freidlin and Wentzell [111, 51]. They invented the idea to use *large deviations in path space* as well as to consider a *Markov chain with exponential small transitions* to model effectively the jumps between different attractors. Let us denote by  $\Gamma$  the set of all path,  $\gamma : [0,T] \rightarrow \mathbb{R}^d$  with arbitrary T. Employing large deviation principles in path space, one can control the probability that the process,  $\{X_t\}$ , stays close to a given path  $\gamma \in \Gamma$  over a time interval [0, T], in the sense that

$$\varepsilon \ln \mathbb{P}\left[\sup_{t \in [0,T]} \|X_t - \gamma_t\| \le \delta\right] = -I(\gamma) + \mathcal{O}_{\varepsilon}(1), \tag{0.3}$$

for  $\delta > 0$  and  $I : \Gamma \to \mathbb{R}_+$  a lower semi-continuous function with compact level sets. For studying a transition between the disjoint neighbourhood A, B of two different local minima, the task is to compute the optimizer of the variational problem  $\inf_{\gamma:A\to B} I(\gamma)$  and to analyze its properties. Notice the analogy, at least to some extend, between this approach and the reaction path considered previously. Various interesting aspects of metastability, e.g. properties of the typical exit path and the asymptotic exponential distribution of exit times, was later proven for random perturbations of dynamical systems of Freidlin-Wentzell type in finite dimensions [52], infinite dimensions [24, 48, 20] and on Riemannian manifolds [107] by means of large deviations in path space. This method has been proven to be robust and rather universal applicable in many different model contexts [77, 92, 93]. For an in-depth overview we refer to a recent monograph by Olivieri and Vares [94].

A limitation of large deviation methods is its precision. It allows to compute for instance mean exit times only up to an multiplicative error of order  $\exp(\mathcal{O}_{\varepsilon}(1))$ . For this reason, it is not possible to resolve the behavior of the process near the saddle point. However, this is an essential task in order to derive rigorously the prefactor in the *Eyring-Kramers formula*, the

### PREFACE

higher-dimensional analog of (0.2). An alternative to the method of large deviations in path space, yielding more precise results beyond logarithmic equivalence, is the *potential theoretic approach to metastability*, systematically developed by Bovier and co-authors, see [**10**, **11**] for a detailed introduction. Its key idea is to express probabilistic quantities of interest in terms of *capacities* and use variational principles to compute the latter. In particular, for reversible diffusion processes of Freidlin-Wentzell type, Bovier, Eckhoff, Gayrard and Klein demonstrated in [**17**] that sharp estimates of capacities can be easily obtained. As a consequence, they gave a first rigorous proof of the prefactor in the classical Eyring-Kramers formula for dimension  $d \ge 1$  and sharp asymptotic estimates of the exponentially small eigenvalues. Later, Berglund and Gentz relaxed the assumptions on the relevant saddle point allowing certain non-degeneracies [**56**].

Recently, Helffer, Nier and Klein have developed a new analytic approach based on *hypoelliptic techniques* initially developed for the regularity analysis of partial differential equations [**63**]. In the reversible diffusion setting and under suitable assumptions on the potential U, they derived rigorously a complete asymptotic expansion of the mean hitting times in powers of  $\varepsilon$  using the so-called *Witten complex* [**62**].

Metastability and model from statistical mechanics. In the sequel, we will give some more points of the enormous work on metastability that has been done in the past 40 years. In contrast to the diffusion setting, where metastable states corresponds to local minima of the potential U, in models coming from statistical mechanics the question of how to characterize metastable states, and how to identify them in a given model context is really an issue.

A first rigorous formulation of metastable states, taking the dynamical aspect into account, dates back to the work of Penrose and Lebowitz in 1971 on the van der Waals limit of Kac potentials [96, 97]. Subsequently, early results on the metastable behavior of the two-dimensional Ising model were obtained by Capocaccia, Cassandro and Olivieri [22].

In [23] Cassandro, Galves, Olivieri and Vares proposed a different approach to metastability that is based on the pathwise analysis of the dynamics. As a first examples, they demonstrated the efficiency of this concept in the Curie-Weiss model describing its dynamical properties. In the sequel, this approach was successfully applied to more realistic models. In the pioneering works [89, 90], Neves and Schonmann studied the metastable behavior of the two-dimensional Ising model under Glauber dynamics in finite volume and very low temperature. This work was later extended to higher dimensions [3, 25], infinite volume but low temperature [32, 33], infinite volume and fixed temperature but vanishing external field [105] and probabilistic cellular automata [28]. In contrast to Glauber dynamics, Kawasaki dynamics is *conservative*, i.e. the total number of particles is fixed. In a series of papers den Hollander et al. [35, 34] and Gaudillière et al. [54] investigated in the nucleation and metastability for particle systems under Kawasaki dynamics in a large box in two and three dimensions at low temperature and low density.

A more recent approach to metastability was developed by Bovier, Eckhoff, Gayrard and Klein [15, 16]. Based on techniques from potential theory, this approach allows to compute sharp estimates for metastable exit times and their asymptotic exponential distribution. Moreover, it establishes quantitatively precise relations between small eigenvalues of the generator associated to the dynamics and mean exit times of metastable domains. In the stochastic Ising model at low temperature, it yields sharp results that go far beyond the logarithmic equivalence obtained by the pathwise approach [19]. This technique was also used to sharpen the previously obtained results on the Kawasaki dynamics in the low temperature regime in a finite box [13] and to prove the first rigorous results in the case of growing boxes [14].

Beside spin systems at very low temperature, the investigation in the dynamical behavior of *disordered mean field systems* is of particular interests. One of the simplest model, from a static

#### AN OUTLINE OF THIS THESIS

point of view, that has been studied intensively in the past, is the random field Curie-Weiss model at finite temperatures. Mathieu and Picco [85] and Fontes, Mathieu and Picco [50] first analyzed the long-time behavior of this model in the case where the random field can take only two values  $\pm \varepsilon$ . By using spectral methods, they obtained the leading exponential asymptotics for metastable exit times. Based on the potential theoretic approach Bovier, Eckhoff, Gayrard and Klein [15] improved the previous results by establishing sharp estimates of transition times between metastable states. They consider the case where the distribution of the random field is discrete but assumes finitely many values. As long as the random field assumes arbitrary values in a finite set, the dynamical behavior of this model can be described effectively by a Markovian dynamics on a lower-dimensional state space. Finally, in two recent papers [6, 5], Bianchi, Bovier and Ioffe analyzed first the metastable behaviour of the Curie-Weiss model at finite temperature in the case of random fields with continuously distributions. By exploiting a dual variational representation of capacities in terms of unit flows, due to Berman and Konsowa [4], they derived in [6] first sharp estimates on averaged metastable exit times including an explicit expression for the prefactor. Secondly, by means of coupling techniques, they showed in [5] that mean metastable exit times are almost constant as functions of the starting configuration within well chosen sets. Additionally, they proved the convergence of the law of normalized metastable exit times to an exponential distribution.

### An outline of this thesis

By now, we have reached a fairly comprehensive understanding of metastable phenomena in at least two different settings: stochastic dynamics of spin systems in the low temperature regime and mean field models at finite temperature. A crucial feature of models at very low temperature is that i) metastable states correspond to local minima of the Hamiltonian and ii) the entropy of paths does not play a role, i.e. the transition between metastable states A, B is realized by microscopic paths in a arbitrary small tube around the optimal microscopic path that is given as the minimizer in the variational problem  $\inf_{\gamma:A\to B} I(\gamma)$ . On the other hand, in stochastic spin systems at finite temperature the entropy of paths really matters. Namely, not only individual microscopic paths have a vanishing small probability but their probability of stay, say, in a tube with radius  $\delta > 0$  around a given microscopic path is very small. For this reason, one has to lump enormously many paths together in order to be able to describe a transition between metastable states. This is usually done by introducing a macroscopic variable and analyzing the dynamical properties of the resulting process. A characteristic feature of mean field models is that the induced dynamics on the coarse-grained level is still Markovian. In such a case, i) metastable states correspond to local minima of the free energy. More important is the fact that ii) by means of the macroscopic variable the dimension of the state space underlying the induced dynamics is diminished, i.e. it remains to study essentially the metastable behavior of a random walk in the free energy landscape. To some extend, this reduction procedure leads to a setting comparable with a low temperature regime where the role of the temperature, however, is replaced by the volume.

The purpose of this thesis is to study metastability for a class of stochastic spin systems at *finite temperature* which are *not* exactly reducible to a low-dimensional model via lumping techniques. Our main objective is to *extend* the potential theoretic approach to metastability for such kind of models by identifying key ingredients for deriving sharp estimates on

- the expected time of a transition from a metastable to a stable state,
- the distribution of the exit time from a metastable state,
- small eigenvalues of the generator.

While the first four chapters of this thesis focus on the general concepts of this approach, a real test of these ideas comes with chapter 5, where we apply this approach to a simple disordered spin system.

Chapter 1 is concerned with some aspects of the profound connection between Markov processes and potential theory that serves as a basis for the potential approach to metastability. The main idea of this approach to express probabilistic objects, e.g. mean hitting times, first in terms of solutions of boundary value problems. Instead of trying to solve approximately the resulting equations, the second idea is to characterize the solution in terms of capacities and equilibrium potentials. Notice that in the setting, we are interested in, the formula for the mean hitting time of a set B when starting from A is only useful if A, B are sufficiently large. However, in such a situation, strict renewal equation breaks down that are crucial to control the equilibrium potential. One of the main objectives in the first chapter is the derive various averaged renewal equations that will be important in later chapters. In order to emphasis the link between probability and potential theory, we give for every statement both a probabilistic and two different analytic proofs.

Chapter 2 is mainly dedicated to establish pointwise estimates of various probabilistic objects. In situations where the process returns very often to the starting set A before escaping to the set B, one expects that the mean hitting time of B is almost constant as a function of a starting configuration  $\sigma \in A$ . In a recent paper [5], Bianchi, Bovier and Ioffe addressed this question for a class of Markov chains with state space  $\{-1, +1\}^N$  by means of coupling and splitting techniques. The main objective in the second chapter is to generalize the coupling construction to a class of Markov chains with state space  $\{1, \ldots, q\}^N$ . In particular, we demonstrate that the splitting techniques, introduced in [5], are rather universal applicable. As a byproduct, we obtain rough estimates on the oscillations of harmonic functions within certain sets. Let us point out that these estimates are the second key ingredient in order to derive sharp bounds on mean hitting times.

In Chapter 3 we introduce the notion of a set of metastable sets. This definition is a natural generalization of the set of metastable points, first introduced in [15], and will be the starting point for our further investigations. Notice that in the context of reversible diffusions a similar generalization of metastable points was already discussed in [10]. One advantage of the potential theoretic approach to metastability is that it suffices to establish sharp bounds on capacities and to derive rough bounds on the corresponding equilibrium potentials in order to compute precisely metastable exit times, see [9, 11]. The later is previously done by means of exact renewal equations. Equipped with the results obtained in the first two chapters we demonstrate how averaged renewal equations combined with the rough estimates on the regularity of harmonic functions can be used to compute precisely the mean exit time from a metastable set and to prove the convergence of normalized metastable exit times to an exponential distribution. Notice that similar ideas were already used in [17] and [5].

In Chapter 4 we investigate the relation between small eigenvalues of the generator and mean hitting times of certain metastable sets. Previously, in the study of reversible diffusions on  $\mathbb{R}^d$  or Markov chains on finite state spaces, a sharp link between metastable points and small eigenvalues of the generator could be established [18, 16]. Let us stress the fact that the method used in [18] is based on regularity properties of eigenfunctions. However, in the setting we are interested in, such sufficient control of the oscillations of eigenfunctions within certain sets is still be missing. For this reason, we invent a different approach that is based on an a posteriori error analysis and pointwise estimates of equilibrium potentials. Assuming that there are at most as many low-lying eigenvalues as metastable sets and taking a non-degeneracy condition into account, this approach allows to associate to each metastable set one simple eigenvalue of the generator. Moreover, every such eigenvalue is equal to the inverse of the mean exit time from this set up to exponentially small errors.

#### ACKNOWLEDGEMENTS

As an application of the theory developed above, we study in the final Chapter 5 the dynamical behavior of the *random field Curie-Weiss-Potts model* at finite temperature and with a continuous distribution of the random field. This model belongs to the class of disordered mean-field spin systems that cannot be reduced exactly to a lower-dimensional system. Since the free energy landscape in the thermodynamic limit is deterministic, from a static point of view, is model is one of the less disordered. The author's contribution in this chapter is to generalize the results that were previously obtained in the study of the random-field Curie-Weiss model, see [**6**].



FIGURE 1. Logical dependence among the chapters

### Acknowledgements

Foremost, I would like to express my sincere gratitude to my thesis advisor, Prof. Dr. Anton Bovier. His lecture on "equilibrium statistical mechanics of ordered and disordered system", I attended as a diploma student, built me a bridge from the world of theoretical physics to the one of probability theory and mathematical physics. While he has given me much freedom to develop my own ideas, I could always asked him for advice. His generous support and his instructive advices have accompanied me all along the way from my early computations of bounds on the spectral gap of the so-called random walk on a dog, beyond the, unfortunately not successful, attempt to analyze the metastable behavior in the east model to the submission of this thesis. Since not every question can be answered with today knowledge, he taught me to judge in advance what might be a good question. I have really enjoyed the time as a member of his group during the last years.

Further, I thank Prof. Dr. Andreas Eberle and Prof. Dr. Patrik Ferrari for giving me the opportunity to take the responsibility for the exercise classes of the lectures "Algorithmische Mathematik 2" and "Angewandte Stochastik" and for various seminars. It was a great experience, and I learnt a lot. In particular, these teaching duties were really helpful during the periods when I got stuck with my own research.

I have greatly enjoyed the very likeable working atmosphere of the Bonn Probability Group, well-attended seminars, illuminating discussions with visitors and much more. I would like to warmly thank my colleagues Dr. Giada Basile, Dr. Nicola Kistler, Adéla Švejda, Daniel Gruhlke, Nikolaus Schweizer, Rene Frings, Shidong Wang for greatly contributing to the cheerful atmosphere. I want to mention and thank in particular Daniel Gruhlke with whom I have spent

### PREFACE

many hours discussing all kinds of problems and progress in the first two chapters of my work. I like to thank as well Nikolaus Schweizer, Dr. Evangelia Petrou and Adéla Švejda for reading parts of my thesis and helping me to enhance considerably the quality of this work.

Furthermore, I thank Prof. Dr. Dmitry Ioffe for agreeing to be the coexaminer of this thesis and Prof. Dr. Reinhold Schneider for being part of the jury.

Finally, I would like to express my gratitude to my parents and my friends for the love and support I constantly received.

Throughout the research on my thesis, I have received financial support from the Weierstrass Institute for Applied Analysis and Stochastics (WIAS) in Berlin, the Institute of Applied Mathematics (IAM) at Bonn University, the German Research Council through the SFB 611 "Singuläre Phänomene und Skalierung in mathematischen Modellen", the International Research Trainings Group "Stochastic Models of Complex Processes" and the Bonn International Graduate School (BIGS). The kind hospitality of both the WIAS and IAM is gratefully acknowledged.

xii

### CHAPTER 1

## Basic ingredients of the potential theoretic approach

In this first chapter, we present some aspects of the profound connection between Markov processes and potential theory. These serve as a basis for the *potential theoretic approach to metastability* developed systematically by Bovier, Eckhoff, Gayrard and Klein in [**15**, **16**]. In later applications, we are mainly interested in the metastable behavior of processes coming from statistical mechanics. Typically, the state space for such kind of processes is large but finite. For this reason, we focus our attention to the study of stochastic processes on *finite state spaces* and in *discrete time* throughout this thesis. Let us emphasize that the potential theoretic approach to metastability is neither limited to finite state spaces nor to processes in discrete time. For instance, in the study of local Kawasaki dynamics in finite volumes with open boundary conditions [**13**] and in volumes that are growing to infinity [**14**] this approach was successfully applied to Markov chains in continuous time. In the context of diffusion processes in  $\mathbb{R}^d$ , the first rigorous proof of the classical *Eyring-Kramers formula* [**17**] and the precise control of the small eigenvalues of the associated generator [**18**, **45**] rely as well on this method.

This chapter is organized as follows. In Section 1.1, after giving a brief introduction to classical potential theory and explaining its relation to probability theory, we specify the setting that is actually used in later applications. In Section 1.2, we treat the Dirichlet problem of finding a function which solves a Feynman-Kac-Poisson equation in a given region and assumes the values of a prescribed function at the boundary. Markov chains can be used to characterize the solution of such a Dirichlet problem and to prove existence and uniqueness. Various applications of this probabilistic representation are given in Section 1.3. A different way to represent the solution of a Dirichlet problem is in terms of the Green function that will be introduced in Section 1.4. As presented in Section 1.5, one major achievement of the potential theoretic approach is the observation that instead of searching for an approximate solution to a Dirichlet problem, the solution can alternatively be represented in terms of capacities which satisfy variational principles. As we will see in Section 1.6 and 1.7, the formulation in terms of the Dirichlet problem. The final Sections 1.8 and 1.9 deal with averaged versions of renewal equations and the almost ultrametricity of capacities which will be important in later chapters.

**Contributions in this chapter.** Most of the material presented in this chapter is of preparatory value for subsequent chapters. In this respect, this chapter does not contain major new contributions to the understanding of the fundamentals of the potential theoretic approach. Nevertheless, the author contributes some, to the best of our knowledge, novel pieces to this subject:

- three different proofs of a lemma that allows to split the mean hitting time of a set *B* starting in *A* into the contributions coming from the returns to the set *A* and the time it takes in average to hit *B* without returning to *A*, see Lemma 1.14;
- three different proofs of an averaged renewal equation for hitting probabilities and for the Laplace transform of hitting times, see Section 1.8.

• a generalization of the statement about almost ultrametricity of capacities, see Lemma 1.26, and a purely analytic proof thereof.

Smaller bits and pieces are:

• a presentation of a simple proof of the Dirichlet principle based on the Cauchy-Schwarz inequality and a slightly generalization of the Thomson principle in terms of super-harmonic functions, see Section 1.6.

### 1.1. Introduction

One of the fundamental problems in potential theory is the classical Dirichlet problem. It originates from the study of physical questions coming from electrostatics. The investigation in existence and uniqueness of solutions to it has a long and venerable history which can be traced back e.g. to the work of Dirichlet [38], Poincaré [98], Kellogg [75], Lebesgue [80]. Various analytic methods have been invented to construct a solution: Schwarz alternating method, variational methods to name only two of them. The idea of Schwarz's alternating method is to decompose first the domain into two or more sub domains having a non-empty intersection. In a second step, similar Dirichlet problems are solved alternately on these sub domains, those values on the additional boundaries are given by the previous solution. In comparison to the Schwarz alternating method, the variational method is based on the fact that the solution of the Dirichlet problem assuming certain values at the boundary is also given as the minimizer of an appropriate energy functional.

Notice that, in general, the Dirichlet problem is not always solvable without assuming a certain regularity of the domain's boundary. This fact was observed at an early stage by considering the example of an punctured sphere or a sharp inward-pointing cups (Lebesgue's thorn).

A first probabilistic solution to the classical Dirichlet problem was given by Kakutani [72]. His pioneering work and the extensive study by Doob [41] and Hunt [67, 68, 69] established the profound connection between potential theory and Markov processes. To some extend these two theories are mathematically equivalent. This allows to translate a proof of a result in one theory into a proof of a corresponding result in the other theory, which we will demonstrate in Section 1.5 and 1.8. There is now a large literature on (transient) potential theory and Markov processes, including books e.g. by Blumenthal and Getoor [8], Port and Stone [101], Karatzas and Shreve [73, chap. 4] and the comprehensive treatment by Doob [43]. A nice introduction into this topic can be found in [102] and [27].

In the context of Markov chains on finite or denumerable state spaces, the connection to (recurrent) potential theory was developed by Doob [42] and Kemeny, Knapp and Snell [76]. While classical potential theory had already been well understood before its connection to Brownian motion was discovered, Markov chains and discrete potential theory were more or less simultaneously studied. Historically, in his famous paper "Random walks and electric currents in networks" [88], Nash-Williams first linked the property of transience or reccurence of an irreducible reversible Markov chain to structural properties of the underlying electric network. While Pólya [100] first proved by using purely probabilistic arguments that a simple random walk on  $\mathbb{Z}^d$  returns to its starting position with probability 1 if and only if  $d \leq 2$ , Nash-Williams showed that this reccurrence holds on graphs if and only if the effective resistance between the starting position and infinity of the corresponding electric network is infinite. Although the discrete potential theory has proven to be a robust tool to study properties of Markov chains, its application attracted new attention with the beautiful elementary book by Doyle and Snell [44]. Since that time various aspects of Markov chains have been studied exploiting the interdependence between probabilistic objects and its analytic counterparts and

2

there is now a large literature on this topic including books e.g. by Woess [114], Telcs [108], Levin, Peres and Wilmer [82], Lyons and Peres [83].

The potential theoretic approach to metastability is based on the observation that most quantities of physical interest can be represented as solutions of certain *Dirichlet problems* with respect to the generator of the dynamics. More important is the fact that the corresponding solutions can be expressed in terms of *capacities* which satisfy *variational principles*. With proper physical insight, these variational principles allow us to derive reasonable upper and lower bounds on capacities.

Since various notions used in the potential theoretic approach are borrowed from electrostatics we will now explain briefly their physical background. For more details, we refer to the classical physical literature on this subject, see [71, 106].

**Electrostatic interpretation.** Many problems in electrostatics appearing in practical applications concern finite regions of space, i.e. bounded connected open subsets  $D \subset \mathbb{R}^d$  having a smooth boundary  $\partial D$  with surface measure S, with or without charges inside D and with predescribed values given by a function on the bounding surfaces  $\partial D$ . The connection between a distribution of free charges q(dx) inside D and the values of the electric displacement D on  $\partial D$  is given by Gauss's law. The latter states that the flux of the electric displacement across the boundary is proportional to the sum of free charges enclosed by D. Hence, by Gauss' divergence theorem,

$$\int_{D} \operatorname{div} \boldsymbol{D}(x) \, \mathrm{d}x = \int_{\partial D} \langle \boldsymbol{D}(x), \boldsymbol{n}(x) \rangle S(\mathrm{d}x) = \int_{D} q(x) \, \mathrm{d}x, \qquad (1.1)$$

where we assume that the signed measure q(dx) has a density, i.e. q(dx) = q(x) dx. Here, n(x) denotes the outwardly directed surface normal of unit length for the surface  $\partial D$  at x. In the absence of a dielectric material, the constitutive relation between the electric field E and the electric displacement D is given by  $D(x) = \varepsilon_0 E(x)$ , where  $\varepsilon_0 > 0$  is the permittivity of free space.

On the other hand, Faraday's law of induction in its differential form relates the curl of the electric field to the time rate of change of the magnetic flux that vanishes in a static situation. Thus, this leads to the Maxwell equations

div 
$$\mathbf{D}(x) = q(x)$$
 and rot  $\mathbf{E}(x) = 0$ ,  $x \in D$  (1.2)

which imply that there exists a potential  $\varphi: \overline{D} \to \mathbb{R}$  such that  $E = -\nabla \varphi$ . One of the common problems in electrostatic arising from (1.2) is the determination of the potential  $\varphi$  in a vacuum given a charge distribution q inside of D and a function u specifying the values of the potential at the boundary, i.e. to solve the Dirichlet problem

$$\begin{cases} \Delta \varphi(x) = -\frac{1}{\varepsilon_0} q(x), & x \in D \\ \varphi(x) = u(x), & x \in \partial D. \end{cases}$$
(1.3)

Originally derived by Green [59] in 1828, an immediate consequence of the divergence theorem applied to  $\varphi \nabla \psi$ , where  $\varphi, \psi \in C^2(\overline{D})$ , is *Green's first identity* 

$$\int_{D} \left( \varphi(x) \Delta \psi(x) + \langle \nabla \varphi(x), \nabla \psi(x) \rangle \right) \mathrm{d}x = \int_{\partial D} \varphi(x) \,\partial_n \psi(x) \, S(\mathrm{d}x)$$

where  $\partial_n \varphi(x) = \langle \nabla \varphi(x), \boldsymbol{n}(x) \rangle$  denotes the normal derivative of  $\varphi$  on  $\partial D$ . Green's second identity is obtained by interchanging  $\varphi$  and  $\psi$  and subtracting, i.e.

$$\int_{D} \left( \varphi(x) \, \Delta \psi(x) \, - \, \psi(x) \, \Delta \varphi(x) \right) \mathrm{d}x \; = \; \int_{\partial D} \left( \varphi(x) \, \partial_n \psi(x) \, - \, \psi(x) \, \partial_n \varphi(x) \right) S(\mathrm{d}x).$$

Resulting from Green's second identity, an integral representation of the solution to (1.3) is given by

$$\varphi(x) = \frac{1}{\varepsilon_0 \,\omega_{d-1}} \int_D q(z) \,G_D(x,z) \,\mathrm{d}z - \frac{1}{\omega_{d-1}} \int_{\partial D} u(z) \,\partial_n G_D(x,z) \,S(\mathrm{d}z) \tag{1.4}$$

where  $\omega_{d-1} = 2 \pi^{d/2} / \Gamma(\frac{d}{2})$  is the surface area of a (d-1)-sphere in  $\mathbb{R}^d$ .  $G_D(x, y)$  is called the Green function that is, for any  $y \in \overline{D}$ , the fundamental solution in distributional sense of  $\Delta G_D(x, y) = -\omega_{d-1} \delta(x - y)$  on D that vanishes on  $\partial D$ .

Of particular interest is the situation where the boundary of the region D consists of two separate conducting surfaces  $\partial A$  and  $\partial B$  which are hold at potential  $\varphi_A = 1$  and  $\varphi_B = 0$ , respectively, while the region D is free of electrical charges. The *capacity*, C, of this capacitor is defined by  $C := Q_A/(\varphi_A - \varphi_B) = Q_A$  where

$$Q_A = -\varepsilon_0 \int_{\partial A} \langle \boldsymbol{E}(x), \boldsymbol{n}(x) \rangle S(\mathrm{d}x) = \varepsilon_0 \int_{\partial A} \partial_n \varphi(x) S(\mathrm{d}x)$$
(1.5)

is the total charge on  $\partial A$ . Notice that Green's first identity immediately gives rise to the following representation of the capacity as a quadratic form

$$C = Q_A = \varepsilon_0 \varphi_A \int_{\partial A} \partial_n \varphi(x) S(\mathrm{d}x) + \varepsilon_0 \varphi_B \int_{\partial B} \partial_n \varphi(x) S(\mathrm{d}x)$$
  
=  $\varepsilon_0 \int_{\partial D} \varphi(x) \partial_n \varphi(x) S(\mathrm{d}x) - \varepsilon_0 \int_D \varphi(x) \Delta \varphi(x) \,\mathrm{d}x = \varepsilon_0 \int_D \|\nabla \varphi(x)\|^2 \,\mathrm{d}x.$ 

The advantage of this representation is that it allows to establish dual variational principles as we will see in Section 1.6.

Setting. In what follows, let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and  $\{\mathcal{F}_t\}_{t\in\mathbb{N}_0}$  be an increasing sequence of sub- $\sigma$ -algebras of  $\mathcal{F}$ , which are fixed. Further, let  $(\mathcal{S}_N, \mathcal{B}_N)$  be a measurable space where  $\mathcal{S}_N$  is assumed to be a finite set. The cardinality of  $\mathcal{S}_N$  depends on an addition parameter N and will diverge as N tends to infinity. Elements of  $\mathcal{S}_N$  are denoted by Greek letters  $\sigma, \eta$  and called configurations. We consider a Markov chain  $\{\sigma(t)\}_{t\in\mathbb{N}_0}$  on  $\mathcal{S}_N$  with transitions probabilities  $p_N(\sigma, \eta)$ . Assume that the dynamics is *irreducible* and *reversible* with respect to a unique invariant measure  $\mu_N$ . In particular, the transition probabilities satisfy the detailed balance condition  $\mu_N(\sigma) p_N(\sigma, \eta) = \mu_N(\eta) p_N(\eta, \sigma)$  for all  $\sigma, \eta \in \mathcal{S}_N$ . By  $L_N$  we denote the generator which acts on functions  $f: \mathcal{S}_N \to \mathbb{R}$  as

$$(L_N f)(\sigma) = \sum_{\eta \in \mathcal{S}_N} p_N(\sigma, \eta) (f(\eta) - f(\sigma)).$$
(1.6)

Further, we denote by  $\mathbb{P}_{\nu}$  the law of the Markov chain given that it starts with initial distribution  $\nu$  and by  $\mathbb{E}_{\nu}$  the expectation with respect to  $\mathbb{P}_{\nu}$ . If the initial distribution is concentrated on a single configuration  $\eta$ , we write simply  $\mathbb{P}_{\eta}$  and  $\mathbb{E}_{\eta}$ , respectively. For any  $A \subset S_N$ , let  $\tau_A$  be the first hitting time of the set *A* after time zero, i.e.

$$\tau_A := \inf \{t > 0 \mid \sigma(t) \in A\}.$$

When the set A is a singleton  $\{\eta\}$  we write simply  $\tau_{\eta}$  instead of  $\tau_{\{\eta\}}$ .

### 1.2. Boundary value problems

In this section, we turn our attention to the discrete analog of the classical *Dirichlet boundary* value problem. That is, given a non-empty subset  $D \subset S_N$ , measurable functions  $g, k: D^c \to \mathbb{R}$ 

and a measurable function  $u: D \to \mathbb{R}$ , the task is to find a (bounded) function  $f: S_N \to \mathbb{R}$ which satisfies

$$\begin{cases} (L_N f)(\sigma) - k(\sigma) f(\sigma) = -g(\sigma), & \sigma \in D^c \\ f(\sigma) = u(\sigma), & \sigma \in D, \end{cases}$$
(1.7)

where  $D^c = S_N \setminus D$ . Provided that such a function f exists, it will be called *solution to the Dirichlet problem*.

While the probabilistic solution of (1.7) in the discrete-time setting is well established in cases where the function k vanishes or is identical to a constant, see for instance [91], we could not find a reference dealing with the probabilistic representation in the discrete-time context for general function k, although we expect that it exists. For this reason and for later reference, we will give below a proof of the corresponding statement.

Based on the deep connection between Markov processes and martingales, the probabilistic method allows to write down immediately a very likely candidate for a solution of (1.7). While uniqueness is obtained by applying Doob's optional stopping theorem to a suitable martingale, the proof of existence is in general more subtle, since the smoothness of a solution nearby and at the boundary of the domain depends itself on the regularity of the boundary. In the context of a discrete-time Markov chain on a finite state space the problem concerning the regularity of a solution at the boundary of the domain is not an issue and the existence can be easily achieved by using the Markov property.

To start with, consider a Markov chain  $\{\sigma(t)\}$  in discrete-time with generator  $L_N$ . Then, for all (bounded) measurable functions f, it holds that

$$f(\sigma(t)) - \sum_{s=0}^{t-1} (L_N f)(\sigma(s))$$
(1.8)

is a martingale with respect to  $\mathbb{P}_{\sigma}$  for all  $\sigma \in S_N$ . The following lemma provides us a useful equivalence to (1.8) being a martingale.

**Lemma 1.1.** Let  $\{\mathcal{F}_t\}$  be a filtration such that the Markov chain  $\{\sigma(t)\}$  on  $\mathcal{S}_N$  is adapted. Let  $f, k : \mathcal{S}_N \to \mathbb{R}$  be (bounded) measurable functions. Further suppose that the function k satisfies  $\min_{\eta \in \mathcal{S}_N} k(\eta) > -1$ . Then, for all  $\sigma \in \mathcal{S}_N$ , (1.8) is a  $\mathbb{P}_{\sigma}$ -martingale if and only if

$$M_{t} = f(\sigma(t)) \prod_{s=0}^{t-1} \frac{1}{1+k(\sigma(s))} + \sum_{s=0}^{t-1} \left(k(\sigma(r)) f(\sigma(r)) - (L_{N}f)(\sigma(r))\right) \prod_{r=0}^{s} \frac{1}{1+k(\sigma(r))}$$
(1.9)

is a  $\mathbb{P}_{\sigma}$ -martingale.

Here we use the convention that empty products, i.e. products where the upper index is smaller than the lower one, are equal 1 and empty sums are equal 0.

Proof. In order to simplify notations and for later reference, we set

$$X_t := f(\sigma(t)) - \sum_{s=0}^{t-1} (L_N f)(\sigma(s))$$
 and  $V_t := \prod_{s=0}^{t-1} \frac{1}{1+k(\sigma(s))}$ 

Notice that the process  $\{V_t\}$  is previsible and, due to the assumption on k, locally bounded. Thus, for all  $\sigma \in S_N$ , the theorem of discrete stochastic integrals implies that  $\{X_t\}$  is a  $\mathbb{P}_{\sigma}$ -martingale if and only  $\{(V \bullet X)_t\}$  is a  $\mathbb{P}_{\sigma}$ -martingale. On the other hand, a straight forward computation shows that

$$(V \bullet X)_t = \sum_{s=1}^t V_s (X_s - X_{s-1}) = \sum_{s=1}^t V_s (f(\sigma(s)) - f(\sigma(s-1)) - (L_N f)(\sigma(s-1)))$$
  
=  $V_t f(\sigma(t)) - V_1 f(\sigma(0)) + \sum_{s=1}^{t-1} (V_s - V_{s+1}) f(\sigma(s)) - \sum_{s=0}^{t-1} V_{s+1} (L_N f)(\sigma(s))$   
=  $M_t - f(\sigma(0)).$  (1.10)

Since  $f(\sigma(0))$  is a constant under  $\mathbb{P}_{\sigma}$ , we conclude the proof of this lemma.

The martingale (1.9) is the key ingredient to prove existence and uniqueness.

**Proposition 1.2.** Let  $k: D^c \to \mathbb{R}$  be a bounded function. Suppose that  $\min_{\eta \in D^c} k(\eta) > -1$  and set  $\kappa := 1 + \min_{\eta \in D^c} k(\eta)$ . Further assume that for some  $\sigma \in D^c$ 

$$\frac{1}{\kappa} \cdot \limsup_{s \to \infty} \left( p_D^s(\sigma, D) \right)^{1/s} < 1$$
(1.11)

where  $p_D(\sigma, \eta)$  denotes the transition probabilities of a sub-Markov chain that is stopped at the arrival of D. Then,

$$f(\sigma) = \mathbb{E}_{\sigma} \left[ u(\sigma(\tau_D)) \prod_{s=0}^{\tau_D - 1} \frac{1}{1 + k(\sigma(s))} + \sum_{s=0}^{\tau_D - 1} g(\sigma(s)) \prod_{r=0}^{s} \frac{1}{1 + k(\sigma(r))} \right]$$
(1.12)

for every  $\sigma \in D^c$  and  $f \equiv u$  on D is a solution of the Dirichlet problem (1.7). In particular, the solution is unique.

**Remark 1.3.** First of all notice that if (1.11) is satisfied for some  $\sigma \in D^c$  and  $\eta \in D$ , then, by irreducibility, it holds simultaneously for all  $\sigma \in D^c$  and  $\eta \in D$ . Moreover, provided that the function k has the property that  $k(\eta) > 0$  for all  $\eta \in D^c$  then (1.11) is immediately satisfied.

*Proof.* First, let us consider a function f that solves (1.7), i.e.  $(L_N f) - kf = -g$  on  $D^c$ . Further, it is seen from Lemma 1.1 that

$$M_t^{\tau_D} := f(\sigma(t \wedge \tau_D)) \prod_{s=0}^{t \wedge \tau_D - 1} \frac{1}{1 + k(\sigma(s))} + \sum_{s=0}^{t \wedge \tau_D - 1} g(\sigma(s)) \prod_{r=0}^s \frac{1}{1 + k(\sigma(r))}$$

is a local martingale up to time  $\tau_D$ . Given a localizing sequence  $\tau_D^n \uparrow \tau_D$  of stopping time, we construct another localizing sequence by

$$\tau_n := \tau_D^n \wedge \inf \{ t > 0 \mid V_t > n \}.$$

$$(1.13)$$

This implies that for each *n* there exists  $K_n < \infty$  such that

$$V_{\tau_n} = \frac{V_{\tau_n - 1}}{1 + k(\sigma(\tau_n - 1))} \le \max_{\eta \in D^c} \frac{n}{1 + k(\eta)} \le K_n,$$
(1.14)

where we exploit the assumption on the function k. In view of the martingale transform (1.10), (1.14) allows us to bound

$$\left| M_t^{\tau_n(\omega)} - M_{t-1}^{\tau_n(\omega)} \right| \leq 2 \max_{\eta \in \mathcal{S}_N} \left| f(\eta) \right| K_n < \infty$$

uniformly for all  $t \in \mathbb{N}_0$  and  $\omega \in \Omega$ . Since the dynamics is assumed to be irreducible and the state space is finite, it holds that  $\mathbb{E}_{\sigma}[\tau_D] < \infty$ . Hence, an application of the optional stopping theorem to the localization sequence  $\tau_n$  reveals

$$f(\sigma) = \mathbb{E}_{\sigma} \left[ u(\sigma(\tau_n)) \prod_{s=0}^{\tau_n - 1} \frac{1}{1 + k(\sigma(s))} + \sum_{s=0}^{\tau_n - 1} g(\sigma(s)) \prod_{r=0}^{s} \frac{1}{1 + k(\sigma(r))} \right],$$
(1.15)

for all  $\sigma \in D^c$ . The assertion (1.12) now follows provided we can interchange the limit  $n \to \infty$ and the expectation value. Notice that the random variable involved in the expectation (1.15) is bounded from above by

$$\max_{\eta \in D} |u(\eta)| + \tau_D(\omega) \cdot \max_{\sigma \in D^c} |g(\eta)| \quad \text{or} \quad \left( \max_{\eta \in D} |u(\eta)| + \frac{1}{1 - \kappa} \cdot \max_{\sigma \in D^c} |g(\eta)| \right) \kappa^{-\tau_D(\omega)},$$

for  $\mathbb{P}_{\sigma}$ -a.a.  $\omega \in \Omega$  depending on whether  $\kappa \geq 1$  or  $\kappa < 1$ . In the later, due to the irreducibility and the assumption (1.11)

$$\mathbb{E}_{\sigma}\left[\kappa^{-\tau_{D}}\right] = \sum_{s=1}^{\infty} \kappa^{-s} p_{D}^{s}(\sigma, D) < \infty, \qquad \forall \sigma \in D^{c}.$$
(1.16)

Hence, we can apply Lebesque's dominated convergence theorem. This proves the uniqueness of representation (1.12). On the other hand, by using the Markov property, a straight forward computation shows that for any  $\sigma \in D^c$ 

$$(P_N f)(\sigma) = \mathbb{E}_{\sigma} [\mathbb{1}_{\tau_D \le 1} u(\sigma(\tau_D))] + (1 + k(\sigma)) \mathbb{E}_{\sigma} \left[ \mathbb{1}_{\tau_D > 1} \mathbb{E}_{\sigma(1)} \left[ u(\sigma(\tau_D)) V_{\tau_D} + \sum_{s=1}^{\tau_D - 1} g(\sigma(s)) V_{s+1} \right] \right] = (1 + k(\sigma)) f(\sigma) - g(\sigma).$$
s complete the proof.

This complete the proof.

**Remark 1.4.** Concerning the function k, a natural question to ask is what is the reason behind the fact that the probabilistic representation (1.12) is infinite as soon as  $k(\sigma) = -1$  for some  $\sigma \in D^c$ . A closer look at (1.11) and (1.16) reveals that the finiteness of (1.12) is related to the fact that  $\kappa$  is within the radius of convergence of the corresponding Neumann series. In order to illustrated this fact, consider for simplicity the Dirichlet problem for vanishing u and constant function k. We denote by  $L_D$  the restriction of the generator to  $D^c$ , i.e.  $L_D(\sigma, \eta) =$  $L_N(\sigma,\eta)$  for  $\sigma,\eta\in D^c$  and by  $P_D$  the corresponding restriction of the transition matrix  $P_N$ . If  $L_D - k \cdot I$  is regular, the analytic solution to (1.7) is given by  $f = (k \cdot I - L_D)^{-1}g$ . On the other hand, provided that  $\frac{1}{1+k} \|P_D\| < 1$ , we can represent the inverse in terms of a Neumann series, i.e.

$$\left(k \cdot I - L_D\right)^{-1} = \frac{1}{1+k} \left(I - \frac{1}{1+k} P_D\right)^{-1} = \sum_{s=0}^{\infty} \left(\frac{1}{1+k}\right)^{-(s+1)} P_D^s$$
(1.17)

The right-hand side of (1.17) correspond to the probabilistic representation, see (1.31).

Remark 1.5. Let us briefly comment on the similarities and differences in the probabilistic representation of a solution to the Dirichlet problem in discrete-time compared to the one in continuous-time. The latter in given by

$$f(\sigma) = \mathbb{E}_{\sigma} \left[ u(\sigma(\tau_D)) e^{-\int_0^{\tau_D} k(\sigma(s)) \, \mathrm{d}s} + \int_0^{\tau_D} g(\sigma(s)) e^{-\int_0^s k(\sigma(r)) \, \mathrm{d}r} \, \mathrm{d}s \right].$$
(1.18)

If  $k \ge 0$  and  $L_N \equiv \frac{1}{2}\Delta$ , the Feyman-Kac formula (1.18) can be interpreted as a *Brownian* motion with killing of particles at the same rate k that is stopped at the arrival of the boundary of D. In particular, the probability that a particle survives up to time t, conditioned on the path  $\{\sigma(s)\}_{0 \le s \le t}$ , is equal to  $\exp(-\int_0^t k(\sigma(s)) ds)$ . In discrete-time, (1.12) suggests that the situation is equivalent to a *random walk with killing* where the conditional survival probability of a particle up to time t is then  $\prod_{s=0}^{t} 1/(1 + k(\sigma(s)))$ .

In the case when  $-1 < k \leq 0$ , (1.12) can be seen as a random walk of particles that branches of with the same rate k, and an analog interpretation can be given to the Feyman-Kac formula in the continuous-time setting. For obvious reasons, the solution may explosion in

finite time such that we have to impose conditions on the function k. Further we want to stress that if k < -2, the probabilistic representation (1.12) is finite and has no counterpart in the continuous-time setting.

**Remark 1.6.** Consider the case that the function k vanishes on  $D^c$ . Then the Dirichlet problem (1.7) is known as *Laplace equation* and *Poisson equation*, respectively, depending on whether either  $g \equiv 0$  or  $u \equiv 0$ . In the case  $k \neq 0$ , (1.7) is also called *Feynman-Kac* or *Feynman-Kac*-*Poisson equation*.

### 1.3. Laplace equation and Poisson equation

Proposition 1.2 is a crucial element in the theory since it allows us to compute many interesting expectation values for Markov chains by solving an appropriate boundary value problem. In the sequel, we will consider various corresponding applications.

The key quantity in the potential theoretic approach is the  $\lambda$ -equilibrium potential,  $h_{A,B}^{\lambda}$ . In order to define it, consider a *capacitor*, (A, B), build up by two disjoint subsets  $A, B \subset S_N$ . Then,  $h_{A,B}^{\lambda}$  is defined as the solution of the Dirichlet problem (1.7) with  $g \equiv 0$  and  $k \equiv e^{\lambda} - 1$ , for  $\lambda \geq 0$ , subject to the boundary condition given by the indicator function  $\mathbb{1}_A$  on the set A. The corresponding boundary value reads

$$\begin{cases} (L_N f)(\sigma) - (e^{\lambda} - 1) f(\sigma) = 0, & \sigma \in (A \cup B)^c \\ f(\sigma) = \mathbb{1}_A(\sigma), & \sigma \in A \cup B. \end{cases}$$
(1.19)

A related quantity is the  $\lambda$ -equilibrium measure,  $e_{AB}^{\lambda}$ , which is defined through

$$e_{A,B}^{\lambda}(\sigma) := -(L_N^{\lambda} h_{A,B}^{\lambda})(\sigma), \qquad \forall \sigma \in \mathcal{S}_N,$$
(1.20)

where we introduce the operator  $L_N^{\lambda} := L_N - (e^{\lambda} - 1)$  to shorten notation. In case  $\lambda = 0$ , we simply write  $h_{A,B}$  and  $e_{A,B}$  instead of  $h_{A,B}^0$  and  $e_{A,B}^0$ .

Clearly, the equilibrium measure is non-vanishing only on the boundary of the sets A and B. An immediate consequence of Proposition 1.2 is that the equation (1.19) has a unique solution for all  $\lambda \geq 0$ . Moreover, in view of (1.12), the function  $h_{A,B}^{\lambda}$  describes the following probabilistic objects. Namely, in the case  $\lambda > 0$  the  $\lambda$ -equilibrium potential,  $h_{A,B}^{\lambda}$ , represents the Laplace transform of the hitting time,  $\tau_A$ , of the set A taking into account that the Markov chain  $\{\sigma(t)\}$  starting in  $\sigma$  gets kill at the arrival of B. For  $\lambda = 0$ , the equilibrium potential has a natural interpretation in terms of *hitting probabilities*. Beside  $h_{A,B}^{\lambda}$  the  $\lambda$ -equilibrium measure has as well a probabilistic representation that can be derived by using the Markov property. All in one, we have for  $\lambda \geq 0$ 

$$\mathbb{E}_{\sigma}\left[e^{-\lambda\tau_{A}}\mathbb{1}_{\tau_{A}<\tau_{B}}\right] = \begin{cases} h_{A,B}^{\lambda}(\sigma), & \sigma \notin A \cup B, \\ 1 - e^{-\lambda} e_{A,B}^{\lambda}(\sigma), & \sigma \in A, \\ -e^{-\lambda} e_{A,B}^{\lambda}(\sigma), & \sigma \in B. \end{cases}$$
(1.21)

Notice that in the case  $\lambda = 0$ , by distinguish two cases, (1.21) implies that for all  $\sigma \in A \cup B$ 

$$e_{A,B}(\sigma) = -(L_N h_{A,B})(\sigma) = (L_N h_{B,A})(\sigma) = -e_{B,A}(\sigma).$$
 (1.22)

A further object of interest is the Laplace transform of the hitting time of a set  $D \subset S_N$ , which is the solution of a Feynman-Kac equation similar to (1.19). More precise, if we denote by  $w_D^{\lambda} = \mathbb{E}_{\sigma} \left[ e^{-\lambda \tau_D} \right]$ , for  $\sigma \in D^c$ , then  $w_D^{\lambda}$  is the solution of the Dirichlet problem (1.7) with  $g \equiv 0, k \equiv e^{\lambda} - 1$ , for  $\lambda \geq 0$ , subject to the boundary condition 1 on the set D.

Let us now turn our attention to the Poisson equation. If the functions k and u vanish, the boundary value problem (1.7) reduces to

$$\begin{cases} (L_N f)(\sigma) = -g(\sigma), & \sigma \in D^c \\ f(\sigma) = 0, & \sigma \in D. \end{cases}$$
(1.23)

In view of Proposition 1.2, the solution of the Poisson equation has the representation

$$f(\sigma) = \mathbb{E}_{\sigma} \left[ \sum_{s=0}^{\tau_D - 1} g(\sigma(s)) \right], \qquad \forall \sigma \in D^c,$$
(1.24)

which can be interpreted as average cost accumulated along a path of the Markov chain,  $\{\sigma(t)\}$ , before hitting the set D. Of particular interest is the mean hitting time of the set D which is obtained by choosing  $g \equiv 1$  on  $D^c$ . Denoting the corresponding solution by  $w_B$  we have

$$w_D(\sigma) = \mathbb{E}_{\sigma}[\tau_D], \quad \forall \sigma \in D^c.$$
(1.25)

Let us now consider two disjoint subsets  $A, B \subset S_N$ . We will be interested in the mean hitting time of A starting the Markov chain in a configuration  $\sigma \notin A \cup B$  that gets killed at the arrival of B. To start with, let us denote by  $w_{A,B}$  the solution of the Poisson equation (1.23) for the choice  $D = A \cup B$  and  $g = h_{A,B}$ . In this case (1.24) reads

$$w_{A,B}(\sigma) = \mathbb{E}_{\sigma} \left[ \sum_{s=0}^{\tau_{A\cup B}-1} h_{A,B}(\sigma) \right], \qquad \forall \sigma \notin A \cup B.$$
(1.26)

The following lemma shows that  $w_{A,B}$  is indeed the object we are interested in.

**Lemma 1.7.** Let  $A, B \subset S_N$  be two disjoint subsets of the finite state space  $S_N$ . Then,

$$w_{A,B}(\sigma) = \mathbb{E}_{\sigma} [\tau_A \, \mathbb{1}_{\tau_A < \tau_B}], \qquad \forall \sigma \notin A \cup B.$$
(1.27)

Proof. Recall that  $h_{A,B}$  is harmonic in  $(A \cup B)^c$ . As a consequence of the discrete martingale problem  $X_t := h_{A,B}(\sigma(t)) - \sum_{s=0}^{t-1} (L_N h_{A,B})(\sigma(s))$  is a martingale and it holds that  $X_t = h_{A,B}(\sigma(t))$  as long as  $t < \tau_{A\cup B}$  provided the Markov chain is started in  $\sigma \notin A \cup B$ . Since the process  $H_t = t$  is previsible and locally bounded,  $M_t := (H \bullet X)_t$  is a martingale that vanishes in zero. In particular,  $M_t$  is a local martingale up to time  $\tau_{A\cup B}$ . By applying the optional stopping theorem with a localizing sequence  $\tau_n \uparrow \tau_{A\cup B}$  of stopping times we obtain

$$0 = \mathbb{E}_{\sigma}\left[(H \bullet X)_{0}\right] = \mathbb{E}_{\sigma}\left[(H \bullet X)_{\tau_{n}}\right] = \mathbb{E}_{\sigma}\left[\tau_{n} h_{A,B}(\sigma(\tau_{n})) - \sum_{s=0}^{\tau_{n}-1} h_{A,B}(\sigma(s))\right].$$
(1.28)

Since  $\mathbb{E}_{\sigma}[\tau_B] < \infty$ , Lebesque's dominated convergence theorem implies that, for all  $\sigma \notin A \cup B$ ,

$$w_{A,B}(\sigma) = \mathbb{E}_{\sigma} \Big[ \tau_{A \cup B} \ h_{A,B} \big( \sigma(\tau_{A \cup B}) \big) \Big] = \mathbb{E}_{\sigma} \big[ \tau_A \, \mathbb{1}_{\tau_A < \tau_B} \big]$$
(1.29)

where we used that the  $h_{A,B}(\sigma) = \mathbb{1}_A(\sigma)$  for  $\sigma \in A \cup B$ . This completes the proof.

### 1.4. Green function and Green's identities

Let us now focus on a specific Feynman-Kac-Poisson equation that is given by the Dirichlet problem (1.7) with the choice of vanishing function u and  $k \equiv e^{\lambda} - 1$ , for  $\lambda \ge 0$ , i.e.

$$\begin{cases} (L_N f)(\sigma) - (e^{\lambda} - 1) f(\sigma) = -g(\sigma), & \sigma \in D^c \\ f(\sigma) = 0, & \sigma \in D. \end{cases}$$
(1.30)

The stochastic representation (1.12) for a solution of (1.30) can be written in the form

$$f(\sigma) = \mathbb{E}_{\sigma} \left[ \sum_{s=0}^{\tau_D - 1} e^{-\lambda(s+1)} g(\sigma(s)) \right] = \sum_{\eta \notin D} G_D^{\lambda}(\sigma, \eta) g(\eta) \quad \forall \sigma \in D^c,$$
(1.31)

where  $G_D^{\lambda} : S_N \times S_N \to \mathbb{R}$  is called the *Green function*, which is defined through

$$G_D^{\lambda}(\sigma,\eta) := \sum_{s=0}^{\infty} e^{-\lambda(s+1)} \mathbb{P}_{\sigma}[\sigma(s) = \eta, \ s < \tau_D], \qquad \forall \sigma, \eta \in D^c$$
(1.32)

and

$$G_D^{\lambda}(\sigma,\eta) := \mathbb{1}_{\sigma=\eta} \quad \forall \sigma, \eta \in D, \quad \text{and} \quad G_D^{\lambda}(\sigma,\eta) := 0, \quad \text{otherwise.}$$
(1.33)

In the case  $\lambda = 0$ , the Green function represents the expected number of visits in  $\eta$  starting the Markov chain,  $\{\sigma(t)\}$ , in  $\sigma$  before hitting the set *D*. Moreover, we denote by

$$p_D^s(\sigma,\eta) := \mathbb{P}_{\sigma}\left[\sigma(s) = \eta, \, s < \tau_D\right] < p_N^s(\sigma,\eta) \tag{1.34}$$

the transition function of a process that is absorbed in D. Notice that  $p_D^s(\sigma, \cdot)$  are subprobabilities distributions. Since the dynamics is assumed to be reversible, we have

$$\mu_N(\sigma) G_D^{\lambda}(\sigma, \eta) = \mu_N(\eta) G_D^{\lambda}(\eta, \sigma).$$
(1.35)

Notice that the  $\lambda$ -equilibrium potential has an immediate representation in terms of the Green function. Namely, for all  $\sigma \in B^c$  it holds that

$$h_{A,B}^{\lambda}(\sigma) = \sum_{\eta \in A} G_B^{\lambda}(\sigma, \eta) e_{A,B}^{\lambda}(\eta).$$
(1.36)

For later reference, the representation of  $w_D$  via the Green function is given by

$$w_D(\sigma) = \sum_{\eta \notin D} G_D(\sigma, \eta).$$
(1.37)

An important formula is the discrete analog of the first and second Green's identity. Let us emphasis the fact that reversibility of the dynamics implies that the (discrete) generator,  $L_N$ , is self-adjoint on  $L^2(S_N, \mu_N)$ . As a consequence, we have

**Proposition 1.8** (First Green's identity). Let  $f, g \in L^2(\mathcal{S}_N, \mu_N)$ . Then,

$$\frac{1}{2} \sum_{\sigma,\eta\in\mathcal{S}_N} \mu_N(\sigma) \, p_N(\sigma,\eta) \left( f(\sigma) - f(\eta) \right) \left( g(\sigma) - g(\eta) \right) = -\sum_{\sigma\in\mathcal{S}_N} \mu_N(\sigma) \, f(\sigma) \left( L_N g \right)(\sigma).$$
(1.38)

*Proof.* By expanding the left-hand side of (1.38) and using the detailed balance condition the assertion is immediate.  $\Box$ 

**Corollary 1.9** (Second Green's identity). Let  $A \subset S_N$  and  $f, g \in L^2(S_N, \mu_N)$ . Then,

$$\sum_{\sigma \in A} \mu_N(\sigma) \left( f(\sigma) \left( L_N g \right)(\sigma) - g(\sigma) \left( L_N f \right)(\sigma) \right)$$
$$= \sum_{\sigma \in A^c} \mu_N(\sigma) \left( g(\sigma) \left( L_N f \right)(\sigma) - f(\sigma) \left( L_N g \right)(\sigma) \right).$$
(1.39)

**Remark 1.10.** In the context of reversible diffusion processes in  $\mathbb{R}^d$ , studied in [17, 18], Green's identities were crucial to derive various estimates. Emphasizing this link, we will refer to (1.39) as second Green's identity, although it is a simple rewriting of  $\langle f, L_N g \rangle_{\mu_N} = \langle L_N f, g \rangle_{\mu_N}$  in the discrete setting.

**Remark 1.11.** While the discrete analogue of Green's first identity is well known and its application can be found at various places in the literature, the strength of Green's second identity, i.e. the systematic exploit of the reversibility, seems to be overlooked in the past.

We turn now our attention to the Feynman-Kac equation which is given by the Dirichlet problem (1.7) with vanishing function g and  $k \equiv e^{\lambda} - 1$ , for  $\lambda \geq 0$ . It reads

$$\begin{cases} (L_N f)(\sigma) - (e^{\lambda} - 1) f(\sigma) = 0, & \sigma \in D^c \\ f(\sigma) = u(\sigma), & \sigma \in D. \end{cases}$$
(1.40)

Our aim is to derive a representation of the solution operator,  $H_D^{\lambda}$ , associated to the Feynman-Kac equation in terms of the Green function to which we will refer to as *Poisson kernel*. In order to do so, suppose that f solves (1.30). Further, for any fixed  $\sigma \in D^c$  an application of the second Green's identity to the functions f and  $G_D^{\lambda}(\cdot, \sigma)$  reveals immediately that

$$f(\sigma) = \sum_{\eta \in D} \frac{\mu_N(\eta)}{\mu_N(\sigma)} \left( L_N^{\lambda} G_D^{\lambda} \right)(\eta, \sigma) u(\eta), \qquad (1.41)$$

where we used that, for all  $\eta \in D^c$ ,  $(L_N^{\lambda}G_D^{\lambda})(\eta, \sigma) = -\mathbb{1}_{\eta=\sigma}$  and  $(L_N^{\lambda}f)(\eta) = 0$ . Recall that  $L_N^{\lambda} = L_N - (e^{\lambda} - 1)$ . In view of the probabilistic representation (1.12), we have

$$H_D^{\lambda}(\sigma,\eta) = \frac{\mu_N(\eta)}{\mu_N(\sigma)} \left( L_N^{\lambda} G_D^{\lambda} \right)(\eta,\sigma) = \mathbb{E}_{\sigma} \Big[ e^{-\lambda \tau_D} \, \mathbb{1}_{\sigma(\tau_D) = \eta} \Big], \qquad \forall \sigma \in D^c, \, \eta \in D, \quad (1.42)$$

while otherwise we set  $H_D^{\lambda}(\sigma, \eta) = 0$ . As an application, we obtain for the  $\lambda$ -equilibrium potential the following representation

$$h_{A,B}^{\lambda}(\sigma) = \sum_{\eta \in A} \frac{\mu_N(\eta)}{\mu_N(\sigma)} \left( L_N^{\lambda} G_{A \cup B}^{\lambda} \right)(\eta, \sigma), \qquad \forall \sigma \notin A \cup B.$$
(1.43)

### 1.5. Connection between equilibrium potential, capacity, and mean hitting times

In this section, we will introduce *the* main object of the potential theoretic approach. Here, we restrict our attention to the case  $\lambda = 0$ . The *capacity* of the capacitor (A, B) with equilibrium potential one on A and zero on B is defined through

$$\operatorname{cap}(A,B) := \sum_{\sigma \in A} \mu_N(\sigma) e_{A,B}(\sigma).$$
(1.44)

Let us denote by  $\langle \cdot, \cdot \rangle_{\mu_N}$  the scalar product in  $L^2(\mathcal{S}_N, \mu_N)$ . The *energy* associated to the pair  $(P_N, \mu_N)$  is defined for any function f on  $\mathcal{S}_N$  by

$$\mathcal{E}(f) := \langle -L_N f, f \rangle_{\mu_N}. \tag{1.45}$$

An immediate consequence of the first Green's identity is that

$$\mathcal{E}(f) = \frac{1}{2} \sum_{\sigma,\eta \in \mathcal{S}_N} \mu_N(\sigma) \, p_N(\sigma,\eta) \left( f(\sigma) - f(\eta) \right)^2. \tag{1.46}$$

Due to the detailed balance condition, the factor 1/2 ensure that each pair of configuration  $(\sigma, \eta)$  contributes to the energy only once. Inspecting the definition of the capacity reveals that  $\operatorname{cap}(A, B) = \mathcal{E}(h_{A,B})$ . In particular, we obtain that

$$\operatorname{cap}(A,B) = \frac{1}{2} \sum_{\sigma,\eta\in\mathcal{S}_N} \mu_N(\sigma) p_N(\sigma,\eta) \left( h_{A,B}(\sigma) - h_{A,B}(\eta) \right)^2.$$
(1.47)

Moreover, the capacity has a probabilistic interpretation in terms of *escape probabilities*, namely  $\mathbb{P}_{\mu_A}[\tau_B < \tau_A] = \operatorname{cap}(A, B)/\mu_N[A]$ , where  $\mu_A(\sigma) = \mu_N[\sigma|A]$ , for  $\sigma \in A$ , stand for the reversible measure conditioned on the set A.

So far, we have seen that the mean hitting time,  $\mathbb{E}_{\sigma}[\tau_B]$ , of a set B, has an analytic counterpart,  $w_B$ , that solves a certain Poisson equation. However, in most applications an exact solution of the corresponding boundary value problem can not be achieved. The first important ingredient of the potential theoretic approach to metastability is a formula for the average mean hitting time when starting the Markov chain in the so called *last exit biased distribution* on a set A that connects it to the capacity and the equilibrium potential. For any two disjoint subsets  $A, B \subset S_N$  this distribution is defined through

$$\nu_{A,B}(\sigma) = \frac{\mu_N(\sigma) \mathbb{P}_{\sigma} [\tau_B < \tau_A]}{\sum_{\eta \in A} \mu_N(\eta) \mathbb{P}_{\eta} [\tau_B < \tau_A]} = \frac{\mu_N(\sigma) e_{A,B}(\sigma)}{\operatorname{cap}(A,B)} \quad \forall \sigma \in A.$$
(1.48)

Notice that  $\nu_{A,B}$  is concentrated on those starting configurations  $\eta \in A$  that are at the boundary of this set.

**Proposition 1.12.** Let  $A, B \subset S_N$  with  $A \cap B = \emptyset$ . Then

$$\mathbb{E}_{\nu_{A,B}}[\tau_B] = \frac{1}{\operatorname{cap}(A,B)} \sum_{\sigma \notin B} \mu_N(\sigma) h_{A,B}(\sigma).$$
(1.49)

**Remark 1.13.** We are typically interested in mean hitting times of the set *B* starting the process in a single configuration  $\sigma \in A$ . For this reason, it appears appealing to replace in (1.49) the subset *A* by the singleton  $\{\sigma\}$ . However, this results in the challenging task of computing precisely the capacity  $cap(\{\sigma\}, B)$  between a configuration and a set, and in most applications this cannot be achieved. On the other hand, we are able to establish sharp estimates on the right-hand side of (1.49), and hence on an averaging of mean hitting times with respect to  $\nu_{A,B}$ , in situations when the sets *A* and *B* are sufficiently large. In the next chapter, we derive pointwise estimates on the mean hitting times by controlling the fluctuations of  $\mathbb{E}_{\sigma}[\tau_B]$  within *A*.

We present three different proofs of this proposition. While the first one, which can be found in [53], uses purely probabilistic arguments, the second proof relies on the Green function representation (1.36). It was first given in [14]. The third one is as well a rather simple proof bases on Green's second identity.

*Proof of Proposition 1.12 via last exit decomposition.* First, let us define the last exit time  $L_{A,B}$  from A before hitting B as

$$L_{A,B} := \sup \left\{ 0 \le t < \tau_B \mid \sigma(t) \in A \right\}.$$

$$(1.50)$$

Then we have for all  $\eta \notin B$ ,

$$\mu_{N}(\eta) \mathbb{P}_{\eta} [\tau_{A} < \tau_{B}] = \mu_{N}(\eta) \mathbb{P}_{\eta} [L_{A,B} > 0]$$

$$= \sum_{s=1}^{\infty} \sum_{\sigma \in A} \mu_{N}(\eta) \mathbb{P}_{\eta} [L_{A,B} = s, \sigma(s) = \sigma]$$

$$= \sum_{s=1}^{\infty} \sum_{\sigma \in A} \mu_{N}(\eta) \mathbb{P}_{\eta} [\sigma(s) = \sigma, s < \tau_{B}] \mathbb{P}_{\sigma} [\tau_{B} < \tau_{A}], \quad (1.51)$$

where we used the Markov property in the last step. By reversibility, we have that

$$\mu_N(\eta) \mathbb{P}_{\eta} \big[ \sigma(s) = \sigma, s < \tau_B \big] = \mu_N(\sigma) \mathbb{P}_{\sigma} \big[ \sigma(s) = \eta, s < \tau_B \big].$$
(1.52)

In view of (1.48), this implies that

$$\mathbb{E}_{\nu_{A,B}}\left[\sum_{s=0}^{\tau_{B}-1}\mathbb{1}_{\sigma(s)=\eta}\right] - \frac{\mu_{N}(\eta) \mathbb{P}_{\eta}[\tau_{B} < \tau_{A}]}{\operatorname{cap}(A,B)} \mathbb{1}_{\eta \in A} = \frac{\mu_{N}(\eta) \mathbb{P}_{\eta}[\tau_{A} < \tau_{B}]}{\operatorname{cap}(A,B)}$$
(1.53)

By summing over all configurations  $\eta$  outside *B*, we obtain (1.49).

*Proof of Proposition 1.12 via Green function.* For any  $\sigma \in A$ , multiply the representation (1.37) of the mean hitting time  $w_B$  by  $\mu_N(\sigma) e_{A,B}(\sigma)$ . Summing over all  $\sigma \in A$  gives

$$\sum_{\sigma \in A} \mu_N(\sigma) e_{A,B}(\sigma) w_B(\sigma) = \sum_{\sigma \in A} \sum_{\eta \notin B} \mu_N(\sigma) G_B(\sigma, \eta) e_{A,B}(\sigma)$$
$$= \sum_{\sigma \in A} \sum_{\eta \notin B} \mu_N(\eta) e_{A,B}(\sigma) G_B(\eta, \sigma) = \sum_{\eta \notin B} \mu_N(\eta) h_{A,B}(\eta),$$
(1.54)

where we used (1.35) in the second step and the representation (1.36) of the equilibrium potential,  $h_{A,B}$ , via the Green function in the third step. Normalizing the measure on the left-hand side yields the assertion.

*Proof of Proposition 1.12 via Green's second identity.* By applying the second Green's identity to the functions  $w_B$  and  $h_{A,B}$ , we obtain

$$\sum_{\sigma \in A} \mu_N(\sigma) e_{A,B}(\sigma) w_B(\sigma) = \sum_{\sigma \notin B} \mu_N(\sigma) h_{A,B}(\sigma).$$
(1.55)

Normalizing the measure on the left-hand side immediately yields the assertion.  $\Box$ 

We will now prove a splitting lemma which will be important in the investigation of the distribution of normalized hitting times.

**Lemma 1.14.** Let  $A, B \subset S_N$  with  $A \cap B = \emptyset$ . Then

$$\mathbb{E}_{\nu_{A,B}}[\tau_B] = \frac{\mathbb{E}_{\mu_A}[\tau_A \, \mathbb{1}_{\tau_A < \tau_B}] + \mathbb{E}_{\mu_A}[\tau_B \, \mathbb{1}_{\tau_B < \tau_A}]}{\mathbb{P}_{\mu_A}[\tau_B < \tau_A]}.$$
(1.56)

We give three proofs, one purely probabilistic and two purely analytic.

*Proof of Lemma 1.14 via strong Markov property.* By using that  $1 = \mathbb{1}_{\tau_B < \tau_A} + \mathbb{1}_{\tau_A < \tau_B}$  together with the strong Markov property, we obtain

$$\mathbb{E}_{\mu_{A}}[\tau_{B}] = \mathbb{E}_{\mu_{A}}[\tau_{B} \mathbb{1}_{\tau_{B} < \tau_{A}}] + \mathbb{E}_{\mu_{A}}\left[\mathbb{1}_{\tau_{A} < \tau_{B}} \mathbb{E}_{\sigma(\tau_{A})}[\tau_{A} + (\tau_{B} - \tau_{A})]\right]$$

$$= \mathbb{E}_{\mu_{A}}[\tau_{B} \mathbb{1}_{\tau_{B} < \tau_{A}}] + \mathbb{E}_{\mu_{A}}[\tau_{A} \mathbb{1}_{\tau_{A} < \tau_{B}}] + \sum_{\eta \in A} \mathbb{P}_{\sigma}[\tau_{A} < \tau_{B}, \sigma(\tau_{A}) = \eta] \mathbb{E}_{\eta}[\tau_{B}].$$

As a consequence of the reversibility of the dynamics, we have that

$$\sum_{\sigma \in A} \mu_N(\sigma) \mathbb{P}_{\sigma} \big[ \tau_A < \tau_B, \, \sigma(\tau_A) = \eta \big] = \mu_N(\eta) \mathbb{P}_{\eta} \big[ \tau_A < \tau_B \big], \qquad \forall \eta \in A.$$
(1.57)

Combining the last two equations and solving for  $\mathbb{E}_{\sigma}[\tau_B]$  yields

$$\frac{1}{\mu_N[A]} \sum_{\eta \in A} \mu_N(\eta) \mathbb{P}_{\eta} [\tau_B < \tau_A] \mathbb{E}_{\eta} [\tau_B] = \mathbb{E}_{\mu_A} [\tau_B \, \mathbb{1}_{\tau_B < \tau_A}] + \mathbb{E}_{\mu_A} [\tau_A \, \mathbb{1}_{\tau_A < \tau_B}]$$
(1.58)

which, together with the definition of the last exit distribution, is equivalent to (1.56). This completes the proof.  $\hfill \Box$ 

*Proof of Lemma 1.14 via Green's function.* Our starting point is the representation of  $w_B$  in terms of the Green function as given in (1.37), which we rewrite as

$$w_B(\sigma) = \sum_{\eta \in A} G_B(\sigma, \eta) + \sum_{\eta \notin A \cup B} G_B(\sigma, \eta) \left( h_{A,B}(\eta) + h_{B,A}(\eta) \right), \tag{1.59}$$

taking advantage of the fact that  $h_{A,B}(\eta) + h_{B,A}(\eta) = \mathbb{P}_{\eta}[\tau_A < \tau_B] + \mathbb{P}_{\eta}[\tau_B < \tau_A] = 1$ , for all  $\eta \notin A \cup B$ . Multiplying both sides with  $\mu_N(\sigma) e_{A,B}(\sigma)$  and summing over all  $\sigma \in A$  gives

$$\sum_{\sigma \in A} \mu_N(\sigma) e_{A,B}(\sigma) w_B(\sigma) = \mu_N[A] + \sum_{\substack{\sigma \in A \\ \eta \notin A \cup B}} e_{A,B}(\sigma) \mu_N(\eta) G_B(\eta,\sigma) \left(h_{A,B}(\eta) + h_{B,A}(\eta)\right),$$
(1.60)

where we used (1.35) as well as the representation, (1.36), of the harmonic function  $h_{A,B}$  via the Green's function. Exploiting that, for all  $\sigma \notin A \cup B$ , the function  $w_{A,B}$  solves the Poisson equation  $(L_N w_{A,B})(\sigma) = -h_{A,B}(\sigma)$  subject to zero boundary condition of  $A \cup B$ , we can replace the harmonic function  $h_{A,B}$  in (1.60) by the left-hand side of the Poisson equation. By applying Green's second identity to the functions  $w_{A,B}$  and  $G_{B^c}(\cdot, \sigma)$  with  $\sigma \in A$  we have

$$\sum_{\eta \notin A \cup B} \mu_N(\eta) G_B(\eta, \sigma) \left( L_N w_{A,B} \right)(\eta) = -\sum_{\eta \in A} \mu_N(\eta) G_B(\eta, \sigma) \left( L_N w_{A,B} \right)(\eta), \quad (1.61)$$

whereas the same holds true when  $w_{A,B}$  is replaced by  $w_{B,A}$ . Carrying out the summation over  $\sigma \in A$  and taking into account (1.36), yields

$$\sum_{\sigma \in A} \mu_N(\sigma) e_{A,B}(\sigma) w_B(\sigma) = \sum_{\eta \in A} \mu_N(\sigma) \left( 1 + (L_N w_{A,B})(\eta) + (L_N w_{B,A})(\eta) \right)$$

$$= \sum_{\eta \in A} \mu_N(\eta) \left( \mathbb{E}_{\eta} [\tau_A \mathbb{1}_{\tau_A < \tau_B}] + \mathbb{E}_{\eta} [\tau_B \mathbb{1}_{\tau_B < \tau_A}] \right).$$
(1.62)

The last step relies on the probabilistic interpretation of the involved analytic quantities and the boundary of *A*. Namely, for all  $\eta \in A$  it holds that  $\mathbb{E}_{\eta}[\tau_B \mathbb{1}_{\tau_B < \tau_A}] = e_{A,B}(\eta) + (L_N w_{B,A})(\eta)$  and  $\mathbb{E}_{\eta}[\tau_A \mathbb{1}_{\tau_A < \tau_B}] = 1 - e_{A,B}(\eta) + (L_N w_{A,B})(\eta)$ . Normalizing the measure on the left-hand side completes the proof.

*Proof of Lemma 1.14 via second Green's formula.* By applying Green's second identity to the functions  $w_{A,B}$  and  $h_{A,B}$  gives

$$\sum_{\sigma \in A} \mu_N(\sigma) \left( L_N w_{A,B} \right)(\sigma) = \sum_{\sigma \notin A \cup B} \mu_N(\sigma) h_{A,B}^2(\sigma),$$
(1.63)

whereas its application to the functions  $w_{B,A}$  and  $h_{A,B}$  leads to

$$\sum_{\sigma \in A} \mu_N(\sigma) \left( L_N w_{B,A} \right)(\sigma) = \sum_{\sigma \notin A \cup B} \mu_N(\sigma) h_{A,B}(\sigma) h_{B,A}(\sigma).$$
(1.64)

Adding up the last two equations, yields

$$\sum_{\sigma \in A} \mu_N(\sigma) \left( \left( L_N w_{A,B} \right)(\sigma) + \left( L_N w_{B,A} \right)(\sigma) \right) = \sum_{\sigma \notin A \cup B} \mu_N(\sigma) h_{A,B}(\sigma).$$
(1.65)

By comparing this equation with (1.55), we immediately recover (1.62) and the remaining part of the proof follows the arguments given in the previous proof.  $\Box$ 

### 1.6. The Dirichlet and Thomson principle: a dual variational principle

The power of Proposition (1.12) is that it represents the mean hitting time in terms of a capacity that is a positive-definite quadratic form, as it can be seen from (1.47). Whenever a object of interests can be expressed as a positive-definite quadratic form, it is often possible to establish dual variational principles from which upper and lower bound for it can be constructed. Notice that many variational methods are based on this tool. Our presentation follows in parts the one given by Lieberstein [64] in the context of partial differential equations.

The bilinear form  $\mathcal{E}(g,h)$  associated to the Dirichlet form  $\mathcal{E}(g)$  is given by

$$\mathcal{E}(g,h) = \frac{1}{2} \sum_{\sigma,\eta\in\mathcal{S}_N} \mu_N(\sigma) p_N(\sigma,\eta) \left(g(\sigma) - g(\eta)\right) \left(h(\sigma) - h(\eta)\right).$$
(1.66)

An immediate consequence of the Cauchy-Schwarz inequality is that

$$\mathcal{E}(g,h)^2 \leq \mathcal{E}(g)\mathcal{E}(h). \tag{1.67}$$

Recall that the equilibrium potential,  $h_{A,B}$ , is defined as the solution of the Laplace equation

$$\begin{cases} (L_N f)(\sigma) = 0, & \sigma \in (A \cup B)^c \\ f(\sigma) = \mathbb{1}_A(\sigma), & \sigma \in A \cup B. \end{cases}$$
(1.68)

If a function  $h \in L^2(\mathcal{S}_N, \mu_N)$  has the property that  $h = \mathbb{1}_A$  on  $A \cup B$ , then by Green's first identity

$$\mathcal{E}(h_{A,B}, h) = -\sum_{\eta \in \mathcal{S}_N} \mu_N(\eta) h(\eta) \left( L_N h_{A,B} \right)(\eta) = \sum_{\eta \in A} \mu_N(\eta) e_{A,B}(\eta)$$
(1.69)

In view of (1.44) and (1.47), for such a function h, by (1.67), we have that  $cap(A, B) \leq \mathcal{E}(h)$ . Since equality is obtains, if we choose for h the equilibrium potential, we have proven the well known

**Proposition 1.15** (Dirichlet principle). For any non-empty disjoint sets  $A, B \subset S_N$ ,

$$\operatorname{cap}(A,B) = \min_{h \in \mathcal{H}_{A,B}} \mathcal{E}(h)$$
(1.70)

where  $\mathcal{H}_{A,B} := \{h: \mathcal{S}_N \to [0,1] \mid h|_A \equiv 1, h|_B \equiv 0\}.$ 

**Remark 1.16.** The importance of the Dirichlet principle is that it yields computable upper bounds for capacities by choosing suitable test functions h. As we will see in chapter 5, with proper physical insight, it is often possible to guess a reasonable test function.

**Remark 1.17.** An immediate corollary of the Dirichlet principle is *Rayleigh's monotonicity law*. It allows to derive a lower bound for capacities by using the monotonicity of the Dirichlet form in the transition probabilities to compare the original process with a simplified one. This is well known in the language of electrical networks, see [44].

Instead of considering a function h which coincides with  $h_{A,B}$  on  $A \cup B$ , suppose that  $h \in L^2(\mathcal{S}_N, \mu_N)$  is super-harmonic on  $B^c$ , i.e.  $(L_N h)(\sigma) \leq 0$  for all  $\sigma \in B^c$ . Then by Green's first identity

$$\mathcal{E}(h_{A,B},h) = -\sum_{\eta \in \mathcal{S}_N} \mu_N(\sigma) h_{A,B}(\sigma) (L_N h)(\sigma) \ge -\sum_{\eta \in A} \mu_N(\sigma) (L_N h)(\sigma), \quad (1.71)$$

where we used in the second step that on  $(A \cup B)^c$  the equilibrium potential is non-negative. Hence, from (1.67), we get that

$$\operatorname{cap}(A,B) \geq \frac{\left(\sum_{\eta \in A} \mu_N(\eta) \left(L_N h\right)(\eta)\right)^2}{\mathcal{E}(h)}$$
(1.72)

where equality is obtain for  $h = h_{A,B}$ . Further, notice that the value of the right-hand side of (1.72) is invariant with respect to multiplication of h by a constant. This allows to choose h is such a way that the numerator of (1.72) is equal to 1. Thus, we have proven a version of Sir William Thomson's (Lord Kelvin's) principle

15

**Proposition 1.18.** For any non-empty disjoint sets  $A, B \subset S_N$ ,

$$\operatorname{cap}(A,B) = \max_{h \in \mathfrak{H}_{A,B}} \frac{\left(\sum_{\eta \in A} \mu_N(\eta) \left(L_N h\right)(\eta)\right)^2}{\mathcal{E}(h)}.$$
(1.73)

where  $\mathfrak{H}_{A,B} := \{h : \mathcal{S}_N \to [0,1] \mid (L_N h)(\eta) \leq 0, \ \forall \eta \notin B \}.$ 

The Thomson principle is well known in the context of random walks and electrical networks. However, in that language it is differently formulated in terms of unit flows, see [44, 83]. For this reason, we will briefly discuss the connection between these formulations.

**Definition 1.19.** Given two non-empty disjoint sets  $A, B \subset S_N$ , and a map  $f: S_N \times S_N \to \mathbb{R}$  such that, for all  $\sigma, \eta \in S_N$ ,  $f(\sigma, \eta) = -f(\eta, \sigma)$  and  $p_N(\sigma, \eta) = 0$  implies  $f(\sigma, \eta) = 0$ .

(a) The map f is a flow from A to B, if it satisfies Kirchhoff's law in  $(A \cup B)^c$ , that is

$$\sum_{\eta \in \mathcal{S}_N} f(\sigma, \eta) = \sum_{\eta \in \mathcal{S}_N} f(\eta, \sigma), \qquad \forall \sigma \in (A \cup B)^c.$$
(1.74)

(b) The map f is a unit flow from A to B, if it is a flow and the strength of f is equal to 1, that is

$$\sum_{\sigma \in A} \sum_{\eta \in \mathcal{S}_N} f(\sigma, \eta) = 1 = \sum_{\eta \in \mathcal{S}_N} \sum_{\sigma \in B} f(\eta, \sigma).$$
(1.75)

With this definition of flows on the network given by  $(p(\sigma, \eta))$  in mind, let us rewrite the bilinear form  $\mathcal{E}(g, h)$  as

$$\mathcal{E}(g,h) = \frac{1}{2} \sum_{\sigma,\eta\in\mathcal{S}_N} \frac{1}{\mu_N(\sigma) p_N(\sigma,\eta)} \left(\mu_N p_N \nabla g\right)(\sigma,\eta) \left(\mu_N p_N \nabla h\right)(\sigma,\eta),$$
(1.76)

where we introduced the notation  $(\mu_N p_N \nabla g)(\sigma, \eta) := \mu_N(\sigma) p_N(\sigma, \eta) (g(\sigma) - g(\eta))$ . Now observe that if the function g is harmonic then  $\mu_N p_N \nabla g$  is a flow. Thus, in view of (1.76), given two flows  $f, \phi$  we define a bilinear form through

$$\mathcal{D}(f,\phi) := \frac{1}{2} \sum_{\sigma,\eta\in\mathcal{S}_N} \frac{1}{\mu_N(\sigma) p_N(\sigma,\eta)} f(\sigma,\eta) \phi(\sigma,\eta), \qquad \mathcal{D}(f) := \mathcal{D}(f,f).$$
(1.77)

In particular,  $\mathcal{D}(\mu_N p_N \nabla h_{A,B}) = \operatorname{cap}(A, B)$ . Since for any unit flow f,

$$\mathcal{D}(\mu_N \, p_N \nabla h_{A,B}, f) = \frac{1}{2} \sum_{\sigma,\eta \in \mathcal{S}_N} \left( h_{A,B}(\sigma) - h_{A,B}(\eta) \right) f(\sigma,\eta) = \sum_{\sigma \in A} \sum_{\eta \in \mathcal{S}_N} f(\sigma,\eta) = 1$$
(1.78)

an application of Cauchy-Schwarz's inequality reveals that  $cap(A, B) \ge 1/\mathcal{D}(f)$ , whereas equality holds for the unit flow which is given in terms of the harmonic function. Thus, we have proven

**Proposition 1.20** (Thomson principle). For any non-empty disjoint sets  $A, B \subset S_N$ ,

$$\operatorname{cap}(A,B) = \max_{f \in \mathfrak{U}_{A,B}} \frac{1}{\mathcal{D}(f)},$$
(1.79)

where  $\mathfrak{U}_{A,B}$  denotes the space of unit flows from A to B.

### 1.7. The Berman-Konsowa principle: a variational principle for the lower bound

The variational principles, we have presented in the previous section, rely on the application of the Cauchy-Schwarz inequality to the bilinear form  $\mathcal{E}(g, h)$  and  $\mathcal{D}(f, \phi)$ , respectively. We will now describe a little-known variational principles for capacities that does not use (1.67). It was first proven by Berman and Konsowa in [4]. Our presentation will follow the arguments first given in [6] and then in [14].

**Definition 1.21.** Given two non-empty disjoint sets  $A, B \subset S_N$ , a map  $f : S_N \times S_N \to [0, \infty)$  such that  $p_N(\sigma, \eta) = 0$  implies  $f(\sigma, \eta) = 0$  for all  $\sigma, \eta \in S_N$ . Then f is a loop-free non-negative unit flow, if it satisfies the following properties

- (i) if  $f(\sigma, \eta) > 0$  then  $f(\eta, \sigma) = 0$ ,
- (ii) f satisfies Kirchhoff's law in  $(A \cup B)^c$ ,
- (iii) the strength of f is equal to 1,
- (iv) any path,  $\gamma$ , from A to B such that  $f(\sigma, \eta) > 0$  for all  $(\sigma, \eta) \in \gamma$  is self-avoiding.

Now, we will show that each loop-free non-negative unit flow f give rise to a probability measure  $\mathbb{P}^f$  on self-avoiding paths. For a given f we define  $\mathcal{F}(\sigma) := \sum_{\eta \in S_N} f(\sigma, \eta)$  which is positive for all  $\sigma \notin B$ . Then  $\mathbb{P}^f$  is the law of a Markov chain  $\{\xi(t)\}$  that is stopped at the arrival of B with initial distribution  $\mathbb{P}^f[\xi(0) = \sigma] = \mathcal{F}(\sigma) \mathbb{1}_A(\sigma)$  and transition probabilities

$$q^{f}(\sigma,\eta) = \frac{f(\sigma,\eta)}{\mathcal{F}(\eta)}, \qquad \sigma \notin B.$$
 (1.80)

Hence, for a path  $\gamma = (\sigma^0, \dots, \sigma^r)$  with  $\sigma^0 \in A$ ,  $\sigma^r \in B$  and  $\sigma^k \in S_N \setminus (A \cup B)$ ,  $\forall k = 1, \dots, r-1$  we have

$$\mathbb{P}^{f}\left[\xi=\gamma\right] = F(\sigma^{0}) \prod_{k=0}^{r-1} \frac{f(\sigma^{k}, \sigma^{k+1})}{F(\sigma^{k})}$$
(1.81)

where we used the convention that 0/0 = 0. By using Kirchhoff's law, we obtain the following representation for the probability that  $\{\xi(t)\}$  passes through an edge  $(\sigma, \eta)$ 

$$\mathbb{P}^{f}\left[\xi \ni (\sigma,\eta)\right] = \sum_{\gamma} \mathbb{P}^{f}\left[\xi = \gamma\right] \mathbb{1}_{(\sigma,\eta)\in\gamma} = f(\sigma,\eta).$$
(1.82)

The equation (1.82) gives rise to the following partition of unity

$$\mathbb{1}_{f(\sigma,\eta)>0} = \sum_{\gamma} \mathbb{P}^f \left[\xi = \gamma\right] \frac{1}{f(\sigma,\eta)} \mathbb{1}_{(\sigma,\eta)\in\gamma}$$
(1.83)

Employing this representation, for every  $h \in L^2(\mathcal{S}_N, \mu_N)$  which satisfy the boundary condition  $h = \mathbb{1}_A$  on  $A \cup B$ , we can bound the Dirichlet form from below by

$$\mathcal{E}(h) \geq \sum_{(\sigma,\eta)} \mu_N(\sigma) p_N(\sigma,\eta) \left( h(\sigma) - h(\eta) \right)^2 \mathbb{1}_{f(\sigma,\eta)>0}$$
  
= 
$$\sum_{\gamma} \mathbb{P}^f \left[ \xi = \gamma \right] \sum_{(\sigma,\eta)\in\gamma} \frac{\mu_N(\sigma) p_N(\sigma,\eta)}{f(\sigma,\eta)} \left( h(\sigma) - h(\eta) \right)^2$$
(1.84)

Let us now minimize the expressions above over all function h with  $h|_A = 1$  and  $h|_B = 0$ . By interchanging the minimum and the sum on the right-hand side of (1.84), we are left with the task to solve optimization problems along one-dimensional path from A to B whose minimizer are explicitly known [11]. As a result, we get

$$\operatorname{cap}(A,B) \geq \sum_{\gamma} \mathbb{P}^{f}\left[\xi = \gamma\right] \left(\sum_{(\sigma,\eta)\in\gamma} \frac{f(\sigma,\eta)}{\mu_{N}(\sigma) p_{N}(\sigma,\eta)}\right)^{-1}.$$
(1.85)

An important loop-free unit flow is the *harmonic flow*,  $f_{A,B}$ , which is defined in terms of the equilibrium potential

$$f_{A,B}(\sigma,\eta) := \frac{\mu_N(\sigma) p_N(\sigma,\eta)}{\operatorname{cap}(A,B)} \left[ h_{A,B}(\sigma) - h_{A,B}(\eta) \right]_+.$$
(1.86)

It is easy to verify that  $f_{A,B}$  satisfies the conditions (i)–(iv). While, (i) is obvious, (ii) is a consequence that  $h_{A,B}$  is harmonic in  $(A \cup B)^c$  and (iii) follows from (1.44). Condition (iv) comes from the fact that the harmonic flow only moves in directions where  $h_{A,B}$  decreases. Since for  $f_{A,B}$  equality is obtained in (1.85)  $\mathbb{P}^f$ -a.s., this proves

**Proposition 1.22.** (Berman-Konsowa principle) Let  $A, B \subset S_N$  be disjoint. Then, with the notations introduced above,

$$\operatorname{cap}(A,B) = \max_{f \in \mathcal{U}_{A,B}} \mathbb{E}^{f} \left[ \left( \sum_{(\sigma,\eta) \in \gamma} \frac{f(\sigma,\eta)}{\mu_{N}(\sigma) p_{N}(\sigma,\eta)} \right)^{-1} \right],$$
(1.87)

where  $U_{A,B}$  denotes the space of all loop-free non-negative unit flows from A to B.

The nice feature of this variational principle is that any flow gives rise to a computable lower bound. In order to obtain upper and lower bounds on the capacity that differs only a little, the strategy is to find first a good approximation of the harmonic function,  $h_{A,B}$ , by a test function h and to construct, in a second step, a test flow f from it. By plugging f into (1.85), the quality of the approximation can be controlled.

**Remark 1.23.** Given a loop-free non-negative unit flow *f*, then, by Jensen's inequality

$$\mathbb{E}^{f}\left[\left(\sum_{(\sigma,\eta)\in\gamma}\frac{f(\sigma,\eta)}{\mu_{N}(\sigma)\,p_{N}(\sigma,\eta)}\right)^{-1}\right] \geq \left(\mathbb{E}^{f}\left[\sum_{(\sigma,\eta)\in\gamma}\frac{f(\sigma,\eta)}{\mu_{N}(\sigma)\,p_{N}(\sigma,\eta)}\right]\right)^{-1} = \frac{1}{\mathcal{D}(f)}.$$
(1.88)

Since the Thomson principle also holds for loop-free non-negative unit flows, this shows that the lower bound on the capacity obtained by the Berman-Konsowa principle provides, in principle, a better approximation compared to the Thomson principle.

### 1.8. Renewal equation

In earlier works, see for example [15, 16], the following renewal equations played a crucial role in order to estimate harmonic functions or to control the Laplace transforms of hitting times. Let  $\sigma, \eta \in S_N$ ,  $\eta \notin B \subset S_N$  and  $u : \mathbb{R} \to \mathbb{R}$ , then

$$\mathbb{P}_{\sigma}[\tau_{\eta} < \tau_{B}] = \frac{\mathbb{P}_{\sigma}[\tau_{\eta} < \tau_{B\cup\sigma}]}{\mathbb{P}_{\sigma}[\tau_{B\cup\eta} < \tau_{\sigma}]} \quad \text{and} \quad \mathbb{E}_{\sigma}\left[e^{-u(\lambda)\tau_{B}}\right] = \frac{\mathbb{E}_{\sigma}\left[e^{-u(\lambda)\tau_{B}} \,\mathbb{1}_{\tau_{B} < \tau_{\sigma}}\right]}{1 - \mathbb{E}_{\sigma}\left[e^{-u(\lambda)\tau_{\sigma}} \,\mathbb{1}_{\tau_{\sigma} < \tau_{B}}\right]} \quad (1.89)$$

where the second equation holds for all  $\lambda$  for which the left-hand side exists. However, as it was pointed out in [6, 5, 14], pointwise renewal equation are of limited use in the general context. Instead of studying the process in a given point  $\sigma$ , we show how an averaged version of the standard renewal argument can be established.

**Lemma 1.24.** Let  $A, B, X \subset S_N$  be mutually disjoint. Then

$$\mathbb{P}_{\nu_{X,A\cup B}}\left[\tau_A < \tau_B\right] = \frac{\mathbb{P}_{\mu_X}\left[\tau_A < \tau_{B\cup X}\right]}{\mathbb{P}_{\mu_X}\left[\tau_{A\cup B} < \tau_X\right]} \le \min\left\{\frac{\operatorname{cap}(X,A)}{\operatorname{cap}(X,B)}, 1\right\}.$$
(1.90)

We give three proofs, one purely probabilistic using strong Markov property and two purely analytic ones using the Green's function and the second Green's identity.

*Proof of Lemma 1.24 via strong Markov property.* By starting the process in the reversible measure  $\mu_X$ , we obtain

$$\mathbb{P}_{\mu_{X}}\left[\tau_{A} < \tau_{B}\right] = \mathbb{P}_{\mu_{X}}\left[\tau_{A} < \tau_{B\cup X}\right] + \sum_{\eta \in X} \mathbb{P}_{\mu_{X}}\left[\tau_{X} < \tau_{A\cup B}, \, \sigma(\tau_{X}) = \eta\right] \mathbb{P}_{\eta}\left[\tau_{A} < \tau_{B}\right] \quad (1.91)$$

As a consequence of the reversibility, we have that

$$\sum_{\sigma \in X} \mu_N(\sigma) \mathbb{P}_{\sigma} \big[ \tau_X < \tau_{A \cup B}, \, \sigma(\tau_X) = \eta \big] = \mu_N(\eta) \mathbb{P}_{\eta} \big[ \tau_X < \tau_{A \cup B} \big].$$
(1.92)

Hence, this allows us to rewrite (1.91) as

$$\frac{1}{\mu_N[X]} \sum_{\sigma \in X} \mu_N(\sigma) \mathbb{P}_{\eta} \left[ \tau_{A \cup B} < \tau_X \right] \mathbb{P}_{\eta} \left[ \tau_A < \tau_B \right] = \mathbb{P}_{\mu_X} \left[ \tau_A < \tau_{B \cup X} \right].$$
(1.93)

Normalizing the measure on the left-hand side (1.93) and comparing the result with (1.48), yields

$$\mathbb{P}_{\nu_{X,A\cup B}}\left[\tau_A < \tau_B\right] = \frac{\mathbb{P}_{\mu_X}\left[\tau_A < \tau_{B\cup X}\right]}{\mathbb{P}_{\mu_X}\left[\tau_{A\cup B} < \tau_X\right]} \le \frac{\mathbb{P}_{\mu_X}\left[\tau_A < \tau_X\right]}{\mathbb{P}_{\mu_X}\left[\tau_B < \tau_X\right]} = \frac{\operatorname{cap}(X,A)}{\operatorname{cap}(X,B)},$$

where we used elementary monotonicity properties in the second step and the definition of capacities in the last step. Of course this bound is useful only if cap(X, A) / cap(X, B) < 1.  $\Box$ 

*Proof of Lemma 1.24 via Green's function.* To start with, let us consider the so called Poisson kernel representation of  $h_{A,B}$ , as defined in (1.43), which reads

$$h_{A,B}(\sigma) = \sum_{\eta \in A \cup B} \frac{\mu_N(\eta)}{\mu_N(\sigma)} h_{A,B}(\eta) \left( L_N G_{A \cup B} \right)(\eta, \sigma).$$
(1.94)

In particular, for any  $\sigma \in X$ , we get that

$$h_{A,B}(\sigma) = \sum_{\eta \in A \cup B} \frac{\mu_N(\eta)}{\mu_N(\sigma)} h_{A,B \cup X}(\eta) \left( L_N G_{A \cup B} \right)(\eta,\sigma).$$
(1.95)

By applying the second Green's identity to the functions  $h_{A,B\cup X}$  and  $G_{A\cup B}(\cdot,\sigma)$ , we obtain

$$\sum_{\eta \in A \cup B} \mu_N(\eta) h_{A,B \cup X}(\eta) \left( L_N G_{A \cup B} \right)(\eta,\sigma) = \sum_{\eta \in X} \mu_N(\eta) e_{B \cup X,A}(\eta) G_{A \cup B}(\eta,\sigma).$$
(1.96)

Hence, combining (1.95) and (1.96) yields

$$h_{A,B}(\sigma) = \sum_{\eta \in X} \frac{\mu_N(\eta)}{\mu_N(\sigma)} e_{B \cup X,A}(\eta) G_{A \cup B}(\eta, \sigma).$$
(1.97)

Thus, if we multiply both sides with  $\mu_N(\sigma) e_{X,A\cup B}(\sigma)$  and sum over all  $\sigma \in X$ , we get

$$\sum_{\sigma \in X} \mu_N(\sigma) e_{X,A \cup B}(\sigma) h_{A,B}(\sigma) = \sum_{\sigma,\eta \in X} \mu_N(\eta) e_{B \cup X,A}(\eta) G_{A \cup B}(\eta,\sigma) e_{X,A \cup B}(\sigma)$$
$$= \sum_{\eta \in X} \mu_N(\eta) e_{B \cup X,A}(\eta),$$
(1.98)

where we used in the second step the representation (1.36) of  $h_{X,A\cup B}$  in terms of the Green function and that  $h_{X,A\cup B}(\eta) = 1$ , for all  $\eta \in X$ . Dividing both sides of (1.98) by  $\operatorname{cap}(X, A\cup B)$  and using the definition of the last exit biased distribution (1.48), we obtain

$$\mathbb{P}_{\nu_{X,A\cup B}}\left[\tau_A < \tau_B\right] = \frac{\sum_{\eta \in X} \mu_N(\eta) e_{B\cup X,A}}{\operatorname{cap}(X, A \cup B)} = \frac{\mathbb{P}_{\mu_X}\left[\tau_A < \tau_{B\cup X}\right]}{\mathbb{P}_{\mu_X}\left[\tau_{A\cup B} < \tau_X\right]} \le \frac{\operatorname{cap}(X, A)}{\operatorname{cap}(X, B)}.$$

This concludes the proof.

*Proof of Lemma 1.24 via second Green's identity.* By applying the second Green's identity to the functions  $h_{A,B}$  and  $h_{X,A\cup B}$ , we get

$$-\sum_{\sigma\in X} \mu_N(\sigma) \left( L_N h_{X,A\cup B} \right)(\sigma) h_{A,B}(\sigma) = \sum_{\sigma\in A} \mu_N(\sigma) \left( L_N h_{X,A\cup B} \right)(\sigma).$$
(1.99)

A further application of the second Green's identity to the functions  $h_{X,A\cup B}$  and  $h_{A,B\cup X}$  reveals that

$$\sum_{\sigma \in A} \mu_N(\sigma) \left( L_N h_{X,A \cup B} \right)(\sigma) = \sum_{\sigma \in X} \mu_N(\sigma) \left( L_N h_{A,B \cup X} \right)(\sigma).$$
(1.100)

Hence, by combing both equations, we obtain

$$\sum_{\sigma \in X} \mu_N(\sigma) e_{X,A \cup B}(\sigma) h_{A,B}(\sigma) = \sum_{\sigma \in X} \mu_N(\sigma) e_{B \cup X,A}(\sigma).$$
(1.101)

By dividing both sides of (1.101) by  $cap(X, A \cup B)$  and using the definition of the last exit biased distribution as well as the probabilistic interpretation of the equilibrium measures  $e_{B\cup X,A}(\eta)$  and  $e_{A\cup B,X}(\eta)$ , allows to deduce (1.90).

Our next goal is to show an averaged version of the renewal equation for the Laplace transform. In the previous sections we already presented the profound connection between potential theory and the Laplace transform. Given two disjoint subsets  $A, B \subset S_N$ , our starting point is a specific probability measure  $\rho_{\lambda}$  on A which is defined for any  $\lambda \geq 0$  by

$$\rho_{\lambda}(\sigma) = \frac{\mu_{N}(\sigma) \left(1 - \mathbb{E}_{\sigma} \left[e^{-\lambda \tau_{A}} \mathbb{1}_{\tau_{A} < \tau_{B}}\right]\right)}{\sum_{\eta \in A} \mu_{N}(\eta) \left(1 - \mathbb{E}_{\eta} \left[e^{-\lambda \tau_{A}} \mathbb{1}_{\tau_{A} < \tau_{B}}\right]\right)} = \frac{\mu_{N}(\sigma) e_{A,B}^{\lambda}(\sigma)}{\sum_{\eta \in A} \mu_{N}(\eta) e_{A,B}^{\lambda}(\eta)}.$$
 (1.102)

Let us point out that  $\rho_{\lambda}$  is a generalization of the last exist biased distribution which can be retrieved from  $\rho_{\lambda}$  by setting  $\lambda = 0$ .

The following renewal equation will be used in Section 3.4.

**Lemma 1.25.** Let  $A, B \subset S_N$  with  $A \cap B \neq \emptyset$  and  $\rho_\lambda$  as defined above. Then

$$\mathbb{E}_{\rho_{\lambda}}\left[e^{-\lambda\tau_{B}}\right] = \frac{\mathbb{E}_{\mu_{A}}\left[e^{-\lambda\tau_{B}} \mathbb{1}_{\tau_{B} < \tau_{A}}\right]}{1 - \mathbb{E}_{\mu_{A}}\left[e^{-\lambda\tau_{A}} \mathbb{1}_{\tau_{A} < \tau_{B}}\right]}.$$
(1.103)

Again, we give three proofs, one purely probabilistic using the strong Markov property and two purely analytic using the Green's function and the second Green's identity, respectively.

*Proof of Lemma 1.25 via strong Markov property.* By using that  $1 = \mathbb{1}_{\tau_B < \tau_A} + \mathbb{1}_{\tau_A < \tau_B}$  and the strong Markov property, we obtain

$$\mathbb{E}_{\mu_{A}}\left[e^{-\lambda\tau_{B}}\right] = \mathbb{E}_{\mu_{A}}\left[e^{-\lambda\tau_{B}} \mathbb{1}_{\tau_{B} < \tau_{A}}\right] + \sum_{\eta \in A} \mathbb{E}_{\mu_{A}}\left[e^{-\lambda\tau_{A}} \mathbb{1}_{\tau_{A} < \tau_{B}} \mathbb{1}_{\sigma(\tau_{A})=\eta}\right] \mathbb{E}_{\eta}\left[e^{-\lambda\tau_{B}}\right].$$
(1.104)

Let us now consider the last term. By reversibility, it holds that

$$\sum_{\sigma \in A} \mu_N(\sigma) \mathbb{E}_{\sigma} \Big[ e^{-\lambda \tau_A} \mathbb{1}_{\tau_A < \tau_B} \mathbb{1}_{X(\tau_A) = \eta} \Big] = \mu_N(\eta) \mathbb{E}_{\eta} \Big[ e^{-\lambda \tau_A} \mathbb{1}_{\tau_A < \tau_B} \Big].$$
(1.105)

Combining (1.105) with (1.104) and solving for  $\mathbb{E}_{\eta}\left[e^{-\lambda\tau_{B}}\right]$  yields

$$\frac{1}{\mu_N[A]} \sum_{\eta \in A} \mu_N(\sigma) \left( 1 - \mathbb{E}_{\eta} \left[ e^{-\lambda \tau_A} \mathbb{1}_{\tau_A < \tau_B} \right] \right) \mathbb{E}_{\eta} \left[ e^{-\lambda \tau_B} \right] = \mathbb{E}_{\mu_A} \left[ e^{-\lambda \tau_B} \mathbb{1}_{\tau_B < \tau_A} \right].$$

Since  $\lambda \geq 0$ , we can normalize the measure on the left-hand side. In view of (1.102), the normalization constant is equal to  $\mu_N[A] - \mu_N[A] \mathbb{E}_{\mu_A}\left[e^{-\lambda \tau_A} \mathbb{1}_{\tau_A < \tau_B}\right]$  and hence, (1.103) follows.

Proof of Lemma 1.25 via Green's function. For any  $\sigma \in A$  consider the Poisson kernel representation of  $w_B^{\lambda}$  given by

$$w_B^{\lambda}(\sigma) = \sum_{\eta \in B} \frac{\mu_N(\eta)}{\mu_N(\sigma)} \left( L_N^{\lambda} G_B^{\lambda} \right)(\eta, \sigma) = \sum_{\eta \in B} \frac{\mu_N(\eta)}{\mu_N(\sigma)} \left( L_N^{\lambda} G_B^{\lambda} \right)(\eta, \sigma) h_{B,A}^{\lambda}(\eta).$$
(1.106)

Applying the second Green's identity to the functions  $h_{B,A}^{\lambda}$  and  $G_{B}^{\lambda}(\cdot, \sigma)$ , we obtain

$$\sum_{\eta \in B} \mu_N(\eta) h_{B,A}^{\lambda}(\eta) \left( L_N^{\lambda} G_B^{\lambda} \right)(\eta, \sigma) = \sum_{\eta \notin B} \mu_N(\eta) G_B^{\lambda}(\eta, \sigma) \left( L_N^{\lambda} h_{B,A}^{\lambda} \right)(\eta).$$
(1.107)

Hence, combining (1.106) and (1.106) and exploiting that  $L_N^{\lambda} h_{B,A}^{\lambda}$  vanishes on  $(A \cup B)^c$ , we get

$$w_B^{\lambda}(\sigma) = \sum_{\eta \in A} \frac{\mu_N(\eta)}{\mu_N(\sigma)} G_B^{\lambda}(\eta, \sigma) \left( L^{\lambda} h_{B,A}^{\lambda} \right)(\eta).$$
(1.108)

Multiplying both sides with  $\mu_N(\sigma) e_{A,B}^{\lambda}(\sigma)$  and summing over all  $\sigma \in A$ , provides

$$\sum_{\sigma \in A} \mu_N(\sigma) e_{A,B}^{\lambda}(\sigma) w_B^{\lambda}(\sigma) = \sum_{\eta, \sigma \in A} \mu_N(\eta) \left( L^{\lambda} h_{B,A}^{\lambda} \right)(\eta) G_B^{\lambda}(\eta, \sigma) e_{A,B}^{\lambda}(\sigma)$$
$$= \sum_{\eta \in A} \mu_N(\eta) \left( L^{\lambda} h_{B,A}^{\lambda} \right)(\eta),$$
(1.109)

where we used in the second step the representation (1.36) of  $h_{A,B}^{\lambda}$  in terms of the Green function and that  $h_{A,B}^{\lambda}(\eta) = 1$ , for all  $\eta \in A$ . By normalizing the measure on the left-hand side and using the probabilistic interpretation (1.21) of  $e_{A,B}^{\lambda}$  concludes the proof.

Proof of Lemma 1.25 via second Green's formula. By applying the second Green's identity to the functions  $h_{A,B}^{\lambda}$  and  $w_{B}^{\lambda}$ , we get

$$\sum_{\sigma \in A} \mu_N(\sigma) \left( L^{\lambda} h_{A,B}^{\lambda} \right)(\sigma) w_B^{\lambda}(\sigma) = -\sum_{\sigma \in B} \mu_N(\sigma) \left( L^{\lambda} h_{A,B}^{\lambda} \right)(\sigma)$$
(1.110)

A further application of the second Green's identity to the functions  $h_{A,B}^{\lambda}$  and  $h_{B,A}^{\lambda}$  yields

$$\sum_{\sigma \in A} \mu_N(\sigma) \left( L^{\lambda} h_{B,A}^{\lambda} \right)(\sigma) = \sum_{\sigma \in B} \mu_N(\sigma) \left( L^{\lambda} h_{A,B}^{\lambda} \right)(\sigma).$$
(1.111)

Hence, by combining these equations, we obtain

$$\sum_{\sigma \in A} \mu_N(\sigma) e_{A,B}^{\lambda}(\sigma) w_B^{\lambda}(\sigma) = -\sum_{\sigma \in A} \mu_N(\sigma) e_{B,A}^{\lambda}(\sigma).$$
(1.112)

Finally, dividing both sides by  $\sum_{\sigma \in A} \mu_N(\sigma) e_{A,B}^{\lambda}(\sigma)$  and exploiting (1.21) as well as the definition of  $\rho_{\lambda}$  we conclude the proof.

## 1.9. Almost ultrametricity of capacities

In order to derive general results under the definition of metastability, the almost ultrametricity of capacities plays a crucial role. This has been noted first in [16]. The proof given there is purely probabilistic and relies on splitting and renewal ideas. In the following, we will present two different proofs of a generalized statement which, in addition, allows to control capacities between different sets. While our first proof, based heavily on analytic arguments, is maybe less intuitive from a probabilistic point of view, it is considerably simpler than the second one which takes advantage of the strong Markov property. The control obtained is crucial for the investigation in the next sections. **Lemma 1.26.** Let  $X, Y, A \subset S_N$  be mutually disjoint. Assume that there exists  $0 < \delta < \frac{1}{C}$ , where

$$C := \max\left\{\frac{\mathbb{P}_{\nu_{X,A}}[\tau_A < \tau_Y]}{\mathbb{P}_{\nu_{X,A\cup Y}}[\tau_A < \tau_Y]}, \frac{\mathbb{P}_{\nu_{Y,A}}[\tau_A < \tau_X]}{\mathbb{P}_{\nu_{Y,X}}[\tau_A < \tau_X]}\right\} = \max\{C_1, C_2\}$$
(1.113)

such that  $cap(X, A) \leq \delta cap(X, Y)$ . Then

$$1 - C_2 \delta \leq \frac{\operatorname{cap}(X, A)}{\operatorname{cap}(Y, A)} \leq \frac{1}{1 - C_1 \delta}.$$
 (1.114)

*Proof of Lemma 1.26 via analytic arguments.* An application of the second Green's identity to the functions  $h_{X,A}$  and  $h_{Y,A}$  reveals

$$\sum_{\sigma \in X} \mu_N(\sigma) \left( Lh_{X,A} \right)(\sigma) h_{Y,A}(\sigma) = \sum_{\sigma \in Y} \mu_N(\sigma) \left( Lh_{Y,A} \right)(\sigma) h_{X,A}(\sigma).$$
(1.115)

\_

Hence, by normalizing the measure on the left-hand side as well as the one on the right-hand side, we immediately obtain

$$1 - \mathbb{P}_{\nu_{Y,A}}\left[\tau_A < \tau_X\right] \leq \frac{\operatorname{cap}(X,A)}{\operatorname{cap}(Y,A)} = \frac{\mathbb{P}_{\nu_{Y,A}}\left[\tau_X < \tau_A\right]}{\mathbb{P}_{\nu_{X,A}}\left[\tau_Y < \tau_A\right]} \leq \frac{1}{1 - \mathbb{P}_{\nu_{X,A}}\left[\tau_A < \tau_Y\right]}.$$
 (1.116)

Now, the assertion of the lemma follows once we have proven that  $\mathbb{P}_{\nu_{X,A\cup Y}}[\tau_A < \tau_Y] \leq \delta$  and  $\mathbb{P}_{\nu_{Y,X}}[\tau_A < \tau_X] \leq \delta$ . From the renewal equation (1.90) we immediately get that

$$\mathbb{P}_{\nu_{X,A\cup Y}}\left[\tau_A < \tau_Y\right] \leq \frac{\operatorname{cap}(X,A)}{\operatorname{cap}(X,Y)} \leq \delta.$$
(1.117)

On the other hand a further application of the seconds Green's identity to the functions  $h_{A,X}$ and  $h_{Y,X}$  shows that

$$\sum_{\sigma \in A} \mu_N(\sigma) \left( Lh_{Y,X} \right)(\sigma) h_{A,X}(\sigma) = \sum_{\sigma \in Y} \mu_N(\sigma) \left( Lh_{A,X} \right)(\sigma) h_{Y,X}(\sigma)$$
(1.118)

which is equivalent to

$$\mathbb{P}_{\nu_{Y,X}}\left[\tau_A < \tau_X\right] = \frac{\operatorname{cap}(X, A)}{\operatorname{cap}(X, Y)} \mathbb{P}_{\nu_{A,X}}\left[\tau_Y < \tau_X\right] \le \delta.$$
(1.119)

This proves (1.114).

*Proof of Lemma 1.26 via probabilistic arguments.* In order to prove the lower bound, consider the Markov process with initial distribution  $\mu_A$ . Then

$$\mathbb{P}_{\mu_{A}}[\tau_{Y} < \tau_{A}] = \mathbb{P}_{\mu_{A}}[\tau_{Y} < \tau_{A}, \tau_{X} < \tau_{A}] + \mathbb{P}_{\mu_{A}}[\tau_{Y} < \tau_{A}, \tau_{A} < \tau_{X}] \\
\leq \mathbb{P}_{\mu_{A}}[\tau_{X} < \tau_{A}] + \sum_{\eta \in Y} \mathbb{P}_{\mu_{A}}[\tau_{Y} < \tau_{A \cup X}, \sigma(\tau_{Y}) = \eta] \mathbb{P}_{\eta}[\tau_{A} < \tau_{X}]. \quad (1.120)$$

By reversibility, we have for all  $\eta \in Y$  that

$$\sum_{\sigma \in A} \mu_N(\sigma) \mathbb{P}_{\sigma} \big[ \tau_Y < \tau_{A \cup X}, \, \sigma(\tau_Y) = \eta \big] = \mu_N(\eta) \mathbb{P}_{\eta} \big[ \tau_A < \tau_{Y \cup X} \big] \leq \mu_N(\eta) \mathbb{P}_{\eta} \big[ \tau_A < \tau_Y \big].$$
(1.121)

Hence, (1.120) can be rewritten as

$$\mathbb{P}_{\mu_{A}}\big[\tau_{Y} < \tau_{A}\big] \leq \mathbb{P}_{\mu_{A}}\big[\tau_{X} < \tau_{A}\big] + \frac{1}{\mu_{N}[A]} \sum_{\eta \in Y} \mu_{N}(\eta) \mathbb{P}_{\eta}\big[\tau_{A} < \tau_{Y}\big] \mathbb{P}_{\eta}\big[\tau_{A} < \tau_{X}\big] \quad (1.122)$$

which is equivalent to

$$\mathbb{P}_{\nu_{Y,A}}\left[\tau_X < \tau_A\right] = \sum_{\eta \in Y} \frac{\mu_N(\eta) \mathbb{P}_{\eta}\left[\tau_A < \tau_Y\right]}{\operatorname{cap}(Y,A)} \mathbb{P}_{\eta}\left[\tau_X < \tau_A\right] \le \frac{\operatorname{cap}(X,A)}{\operatorname{cap}(Y,A)}.$$
 (1.123)

#### 1.9. ALMOST ULTRAMETRICITY OF CAPACITIES

By a similar computation, we derive that  $\mathbb{P}_{\nu_{Y,X}}[\tau_A < \tau_X] \leq \operatorname{cap}(X, A) / \operatorname{cap}(Y, X)$ . Hence,

$$\frac{\operatorname{cap}(X,A)}{\operatorname{cap}(Y,A)} \ge 1 - \mathbb{P}_{\nu_{Y,A}}\left[\tau_A < \tau_X\right] \ge 1 - C_2 \frac{\operatorname{cap}(X,A)}{\operatorname{cap}(Y,X)} \ge 1 - C_2 \,\delta, \tag{1.124}$$

where we used that  $cap(X, A) \leq \delta cap(X, Y)$  by assumption. To obtain an upper bound on the ratio of cap(X, A) and cap(Y, A), we repeat the computation leading to (1.123) with Y and X interchanged. This gives

$$\mathbb{P}_{\nu_{X,A}}\left[\tau_Y < \tau_A\right] \leq \frac{\operatorname{cap}(Y,A)}{\operatorname{cap}(X,A)}.$$
(1.125)

On the other hand, the averaged renewal equation (1.90) implies

$$\mathbb{P}_{\nu_{X,B\cup M}}\left[\tau_B < \tau_M\right] \leq \frac{\operatorname{cap}(X,B)}{\operatorname{cap}(M,B)} \leq \delta.$$
(1.126)

Hence, by combing (1.126) with (1.125) we finally get that

$$\frac{\operatorname{cap}(X,A)}{\operatorname{cap}(Y,A)} \leq \left(1 - \mathbb{P}_{\nu_{X,A}}\left[\tau_A < \tau_Y\right]\right)^{-1} \leq \left(1 - C_1 \mathbb{P}_{\nu_{X,A\cup Y}}\left[\tau_A < \tau_Y\right]\right)^{-1} \leq \left(1 - C_1 \delta\right)^{-1}.$$
(1.127)

This completes the proof.

**Lemma 1.27.** Let  $X, A \subset S_N$  be disjoint and  $Y \subsetneq X$ . Assume that there exists  $0 < \delta < \frac{1}{C}$  with  $C = \mathbb{P}_{\nu_{X,A}} \left[ \tau_A < \tau_Y \right] / \mathbb{P}_{\nu_{X \setminus Y,Y}} \left[ \tau_A < \tau_Y \right]$  such that  $\operatorname{cap}(Y, A) \leq \delta \operatorname{cap}(Y, X \setminus Y)$ . Then

$$1 - C\delta \leq \frac{\operatorname{cap}(Y, A)}{\operatorname{cap}(X, A)} \leq 1.$$
(1.128)

*Proof.* By applying the second Green's identity to the functions  $h_{Y,A}$  and  $h_{X,A}$  we get

$$\sum_{\sigma \in Y} \mu_N(\sigma) \left( Lh_{Y,A} \right)(\sigma) = \sum_{\sigma \in X} \mu_N(\sigma) \left( Lh_{X,A} \right)(\sigma) h_{Y,A}(\sigma).$$

Using the property that the harmonic function  $h_{Y,A}$  is bounded from above by one as well as the definition of capacities, the upper bound in (1.128) is immediate. Moreover, by replacing  $h_{Y,A}(\sigma)$  by  $\mathbb{P}_{\sigma}[\tau_Y < \tau_A]$  for all  $\sigma \in Y$ , provides the following lower bound

$$\frac{\operatorname{cap}(Y,A)}{\operatorname{cap}(X,A)} \geq \mathbb{P}_{\nu_{X,A}} \big[ \tau_Y < \tau_A \big] = 1 - C \mathbb{P}_{\nu_{X\setminus Y,Y}} \big[ \tau_A < \tau_Y \big].$$

Hence, it remains to show that  $\mathbb{P}_{\nu_{X \setminus Y,Y}}[\tau_A < \tau_Y] \leq \delta$ . Now, a further application of the second Green's identity to the functions  $h_{X \setminus Y,Y}$  and  $h_{A,Y}$  shows that

$$\sum_{\sigma \in X \setminus Y} \mu_N(\sigma) \, h_{A,Y}(\sigma) \left( Lh_{X \setminus Y,Y} \right)(\sigma) = \sum_{\sigma \in A} \, \mu_N(\sigma) \, h_{X \setminus Y,Y}(\sigma) \left( Lh_{A,Y} \right)(\sigma)$$

which is equivalent to

$$\mathbb{P}_{\nu_{X\setminus Y,Y}}\left[\tau_A < \tau_Y\right] = \frac{\operatorname{cap}(Y,A)}{\operatorname{cap}(Y,X\setminus Y)} \mathbb{P}_{\nu_{A,Y}}\left[\tau_{X\setminus Y} < \tau_Y\right] \leq \delta.$$

This completes the proof of (1.128).

**Corollary 1.28.** Let  $X, Y, A \subset S_N$  be mutually disjoint,  $C_2$  be given via (1.113) and consider  $0 < \delta < \frac{1}{1+C_2}$ . Then,

$$\operatorname{cap}(X,A) \ge \delta \min\left\{\operatorname{cap}(X,Y), \operatorname{cap}(Y,A)\right\}.$$
(1.129)

*Proof.* By contradiction. Assume that  $cap(X, A) < \delta \min \{cap(X, Y), cap(Y, A)\}$ . This implies in particular that  $cap(X, A) < \delta cap(X, Y)$ . Hence, by Lemma 1.26 it holds that

$$\operatorname{cap}(Y,A) \leq \frac{1}{1 - C_2 \,\delta} \, \operatorname{cap}(X,A) \leq \frac{\delta}{1 - C_2 \,\delta} \, \operatorname{cap}(Y,A)$$

which is in contradiction with the assumption  $\delta < \frac{1}{1+C_2}.$
# CHAPTER 2

# Coupling arguments and pointwise estimates

In the previous chapter we have presented the first important ingredient of the potential theoretic approach to metastability. Namely, a formula that relates the averaged mean hitting time of a set *B* when the process is started in a specific measure,  $\nu_{A,B}$ , on the set *A* with the capacity between these sets and the corresponding equilibrium potential. An natural question that comes to mind is whether the mean hitting time of *B* really depend on  $\nu_{A,B}$ .

This chapter has three main objectives. The first is to couple two versions of the Markov chain  $\{\sigma(t)\}$ , that starts in different configurations. For later applications the coupling should be constructed in such a way that allows to control the probability that both chains have be merged until a given time T. For this purpose, we first describe in Section 2.1 a general setting in which our methods can be applied. Notice that this setting is slightly more specific compared to the one considered in the previous chapter. Afterwards, we present in Section 2.2 the actual construction of the coupling. Let us point out that this tool is crucial in the later investigations. In Section 2.3 we address the second objective. We prove an estimate on the oscillations of harmonic functions within mesoscopic sets. Although the resulting bounds on equilibrium potentials are quiet rough, they are sufficient to compute precisely e.g. averaged metastable exit times in the next chapter. Let us point out that a rough estimate of the regularity of harmonic functions within mesoscopic sets is the second important ingredient of our approach. The third objective of this chapter is to derive sharp of mean hitting times and its Laplace transform. In situations where the Markov chain returns often to a small neighborhood of its starting point  $\sigma \in A$  before escaping to the set B, one expects that mean exit times are almost constant as functions on the starting configuration. By taking advantage of this kind of recurrence property, in Section 2.4 we decompose the path of a Markov chain into cycles, and we prove as a first step a lower bound for the probability that the Markov chain returns to Abefore escaping to the set B. As presented in Section 2.4 this cycle decomposition is the tool in order to establish in a second step the desired sharp estimates.

**Contributions in this chapter.** In a recent paper [5], Bianchi, Bovier and Ioffe studied some aspects concerning the regularity of harmonic functions and mean hitting times for a certain class of Markov chains with state space  $\{-1,1\}^N$ . In particular, they introduced the strategy to consider a cycle decomposition. One intention of the present chapter is to continue these investigations and to extend the results to a class of Markov chains on  $S_0^N$  where  $S_0$  is a finite set. The author contributes the following novel pieces:

- a different proof for an optimal coupling of two measures on a finite set, see 2.3, and an extension of the coupling construction;
- a simple proof based on coupling methods that allows to control the tail probability of involved coin tosses, see Lemma 2.6;
- a model independent proof that the return probability is close to one, see Lemma 2.11.

## 2.1. Introduction

Coupling methods are well-known in all branches of probability theory and many textbooks on stochastic processes include this useful tool [**110**, **82**]. Since we use this method heavily in this chapter, it might be a good idea to recall its definition.

**Definition 2.1** (Coupling). A coupling of two probability distributions  $\mu$  and  $\nu$  is a pair of random variables (X, Y) defined on a single probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  such that the marginal distribution of X is  $\mu$  and the marginal distribution of Y is  $\nu$ . That is, a coupling (X, Y) satisfies  $\mathbb{P}[X = \sigma] = \mu(\sigma)$  and  $\mathbb{P}[Y = \eta] = \nu(\eta)$ .

In particular, we will consider a Markovian coupling of two Markov chains  $\{\sigma(t)\}$  and  $\{\eta(t)\}$ on  $S_N$ , having the same transition probabilities  $p_N(\sigma, \eta)$ , which start in two different initial distributions. That is, a Markov chain  $\{(\sigma(t), \eta(t))\}$  on  $S_N \times S_N$  such that the marginal processes have transition probabilities  $p_N(\sigma, \eta)$ . Of course, such a coupling will only be useful if we introduce a coupling mechanism that connects both Markov chains in a non-trivial way.

Setting. We consider a Markov chain  $\{\sigma(t)\}_{t\in\mathbb{N}_0}$  on a finite state space  $S_N = \{1, \ldots, q\}^N$  with transition probabilities denoted by  $p_N$ . Here, N is a large parameter. We assume that the process is *irreducible* and *reversible* with respect to the unique stationary (Gibbs) measure  $\mu_N$ . The transition probabilities  $p_N$  are such that the chain should evolve by selecting at each step a site  $i \in \{1, \ldots, N\}$  uniformly at random and setting the corresponding spin variable to a randomly chosen  $r \in S_0 = \{1, \ldots, q\}$  according to the distribution  $Np_N(\sigma, \sigma^{i, \cdot})$  on  $S_0$ . Elements of  $S_0$  will be called *colors*. Moreover, we assume that there exists  $\alpha > 0$  such that

$$N p_N(\sigma, \sigma^{i,r}) \ge \alpha \quad \forall \sigma \in \mathcal{S}_N, \quad i \in \{1, \dots, N\} \text{ and } r \in \mathcal{S}_0.$$
 (2.1)

Here, the configuration  $\sigma^{i,r}$  is obtained from  $\sigma$  by replacing the color  $\sigma_i$  at site *i* through  $r \in S_0$ . Again, we denote by  $L_N$  the discrete generator which acts on functions  $f : S_N \to \mathbb{R}$  as

$$(L_N f)(\sigma) = \sum_{\eta \in S_N} p_N(\sigma, \eta) (f(\eta) - f(\sigma)).$$

Moreover, for a given subset  $X \subset S_N$  we introduce the notation  $\mu_X(\sigma) = \mu_N[\sigma | X]$  for  $\sigma \in X$  to denote the reversible measure,  $\mu_N$ , conditioned on X.

Given a sequence of partitions  $\{\Lambda^n\}_{n\in\mathbb{N}} \equiv \{\Lambda_1^n, \ldots, \Lambda_{k_n}^n\}_{n\in\mathbb{N}}$  of  $\{1, \ldots, N\}$  having the property that  $\Lambda^{n+1}$  is a refinement of  $\Lambda^n$ , we consider a family of maps  $\boldsymbol{\varrho}^n : S_N \to \Gamma^n \subset \mathbb{R}^{k_n \cdot q}$ 

$$\boldsymbol{\varrho}^{n}(\sigma) := \sum_{k=1}^{k_{n}} e^{k} \otimes \frac{1}{N} \sum_{i \in \Lambda_{k}^{n}} \delta_{\sigma_{i}}$$
(2.2)

where  $e^k \in \mathbb{R}^{k_n}$  denotes the coordinate vector in  $\mathbb{R}^{k_n}$  and  $\delta_x$  is the point-mass at  $x \in \mathbb{R}$ . One may think of the maps  $\rho^n$  as a vector of averages of *microscopic variables*  $\sigma_i$  over blocks of *mesoscopic sizes* which are decreasing in n.

Notice that the image process  $\{ \boldsymbol{\varrho}^n(\sigma(t)) \}$  on  $\Gamma^n$  is in general not Markovian. However, there is a canonical Markov process  $\{ \boldsymbol{\varrho}^n(t) \}$  with state space  $\Gamma^n$  which is reversible with respect to the measure  $\mathcal{Q}^n := \mu_N \circ (\boldsymbol{\varrho}^n)^{-1}$  having the property that  $\boldsymbol{\varrho}^n(t) = \boldsymbol{\varrho}^n(\sigma(t))$  in law whenever  $\{ \boldsymbol{\varrho}^n(\sigma(t)) \}$  is a Markov process. Its transition probabilities are given by

$$r^{n}(\boldsymbol{x},\boldsymbol{y}) := \frac{1}{\mathcal{Q}^{n}(\boldsymbol{x})} \sum_{\sigma \in \mathcal{S}^{n}[\boldsymbol{x}]} \mu_{N}(\sigma) \sum_{\eta \in \mathcal{S}^{n}[\boldsymbol{y}]} p_{N}(\sigma,\eta), \qquad \boldsymbol{x},\boldsymbol{y} \in \Gamma^{n}.$$
(2.3)

To simplify notation, we denote by  $\mathcal{S}^n[x] := (\varrho^n)^{-1}(x)$  the set-valued preimage of  $\varrho^n$ .

Now, the key assumption is to choose the sequence of partitions  $\{\Lambda^n\}$  in such a way that the following two conditions holds true:

#### 2.1. INTRODUCTION

(A.1) There exists  $\varepsilon(n) \downarrow 0$  as  $n \uparrow \infty$  such that for any  $x, y \in \Gamma^n$  with  $r^n(x, y) > 0$ ,

$$\max_{\boldsymbol{\sigma}\in\mathcal{S}^{n}[\boldsymbol{x}],\,\boldsymbol{\eta}\in\mathcal{S}^{n}[\boldsymbol{y}]}\left|\frac{p_{N}(\boldsymbol{\sigma},\boldsymbol{\eta})\left|\{\boldsymbol{\eta}\in\mathcal{S}^{n}[\boldsymbol{y}]\mid d_{\mathrm{H}}(\boldsymbol{\sigma},\boldsymbol{\eta})=1\}\right|}{r^{n}(\boldsymbol{x},\boldsymbol{y})}-1\right| \leq \frac{\varepsilon(n)}{3},\tag{2.4}$$

(A.2) If  $\boldsymbol{\varrho}^n(\sigma) = \boldsymbol{\varrho}^n(\eta)$  and  $\sigma_i = \eta_i$ , for some *i*, then  $p_N(\sigma, \sigma^{i,r}) = p_N(\eta, \eta^{i,r})$  for all  $r \in \mathcal{S}$ .

Let us remark that, provided both conditions above are satisfied, the Markov process  $\{ \boldsymbol{\varrho}^n(t) \}$  can be seen as a good approximation of the process  $\{ \boldsymbol{\varrho}^n(\sigma(t)) \}$ .

Our starting point in the study of the dynamical behavior of  $\{\sigma(t)\}$  will be the following notion of *mimics a metastable situation on level n*.

**Definition 2.2** (Metastable situation). Let us consider two disjoint subsets  $A, B \subset S_N$  for which there exists  $n \in \mathbb{N}$  such that A and B are given as the preimage of some  $A, B \subset \Gamma^n$  under  $\varrho^n$ . We say that A, B mimics a metastable situation of level n if there exists constants 0 < c < C such that

(i) for all 
$$X = S^n[x]$$
 with  $x \in A$ 

$$\mathbb{P}_{\mu_X}[\tau_B < \tau_X] \le e^{-CN}, \tag{2.5}$$

(ii) for all 
$$X = S^n[\boldsymbol{x}], Y = S^n[\boldsymbol{y}]$$
 with  $\boldsymbol{x} \neq \boldsymbol{y} \in \boldsymbol{A}$   
$$\mathbb{P}_{\mu_X}[\tau_Y < \tau_X] \geq e^{-cN}.$$

In the sequel, we will make the dependence on n explicit whenever we want to stress it. Otherwise, we will drop the superscript n, identify  $k_n \equiv n$  and refer to the generic partition  $\Lambda_1, \ldots, \Lambda_n$ .

The challenge. In the context of stochastic spin systems, we are typically faced this the following situation. By (2.5), the Markov chain  $\{\sigma(t)\}$  starting in  $\sigma \in A$  will spent an exponential large amount of time in the neighborhood of A. On the other hand, the set A contains an exponentially large number of configurations and for any given configurations  $\sigma, \eta \in A$ , the probability of the Markov chain that starts in  $\sigma$  to hit  $\eta$  before entering in the set B is exponential small as  $N \to \infty$ . Hence,  $\{\sigma(t)\}$  visits not more than a small fraction of A. In particular, the "time of thermalization", i.e. the typical time of two consecutive visits of a single configuration  $\sigma \in A$  before entering in the set B, is of the same order than the metastable exit times. Thus, we can not expect that the chain equilibrates in a neighborhood of A before reaching B. In a recent paper [5], Bianch, Bovier and Ioffe have presented a strategy to derive pointwise estimates of mean hitting times in such a situation for a certain class of Ising-spin system, i.e.  $S_N = \{-1, 1\}^N$ . Their approach is based on an explicit construction of a coupling of two Markov chains  $\{\eta(t)\}$  and  $\{\sigma(t)\}$  starting in different configurations  $\sigma, \eta \in A$ .

In what follows, we let us present the problem, we are faced with, in the construction of a coupling and sketch the main ideas and arguments to overcome these difficulties.

Let  $\sigma$  and  $\eta$  be two configurations with  $\rho^n(\sigma) = \rho^n(\eta)$ . We will construct a coupling  $\{(\sigma(t), \eta(t))\}$  of two microscopic Markov chains starting from initial configuration  $\sigma$  and  $\eta$ , respectively, such that  $\{\sigma(t)\}$  as well as  $\{\eta(t)\}$  are versions of the original Markov chain with transition matrix  $P_N$ . For the application, we have in mind, the coupling should have the property that the Hamming distance between  $\sigma(t)$  and  $\eta(t)$  is non-increasing as long as both microscopic chains have the same mesoscopic value. However, due to the assumption (A.2), when  $\sigma, \eta$  and  $i, j \in \Lambda_k$ , for any k, are such that  $\rho^n(\sigma) = \rho^n(\eta)$  and  $\sigma_i = \eta_j$  then

$$p_N(\sigma, \sigma^{i,r}) = p_N(\eta, \eta^{j,r}) \quad \iff \quad i = j.$$
 (2.7)

Hence, we cannot find a coupling which assures that both chains maintain always the same mesoscopic value. Whenever, for the first time one chain accept the proposed single site update while the other one does not, we will use the independent coupling afterwards.

(2.6)

On the other hand, a closer look at (2.4) reveals that the difference between the probabilities is quite small, in the sense that there exists n large enough such that for all  $\sigma, \eta$  with  $\rho^n(\sigma) = \rho^n(\eta)$  and  $i, j \in \Lambda_k$ , for any  $k \in \{1, \ldots, n\}$ , with  $\sigma_i = \eta_j$ , it holds that

$$\frac{p_N(\eta, \eta^{j,r})}{p_N(\sigma, \sigma^{i,r})} \le e^{\varepsilon(n)}, \quad \forall r \in \mathcal{S}_0.$$
(2.8)

Clearly, our strategy should be to merge both microscopic chains as long as their mesoscopic values coincide. But notice that such a procedure may involve an implicit sampling of  $\eta$ -paths which may distort their statistical properties.

The main idea to overcome this difficulty is to separate path properties of the  $\eta$ -chain from the probability to decide whether  $\sigma(t)$  maintain the same mesoscopic value as  $\eta(t)$ . Moreover, by decomposing the trajectory of  $\{\sigma(t)\}$  into cycles, we can iterate the coupling using each time an independent copy of the  $\eta$ -chain.

## 2.2. Construction of the Coupling

Let  $\sigma$  and  $\eta$  be two configurations with  $\rho^n(\sigma) = \rho^n(\eta)$ . Our first objective in this section is to explain how to couple the probability distributions  $N p_N(\eta, \eta^{i,\cdot})$  and  $N p_N(\sigma, \sigma^{j,\cdot})$  on  $S_0$ where i, j are chosen from some  $\Lambda_k$  such that  $\sigma_i = \eta_j$ . Taking advantage of (2.8), the coupling (X, Y) is constructed in such a way that we can decide in advance by tossing a coin whether Y realizes at least the same value as X.

Note that the actual construction of the coupling is a modification of the *optimal coupling* between two probability measures on a finite state space, as given in [82], which was already used in the easier case  $|S_0| = 2$  in [5].

**Lemma 2.3.** Let  $\mu, \nu \in \mathcal{M}_1(\mathcal{S}_0)$ . Suppose that there exits  $\delta \in (0, 1)$  such that  $\delta \mu(r) \leq \nu(r)$  for all  $r \in \mathcal{S}_0$ . Then there exists an optimal coupling (X, Y) of  $\mu$  and  $\nu$  with the additional property that there exists a Bernoulli- $\delta$ -distributed random variable U independent of X such that

$$\mathbb{P}[Y = s \mid U = 1, X = r] = \mathbb{1}_{r=s}.$$
(2.9)

*Proof.* To start with, choose a color  $r \in S_0$  randomly according to  $\mu$  and set X = r. We use the following procedure to generate Y. Since X = r implies that  $\mu(r) > 0$ 

$$p(r) = \frac{\mu(r) \wedge \nu(r) - \delta \mu(r)}{(1 - \delta) \,\mu(r)}$$
(2.10)

is well defined. Now, flip a coin with probability of heads equal to  $\delta$ . If the coin comes up heads, then set Y = r. Otherwise, flip another coin with probability of heads equal to p(r). If the second coin comes up heads, set again Y = r. If it comes up tails, choose Y according to the probability distribution  $\gamma_r$  on  $S_0 \setminus \{r\}$  given by

$$\gamma_r(s) = \frac{(\nu(s) - \mu(s)) \mathbb{1}_{\nu(s) \ge \mu(s)}}{\sum_{z \in \mathcal{S}_0 \setminus \{r\}} (\nu(z) - \mu(z)) \mathbb{1}_{\nu(z) \ge \mu(z)}}.$$
(2.11)

By this construction the additional property in the statement of the lemma is immediate. It remains to check that *Y* is indeed distributed according to  $\nu$ . Hence, consider

$$\mathbb{P}[Y=r] = \delta \mu(r) + (1-\delta) \mu(r) p(r) + \sum_{s \in S_0 \setminus \{r\}} (1-\delta) \mu(s) (1-p(s)) \gamma_s(r)$$
  
=  $\mu(r) \wedge \nu(r) + (\nu(r) - \mu(r) \wedge \nu(r)) \sum_{s \in S_0 \setminus \{r\}} \frac{(\mu(s) - \nu(s)) \mathbb{1}_{\mu(s) \ge \nu(s)}}{\sum_{z \in S_0 \setminus \{s\}} (\nu(z) - \mu(z)) \mathbb{1}_{\nu(z) \ge \mu(z)}}$   
(2.12)

Expressing the denominator in (2.12) in terms of the total variation distance reveals that

$$\sum_{\substack{z \in S_0 \setminus \{s\}\\\nu(z) \ge \mu(z)}} (\nu(z) - \mu(z)) = \|\mu - \nu\|_{\mathrm{TV}} - (\nu(s) - \mu(s)) \mathbb{1}_{\nu(s) \ge \mu(s)}.$$
(2.13)

Thus, by distinguishing two cases, this implies that the sum in (2.12) equals

$$\sum_{\in \mathcal{S}_0 \setminus \{r\}} \frac{(\mu(s) - \nu(s)) \,\mathbb{1}_{\mu(s) \ge \nu(s)}}{\|\mu - \nu\|_{\mathrm{TV}}} = 1 - \frac{(\mu(r) - \nu(r)) \,\mathbb{1}_{\mu(r) \ge \nu(r)}}{\|\mu - \nu\|_{\mathrm{TV}}}$$
(2.14)

But notice that, since  $\nu(r) - \mu(r) \wedge \nu(r) = (\nu(r) - \mu(r)) \mathbb{1}_{\nu(r) \ge \mu(r)}$ , the last term in (2.14) vanishes. Hence the probability in (2.12) equals  $\nu(r)$ , as desired. In order to prove the optimality of the coupling, it remains to show that  $P[X \ne Y] = \|\mu - \nu\|_{TV}$ . Note that the coupling procedure implies that

$$\mathbb{P}[X \neq Y] = 1 - \sum_{r \in S_0} \left( \delta \,\mu(r) + (1 - \delta) \,\mu(r) \,p(r) \right) = 1 - \sum_{r \in S_0} \,\mu(r) \wedge \nu(r) = \|\mu - \nu\|_{\mathrm{TV}},$$
(2.15)

where we used in the last step that

s

$$\sum_{r \in \mathcal{S}_0} \mu(r) \wedge \nu(r) = \sum_{\substack{r \in \mathcal{S}_0 \\ \mu(r) < \nu(r)}} \mu(r) + \sum_{\substack{r \in \mathcal{S}_0 \\ \mu(r) \ge \nu(r)}} \nu(r) = 1 - \|\mu - \nu\|_{\mathrm{TV}}.$$
 (2.16)

This completes the proof of the lemma.

Let us now explain one step of a coupling that was used by [5] in the study of the random field Curie-Weiss model q = 2. Note that the coupling mechanism was originally invented in [81]. Recall that we consider  $\sigma, \eta$  with  $\rho^n(\sigma) = \rho^n(\eta)$ .

To update both configurations pick a site  $i \in \{1, ..., N\}$  uniformly at random. Whenever  $\eta_i$ and  $\sigma_i$  coincides, (2.7) guaranties that there is a coupling assuring that the  $\sigma$ -chain maintain the same mesoscopic values as the  $\eta$ -chain. On the other hand, if for some k and  $i \in \Lambda_k$  the spin variables  $\eta_i$  and  $\sigma_i$  differs, the idea is to choose j uniformly at random among all sites in  $\Lambda_k$  having the property that  $\eta_j \neq \sigma_j$  and  $\eta_i = \sigma_j$  and to apply Lemma 2.3 to the probability distributions  $N p_N(\eta, \eta^{i, \cdot})$  and  $N p_N(\sigma, \sigma^{j, \cdot})$ . By tossing a coin with success probability  $e^{-\varepsilon(n)}$ , the coupling, used in Lemma 2.3, allows to decide in advance and independently on what happens to the  $\eta$ -chain whether both chains at least maintain the property having the same mesoscopic value.

Let us emphasis the fact that, provided the coin comes up heads, the Hamming distance between  $\eta$  and  $\sigma$  decreases by one whenever a site *i* is updated where  $\sigma_i \neq \eta_i$ . This allows to define an event, depending only on the filtration induced by the  $\eta$ -chain and the outcome of the coin tosses, which guaranties the coalescence of both chains. However, this property is no longer true in the general case  $q \geq 3$ , although the coupling, described above, could also be applied there.

In order to overcome this problem, we modify the coupling above by taking into account a family of bijections  $\pi_t$  between the sites in the configurations  $\eta(t)$  and  $\sigma(t)$ , respectively. Namely, for the initial configurations  $\eta$  and  $\sigma$  let  $\pi_0$  be a permutation of  $\{1, \ldots, N\}$  with the property that

- (i)  $\pi_0(i) = i$ , if and only if,  $\sigma_i = \eta_i$ ,
- (ii)  $\pi_0(i) = j$ , if and only if, there exists  $k \in \{1, ..., n\}$  such that  $i, j \in \Lambda_k$  and

 $\eta_i \neq \sigma_i \land \eta_j \neq \sigma_j \land \eta_i = \sigma_j.$ 



FIGURE 1. Illustration of an update step of  $\eta(t)$  and  $\sigma(t)$ . a) Possible choice of a bijection between sites in the  $\eta$ - and  $\sigma$ -chain and its update in cases (2). b) Corresponding graphical representation in terms of  $\eta(t)$  and  $\pi_t$ . On the event that all coin tosses involved in the construction come up heads, a merging of the  $\eta$ - and  $\sigma$ -chain is indicated by the fact that  $\pi_t$  is the identity, i.e.  $\pi_t$  is represented only by self-loops.

Since we have assumed that  $\rho^n(\sigma) = \rho^n(\eta)$ ,  $\pi_0$  is well defined. Moreover, for any t > 1, the construction of  $\pi_t$  depends only on the filtration induced by the Markov chain started in  $\eta$ . In order to see this, suppose  $\pi_t$  is already given and at time t + 1 site i is drawn. If  $\pi_t(i) = i$  set  $\pi_{t+1} = \pi_t$ . Otherwise, we have to distinguish the following four cases. If

- (1)  $\eta_i(t+1) \neq \eta_{\pi_t(i)}(t)$  and  $\eta_i(t+1) \neq \eta_{\pi_t^{-1}(i)}(t)$  set  $\pi_{t+1} = \pi_t$ . (2)  $\eta_i(t+1) = \eta_{\pi_t(i)}(t)$  and  $\eta_i(t+1) \neq \eta_{\pi_t^{-1}(i)}(t)$  set

$$\pi_{t+1}(\pi_t(i)) = \pi_t(i), \quad \pi_{t+1}(i) = \pi_t(\pi_t(i)), \quad \pi_{t+1}(j) = \pi_t(j), \ \forall j \neq i, \pi_t(i).$$

(3)  $\eta_i(t+1) \neq \eta_{\pi_t(i)}(t)$  and  $\eta_i(t+1) = \eta_{\pi_t^{-1}(i)}(t)$  set

$$\pi_{t+1}(i) = i, \quad \pi_{t+1}(\pi_t^{-1}(i)) = \pi_t(i), \quad \pi_{t+1}(j) = \pi_t(j), \ \forall j \neq i, \pi_t^{-1}(i).$$

(4) 
$$\eta_i(t+1) = \eta_{\pi_t(i)}(t)$$
 and  $\eta_i(t+1) = \eta_{\pi_t^{-1}(i)}(t)$  set

$$\pi_{t+1}(i) = i, \quad \pi_{t+1}(\pi_t(i)) = \pi_t(i), \quad \pi_{t+1}(\pi_t^{-1}(i)) = \pi_t(\pi_t(i)),$$

and  $\pi_{t+1}(j) = \pi_t(j)$  for all  $j \neq i, \pi_t(i), \pi_t^{-1}(i)$ .

An illustration of construction of  $\pi_{t+1}$  from  $\pi_t$  is given in Figure 2.2.

Now, we describe a coupling of two Markov processes starting from the initial configuration  $\eta$  and  $\sigma$ . Recall that  $\eta$  and  $\sigma$  satisfy  $\varrho^n(\eta) = \varrho^n(\sigma)$ . There are two parameters M and T involved those values will be quantified later on. Let  $\{V_0, \ldots, V_{M-1}\}$  be a sequence of i.i.d. Bernoulli random variables with

$$\mathbb{P}\left[V_i=1\right] = 1 - \mathbb{P}\left[V_i=0\right] = e^{-\varepsilon(n)}.$$
(2.17)

Now, set  $\eta(0) = \eta$ ,  $\sigma(0) = \sigma$ ,  $M_0 = 0$  and  $\xi = 0$ . The coupling during the first *T*-steps of its life is given by

for  $t = 0, 1, \dots, T - 1$  do

if  $\xi = 0 \land M_t < M$  then

Choose  $i \in \{1, \ldots, N\}$  uniform at random.

if  $\pi_t(i) = i$  then

Choose  $s \in S_0$  at random according to the distribution  $N p_N(\eta, \eta^{i,s})$ .

Set

$$\sigma_j(t+1) = \begin{cases} \sigma_j(t), & j \neq i \\ s, & j = i \end{cases} \quad \text{and} \quad \eta_j(t+1) = \begin{cases} \eta_j(t), & j \neq i \\ s, & j = i \end{cases}.$$
  
Set  $M_{t+1} = M_t$  and  $\pi_{t+1} = \pi_t$ .

else

Apply the optimal coupling, describe in Lemma 2.3, to  $N p_N(\eta(t), \eta(t)^{i, \cdot})$  and  $N p_N(\sigma(t), \sigma(t)^{\pi_t(i), \cdot})$ , where  $V_{M_t}$  decides if both chains maintain at least the same mesoscopic value.

Set 
$$M_{t+1} = M_t + 1$$
.

if  $V_{M_t} = 1$  then

Generate  $\pi_{t+1}$  as described above.

else

```
Set \xi = 1.
```

end if

end if

else

Use the independent coupling to update  $\eta(t)$  and  $\sigma(t)$ .

end if

# end for

For later reference, we denote by  $\{I_t\}_{t=0}^{T-1}$  the family of independent and uniformly on  $\{1, \ldots, N\}$  distributed random variables describing the random experiment of choosing a site in the algorithm. Moreover, notice that  $\{M_t\}$  is a process that increases by one each time a new coin  $V_i$  is used in the coupling whereas the value  $\xi = 1$  indicates that a coin comes up tails.

**Lemma 2.4.** Let  $\mathbb{P}_{\sigma,\eta}$  be the joint distribution of the processes  $\{\eta(t)\}$ ,  $\{\sigma(t)\}$  and  $\{V_i\}$ . Then, for all  $t \leq T$ , the construction above is a coupling of two Markov chains starting in  $\eta$  and  $\sigma$  with transition matrix  $P_N$ .

*Proof.* The assertion is obvious as soon as  $\xi = 1$  or  $M_t \ge M$  for some t < T, since both chains are updated independently. Hence, it remains to check that the statement holds true when  $\xi = 0$  and  $M_t < M$ . Due to the bijection, each lattice site is drawn in both chains with probability 1/N. Moreover, in view of (2.7) and Lemma 2.3, we can conclude the statement of the lemma.

In this construction, the Hamming distance between  $\eta(t)$  and  $\sigma(t)$  is non-increasing as long as  $\xi = 0$  and  $M_t < M$ . Provided that for some t < T these conditions still hold and both chains coalesce,  $\eta(t) = \sigma(t)$ , then the two dynamics stay automatically together until T. As it was pointed out in [5], conditioned on such a situation involves an implicit sampling of  $\eta$ -paths which may lead to a distortion of their statistical properties. In order to avoid an implicit sampling of paths, we will define two *independent* events which allows to conclude that a merging of the Markov chains started in  $\eta$  and  $\sigma$ , respectively, has taken place whenever both events happen simultaneously. This strategy was successfully demonstrated in [5]. Hence, let us consider

(i) the event that all coin tosses has come up heads, i.e.

$$\mathcal{A} := \{ V_i = 1, \quad \forall i \in \{0, \dots, M - 1\} \}.$$
(2.18)

(ii) an event depending only on the random variables  $\eta(t)$  for t < T. In order to define it, notice that for any realization of  $\{\eta(t)\}$  and given  $\pi_0$ , which is determined by the initial configuration  $\sigma$ , we can uniquely construct  $\{\pi_t\}$ . Moreover, we introduce stopping times,  $\mathfrak{t}_i$ , through

$$\mathfrak{t}_i := \inf \left\{ t \ge 0 \mid \eta_i(t+1) \neq \eta_i(t) \land \eta_i(t+1) = \eta_{\pi_t^{-1}(i)}(t) \text{ if } \pi_t^{-1} \neq i \right\}$$

and set  $\mathfrak{t} := \max_{1 \le i \le N} \mathfrak{t}_i$ . By construction,  $\mathfrak{t}_i$  is the first time that the spin at site *i* has been flipped to a specific or an arbitrary color depending on  $\{\pi_t\}$ . Further define the random variable

$$\mathcal{N} := \sum_{i=1}^{N} \sum_{t=0}^{t_i} \mathbb{1}_{I_t = i}.$$
(2.19)

representing the total number of flipping attempts until time  $\mathfrak t.$  Finally, let us consider the events

$$\mathcal{B} := \{\tau_B \ge T\} \cap \{\mathfrak{t} < T\} \cap \{\mathcal{N} \le M\}.$$
(2.20)

An important observation is the following

**Lemma 2.5.** For any given value of the parameters T and M, on  $A \cap B$  a merging of  $\{\eta(t)\}$  and  $\{\sigma(t)\}$  has taken place, in the sense that

$$\mathcal{A} \cap \mathcal{B} \subset \{\eta(T) = \sigma(T)\}.$$
(2.21)

*Proof.* On the event  $A \cap B$ , the process  $\{\eta(t)\}$  has not reached B by time T but all spins have been flipped at least once to an arbitrary or to a specific color depending on  $\{\pi_t\}$ . Notice that whenever a site i where  $\eta_i \neq \sigma_i$  is flipped to the color specified by  $\{\pi_t\}$ , the corresponding spins become aligned because on  $A \cap B$  all coin tosses come up heads. Since  $\mathfrak{t} < T$  this implies that  $\eta(T) = \sigma(T)$ .

Further, by taking advantage of the property (2.1), we can establish a bound on the tail probability of the random variable N.

**Lemma 2.6.** Let  $c_1 > 1 + \alpha^{-1}$  and set  $M = c_1 N$ . Then, it holds that

$$\mathbb{P}_{\eta}\left[\mathcal{N} > M\right] \leq e^{-c_2 N}. \tag{2.22}$$

with a constant  $c_2 = -c_1 \ln (1 - \alpha^2) - \ln (1 + \alpha)$ .

*Proof.* In view of (2.1), the idea behind the proof is to construct on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  a coupling between the Markov chain  $\{\eta(t)\}$  and a sequence  $\{\omega(t)\}$  of i.i.d. Bernoulli random variables in such a way that we can define a random variable  $\mathfrak{N}$  depending only on the natural filtration induced by  $\{\omega(t)\}$  with the property that  $\mathcal{N} \leq \mathfrak{N}$ . Clearly, by bounding the tail probability of  $\mathfrak{N}$  from above, we obtain an upper bound on the tail probability  $\mathcal{N}$ .

Let us now describe one step of the actual coupling construction. At time t + 1  $(t \ge 0)$  choose a site  $i \in \{1, ..., N\}$  uniformly at random and toss a coin with probability of heads equal to  $\alpha$ . Depending on  $\pi_t$ , we have to distinguish two cases:

(1)  $\pi_t(i) \neq i$ . If the coin comes up heads, set  $\omega(t+1) = 1$  and  $\eta_i(t+1)$  to the color given by  $\eta_{\pi^{-1}(i)}(t)$ . Otherwise, set  $\omega(t+1) = 0$  and draw a  $r \in S_0$  according to the distribution

$$\gamma_{i,t}(r) = \begin{cases} \frac{N p_N(\eta(t), \eta(t)^{i,r}) - \alpha}{1 - \alpha}, & r = \eta_{\pi_t^{-1}(i)}(t) \\ \frac{N p_N(\eta(t), \eta(t)^{i,r})}{1 - \alpha}, & r \neq \eta_{\pi_t^{-1}(i)}(t), \end{cases}$$
(2.23)

(2)  $\pi_t(i) = i$ . If the coin comes up heads, set  $\omega(t+1) = 1$  and choose a color  $r \in S_0 \setminus \{\eta_i(t)\}$ uniformly at random. Otherwise, set  $\omega(t+1) = 0$  and draw a  $r \in S_0$  according to the distribution

$$\gamma_{i,t}(r) = \begin{cases} \frac{N p_N(\eta(t), \eta(t)^{i,r})}{1 - \alpha}, & r = \eta_i(t) \\ \frac{N p_N(\eta(t), \eta(t)^{i,r}) - \frac{\alpha}{q - 1}}{1 - \alpha}, & r \neq \eta_i(t). \end{cases}$$
(2.24)

It is easy to check that the construction above is a coupling  $\{(\eta(t), \omega(t))\}$  of the Markov chain  $\{\eta(t)\}_{t\geq 1}$  and a sequence  $\{\omega(t)\}_{t\geq 1}$  of independent Bernoulli- $\alpha$ -distributed random variables. Further, let us introduce the stopping time

$$\mathfrak{N} := \inf\left\{s \ge 1 \mid \sum_{t=1}^{s} \omega(t) = N\right\} - N.$$
(2.25)

Notice that  $\mathfrak{N}$  is negative binomial distributed to the parameters N and  $\alpha$  and, due to the construction above, it holds that  $\mathcal{N} \leq \mathfrak{N} + N$ . Hence, by standard large deviation estimates, we obtain

$$\mathbb{P}_{\eta}\left[\mathcal{N} > c_1 N\right] \leq \mathbb{P}\left[\mathfrak{N} > (c_1 - 1)N\right] \leq \mathrm{e}^{-NI(c_1 - 1)} \leq \mathrm{e}^{-c_2 N}.$$
(2.26)

The last inequality uses the fact that the entropy I is convex and for all  $c > \alpha^{-1}$  it holds that

$$I(c) = c \ln \frac{c}{1+c} - c \ln (1-\alpha) - \ln (1+c) - \ln \alpha \ge -\ln (1-\alpha^2) c - \ln (1+\alpha) > 0.$$
  
is completes the proof.

This completes the proof.

# 2.3. Bounds on harmonic functions and local recurrence

As a first application of the coupling construction, we will show a statement about the regularity of harmonic functions within a given set of configurations that is determined by the same mesoscopic value.

**Proposition 2.7.** For any  $n \in \mathbb{N}$ , let us consider two disjoint subsets  $X, Y \subset \Gamma^n$  and denote by X, Y their preimages under  $\varrho^n$ . Further, let  $x \in \Gamma^n$  choose  $c_1 > 1 + \alpha^{-1}$ . Then,

$$\mathbb{P}_{\sigma}\left[\tau_{X} < \tau_{Y}\right] \geq e^{-\varepsilon(n) c_{1}N} \left(\mathbb{P}_{\eta}\left[\tau_{X} < \tau_{Y}\right] - e^{-c_{2}N}\right), \qquad \forall \sigma, \eta \in \mathcal{S}^{n}[\boldsymbol{x}],$$
(2.27)

where  $c_2 \equiv c_2(c_1)$  depends linearly on  $c_1$ .

*Proof.* We will use the coupling construction where we choose the involved parameters  $T = \infty$ and  $M = c_1 N$ . Further, consider the following two events

$$\mathcal{B}_1 := \{\tau_X < \mathfrak{t}\} \cap \{\mathcal{N} \le M\} \text{ and } \mathcal{B}_2 := \{\tau_X \ge \mathfrak{t}\} \cap \{\mathcal{N} \le M\}.$$
(2.28)

Analog to Lemma 2.5, on the event  $\mathcal{A} \cap \mathcal{B}_2$ , the process  $\{\eta(t)\}$  has not reached X by time t and at time t the spins at any site i with  $\eta_i \neq \sigma_i$  become aligned. Since the corresponding coin tosses come up heads, this implies that  $\eta(\mathfrak{t}) = \sigma(\mathfrak{t})$  and hence  $\tau_X^{\eta} = \tau_X^{\sigma}$ . Moreover, on the event  $\{\tau_X^{\eta} < \tau_Y^{\eta}\}$ , it holds that  $\tau_Y^{\eta} = \tau_Y^{\sigma}$ , since X, Y are given as preimages under  $\varrho^n$ .

On the event  $\mathcal{A} \cap \mathcal{B}_1$ , the process  $\{\eta(t)\}$  reaches X before time t. Since until time t all necessary coin tosses come up heads, we can conclude that  $\rho^n(\eta(t)) = \rho^n(\sigma(t))$  for all  $t \leq \tau_X$ .

Therefore, due to the assumption on X, Y, it holds that  $\tau_X^{\eta} = \tau_X^{\sigma}$  and on the event  $\{\tau_X^{\sigma} < \tau_Y^{\sigma}\}$  the  $\eta$ -chain can not reach Y before time  $\tau_X^{\eta}$ . Thus,

$$\mathbb{P}_{\sigma}\left[\tau_{X} < \tau_{Y}\right] \geq \mathbb{P}_{\sigma,\eta}\left[\tau_{X}^{\sigma} < \tau_{Y}^{\sigma}, \ \mathcal{A} \cap \mathcal{B}_{1}\right] + \mathbb{P}_{\sigma,\eta}\left[\tau_{X}^{\sigma} < \tau_{Y}^{\sigma}, \ \mathcal{A} \cap \mathcal{B}_{2}\right] \\
= \mathbb{P}_{\sigma,\eta}\left[\tau_{X}^{\eta} < \tau_{Y}^{\eta}, \ \mathcal{A} \cap \mathcal{B}_{1}\right] + \mathbb{P}_{\sigma,\eta}\left[\tau_{X}^{\eta} < \tau_{Y}^{\eta}, \ \mathcal{A} \cap \mathcal{B}_{2}\right] \\
\geq e^{-\varepsilon(n)M}\left(\mathbb{P}_{\eta}\left[\tau_{X} < \tau_{Y}\right] - \mathbb{P}_{\eta}\left[\mathcal{N} > M\right]\right).$$
(2.29)

Taking into account (2.22) concludes the proof of the lemma.

**Remark 2.8.** Instead of (2.29), one can also consider the following lower bound which is based on the *h*-transform

$$\mathbb{P}_{\sigma}[\tau_X < \tau_Y] \geq e^{-\varepsilon(n)M} \mathbb{P}_{\eta}[\tau_X < \tau_Y] \left(1 - \mathbb{P}_{\eta}^h[\mathcal{N} > M]\right).$$

In order to apply Lemma 2.6 to bound  $\mathbb{P}_{\eta}^{h}[\mathcal{N} > M]$ , it remains to show that  $N p_{N}^{h}(\sigma, \eta)$  is bounded from below by a constant which is independent of N for all  $\sigma, \eta$  with  $d_{\mathrm{H}}(\sigma, \eta) \leq 1$ . However, the only bound we can achieve is

$$N p_N^h(\sigma, \eta) = \frac{1}{\mathbb{P}_{\sigma}[\tau_X < \tau_Y]} N p_N(\sigma, \eta) \mathbb{P}_{\eta}[\tau_X < \tau_Y] \ge \frac{\mathbb{P}_{\eta}[\tau_X < \tau_Y]}{\mathbb{P}_{\sigma}[\tau_X < \tau_Y]} \alpha \ge \frac{\alpha^2}{N}$$

where we used in the last step that  $\mathbb{P}_{\eta}[\tau_X < \tau_Y] \ge p_N(\eta, \sigma) \mathbb{P}_{\sigma}[\tau_X < \tau_Y].$ 

**Corollary 2.9.** Suppose that A, B mimics a metastable situation on level n. Let  $c_1 > 1 + \alpha^{-1}$  and assume that n satisfy  $\varepsilon(n)c_1 < C$ . Then, there exists  $c_3 > 0$  such that, for N large enough,

$$\mathbb{P}_{\eta}[\tau_B < \tau_A] \leq e^{-c_3 N}, \qquad \forall \eta \in A.$$
(2.30)

*Proof.* Recall that the disjoint subsets  $A, B \subset S_N$  are given as preimages of  $A, B \subset \Gamma^n$  under  $\varrho^n$ . Further, for any  $\eta \in A$  there exists  $x \in A$  such that  $\eta \in X = S^n[x]$ . Hence, by combining (2.5) and (2.27), we obtain

$$\mathbb{P}_{\eta}\left[\tau_{B} < \tau_{X}\right] \leq e^{\varepsilon(n)c_{1}N} \mathbb{P}_{\mu_{X}}\left[\tau_{B} < \tau_{X}\right] + e^{-c_{2}N} \leq e^{-N\left(C-\varepsilon(n)c_{1}\right)} + e^{-c_{2}N} \leq e^{-c_{3}N},$$
(2.31)

with a constant  $c_3 > 0$ . Since  $\mathbb{P}_{\eta}[\tau_B < \tau_A] \leq \mathbb{P}_{\eta}[\tau_B < \tau_X]$ , we conclude the statement.  $\Box$ 

# 2.4. Cycle decomposition of $\sigma$ -paths

A further reason for introducing the likely event  $\mathcal{B}$  is that it does not distort the hitting time of the Markov chain  $\{\eta(t)\}$ . This is the statement of the following lemma.

**Lemma 2.10.** Uniformly for all  $\eta \in A$ , there exists a constant  $c_4 > 0$ , independent of n, such that for N large enough

$$\mathbb{P}_{\eta}\left[\mathcal{B}^{c}\right] \leq \mathrm{e}^{-c_{4}N} \tag{2.32}$$

and

$$\mathbb{E}_{\eta}[\tau_B \,\mathbb{1}_{\mathcal{B}}] \geq \mathbb{E}_{\eta}[\tau_B] \,\Big(1 - \mathrm{e}^{-c_4 N}\Big). \tag{2.33}$$

*Proof.* In order to prove this lemma, we follow in parts the proof of [5, Lemma 3.3]. We will use the coupling constructed above where we choose  $T = N^{\kappa}$ , for some  $\kappa > 2$ . To start with, notice that the event  $\mathcal{B}^c$  can be written as

$$\mathcal{B}^{c} = \{\tau_{B} \leq N^{\kappa}\} \cup \{\tau_{B} > N^{\kappa}\} \cap \{\mathfrak{t} \geq N^{\kappa}\} \cup \{\tau_{B} > N^{\kappa}\} \cap \{\mathfrak{t} < N^{\kappa}\} \cap \{\mathcal{N} > M\}.$$
(2.34)

In view of (2.34), there are three terms to bound in the equations (2.32) and (2.33). Bounding the first term of (2.33) is quite simple. Namely,

$$\mathbb{E}_{\eta} \big[ \tau_B \, \mathbb{1}_{\tau_B < N^{\kappa}} \big] \leq N^{\kappa} \, \mathbb{P}_{\eta} \big[ \tau_B < N^{\kappa} \big]. \tag{2.35}$$

In order to reach B, we exploit the fact that the Markov chain  $\{\eta(t)\}$  has to make one final excursion without returning to the starting set A and there are at most  $N^{\kappa}$  attempts. This implies that

$$\mathbb{P}_{\eta}\left[\tau_{B} \leq N^{\kappa}\right] \leq N^{\kappa} \max_{\sigma \in A} \mathbb{P}_{\sigma}\left[\tau_{B} < \tau_{A}\right] \leq N^{\kappa} e^{-c_{3}N}$$
(2.36)

where we used in the last step the local recurrence property (2.30). The second term of (2.32) can be easily bounded by

$$\mathbb{P}_{\eta}\big[\tau_B > N^{\kappa}, \mathfrak{t} \ge N^{\kappa}\big] \le \mathbb{P}_{\eta}\big[\mathfrak{t} \ge N^{\kappa}\big].$$
(2.37)

By using the strong Markov property we can split the paths at time  $N^{\kappa}$ . By proceeding in such a way we obtain for the second term of (2.33)

$$\mathbb{E}_{\eta} \left[ \tau_B \, \mathbb{1}_{\tau_B > N^{\kappa}} \, \mathbb{1}_{\mathfrak{t} \ge N^{\kappa}} \right] \leq \left( N^{\kappa} + \max_{\sigma \in \mathcal{S}_N} \, \mathbb{E}_{\sigma} \left[ \tau_B \right] \right) \, \mathbb{P}_{\eta} \left[ \mathfrak{t} \ge N^{\kappa} \right] \leq \, \mathrm{e}^{2c_5 N} \, \mathbb{P}_{\eta} \left[ \mathfrak{t} \ge N^{\kappa} \right]. \tag{2.38}$$

Notice that, in order to obtain a rough upper bound on the mean hitting time  $\mathbb{E}_{\sigma}[\tau_B]$ , it is enough to bound the corresponding capacity  $\operatorname{cap}(\sigma, B)$  from below. Following [15, 11], there exists a constant  $c_5 > 0$ , independent of n, such that  $\operatorname{cap}(\sigma, B) \ge e^{-c_5 N}$ . Thus, it remains to show that the probability  $\mathbb{P}_{\eta}[\mathfrak{t} \ge N^{\kappa}]$  is super-exponentially small. However, at each step the probability to flip a particular spin is bounded from below by  $\alpha/N$ , which implies that

$$\mathbb{P}_{\eta}\left[\mathfrak{t} \ge N^{\kappa}\right] \le N\left(1 - \frac{\alpha}{N}\right)^{N^{\kappa}} \le e^{-c_6 N^{\kappa-1}}.$$
(2.39)

Finally, we can bound the third term by

$$\mathbb{P}_{\eta}\big[\tau_B > N^{\kappa}, \, \mathfrak{t} < N^{\kappa}, \, \mathcal{N} > M\big] \leq \mathbb{P}_{\eta}\big[\mathcal{N} > M\big]$$
(2.40)

and, proceeding similarly as in the treatment of the second term, we obtain

$$\mathbb{E}_{\eta} \left[ \tau_B \, \mathbb{1}_{\tau_B > N^{\kappa}} \, \mathbb{1}_{\mathfrak{t} < N^{\kappa}} \, \mathbb{1}_{\mathcal{N} > M} \right] \leq \left( N^{\kappa} + \max_{\sigma \in \mathcal{S}_N} \, \mathbb{E}_{\sigma} \big[ \tau_B \big] \right) \, \mathbb{P}_{\eta} \big[ \mathcal{N} > M \big] \leq e^{2c_5 N} \, \mathbb{P}_{\eta} \big[ \mathcal{N} > M \big].$$
(2.41)

For  $M = c_1 N$  and  $c_1 > 1 + \alpha^{-1}$  Lemma 2.6 implies that  $\mathbb{P}_{\eta}[N > c_1 N] \leq e^{-c_2 N}$ . Since  $c_2$  increases linearly with  $c_1$ , the right-hand side of (2.41) is exponentially small, if we choose  $c_1$  such that  $c_2 > 2 c_5$ . Thus, by combining all estimates above we conclude the proof of the lemma.

By Lemma 2.5, we know that the Markov chains  $\{\sigma(t)\}$  and  $\{\eta(t)\}$  has merged on the event  $\mathcal{A} \cap \mathcal{B}$  by time T. While the probability of the event  $\mathcal{B}$  is close to one, the probability of the event  $\mathcal{A}$  is small, namely  $e^{-\varepsilon(n)M}$ . Hence, the occurrence of  $(\mathcal{A} \cap \mathcal{B})^c$  is rather likely. However, in view of (2.5) and (2.30), respectively, the  $\sigma$ -chain will return to the set A after time T with a probability close to one. This allows to decompose  $\{\sigma(t) : t \in [0, \tau_B^{\sigma}]\}$  into cycles.

Let us emphasis that the cycles decomposition, given below, goes along the lines of the construction originally presented in [5]. To start with, we define for any  $\eta \in A$  the following stopping times

$$\mathfrak{s}_{-1} = 0, \qquad \mathfrak{s}_k := \inf \{ t > \mathfrak{s}_{k-1} + T \mid \sigma(t) \in \varrho^n(\eta) \}, \qquad \forall k \in \mathbb{N}_0,$$
 (2.42)

as well as the events

$$\mathcal{D}^k := \{\mathfrak{s}_k < \tau_B^\sigma\}, \qquad \forall k \in \mathbb{N}_0.$$
(2.43)

The cycle decomposition of  $\{\sigma(t) : t \in [0, \tau_B^{\sigma}]\}$  is based on a collection of independent copies of Markov chains starting in  $\eta$ ,  $\{\eta^k = \{\eta^k(t) : t \in [0, \tau_B^{\eta,k}]\}\}_k$ , and on a collection of

i.i.d. stacks of coins  $\{V^k \equiv \{V_0^k, \dots, V_{M-1}^k\}\}$ . The corresponding events  $\{\mathcal{A}^k\}$  and  $\{\mathcal{B}^k\}$  are well defined and independent.

To start with, we focus first of all on the case  $\rho^n(\sigma) = \rho^n(\eta)$ . In the first cycle we apply the coupling construction to the Markov chains  $\{\sigma(t)\}$  starting in  $\sigma$  and  $\{\eta^0(t)\}$  up to time T. On the event  $\mathcal{A}^0 \cap \mathcal{B}^0$  we can update  $\sigma(t)$  and  $\eta(t)$  in such a way that  $\sigma(t) = \eta(t)$  for all t > T. In this case the cycle decomposition terminates. On the event  $(\mathcal{A}^0 \cap \mathcal{B}^0)^c$ , we update  $\sigma(t)$  and  $\eta(t)$  independently for all t > T. If  $\mathcal{D}^0$  occurs, the first cycle ends at the random time  $\mathfrak{s}_0$  whereas if  $(\mathcal{D}^0)^c$  happens, the cycle decomposition terminates as well, and a merging has not occurred.

Provided that the cycle decomposition has not terminated so far, in the (k + 1)th cycle for (k > 1), we consider the coupling  $\{(\sigma(\mathfrak{s}_k + t), \eta^k(t))\}_{t=0}^T$  of the Markov chain  $\{\sigma(t)\}$  starting from the configuration  $\sigma(\mathfrak{s}_k)$  and an independent copy  $\{\eta^k(t)\}$  of the  $\eta$ -chain starting in  $\eta$  using the independent sequence of coin tosses  $\{V_1^k, \ldots, V_{M-1}^k\}$  in the coupling construction.

As a consequence, we arrive at the following decomposition of the hitting time  $\tau_B^{\sigma}$  in terms of the (independent) hitting times  $\tau_B^{\eta,k}$ 

$$\tau_{B}^{\sigma} = \sum_{k=0}^{\infty} \left( \left( \mathfrak{s}_{k-1} + \tau_{B}^{\eta,k} \right) \mathbb{1}_{\mathcal{A}^{k}} \mathbb{1}_{\mathcal{B}^{k}} + \tau_{B}^{\sigma} \left( 1 - \mathbb{1}_{\mathcal{D}^{k}} \right) \left( 1 - \mathbb{1}_{\mathcal{A}^{k}} \mathbb{1}_{\mathcal{B}^{k}} \right) \right) \prod_{l=0}^{k-1} \mathbb{1}_{\mathcal{D}^{l}} \left( 1 - \mathbb{1}_{\mathcal{A}^{l}} \mathbb{1}_{\mathcal{B}^{l}} \right).$$

$$(2.44)$$

In the case  $\rho^n(\sigma) \neq \rho^n(\eta)$  we cannot directly apply our coupling construction. Instead, we will first update  $\sigma(t)$  and  $\eta(t)$  independently until time  $\mathfrak{s}_0$ , provided that the event  $\mathcal{D}^0$  occurs, and we use the cycle decomposition, as described above, afterwards. Let us emphasis that in the definition of  $\mathfrak{s}_0$  the involved parameter T is equal to zero. This implies that we have to replace (2.44) by

$$\tau_{B}^{\sigma} = \tau_{B}^{\sigma} \left( 1 - \mathbb{1}_{\mathcal{D}^{0}} \right) \\ + \sum_{k=1}^{\infty} \left( \left( \mathfrak{s}_{k-1} + \tau_{B}^{\eta,k} \right) \mathbb{1}_{\mathcal{A}^{k}} \mathbb{1}_{\mathcal{B}^{k}} + \tau_{B}^{\sigma} \left( 1 - \mathbb{1}_{\mathcal{D}^{k}} \right) \left( 1 - \mathbb{1}_{\mathcal{A}^{k}} \mathbb{1}_{\mathcal{B}^{k}} \right) \right) \mathbb{1}_{\mathcal{D}^{0}} \prod_{l=1}^{k-1} \mathbb{1}_{\mathcal{D}^{l}} \left( 1 - \mathbb{1}_{\mathcal{A}^{l}} \mathbb{1}_{\mathcal{B}^{l}} \right)$$

$$(2.45)$$

In (2.44) and (2.45) we used the convention that products with a negative number of terms are equal to one.

**Lemma 2.11.** Suppose that  $A, B \subset S_N$  mimic a metastable situation on level n. If n satisfies  $\varepsilon(n) c_1 < (C - c)/2$  for some  $c_1 > 1 + \alpha^{-1}$ , then there exists a constant  $c_7 > 0$  such that for all  $\sigma, \eta \in A$  and N large enough

$$\mathbb{P}_{\sigma}\left[\mathcal{D}^{0}\right] \geq 1 - \mathrm{e}^{-c_{7}N}. \tag{2.46}$$

where  $\mathcal{D}^0$  depend on the configuration  $\eta$ .

*Proof.* Let us first consider the case when  $\rho^n(\sigma) = \rho^n(\eta)$ . To start with, fix  $\delta = (C - c)/2$  and set  $X = (\rho^n)^{-1} \circ \rho^n(\eta) \subset A$ . Further, define

$$\boldsymbol{X}_{\delta} := \left\{ \boldsymbol{z} \in \Gamma^{n} \mid \mathbb{P}_{\mu_{X}} \left[ \tau_{Z} < \tau_{X} \right] > \mathrm{e}^{-\delta N}, \ Z = \mathcal{S}^{n}[\boldsymbol{z}] \right\}$$

and set  $X_{\delta} = S^n[\mathbf{X}_{\delta}] \cup X$ . Using the definition of  $\mathcal{D}^0$  together with the strong Markov property we obtain that

$$\mathbb{P}_{\sigma}\big[\mathfrak{s}_{0} < \tau_{B}\big] \geq \sum_{\xi \in X_{\delta}} \mathbb{P}_{\sigma}\big[\sigma(N^{\kappa}) = \xi\big] \mathbb{P}_{\xi}\big[\tau_{X} < \tau_{B}\big] \geq \min_{\xi \in X_{\delta}} \mathbb{P}_{\xi}\big[\tau_{X} < \tau_{B}\big] \mathbb{P}_{\sigma}\big[\sigma(N^{\kappa}) \in X_{\delta}\big].$$
(2.47)

Hence, there remain two terms which we have to bound from below. We will first focus on the last term in (2.47). For any  $Z = S^n[z]$  where  $z \notin X_{\delta}$  it holds that  $\mathbb{P}_{\mu_X}[\tau_Z < \tau_X] \leq e^{-\delta N}$ . Moreover,

$$\mathbb{P}_{\sigma}\big[\sigma(N^{\kappa}) \in Z\big] \leq \mathbb{P}_{\sigma}\big[\tau_{Z} \leq N^{\kappa}\big] \leq N^{\kappa} \max_{\sigma \in X} \mathbb{P}_{\sigma}\big[\tau_{Z} < \tau_{X}\big],$$
(2.48)

where we used in the last step that the Markov chain  $\{\sigma(t)\}$  has to make one final excursion without returning to the starting set X and that there are at most  $N^{\kappa}$  attempts. On the other hand, Proposition 2.7 implies that

$$\mathbb{P}_{\sigma}\left[\tau_{Z} < \tau_{X}\right] \leq e^{\varepsilon(n)c_{1}N} \mathbb{P}_{\mu_{X}}\left[\tau_{Z} < \tau_{X}\right] + e^{-c_{2}N} \leq e^{-N(\delta - \varepsilon(n)c_{1})} + e^{-c_{2}N}.$$
(2.49)

Hence, by combining the last two estimates we obtain

$$\mathbb{P}_{\sigma}\left[\sigma(N^{\kappa}) \in X_{\delta}\right] = 1 - \sum_{\boldsymbol{z} \notin \boldsymbol{X}_{\delta}} \mathbb{P}_{\sigma}\left[\sigma(N^{\kappa}) \in \mathcal{S}^{n}[\boldsymbol{z}]\right] \geq 1 - N^{nq+\kappa} \left(e^{-N(\delta - \varepsilon(n)c_{1})} + e^{-c_{2}N}\right).$$
(2.50)

Let us now focus on the first factor of the right-hand side of (2.47) for any arbitrary  $\xi \in X_{\delta}$ . Suppose that  $\xi \in Z = S^n[z]$  with  $z \in X_{\delta}$ , then from (1.116) we have

$$\mathbb{P}_{\nu_{Z,X}}\left[\tau_B < \tau_X\right] = \frac{\operatorname{cap}(B,X)}{\operatorname{cap}(Z,X)} \mathbb{P}_{\nu_{B,X}}\left[\tau_Z < \tau_X\right] \le \frac{\mathbb{P}_{\mu_X}\left[\tau_B < \tau_X\right]}{\mathbb{P}_{\mu_X}\left[\tau_Z < \tau_X\right]} \le e^{-N(C-\delta)}, \quad (2.51)$$

where we used (2.5) in the last step. Again, Proposition 2.7 implies that

$$\mathbb{P}_{\xi}\left[\tau_{B} < \tau_{X}\right] \leq e^{\varepsilon(n)c_{1}N} \mathbb{P}_{\nu_{Z,X}}\left[\tau_{B} < \tau_{X}\right] + e^{-c_{2}N} \leq e^{-N(C-\delta-\varepsilon(n)c_{1})} + e^{-c_{2}N}, \quad (2.52)$$

whereas in the case when  $\xi \in X$  we get that

$$\mathbb{P}_{\xi}\left[\tau_{B} < \tau_{X}\right] \leq e^{\varepsilon(n)c_{1}N} \mathbb{P}_{\mu_{X}}\left[\tau_{B} < \tau_{X}\right] + e^{-c_{2}N} \leq e^{-N(C-\varepsilon(n)c_{1})} + e^{-c_{2}N}.$$
(2.53)

Thus, combining the last two estimates implies

$$\min_{\xi \in X_{\delta}} \mathbb{P}_{\xi} \big[ \tau_X < \tau_B \big] \ge 1 - \Big( \mathrm{e}^{-N(C - \delta - \varepsilon(n)c_1)} + \mathrm{e}^{-c_2 N} \Big).$$
(2.54)

It remains to consider the case that  $\rho^n(\sigma) \neq \rho^n(\eta)$  for any  $\sigma, \eta \in A$ . Here, we denote by X, Y the set-valued preimage of  $\boldsymbol{x} = \rho^n(\eta)$  and  $\boldsymbol{y} = \rho^n(\sigma)$  under  $\rho^n$ . Recall that in this case the event  $\mathcal{D}^0$  is given by  $\{\tau_X < \tau_B\}$ . Hence, analog to (2.51) we have that

$$\mathbb{P}_{\nu_{Y,X}}\left[\tau_B < \tau_X\right] = \frac{\operatorname{cap}(B,X)}{\operatorname{cap}(Y,X)} \mathbb{P}_{\nu_{B,X}}\left[\tau_Y < \tau_X\right] \le \frac{\mathbb{P}_{\mu_X}\left[\tau_B < \tau_X\right]}{\mathbb{P}_{\mu_X}\left[\tau_Y < \tau_X\right]} \le e^{-N(C-c)}, \quad (2.55)$$

where we used (2.6) in the last step. By Proposition 2.7 we obtain

$$\mathbb{P}_{\sigma}\left[\tau_{B} < \tau_{X}\right] \leq e^{\varepsilon(n)c_{1}N} \mathbb{P}_{\nu_{Y,X}}\left[\tau_{B} < \tau_{X}\right] + e^{-c_{2}N} \leq e^{-N(C-c-\varepsilon(n)c_{1})} + e^{-c_{2}N}$$
(2.56)

which implies that

$$\mathbb{P}_{\sigma}[\tau_X < \tau_B] \ge 1 - \left(e^{-N(C - c - \varepsilon(n)c_1)} + e^{-c_2N}\right).$$
(2.57)

By assumption  $\varepsilon(n) c_1 < (C - c)/2$ . Hence, in view of (2.50), (2.54) and (2.57), there exists  $c_7 > 0$  such that, for large enough N, (2.46) holds.

## 2.5. Bounds on the mean hitting time and the Laplace transform

Finally, in a situation that mimics to be metastable, we prove the following pointwise estimate on metastable times.

**Theorem 2.12.** Suppose that  $A, B \subset S_N$  mimics a metastable situation on level n, and assume that n satisfies  $\varepsilon(n) c_1 < (C - c)/4$  for a given  $c_1 > 1 + \alpha^{-1}$ . Then, there exists  $c_0 > 0$  such that, for N large enough,

$$\max_{\sigma,\eta\in A} \left| \frac{\mathbb{E}_{\sigma}[\tau_B]}{\mathbb{E}_{\eta}[\tau_B]} - 1 \right| \leq e^{-c_0 N}.$$
(2.58)

*Proof.* Let  $\mathbb{E}_{\sigma,\eta}$  denote the expectation with respect to  $\mathbb{P}_{\sigma,\eta}$  on the enlarged probability space. Provided that  $\varrho^n(\sigma) = \varrho^n(\eta)$  for  $\sigma, \eta \in A$ , by (2.44) we have

$$\mathbb{E}_{\sigma}[\tau_{B}] \geq \sum_{k=0}^{\infty} \mathbb{E}_{\sigma,\eta} \left[ \tau_{B}^{\eta,k} \, \mathbb{1}_{\mathcal{A}^{k}} \, \mathbb{1}_{\mathcal{B}^{k}} \prod_{l=0}^{k-1} \, \mathbb{1}_{\mathcal{D}^{l}} \left( 1 - \mathbb{1}_{\mathcal{A}^{l}} \, \mathbb{1}_{\mathcal{B}^{l}} \right) \right]$$
(2.59)

Further, let  $\mathcal{F}_{\mathfrak{s}_k}$  denote the  $\sigma$ -algebra generated by all the events and trajectories  $\mathcal{A}^l, \mathcal{B}^l, \mathcal{D}^l, \eta^l$ and  $\{\sigma(t) : t \in (\mathfrak{s}_{l-1}, \mathfrak{s}_l)\}$  for  $l \leq k$ . Due to the independence of the copies  $\{\eta^l, V^l\}$  it holds that

$$\mathbb{E}_{\sigma,\eta} \left[ \tau_B^{\eta,k} \, \mathbb{1}_{\mathcal{A}^k} \, \mathbb{1}_{\mathcal{B}^k} \, \big| \, \mathcal{F}_{\mathfrak{s}_{k-1}} \right] = \mathbb{E}_{\eta} \left[ \tau_B \, \mathbb{1}_{\mathcal{A}} \, \mathbb{1}_{\mathcal{B}} \right] = e^{-\varepsilon(n)M} \, \mathbb{E}_{\eta} \left[ \tau_B \, \mathbb{1}_{\mathcal{B}} \right]$$
(2.60)

and for all l < k

$$\mathbb{E}_{\sigma,\eta} \left[ \mathbb{1}_{\mathcal{D}^{l}} \left( 1 - \mathbb{1}_{\mathcal{A}^{l}} \mathbb{1}_{\mathcal{B}^{l}} \right) \middle| \mathcal{F}_{\mathfrak{s}_{l-1}} \right] \geq \mathbb{E}_{\eta} \left[ 1 - \mathbb{1}_{\mathcal{A}} \mathbb{1}_{\mathcal{B}} \right] - \max_{\sigma' \in A} \left( 1 - \mathbb{P}_{\sigma'} \left[ \mathcal{D}^{0} \right] \right) \\ \geq 1 - e^{-\varepsilon(n)M} - e^{-c_{7}N}$$
(2.61)

where we used (2.46) in the last step. By choosing  $M = c_1 N$ , as in Lemma 2.6,  $T = N^{\kappa}$  for some  $\kappa > 2$  and  $c_1 > 1 + \alpha^{-1}$  large enough to ensure that  $c_2 > 2c_5$ , then (2.22) and (2.33) implies that

$$\mathbb{E}_{\sigma} \left[ \tau_B \right] \geq \mathbb{E}_{\eta} \left[ \tau_B \, \mathbb{1}_{\mathcal{B}} \right] e^{-\varepsilon(n)c_1 N} \sum_{k=0}^{\infty} \left( 1 - e^{-\varepsilon(n)c_1 N} - e^{-c_7 N} \right)^k \\ \geq \mathbb{E}_{\eta} \left[ \tau_B \right] \frac{1 - e^{-c_4 N}}{1 + e^{-N(c_7 - \varepsilon(n)c_1)}}.$$
(2.62)

If  $\rho^n(\sigma) \neq \rho^n(\eta)$  for  $\sigma, \eta \in A$ , by using (2.45), an analog computation reveals that

$$\mathbb{E}_{\sigma} \left[ \tau_{B} \right] \geq \mathbb{E}_{\eta} \left[ \tau_{B} \mathbb{1}_{\mathcal{B}} \right] e^{-\varepsilon(n)c_{1}N} \left( 1 - e^{-c_{7}N} \right) \sum_{k=0}^{\infty} \left( 1 - e^{-\varepsilon(n)c_{1}N} - e^{-c_{7}N} \right)^{k}$$
$$\geq \mathbb{E}_{\eta} \left[ \tau_{B} \right] \frac{1 - e^{-c_{4}N} - e^{-c_{7}N}}{1 + e^{-N(c_{7}-\varepsilon(n)c_{1})}}.$$
(2.63)

If  $\varepsilon(n)c_1 < (C-c)/4$ , we can choose  $c_7$  in such a way that  $c_7 > (C-c)/4$ . Hence, there exists a constant  $c_0 > 0$  such that  $\mathbb{E}_{\sigma}[\tau_B] \ge \mathbb{E}_{\eta}[\tau_B] (1 - e^{-c_0 N})$  for all  $\sigma, \eta \in A$  and large enough N. This concludes the proof.

As a further application of the coupling construction and the cycle decomposition we will prove a pointwise estimate on the Laplace transform of  $\tau_B$ .

**Proposition 2.13.** Suppose  $A, B \subset S_N$  mimic a metastable situation on level n. If n satisfy  $\varepsilon(n) c_1 < (C - c)/4$  for some  $c_1 > 1 + \alpha^{-1}$  sufficiently large, then there exists  $c_8 > 0$  such that for all  $\lambda > 0$  and N large enough

$$\max_{\sigma,\eta\in A} \left| \frac{\mathbb{E}_{\sigma}\left[ e^{-\frac{\lambda}{T}\tau_{B}} \right]}{\mathbb{E}_{\eta}\left[ e^{-\frac{\lambda}{T}\tau_{B}} \right]} - 1 \right| \leq e^{2\lambda} e^{-c_{8}N},$$
(2.64)

where  $T = \mathbb{E}_{\nu_A}[\tau_B]$  for some probability measure  $\nu_A$  on A.

*Proof.* Let us point out that the proof of this proposition goes along the lines of [5, Prop. 3.4]. Analog to (2.44) we have, for any  $\sigma, \eta \in A$  with  $\varrho^n(\sigma) = \varrho^n(\eta)$ ,

$$e^{-\frac{\lambda}{T}\tau_{B}^{\sigma}} = \sum_{k=0}^{\infty} e^{-\frac{\lambda}{T}(\mathfrak{s}_{k-1}+\tau_{B}^{\eta,k})} \mathbb{1}_{\mathcal{A}^{k}} \mathbb{1}_{\mathcal{B}^{k}} \prod_{l=0}^{k-1} \mathbb{1}_{\mathcal{D}^{l}} \left(1-\mathbb{1}_{\mathcal{A}^{l}} \mathbb{1}_{\mathcal{B}^{l}}\right) + \sum_{k=0}^{\infty} e^{-\frac{\lambda}{T}\tau_{B}^{\sigma}} \left(1-\mathbb{1}_{\mathcal{D}^{k}}\right) \left(1-\mathbb{1}_{\mathcal{A}^{k}} \mathbb{1}_{\mathcal{B}^{k}}\right) \prod_{l=0}^{k-1} \mathbb{1}_{\mathcal{D}^{l}} \left(1-\mathbb{1}_{\mathcal{A}^{l}} \mathbb{1}_{\mathcal{B}^{l}}\right).$$

This implies that

$$\mathbb{E}_{\sigma}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}}\right] \leq \sum_{k=0}^{\infty} \mathbb{E}_{\sigma,\eta}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}^{\eta,k}} \mathbb{1}_{\mathcal{A}^{k}}\mathbb{1}_{\mathcal{B}^{k}} \prod_{l=0}^{k-1} \mathbb{1}_{\mathcal{D}^{l}}\left(1-\mathbb{1}_{\mathcal{A}^{l}}\mathbb{1}_{\mathcal{B}^{l}}\right)\right] \\
+ \sum_{k=0}^{\infty} \mathbb{E}_{\sigma,\eta}\left[\left(1-\mathbb{1}_{\mathcal{D}^{k}}\right)\left(1-\mathbb{1}_{\mathcal{A}^{k}}\mathbb{1}_{\mathcal{B}^{k}}\right) \prod_{l=0}^{k-1} \mathbb{1}_{\mathcal{D}^{l}}\left(1-\mathbb{1}_{\mathcal{A}^{l}}\mathbb{1}_{\mathcal{B}^{l}}\right)\right].$$
(2.65)

Analog to (2.60) and (2.61), we obtain, due to the independence of  $\{\eta^l, V^l\}$ ,

$$\mathbb{E}_{\sigma,\eta}\left[e^{-\frac{\lambda}{T}\tau_{B}^{\eta,k}}\mathbb{1}_{\mathcal{A}^{k}}\mathbb{1}_{\mathcal{B}^{k}}\,\middle|\,\mathcal{F}_{\mathfrak{s}_{k-1}}\right] = \mathbb{E}_{\eta}\left[e^{-\frac{\lambda}{T}\tau_{B}}\mathbb{1}_{\mathcal{A}}\mathbb{1}_{\mathcal{B}}\right] \leq e^{-\varepsilon(n)M} \mathbb{E}_{\eta}\left[e^{-\frac{\lambda}{T}\tau_{B}}\right]$$
(2.66)  
or all  $l < k$ 

and for all l < k

$$\mathbb{E}_{\sigma,\eta} \left[ \mathbb{1}_{\mathcal{D}^{l}} \left( 1 - \mathbb{1}_{\mathcal{A}^{l}} \mathbb{1}_{\mathcal{B}^{l}} \right) \middle| \mathcal{F}_{\mathfrak{s}_{l-1}} \right] \leq \mathbb{E}_{\eta} \left[ 1 - \mathbb{1}_{\mathcal{A}} \mathbb{1}_{\mathcal{B}} \right] = 1 - e^{-\varepsilon(n)M} \mathbb{P}_{\eta} \left[ \mathcal{B} \right].$$
(2.67)

By choosing  $M = c_1 N$ , as in Lemma 2.6, and  $T = N^{\kappa}$  for some  $\kappa > 2$  and taking into account (2.46), we obtain for the second term in (2.65) the following upper bound

$$e^{-c_7 N} \sum_{k=0}^{\infty} \left( 1 - e^{-\varepsilon(n)c_1 N} \left( 1 - e^{-c_4 N} \right) \right)^k \le 2 e^{-N(c_7 - \varepsilon(n)c_1)}$$

for N large enough. Since  $c_7 - \varepsilon(n)c_1 > (C - c)/4 - \varepsilon(n)c_1 > 0$  this upper bound is exponential small in N. Combining all the estimates above, we arrive at

$$\mathbb{E}_{\sigma}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}}\right] \leq \mathbb{E}_{\eta}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}}\right] \mathrm{e}^{-\varepsilon(n)c_{1}N} \sum_{k=0}^{\infty} \left(1 - \mathrm{e}^{-\varepsilon(n)c_{1}N}\left(1 - \mathrm{e}^{-c_{4}N}\right)\right)^{k} + 2 \,\mathrm{e}^{-N(c_{7}-\varepsilon(n)c_{1})} \\
\leq \mathbb{E}_{\eta}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}}\right] \left(1 + 2 \,\mathrm{e}^{-c_{4}N}\right) + 2 \,\mathrm{e}^{-N(c_{7}-\varepsilon(n)c_{1})}.$$
(2.68)

In the case when  $\rho^n(\sigma) = \rho^n(\eta)$  for  $\sigma, \eta \in A$ , by a similar computation we obtain

$$\mathbb{E}_{\sigma}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}}\right] \leq \mathbb{E}_{\eta}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}}\right]\left(1+2\,\mathrm{e}^{-c_{4}N}\right)+2\,\mathrm{e}^{-N(c_{7}-\varepsilon(n)c_{1})}+\mathrm{e}^{-c_{7}N}.$$
(2.69)

By Jensen's inequality and (2.58), for every  $\eta \in A$  it holds that

$$\mathbb{E}_{\eta}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}}\right] \geq \mathrm{e}^{-\frac{\lambda}{T}\mathbb{E}_{\eta}[\tau_{B}]} \geq \mathrm{e}^{-\lambda(1+\mathrm{e}^{-c_{0}N})} \geq \mathrm{e}^{-2\lambda}$$
(2.70)

This implies that there exists  $c_8 > 0$  such that  $\mathbb{E}_{\sigma} \left[ e^{-\frac{\lambda}{T} \tau_B} \right] \leq \mathbb{E}_{\eta} \left[ e^{-\frac{\lambda}{T} \tau_B} \right] \left( 1 + e^{2\lambda} e^{-c_8 N} \right)$  for N large enough.

# CHAPTER 3

# Metastability

We turn now to a characterization of metastability for a class of reversible Markov chains. Our aim is to derive sharp estimates on metastable exit times both on the level of expected values and Laplace transforms by using the tools we have presented in the previous two chapters. For this reason, we restrict ourselves to the setting introduced in Chapter 2.

Starting point of our further investigations is the definition of a set of metastable sets. As a first step, we coarse-grain the state space by introducing a family of mesoscopic variables which depend on an additional parameter n controlling the coarsening level. The key idea behind our definition is that, for a fixed n, these sets are defined as preimages under the mesoscopic map of certain single points in the coarse-grained space. In doing so, the mesoscopic points are chosen in such a way to ensure that the process with initial distribution given by the reversible measure conditioned on one of the corresponding sets returns very often to it before escaping to a "more stable" set.

The present chapter is organized as follows. In Section 3.1 we start with reviewing some of the different definitions and approaches of metastability that have been introduced in the past. Afterwards, we present a definition of a set of metastable sets that will be used in the remaining part of this thesis. As an immediate consequence of this definition combined with a rough control on the regularity of harmonic functions, we show in Section 3.2 the existence of an ultra-metric on the set of metastable sets in the limit when N tends to infinity. As we will further show in Section 3.3, this property allows us to derive various bounds on harmonic functions and to prove sharp estimates for averaged mean hitting times. In the final Section 3.4 we prove the convergence of normalized metastable exit times to an exponential distribution. In the proof we use pointwise estimates of the Laplace transforms of metastable exit times as well as averaged renewal equations.

**Contributions in this chapter.** In the past ten years, substantially progress has been made in the systematical development of the potential theoretic approach and its application by Bovier and coworkers. A crucial observation in this theory is that rough bounds on equilibrium potentials obtained by exact renewal equations suffice to derive sharp estimates on metastable characteristics. In this respect, the main contribution in the present chapter is to work out the details that in situations where exact renewal equations cannot be established it suffices to combine averaged renewal equations with rough pointwise estimates of harmonic functions to prove very accurate results. Beyond that, the author contributes the following bits and pieces:

- a proof how pointwise estimates together with Lemma 1.26 can be used to establish an asymptotic ultra-metric on the set of metastable sets, see Lemma 3.5;
- a proof of various bounds on the equilibrium potential,  $h_{A,B}$ , see Lemma 3.8, that serves as the basis to prove sharp results for averaged mean hitting times;
- a proof of a bound on the λ-equilibrium potential, h<sup>λ</sup><sub>A,B</sub>, that shows that h<sup>λ</sup><sub>A,B</sub> is equal to one in a neighborhood of the set A up to a small correction, see Lemma 3.17.

#### 3. METASTABILITY

## 3.1. Introduction

In the diffusion setting, metastable sets are considerable easy to identify. Namely, they correspond to suitable chosen neighborhoods of local minima of the potential landscape. The dynamical aspects of metastability in such a context of randomly perturbed dynamical systems was first analyzed with mathematical rigor in the seminal work by Freidlin and Wentzell [51]. In models coming from statistical mechanics the question of how to characterize metastable states, and how to identify them in a given model context is really an issue.

**Characterization of metastability: A partial review.** Lebowitz and Penrose introduced a first dynamical approach to metastability that describes the metastable behavior of the systems in terms of *evolution of ensembles*. Their theory takes into account dynamical and statical properties of the system. They characterized *metastable thermodynamical states* by the following properties [**96**, **97**]:

- (i) only one thermodynamical phase is present;
- (ii) if the system is isolated the lifetime of a metastable state is very large, i.e. a system that starts in this state is likely to take a long time to get out;
- (iii) the escape from a metastable state is an irreversible process, i.e. once the system has gotten out, it is unlikely to return.

Further, they invented the idea to define a metastable state by means of *restricted ensembles*,  $\mu_X := \mu[\cdot|X]$ , where  $\mu$  is the equilibrium Gibbs measure of the system and X is a subset of the state space. The main task is to choose X is such a way that first  $\mu_X$  describes a pure phase and secondly the escape rate from X at time zero is small. The latter implies that condition (ii) is satisfied provided that the escape rate is maximal at time zero. A necessary condition to satisfy the third criterion is to ensure that  $\mu[X]$  is negligible. For an excellent review of this method we refer to the recent monograph by Olivieri and Vares [94].

Acting on the assumption that metastability is characterized by the existence of at least two different time scales for the evolution, Davies [29, 30, 31] suggested to analyze the small eigenvalues and to construct metastable sets from the corresponding eigenfunctions. Based on the observation that a Markov process, exhibiting a metastable behavior, is *almost reducible*, we can view the original, irreducible Markov process as a perturbation of a reducible one. Assuming that the process is reducible, the theorem of Perron-Frobenius implies that the generator has a degenerate eigenvalue zero whose multiplicity is equal to the number of different metastable sets. Moreover, the corresponding eigenfunctions are given as indicator functions on these sets. Provided that the perturbation is sufficiently small, this leads to a cluster of small eigenvalues that is separated by a gap from the rest of the spectrum. This approach has been continued by Gaveau and Schulman [55] and more recently by Huisinga, Meyn and Schütte [66]. In particular, in the study of metastable chemical conformations of bio-molecules Deuflhard et al. introduced a numerical algorithm to determine metastable sets [36, 37].

From the theoretical point of view, the spectral signature of metastability is maybe the most elegant and complete way to characterize this phenomenon in Markov processes. However, in the majority of applications coming from statistical mechanics it is very hard to determine both the small eigenvalues and the corresponding eigenvectors explicitly.

In a recent paper [7], Bianchi and Gaudillière presented an approach to metastability that combines aspects of the idea of the evolution of ensembles and spectral properties. Instead of considering the restricted ensemble  $\mu_X$ , they suggested to characterize metastable states by the quasi-stationary measure  $\mu_X^*$  that is the left eigenvector to the principle Dirichlet eigenvalue of the generator subject to Dirichlet boundary conditions on  $X^c$ . The advantage of choosing  $\mu_X^*$  as the initial distribution of the Markov process is that the exit times of the set X are exponential distributed. Provided that the ratio between the principle Dirichlet eigenvalue and the principle eigenvalue with respect to the reflected process can be controlled, they established sharp estimates on the mean exit times.

As it was pointed out in [23], the approach of evolution of ensembles has the drawback that by studying the time evolution of probability distributions it is difficult to distinguish between a smooth but very slow relaxation towards equilibrium and a typical metastable behavior. The latter is characterized by the fact that the process is apparently stationary for a long period until it undergoes a rapid transition at some randomly distributed point in time leading the system to a different state that again seems be to stationary. Based on this fact Cassandro, Galves, Olivieri and Vares proposed a *pathwise approach to metastability* that uses time averages along trajectories to single out properties of typical paths as well as their statistics [23, 94]. The main tools in their approach are large deviation methods in path space and model reduction techniques. The latter, originally introduced by Freidlin and Wentzell in the context of randomly perturbed dynamical systems, allows to reduce the problem of describing the long-time behavior of a process in terms of a *Markov chain with exponentially small transition probabilities* between metastable sets. This approach has been proven to be robust and rather universally applicable in various model contexts.

While the pathwise approach to metastability is able to yield detailed informations, for instance on the typical exit path, its precision to predict, e.g. the mean exit times, is limited to logarithmic equivalence.

More recently, Bovier, Gayrard, Eckhoff and Klein introduced a *potential theoretic approach* to metastability [15, 16] that was systematically developed in the sequel. Their starting point is the definition of a set of metastable points,  $\mathcal{M}$ , that is characterized by the ratio of two escape probabilities. More precisely, for any  $\sigma \in \mathcal{M}$  the probability to escape from  $\sigma$  to the remaining metastable points  $\mathcal{M} \setminus \{\sigma\}$  should be much smaller compared to the probability to reach  $\mathcal{M}$  starting from some arbitrary point in the state space outside  $\mathcal{M}$  before returning to it. The key idea of this approach is to express quantities of interest in terms of capacities and use variational principles to compute the latter. Having identified the set of metastable points, this approach allows to prove sharp estimates on mean exit times, to analyze precisely the low-lying spectrum of the corresponding generator and to control the deviation of the law of metastable exit times from the exponential distribution for a wide class of *reversible* Markov processes. Thereby, strict renewal equations play an important role to derive various estimates. For an introduction to the potential theoretic approach to metastability, we refer to [11].

As it was already pointed out in [9], the definition of metastable points relies crucially on the fact that the time to reach the set of metastable points is small compared to the transition time between different elements of this set. However, in the context of reversible diffusion processes or spin systems at finite temperature the probability to hit a given point in finite time is either zero or exponentially small. For diffusion processes, this problem can be overcome by considering around each metastable point, x, a small ball,  $B_{\varepsilon}(x)$ , of radius  $\varepsilon > 0$ . In this spirit, a definition of a set of metastable points was introduced in [10]. Based on Harnack and Hölder inequalities, apriori estimates on the local regularity of solutions of boundary value problems with respect to the generator of the diffusion process yield that the typical oscillations of such functions within  $B_{\varepsilon}(x)$  are bounded by some polynomial in  $\varepsilon$ . Provided that  $\varepsilon$  is chosen appropriately, this allows to establish strict renewal equations e.g. for harmonic functions or Laplace transforms of mean hitting times [17, 18].

Metastable sets and local valleys. Our starting point in the study of the dynamical behavior of the Markov chain  $\{\sigma(t)\}$  is the following definition of a *set of metastable sets*.

#### 3. METASTABILITY

**Definition 3.1.** Let  $\mathcal{M}^n \equiv \mathcal{M}^n_N$  be a set of disjoint subsets of  $\mathcal{S}_N$ . Assume that there exists  $n \in \mathbb{N}$  such that  $n \ll N$  and for every  $A \in \mathcal{M}^n$  there exists  $a \in \Gamma^n$  such that A is given as the set-valued preimage of a under  $\varrho^n$ , i.e.  $A = \mathcal{S}^n[a]$ . Further, let  $\mathcal{S}^n_N := \{\mathcal{S}^n[x] \subset \mathcal{S}_N \mid x \in \Gamma^n\}$ . Then,  $\mathcal{M}^n$  is called a *set of metastable sets* if there exists  $\mathfrak{C} \equiv \mathfrak{C}(n) > 0$  such that

$$\frac{\max_{A \in \mathcal{M}^n} \mathbb{P}_{\mu_A} \left[ \tau_{\mathcal{M}^n \setminus A} < \tau_A \right]}{\min_{X \in \mathcal{S}_N^n \setminus \mathcal{M}^n} \mathbb{P}_{\mu_X} \left[ \tau_{\mathcal{M}^n} < \tau_X \right]} \le e^{-\mathfrak{C}N}.$$
(3.1)

Here,  $\mu_X(\sigma) = \mu_N[\sigma | X]$  for  $\sigma \in X \subset S_N$  stands for the reversible measure,  $\mu_N$ , conditioned on the set *X*.

**Remark 3.2.** The definition above is a natural generalization of the one for metastable points given in [9, 11].

**Remark 3.3.** Depending on the choice of  $\mathfrak{C}$ , we may find different sets  $\mathcal{M}^n$ . Let us emphasize the fact that, in the considered setting, the cardinality of  $\mathcal{M}^n$  grows at most polynomial with N. Moreover, the equation (3.1) has the pleasant feature that in the choice of the elements of  $\mathcal{M}^n$  we have to take into account only sufficiently "deep minima". This allow us to deal with situations where an exponential large number (in N) of "shallow local minima" are present.

**Remark 3.4.** In order to compute precisely e.g. metastable exit times, in the approach presented below we need as an input some estimates of the oscillations of harmonic functions within mesoscopic sets. However, in view of (3.1), it suffice to control these oscillations on exponential scales.

We associate to each  $M \in \mathcal{M}^n$  a local valley on mesoscopic level n

$$\mathcal{V}^{n}(M) := \left\{ X \subset \mathcal{S}_{N} \mid \exists \boldsymbol{x} \in \Gamma^{n} : X = \mathcal{S}^{n}[\boldsymbol{x}], \quad \mathbb{P}_{\mu_{X}}[\tau_{M} \leq \tau_{X}] \geq \mathbb{P}_{\mu_{X}}[\tau_{\mathcal{M}^{n} \setminus M} \leq \tau_{X}] \right\}.$$

Notice that the sets  $\mathcal{V}^n(M)$  are not necessary disjoint. By using arguments similar to the ones given in [16] in the case of points, we will show, see Corollary 3.7, that the set of sets which belong to more than one local valley has a very small mass under  $\mu_N$ . Since the conditions above do not uniquely determine  $\mathcal{M}^n$ , it will be reasonable to choose  $\mathcal{M}^n$  in such a way that for all  $M \in \mathcal{M}^n$ 

$$\mu_N[X] \le \mu_N[M], \quad \forall X \in \mathcal{V}^n(M).$$
(3.2)

Furthermore, for every  $A \in \mathcal{M}^n$  we denote by  $\mathcal{M}^n(A) := \{X \in \mathcal{M}^n \mid \mu_N[X] > \mu_N[A]\}$  the subset of "deeper" metastable sets.

# 3.2. Ultrametricity

Let us start with studying the capacity between different metastable sets. More precisely, for an arbitrary  $M \in \mathcal{M}^n$  and  $B \subset \mathcal{M}^n \setminus M$  our aim is to show that if  $X \in \mathcal{V}^n(M)$  then either the escape probability  $\mathbb{P}_{\mu_X}[\tau_B < \tau_X]$  can be bounded from below, or the capacity  $\operatorname{cap}(X, B)$ is essentially the same as  $\operatorname{cap}(M, B)$ . The strategy to prove such a statement relies on the ultrametricity of capacities, as specified in Lemma 1.26. In order to verify the assumptions given there, we would like to take advantage of Proposition 2.7. However, the bounds obtained on harmonic functions are unpleasant to derive a suitable upper bound on (1.113). For this reason, our starting point in the proof is instead of (1.114) the equation (1.116).

**Lemma 3.5.** Let  $M \in \mathcal{M}^n$ ,  $B \subset \mathcal{M}^n \setminus M$  and  $X \in \mathcal{V}^n(M) \setminus M$ . Further, let  $c_1 > 1 + \alpha^{-1}$  and choose  $\delta_1 > \varepsilon(n)c_1$ . Suppose that  $\mathfrak{C} - \delta_1 > 0$ , then either

$$\mathbb{P}_{\mu_X}\left[\tau_B < \tau_X\right] \geq \frac{1}{2} e^{N(\mathfrak{C} - \delta_1)} \max_{A \in \mathcal{M}^n} \mathbb{P}_{\mu_A}\left[\tau_{\mathcal{M}^n \setminus A} < \tau_A\right]$$
(3.3)

#### 3.2. ULTRAMETRICITY

or

$$1 - e^{-N(\delta_1 - \varepsilon(n)c_1)} - e^{-c_2N} \le \frac{\operatorname{cap}(X, B)}{\operatorname{cap}(M, B)} \le \left(1 - e^{-N(\delta_1 - \varepsilon(n)c_1)} - e^{-c_2N}\right)^{-1}$$
(3.4)

**Remark 3.6.** Since  $\varepsilon(n)$  is decreasing in n, by choosing n in the definition of metastable sets large enough, we can ensure that  $\mathfrak{C} - \delta_1 > 0$ .

*Proof.* First of all note that the definition of a local valley on mesoscopic level n implies that for all  $X \in \mathcal{V}^n(M)$ 

$$\mathbb{P}_{\mu_X}\left[\tau_{\mathcal{M}^n} < \tau_X\right] \leq \mathbb{P}_{\mu_X}\left[\tau_M < \tau_X\right] + \mathbb{P}_{\mu_X}\left[\tau_{\mathcal{M}^n \setminus M} < \tau_X\right] \leq 2 \mathbb{P}_{\mu_X}\left[\tau_M < \tau_X\right].$$
(3.5)

Hence, together with the definition of metastable sets, see (3.1), we get that

$$\max_{A \in \mathcal{M}^n} \mathbb{P}_{\mu_A} \left[ \tau_{\mathcal{M}^n \setminus A} < \tau_A \right] \leq e^{-\mathfrak{C}N} \mathbb{P}_{\mu_X} \left[ \tau_{\mathcal{M}^n} < \tau_X \right] \leq 2 e^{-\mathfrak{C}N} \mathbb{P}_{\mu_X} \left[ \tau_M < \tau_X \right]$$
(3.6)

Let us now assume that  $\operatorname{cap}(X, B) > e^{-\delta_1 N} \operatorname{cap}(X, M)$ . Then, the combination of the two equations above yields immediately (3.3). Hence, we are left with considering the remaining case where  $\operatorname{cap}(X, B) \leq e^{-\delta_1 N} \operatorname{cap}(X, M)$ . Here, our starting point is (1.116)

$$1 - \mathbb{P}_{\nu_{M,B}}\left[\tau_B < \tau_X\right] \leq \frac{\operatorname{cap}(X,B)}{\operatorname{cap}(M,B)} \leq \left(1 - \mathbb{P}_{\nu_{X,B}}\left[\tau_B < \tau_M\right]\right)^{-1}$$
(3.7)

Our strategy is the following: with the help of Proposition 2.7 we replace the initial distribution,  $\nu_{M,B}$ , on the left-hand side of (3.7) by the distribution  $\nu_{M,X}$  which has the advantage that by a computation analog to (1.116) combined with the assumption  $\operatorname{cap}(X, B) \leq e^{-\delta_1 N} \operatorname{cap}(X, M)$  we obtain immediately that

$$\mathbb{P}_{\nu_{M,X}}\left[\tau_B < \tau_X\right] \leq \frac{\operatorname{cap}(B,X)}{\operatorname{cap}(M,X)} \leq e^{-\delta_1 N}.$$
(3.8)

On the right-hand side of (3.7) we proceed similarly and replace  $\nu_{X,B}$  by  $\nu_{X,B\cup M}$ . Then, an application of the averaged renewal equation for capacities (1.90) reveals that

$$\mathbb{P}_{\nu_{X,B\cup M}}\left[\tau_B < \tau_M\right] \leq \frac{\operatorname{cap}(X,B)}{\operatorname{cap}(X,M)} \leq e^{-\delta_1 N}.$$

Let us now describe how the actual replacement of the initial distributions is done. Since M, B as well as X are given as preimages under  $\rho^n$  of certain points in  $\Gamma^n$ , Proposition 2.7 is applicable. Hence, for all  $\sigma, \eta \in M$ ,

$$\mathbb{P}_{\sigma}[\tau_B < \tau_X] \leq \mathrm{e}^{\varepsilon(n)c_1 N} \mathbb{P}_{\eta}[\tau_B < \tau_X] + \mathrm{e}^{-c_2 N}.$$
(3.9)

By multiplying both sides with  $\nu_{M,X}(\eta)$  and summing over all  $\eta \in M$ , we obtain the following upper bound

$$\mathbb{P}_{\sigma}\left[\tau_{B} < \tau_{X}\right] \leq e^{\varepsilon(n)c_{1}N} \mathbb{P}_{\nu_{M,X}}\left[\tau_{B} < \tau_{X}\right] + e^{-c_{2}N} \leq e^{-N(\delta_{1}-\varepsilon(n)c_{1})} + e^{-c_{2}N}, \quad (3.10)$$

uniformly for all  $\sigma \in M$ . By a similar computation we have that

$$\mathbb{P}_{\sigma}\left[\tau_{B} < \tau_{M}\right] \leq e^{\varepsilon(n)c_{1}N} \mathbb{P}_{\nu_{X,B\cup M}}\left[\tau_{B} < \tau_{M}\right] + e^{-c_{2}N} \leq e^{-N(\delta_{1}-\varepsilon(n)c_{1})} + e^{-c_{2}N}, \quad (3.11)$$

uniformly for all  $\sigma \in X$ . Hence, by combining (3.10) and (3.11) with (3.7) the estimate (3.4) follows.

A simple corollary of Lemma 3.5 shows that sets  $X = S^n[x]$  that belong to more than one local valley,  $\mathcal{V}^n(M)$ , have a vanishing mass under the invariant measure  $\mu_N$ .

**Corollary 3.7.** Consider two distinct metastable sets  $A, B \in \mathcal{M}^n$  with  $\mathcal{V}^n(A) \cap \mathcal{V}^n(B) \neq \emptyset$ . Further, let  $c_1 > 1 + \alpha^{-1}$  and  $\delta_1 > \varepsilon(n)c_1$ . If  $\mathfrak{C} - \delta_1 > 0$ , then for all  $X \in \mathcal{V}^n(A) \cap \mathcal{V}^n(B)$ 

$$\mu_N[X] \le 2 e^{-N(\mathfrak{C} - \delta_1)} \min\{\mu_N[A], \mu_N[B]\}.$$
(3.12)

*Proof.* Suppose that  $\mu_N[B] < \mu_N[A]$ . Since  $X \in \mathcal{V}^n(A) \cap \mathcal{V}^n(B)$ , (3.6) implies that

$$\mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right] \leq 2 e^{-\mathfrak{C}N} \mathbb{P}_{\mu_{X}}\left[\tau_{A} < \tau_{X}\right] \text{ and } \mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right] \leq 2 e^{-\mathfrak{C}N} \mathbb{P}_{\mu_{X}}\left[\tau_{B} < \tau_{X}\right].$$
(3.13)

## In particular, we have that

$$\operatorname{cap}(A, B) \leq 2 \operatorname{e}^{-\mathfrak{C}N} \frac{\mu_N[A]}{\mu_N[X]} \operatorname{cap}(B, X).$$

In the remaining part of the proof we have to consider two different cases. Let us first of all assume that  $2 e^{-\mathfrak{C}N} \mu_N[A] / \mu_N[X] \le e^{-\delta_1 N}$ . Then, by Lemma 3.5, we obtain that

$$\operatorname{cap}(A, X) \leq \left(1 - \mathrm{e}^{-N(\delta_1 - \varepsilon(n)c_1)} - \mathrm{e}^{-c_2 N}\right)^{-1} \operatorname{cap}(A, B).$$
 (3.14)

Hence,

$$\mu_{N}[X] \leq \left(1 - \mathrm{e}^{-N(\delta_{1} - \varepsilon(n)c_{1})} - \mathrm{e}^{-c_{2}N}\right)^{-1} \frac{\mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right]}{\mathbb{P}_{\mu_{X}}\left[\tau_{A} < \tau_{X}\right]} \mu_{N}[A]$$

$$\leq 2 \left(1 - \mathrm{e}^{-N(\delta_{1} - \varepsilon(n)c_{1})} - \mathrm{e}^{-c_{2}N}\right)^{-1} \mathrm{e}^{-\mathfrak{C}N} \mu_{N}[A], \qquad (3.15)$$

where we used in the first step that  $\operatorname{cap}(A, X) = \mu_N[X] \mathbb{P}_{\mu_X}[\tau_A < \tau_X]$ , while in the second step we took advantage of (3.13). On the other hand, if  $2e^{-\mathfrak{C}N}\mu_N[A]/\mu_N[X] > e^{-\delta_1N}$  the assertion of the corollary follows easily.

### 3.3. Sharp estimates on mean hitting times

In this section we demonstrate how the averaged renewal equation (1.90) and the almost ultrametricity of capacities (1.114) combined with the coupling method can be used to prove various estimates on harmonic functions. In particular, we derive a precise formula for mean hitting times of metastable sets. The control obtained here is crucial for the investigation of the distribution of metastable exit times in Section 3.4.

Let  $\mathcal{M}^n$  be a set of metastable sets and  $\mathfrak{C}$  as defined in Definition 3.1. Let us now choose some  $A \in \mathcal{M}^n$  and let  $B \subset \mathcal{M}^n(A)$ . Moreover, we may rewrite

$$\mathbb{E}_{\nu_{A,B}}[\tau_B] = \sum_{\sigma \in A} \nu_{A,B}(\sigma) \mathbb{E}_{\sigma}[\tau_B] = \frac{1}{\operatorname{cap}(A,B)} \sum_{\boldsymbol{x} \in \Gamma^n} \sum_{\sigma \in \mathcal{S}^n[\boldsymbol{x}]} \mu_N(\sigma) h_{A,B}(\sigma).$$
(3.16)

Our aim is to show that, under a non-degeneracy condition that will be specified below, the sum on the right-hand side of (3.16) is of the order  $\mu_N[A]$ .

The following lemma provides a necessary control over the equilibrium potential,  $h_{A,B}$ , within the different local valleys.

**Lemma 3.8.** Let  $A \in \mathcal{M}^n$  and  $B \subset \mathcal{M}^n \setminus A$ . Further, suppose that  $c_1 > 1 + \alpha^{-1}$  is chosen in such a way that  $e^{-c_2N} < \mathbb{P}_{\mu_A}[\tau_B < \tau_A]$ . Then,

(i) for every  $M \in B$  and  $X \in \mathcal{V}^n(M) \setminus M$ 

$$\sum_{\sigma \in X} \mu_N(\sigma) h_{A,B}(\sigma) \leq 2 e^{-N(\mathfrak{C}-\varepsilon(n)c_1)} \left(1 - 2 e^{-\mathfrak{C}N}\right)^{-1} \mu_N[A],$$
(3.17)

(ii) for 
$$X \in \mathcal{V}^{n}(A) \setminus A$$
  

$$\sum_{\sigma \in X} \mu_{N}(\sigma) \left(1 - h_{A,B}(\sigma)\right) \leq 2 e^{-N(\mathfrak{C} - \varepsilon(n)c_{1})} \left(1 - 2 e^{-\mathfrak{C}N}\right)^{-1} \mu_{N}[A], \qquad (3.18)$$

46

(iii) for every  $M \in \mathcal{M}^n \setminus (\{A\} \cup B)$  and  $X \in \mathcal{V}^n(M)$  either

$$\sum_{\sigma \in X} \mu_N(\sigma) h_{A,B}(\sigma) \leq 2 e^{-N(\mathfrak{C} - \delta_1 - \varepsilon(n)c_1)} \left(1 - 2 e^{-N(\mathfrak{C} - \delta_1)}\right)^{-1} \mu_N[A] \quad (3.19)$$
  
or  
$$\sum_{\sigma \in X} \mu_N(\sigma) h_{A,B}(\sigma) \leq a(n,N) e^{\varepsilon(n)c_1N} \left(1 - a(n,N) \frac{e^{-c_2N}}{\mathbb{P}_{\mu_M}\left[\tau_B < \tau_M\right]}\right)^{-1} \frac{\mathbb{P}_{\mu_A}\left[\tau_B < \tau_A\right]}{\mathbb{P}_{\mu_M}\left[\tau_B < \tau_M\right]} \mu_N[A],$$
  
$$(3.20)$$

where  $a(n, N)^{-1} = 1 - 2 e^{-N(\delta_1 - \varepsilon(n)c_1)}$  and  $0 < \delta_1 < \mathfrak{C}$ . Moreover, the same statement holds for  $1 - h_{A,B}$  except that the term  $\mathbb{P}_{\mu_M}[\tau_B < \tau_M]$  in (3.20) has to be replaced by the probability  $\mathbb{P}_{\mu_M}[\tau_A < \tau_M]$ .

*Proof.* The idea behind the proof is based on the representation (1.116) and pointwise estimates on harmonic functions (2.27). In view of the definition of metastable sets (3.1), we obtain first of all that  $\mathbb{P}_{\mu_A}[\tau_B < \tau_A] \leq e^{-\mathfrak{C}N} \mathbb{P}_{\mu_X}[\tau_{\mathcal{M}^n} < \tau_X]$ . Hence, for every  $M \in \mathcal{M}^n$  and  $X \in \mathcal{V}^n(M) \setminus M$  it holds that

$$\mathbb{P}_{\mu_A}\left[\tau_B < \tau_A\right] \leq 2 e^{-\mathfrak{C}N} \mathbb{P}_{\mu_X}\left[\tau_M < \tau_X\right]$$
(3.21)

where we used (3.6).

(i) and (ii). Due to the fact that the proof of (3.17) and (3.18) are completely similar, we will give a detailed proof only for (i). Let us fix some  $M \in B$ . Since  $h_{A,B}(\sigma) = 0$  for all  $\sigma \in B$ , it remains to consider all  $X \in \mathcal{V}^n(M) \setminus M$ . Recall that, by definition, the metastable sets are preimages under  $\varrho^n$  of some points in  $\Gamma^n$ . Hence, in view of (2.27), we have that

$$\mathbb{P}_{\sigma}[\tau_B < \tau_X] \geq e^{-\varepsilon(n)c_1 N} \left( \mathbb{P}_{\eta}[\tau_B < \tau_X] - e^{-c_2 N} \right)$$
(3.22)

for all  $\sigma, \eta \in X$ . By multiplying both sides with  $\mu_N(\eta)$  and summing over all  $\eta \in X$ , we obtain the following pointwise estimate

$$\mathbb{P}_{\sigma}\left[\tau_{B} < \tau_{X}\right] \geq e^{-\varepsilon(n)c_{1}N} \frac{\operatorname{cap}(B,X)}{\mu_{N}[X]} \left(1 - \frac{e^{-c_{2}N}}{\mathbb{P}_{\mu_{X}}\left[\tau_{B} < \tau_{X}\right]}\right).$$
(3.23)

On the other hand, a computation analog to (1.116) shows that

$$\mathbb{P}_{\nu_{X,B}}\left[\tau_A < \tau_B\right] \leq \frac{\operatorname{cap}(A,B)}{\operatorname{cap}(X,B)} = \frac{\mathbb{P}_{\mu_A}\left[\tau_B < \tau_A\right]}{\mathbb{P}_{\mu_X}\left[\tau_B < \tau_X\right]} \frac{\mu_N[A]}{\mu_N[X]}.$$
(3.24)

By using (3.23), we can replace the last exit biased distribution on the left-hand side of (3.24) through the conditional reversible measure  $\mu_X$ . Thus,

$$\sum_{\sigma \in X} \mu_N(\sigma) h_{A,B}(\sigma) \leq e^{\varepsilon(n)c_1 N} \left( 1 - \frac{e^{-c_2 N}}{\mathbb{P}_{\mu_X} \left[ \tau_B < \tau_X \right]} \right)^{-1} \frac{\mathbb{P}_{\mu_A} \left[ \tau_B < \tau_A \right]}{\mathbb{P}_{\mu_X} \left[ \tau_B < \tau_X \right]} \mu_N[A]$$
(3.25)

$$\leq 2 e^{-N(\mathfrak{C}-\varepsilon(n)c_1)} \left(1-2 e^{-\mathfrak{C}N}\right)^{-1} \mu_N[A], \qquad (3.26)$$

where we used in the last step that  $\mathbb{P}_{\mu_A}[\tau_B < \tau_A] \leq 2 e^{-\mathfrak{C}N} \mathbb{P}_{\mu_X}[\tau_B < \tau_X]$  as well as the assumption on the choice of  $c_2$ . This concludes the proof of (i) and (ii).

(iii) Let  $X \in \mathcal{V}^n(M)$  where  $M \in \mathcal{M}^n \setminus \{A\} \cup B$ . If  $X \neq M$ , an application of (3.3) for some  $0 < \delta_1 < \mathfrak{C}$  reveals that  $\mathbb{P}_{\mu_A}[\tau_B < \tau_A] \le 2 e^{-N(\mathfrak{C} - \delta_1)} \mathbb{P}_{\mu_X}[\tau_B < \tau_X]$ . Hence, by (3.25),

$$\sum_{\sigma \in X} \mu_N(\sigma) h_{A,B}(\sigma) \leq 2 e^{-N(\mathfrak{C} - \delta_1 - \varepsilon(n)c_1)} \left( 1 - 2 e^{-N(\mathfrak{C} - \delta_1)} \right)^{-1} \mu_N[A].$$
(3.27)

3. METASTABILITY



FIGURE 1. Illustration of various different landscapes that are excluded by the non-degeneracy condition.

On the other hand, from (3.4), it follows that we can bound the probability  $\mathbb{P}_{\mu_X} [\tau_B < \tau_X]$  from below by  $a(n, N)^{-1} \mathbb{P}_{\mu_M} [\tau_B < \tau_M]$ . By plugging this estimate into (3.25), we obtain (3.20) for  $X \in \mathcal{V}^n(M) \setminus M$ . Notice that (3.20) follows immediately from the equation (3.25) in the case when X = M. The completes the proof.

In view of Lemma 3.8, we can derive precise expressions for mean hitting times (3.16) that involve only capacities and the invariant measure of local valleys. For simplicity, let us take into account the following additional *non-degeneracy* condition:

**Definition 3.9.** Let  $A \in \mathcal{M}^n$  and  $B \subset \mathcal{M}^n \setminus A$ . We say that the non-degeneracy condition is satisfied, if there exists a  $0 < \delta < \mathfrak{C}$  such that either

$$\mu_N[M] < e^{-\delta N} \mu_N[A] \quad \text{or} \quad \mathbb{P}_{\mu_A}[\tau_B < \tau_A] < e^{-\delta N} \mathbb{P}_{\mu_M}[\tau_B < \tau_M] \quad (3.28)$$

for all  $M \in \mathcal{M}^n \setminus (A \cup B)$ .

We can now easily prove the following

**Theorem 3.10.** For  $A \in \mathcal{M}^n$  and  $B \subset \mathcal{M}^n(A)$  assume that the non-degeneracy condition for some  $0 < \delta < \mathfrak{C}$  is satisfied. If  $n \ll N$  is such that  $\varepsilon(n)c_1 < \min\{\delta, \mathfrak{C} - \delta\}$  where  $c_1$  is chosen in such a way that  $e^{-c_2N} < \mathbb{P}_{\mu_A}[\tau_B < \tau_A]$ , then there exists  $c_9 > 0$  such that for N large enough

$$\mathbb{E}_{\nu_{A,B}}[\tau_B] = \frac{\mu_N \left[\mathcal{V}^n(A)\right]}{\operatorname{cap}(A,B)} \left(1 + e^{-c_9 N}\right).$$
(3.29)

*Proof.* For  $\delta$  and n as specified above, let us define

$$\mathcal{U}^n_{\delta} := \Big\{ X \subset \mathcal{S}_N \ \Big| \ \exists \, \boldsymbol{x} \in \Gamma^n : X = \mathcal{S}^n[\boldsymbol{x}], \ \mu_N[X] \ge \mu_N[A] \, \mathrm{e}^{-\delta N} \Big\}.$$

An obvious consequence is the following

Lemma 3.11. With the notations introduced above, it holds that

$$\sum_{X \notin \mathcal{U}_{\delta}^{n}} \sum_{\sigma \in X} \mu_{N}(\sigma) h_{A,B}(\sigma) \leq N^{nq} e^{-\delta N} \mu_{N}[A].$$
(3.30)

It remains to control the harmonic function  $h_{A,B}$  for all configurations  $\sigma \in X$  where  $X \in \mathcal{U}_{\delta}$ . For this purpose, let us rewrite the neighborhood  $\mathcal{U}_{\delta}^n$  as

$$\mathcal{U}^n_\delta = \bigcup_{M \in \mathcal{M}^n} \mathcal{U}^n_\delta(M)$$

where  $\mathcal{U}_{\delta}^{n}(M) := \mathcal{U}_{\delta}^{n} \cap \mathcal{V}^{n}(M)$  for  $M \in \mathcal{M}^{n}$ . Notice that for  $M, M' \in \mathcal{M}^{n}$  it may happen that  $\mathcal{U}_{\delta}^{n}(M) \cap \mathcal{U}_{\delta}^{n}(M') \neq \emptyset$ . We begin with controlling  $h_{A,B}$  on  $\mathcal{U}_{\delta}^{n}(M)$  for all  $M \in \{A\} \cup B$ . In view of Lemma 3.8 it is immediate that, for N large enough,

$$\sum_{X \in \mathcal{U}_{\delta}^{n}(A)} \sum_{\sigma \in X} \mu_{N}(\sigma) \left( 1 - h_{A,B}(\sigma) \right) \leq N^{nq} e^{-N(\mathfrak{C} - \varepsilon(n)c_{1})} \mu_{N}[A]$$
(3.31)

while for all  $M \in B$ 

$$\sum_{X \in \mathcal{U}_{\delta}^{n}(A)} \sum_{\sigma \in X} \mu_{N}(\sigma) \left( 1 - h_{A,B}(\sigma) \right) \leq N^{nq} e^{-N(\mathfrak{C} - \varepsilon(n)c_{1})} \mu_{N}[A].$$
(3.32)

As a next step, let us consider the contribution to (3.16) coming from the neighborhood  $\mathcal{U}^n_{\delta}(M)$  for  $M \in \mathcal{M}^n \setminus (\{A\} \cup B)$ . For this purpose, we set  $\delta_1 = \delta$  in Lemma 3.8 (iii). Then, in the case when  $\operatorname{cap}(X, B) > e^{-\delta N} \operatorname{cap}(X, M)$ , (3.19) implies that

$$\sum_{X \in \mathcal{U}_{\delta}^{n}(M)} \sum_{\sigma \in X} \mu_{N}(\sigma) h_{A,B}(\sigma) \leq N^{nq} e^{-N(\mathfrak{C}-\delta-\varepsilon(n)c_{1})} \mu_{N}[A].$$
(3.33)

Hence, it remains to consider the case  $\operatorname{cap}(X, B) \leq e^{-\delta N} \operatorname{cap}(X, M)$ . Let us point out that, as a consequence of the non-degeneracy condition,  $\mathbb{P}_{\mu_A} \left[ \tau_B < \tau_A \right] \leq e^{-\delta N} \mathbb{P}_{\mu_M} \left[ \tau_B < \tau_M \right]$ . On the other hand, by assumption, it holds that  $e^{-c_2 N} < \mathbb{P}_{\mu_A} \left[ \tau_B < \tau_A \right]$ . Therefore, in view of (3.20),

$$\sum_{X \in \mathcal{U}_{\delta}^{n}(M)} \sum_{\sigma \in X} \mu_{N}(\sigma) h_{A,B}(\sigma) \leq N^{nq} e^{-N(\delta - \varepsilon(n)c_{1})} \mu_{N}[A].$$
(3.34)

Now, the proof of (3.29) is an immediate consequence of (3.16) combined with (3.30) and (3.32) - (3.34). Namely, equation (3.30) together with (3.32), (3.33) an (3.34) provide the following upper bound

$$\sum_{\sigma \in \mathcal{S}_{N}} \mu_{N}(\sigma) h_{A,B}(\sigma)$$

$$\leq \sum_{X \in \mathcal{V}^{n}(A)} \sum_{\sigma \in X} \mu_{N}(\sigma) + N^{nq} \left( e^{-\delta N} + e^{-c_{10}(n)N} + |\mathcal{M}^{n}| e^{-c_{11}(n)N} \right) \mu_{N}[A]$$

$$\leq \mu_{N} \left[ \mathcal{V}^{n}(A) \right] + N^{nq} \left( e^{-\delta N} + e^{-c_{10}(n)N} + |\mathcal{M}^{n}| e^{-c_{11}(n)N} \right) \mu_{N}[A], \quad (3.35)$$

where  $c_{10}(n) := \mathfrak{C} - \varepsilon(n)c_1$  and  $c_{11}(n) := \min\{\delta, \mathfrak{C} - \delta\} - \varepsilon(n)c_1$ , while, by employing (3.30) and (3.31), the corresponding lower bound is given through

$$\sum_{\sigma \in \mathcal{S}_{N}} \mu_{N}(\sigma) h_{A,B}(\sigma) \geq \sum_{X \in \mathcal{V}^{n}(A)} \sum_{\sigma \in X} \mu_{N}(\sigma) - \sum_{X \in \mathcal{V}^{n}(A)} \sum_{\sigma \in X} \mu_{N}(\sigma) \left(1 - h_{A,B}(\sigma)\right)$$
$$\geq \mu_{N} \left[\mathcal{V}^{n}(A)\right] - N^{nq} \left(e^{-\delta N} + e^{-c_{10}(n)N}\right) \mu_{N}[A].$$
(3.36)

Hence, there exists  $c_9 > 0$  such that for N large enough the assertion (3.29) follows.

**Corollary 3.12.** For  $A \in \mathcal{M}^n$  and  $B = \mathcal{M}^n(A)$  assume that the non-degeneracy condition is satisfied. Further, suppose that  $n \ll N$  is such that  $\varepsilon(n) c_1 < \mathfrak{C}/4$  where  $c_1$  is chosen in such a way that  $e^{-c_2N} < \mathbb{P}_{\mu_A}[\tau_B < \tau_A]$ . Then the conclusions of Theorem 3.10 also hold. Moreover,

$$\mathbb{E}_{\sigma}[\tau_B] = \mathbb{E}_{\nu_{A,B}}[\tau_B] \left( 1 + \mathcal{O}(\mathrm{e}^{-c_0 N}) \right), \quad \forall \sigma \in A.$$
(3.37)

*Proof.* By inspecting the proof of Theorem 3.10, we see that in the case when  $B = \mathcal{M}^n(A)$  it suffices to ensure that  $\varepsilon(n)c_1 < \mathfrak{C}$ . Further, recall that for any metastable sets  $A \in \mathcal{M}^n$  there exists  $a \in \Gamma^n$  such that  $A = S^n[a]$ . Thus, for every  $A \in \mathcal{M}^n$  and  $B \subset \mathcal{M}^n \setminus A$  the condition (2.5) follows from (3.1), while (2.6) is redundant. Thus, as soon as  $\varepsilon(n)c_1 < \mathfrak{C}/4$ , Theorem 2.12 implies the pointwise estimate.

#### 3. METASTABILITY

## 3.4. Asymptotic exponential distribution

The content of this section is to show how the pointwise estimates on the Laplace transform of metastable exit times, which we obtained via coupling methods, combined with the averaged renewal equation (1.103) can be used to prove the convergence of the normalized metastable exit times to an exponential distribution, in cases where Theorem 3.10 applies.

**Theorem 3.13.** For  $A \in \mathcal{M}^n$  and  $B \subset \mathcal{M}^n(A)$  assume that the non-degeneracy condition is satisfied. Further, assume that  $n \ll N$  is such that  $\varepsilon(n)c_1 < \min\{\delta, \mathfrak{C}/4\}$  where  $\delta$  is specified in (3.28) and  $c_1$  is chosen in such a way that  $e^{-c_2N} < \mathbb{P}_{\mu_A}[\tau_B < \tau_A]$ . Then, for all  $t \in \mathbb{R}_+$ ,

$$\mathbb{P}_{\sigma}[\tau_B/\mathbb{E}_{\sigma}[\tau_B] > t] \longrightarrow e^{-t}, \quad as \quad N \to \infty$$
(3.38)

for all  $\sigma \in A$ .

In view of Proposition 2.13, the main step in the proof is to show the convergence of the Laplace transform  $\mathbb{E}_{\rho_{\lambda}}\left[e^{-\frac{\lambda}{T}\tau_{B}}\right]$  as N tends to  $\infty$ , where the probability measure  $\rho_{\lambda}$  is defined in (1.102) and  $T \equiv \mathbb{E}_{\nu_{A,B}}[\tau_{B}]$ . By Lemma 1.25, our starting point is

$$\mathbb{E}_{\rho_{\lambda}}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}}\right] = \frac{\mathbb{E}_{\mu_{A}}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{B}} \mathbb{1}_{\tau_{B} < \tau_{A}}\right]}{1 - \mathbb{E}_{\mu_{A}}\left[\mathrm{e}^{-\frac{\lambda}{T}\tau_{A}} \mathbb{1}_{\tau_{A} < \tau_{B}}\right]}.$$
(3.39)

As a result of the representation (3.39) and the continuity theorem for Laplace transforms, see [49, XIII.1 Theorem 2], Theorem 3.13 will follow from (2.64) and (2.58), respectively, once we have proven the following lemma.

**Lemma 3.14.** Under the assumptions of Theorem 3.13, for any  $\lambda \ge 0$ 

$$\lim_{N \to \infty} \frac{\mathbb{E}_{\mu_A} \left[ e^{-\frac{\lambda}{T} \tau_B} \mathbb{1}_{\tau_B < \tau_A} \right]}{1 - \mathbb{E}_{\mu_A} \left[ e^{-\frac{\lambda}{T} \tau_A} \mathbb{1}_{\tau_A < \tau_B} \right]} = \frac{1}{1 + \lambda}.$$
(3.40)

Proof. The proof of this lemma comprises the following three steps.

STEP 1. First of all, we will prove a crucial bound to which we refer to as Uphill Lemma.

**Lemma 3.15.** [5, Lemma 4.3] Under the assumptions of Theorem 3.13, there exists  $c_{12} > 0$  and  $N_0 \in \mathbb{N}$  such that

$$\mathbb{E}_{\mu_A} \left[ \tau_B \, \mathbb{1}_{\tau_B < \tau_A} \right] \leq e^{-c_{12}N} \, \mathbb{E}_{\mu_A} \left[ \tau_A \, \mathbb{1}_{\tau_A < \tau_B} \right], \qquad \forall N \ge N_0. \tag{3.41}$$

Proof. To start with, we use the fact that

$$\mathbb{E}_{\mu_A}[\tau_{A\cup B}] = \mathbb{E}_{\mu_A}[\tau_A \mathbb{1}_{\tau_A < \tau_B}] + \mathbb{E}_{\mu_A}[\tau_B \mathbb{1}_{\tau_B < \tau_A}].$$
(3.42)

Notice that, for  $\sigma \in A$ ,

$$\mathbb{E}_{\sigma} \big[ \tau_A \, \mathbb{1}_{\tau_A < \tau_B} \big] = \mathbb{P}_{\sigma} \big[ \tau_A < \tau_B \big] + \big( L w_{A,B} \big) (\sigma). \tag{3.43}$$

An application of the second Green's identity to the functions  $w_{A,B}$  and  $h_{A,B}$  reveals that

$$\sum_{\sigma \in A} \mu_N(\sigma) \left( Lw_{A,B} \right)(\sigma) = \sum_{\sigma \notin A \cup B} \mu_N(\sigma) \left( h_{A,B}(\sigma) \right)^2.$$
(3.44)

Hence,

$$\mathbb{E}_{\mu_{A}}\left[\tau_{A} \mathbb{1}_{\tau_{A} < \tau_{B}}\right] = \mathbb{P}_{\mu_{A}}\left[\tau_{A} < \tau_{B}\right] + \frac{1}{\mu_{N}[A]} \sum_{\sigma \notin A \cup B} \mu_{N}(\sigma) \left(h_{A,B}(\sigma)\right)^{2}.$$
(3.45)

On the other hand, by an analog procedure, one shows that

$$\mathbb{E}_{\mu_A}\left[\tau_{A\cup B}\right] = 1 + \frac{1}{\mu_N[A]} \sum_{\sigma \in A} \mu_N(\sigma) \left(Lw_{A\cup B}\right)(\sigma) = 1 + \frac{1}{\mu_N[A]} \sum_{\sigma \notin A \cup B} \mu_N(\sigma) h_{A,B}(\sigma).$$
(3.46)

Thus, taking into account (3.42) we can conclude that

$$\mathbb{E}_{\mu_A}\left[\tau_B \,\mathbbm{1}_{\tau_B < \tau_A}\right] = \frac{\operatorname{cap}(A, B)}{\mu_N[A]} + \frac{1}{\mu_N[A]} \sum_{\sigma \notin A \cup B} \mu_N(\sigma) \,h_{A,B}(\sigma) \,h_{B,A}(\sigma). \tag{3.47}$$

Due to (3.1), the first term on the right-hand side is bounded from above by  $e^{-\mathfrak{C}N}$ . Further, from Lemma 3.8 together with the non-degeneracy condition we get that the second term is as well exponentially small compared to  $\mu_N[A]$ . Moreover, since  $\mathbb{E}_{\mu_A}[\tau_{A\cup B}] \ge 1$  we conclude that  $\mathbb{E}_{\mu_A}[\tau_A \mathbb{1}_{\tau_A < \tau_B}]$  is exponentially close to one. Hence, choosing  $c_{12} > 0$  and  $N_0 \in \mathbb{N}$  appropriately, we deduce (3.41).

Corollary 3.16. With the notations introduced above, it holds that

$$\mathbb{E}_{\nu_{A,B}}\left[\tau_{B}\right] = \frac{\mathbb{E}_{\mu_{A}}\left[\tau_{A} \ \mathbb{1}_{\tau_{A} < \tau_{B}}\right]}{\mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right]} \left(1 + e^{-c_{12}N}\right), \qquad \forall N \ge N_{0}.$$
(3.48)

*Proof.* As an immediate consequence of Lemma 1.14 combined with (3.41), we obtain (3.48).

STEP 2. Using that,  $1 - x \le e^{-x} \le 1$  for  $x \ge 0$ , it follows that the numerator in the renewal equation (3.39) can be written as

$$\mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right] - \frac{\lambda}{T} \mathbb{E}_{\mu_{A}}\left[\tau_{B} \,\mathbb{1}_{\tau_{B} < \tau_{A}}\right] \leq \mathbb{E}_{\mu_{A}}\left[e^{-\frac{\lambda}{T}\tau_{B}} \,\mathbb{1}_{\tau_{B} < \tau_{A}}\right] \leq \mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right].$$
(3.49)

In view of the Uphill Lemma and (3.48), we get

$$\mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right]\left(1 - \lambda e^{-c_{11}N}\right) \leq \mathbb{E}_{\mu_{A}}\left[e^{-\frac{\lambda}{T}\tau_{B}} \mathbb{1}_{\tau_{B} < \tau_{A}}\right] \leq \mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right].$$
(3.50)

Let us now turn to the denominator in (3.39). As a consequence of the second Green's identity applied to the function  $h_{A,B}$  and  $h_{A,B}^{\lambda}$ , we may rewrite it as

$$1 - \mathbb{E}_{\mu_{A}} \left[ e^{-\frac{\lambda}{T}\tau_{A}} \mathbb{1}_{\tau_{A} < \tau_{B}} \right] = \mathbb{P}_{\mu_{A}} \left[ \tau_{B} < \tau_{A} \right] e^{-\frac{\lambda}{T}} + \frac{1 - e^{-\frac{\lambda}{T}}}{\mu_{N}[A]} \sum_{\sigma \notin B} \mu_{N}(\sigma) h_{A,B}(\sigma) h_{A,B}^{\lambda}(\sigma).$$

$$(3.51)$$

Taking into account the representation (1.49) and using the fact that  $h_{A,B}^{\lambda}(\sigma) \leq h_{A,B}(\sigma) \leq 1$  for all  $\sigma \in S_N$  we get

$$1 - \mathbb{E}_{\mu_A} \left[ e^{-\frac{\lambda}{T} \tau_A} \, \mathbb{1}_{\tau_A < \tau_B} \right] \leq \mathbb{P}_{\mu_A} \left[ \tau_B < \tau_A \right] \left( 1 + \lambda \right). \tag{3.52}$$

Hence, our remaining task is to derive a lower bound that is exponentially close to the upper bound. In order to do so, let us restrict the summation on the right-hand side of (3.51) to the subset  $A_{\delta} \subset S^n$  that is given by the union of all sets in  $\mathcal{U}_{\delta}^n(A)$ . Moreover,  $\delta$  it given by the non-degeneracy condition. This yields

$$1 - \mathbb{E}_{\mu_{A}} \left[ e^{-\frac{\lambda}{T} \tau_{A}} \mathbb{1}_{\tau_{A} < \tau_{B}} \right] \\ \geq \mathbb{P}_{\mu_{A}} \left[ \tau_{B} < \tau_{A} \right] e^{-\frac{\lambda}{T}} \left( 1 + \frac{\lambda}{T \operatorname{cap}(A, B)} \sum_{\sigma \in A_{\delta}} \mu_{N}(\sigma) h_{A, B}(\sigma) h_{A, B}^{\lambda}(\sigma) \right)$$
(3.53)

where we used that  $e^x - 1 \ge x$  for all  $x \ge 0$ .

STEP 3. We can conclude the proof of the Lemma 3.14, ones we have proven the following estimate.

**Lemma 3.17.** Under the assumptions of Theorem 3.13 there exists  $c_{13} > 0$  and  $N_0 \in \mathbb{N}$  such that

$$\sum_{\sigma \in A_{\delta}} \mu_N(\sigma) h_{A,B}(\sigma) h_{A,B}^{\lambda}(\sigma) \geq \sum_{\sigma \notin B} \mu_N(\sigma) h_{A,B}(\sigma) \left(1 - \lambda e^{-c_{13}N}\right), \quad \forall N \geq N_0.$$
(3.54)

*Proof.* The idea behind the proof is to show that  $h_{A,B}^{\lambda}$  and  $h_{A,B}$  are almost of the same size on the subset  $A_{\delta} \subset S_N$ . To start with, we rewrite (3.54) as

$$\sum_{\sigma \in A_{\delta}} \mu_{N}(\sigma) h_{A,B}(\sigma) h_{A,B}^{\lambda}(\sigma)$$

$$\geq \sum_{\sigma \in A_{\delta}} \mu_{N}(\sigma) h_{A,B}(\sigma)^{2} - \sum_{\sigma \in A_{\delta}} \mu_{N}(\sigma) \left( h_{A,B}(\sigma) - h_{A,B}^{\lambda}(\sigma) \right).$$
(3.55)

In view of Lemma 3.8 and 3.11, we immediately obtain, for N large enough, that

$$\sum_{\sigma \in A_{\delta}} \mu_{N}(\sigma) h_{A,B}(\sigma)^{2} \\ \geq \sum_{\sigma \notin B} \mu_{N}(\sigma) h_{A,B}(\sigma) \left( 1 - N^{nq} \left( e^{-\delta N} + 2 e^{-c_{10}(n)N} + |\mathcal{M}^{n}| e^{-c_{11}(n)N} \right) \right).$$
(3.56)

Let us point out that  $c_{10}(n), c_{11}(n) > 0$  provided that n is chosen large enough such that  $\varepsilon(n)c_1 < \min\{\delta, \mathfrak{C} - \delta\}$ . Exploiting the probabilistic interpretation of  $h_{A,B}$  and  $h_{A,B}^{\lambda}$ , gives

$$h_{A,B}(\sigma) - h_{A,B}^{\lambda}(\sigma) = \mathbb{E}_{\sigma} \left[ \left( 1 - e^{-\frac{\lambda}{T}\tau_A} \right) \mathbb{1}_{\tau_A < \tau_B} \right] \le \frac{\lambda}{T} \mathbb{E}_{\sigma} \left[ \tau_A \, \mathbb{1}_{\tau_A < \tau_B} \right]$$
(3.57)

for all  $\sigma \in A_{\delta} \setminus A$ , whereas the difference between  $h_{A,B}$  and  $h_{A,B}^{\lambda}$  vanishes on A. Hence,

$$\sum_{\sigma \in A_{\delta}} \mu_N(\sigma) \left( h_{A,B}(\sigma) - h_{A,B}^{\lambda}(\sigma) \right) \leq \frac{\lambda}{T} \sum_{X \in \mathcal{U}_{\delta}^n(A) \setminus A} \sum_{\sigma \in X} \mu_N(\sigma) w_{A,B}(\sigma).$$
(3.58)

On the other hand, let us remark that for any  $X \subset S_N$ , an application of the second Green's identity to the functions  $w_{A,B}$  and  $h_{X,A\cup B}$  yields

$$-\sum_{\sigma \in X} \mu_N(\sigma) \left( Lh_{X,A\cup B} \right)(\sigma) w_{A,B}(\sigma) = \sum_{\sigma \notin A\cup B} \mu_N(\sigma) h_{X,A\cup B}(\sigma) h_{A,B}(\sigma).$$
(3.59)

Hence, by dividing both sides by  $cap(X, A \cup B)$ , we obtain

$$\mathbb{E}_{\nu_{X,A\cup B}}\left[\tau_{A}\,\mathbb{1}_{\tau_{A}<\tau_{B}}\right] = \frac{1}{\operatorname{cap}(X,A\cup B)} \sum_{\sigma\notin A\cup B} \mu_{N}(\sigma)\,h_{X,A\cup B}(\sigma)\,h_{A,B}(\sigma),\tag{3.60}$$

where  $\nu_{X,A\cup B}$  is the last exit biased distribution on *X*. In view of (2.27), we have for any  $X \in \mathcal{U}^n_{\delta}(A) \setminus A$  and all  $\sigma, \eta \in X$ , that

$$\mathbb{P}_{\sigma}\left[\tau_{A\cup B} < \tau_{X}\right] \geq e^{-\varepsilon(n)c_{1}N} \left(\mathbb{P}_{\eta}\left[\tau_{A\cup B} < \tau_{X}\right] - e^{-c_{2}N}\right).$$
(3.61)

By multiplying both sides with  $\mu_N(\eta)$  and summing over all  $\eta \in X$ , we obtain the following pointwise estimate

$$\mathbb{P}_{\sigma}\left[\tau_{A\cup B} < \tau_{X}\right] \geq e^{-\varepsilon(n)c_{1}N} \frac{\operatorname{cap}(A \cup B, X)}{\mu_{N}[X]} \left(1 - 2e^{-\mathfrak{C}N}\right),$$
(3.62)

since (3.1) implies that  $\mathbb{P}_{\mu_A}[\tau_B < \tau_A] \leq 2 e^{-\mathfrak{C}N} \mathbb{P}_{\mu_X}[\tau_{A\cup B} < \tau_X]$  for all  $X \in \mathcal{U}^n_{\delta}(A) \setminus A$ . Further, we used that  $c_2$  is chosen in such a way that  $e^{-c_2N} < \mathbb{P}_{\mu_A}[\tau_B < \tau_A]$ . Hence, by combining the pointwise estimate with (3.60) yields

$$\sum_{\sigma \in X} \mu_N(\sigma) w_{A,B}(\sigma) \leq e^{\varepsilon(n)c_1 N} \left( 1 - 2 e^{-\mathfrak{C}N} \right)^{-1} \frac{\mu_N[X]}{\operatorname{cap}(X, A \cup B)} \sum_{\sigma \in \mathcal{S}_N} \mu_N(\sigma) h_{A,B}(\sigma)$$
$$\leq \frac{2 e^{-N(\mathfrak{C} - \varepsilon(n)c_1)}}{\mathbb{P}_{\mu_A}[\tau_B < \tau_A]} \left( 1 - 2 e^{-\mathfrak{C}N} \right)^{-1} \sum_{\sigma \in \mathcal{S}_N} \mu_N(\sigma) h_{A,B}(\sigma).$$
(3.63)

Finally, by plugging (3.63) into (3.58), we obtain

$$\sum_{\sigma \in A_{\delta}} \mu_{N}(\sigma) \left( h_{A,B}(\sigma) - h_{A,B}^{\lambda}(\sigma) \right) \leq \lambda 4 N^{nq} e^{-N(\mathfrak{C} - \varepsilon(n)c_{1})} \sum_{\sigma \notin B} \mu_{N}(\sigma) h_{A,B}(\sigma)$$
(3.64)

for N large enough. Thus, by combining the estimates above concludes the proof.

From Lemma 3.17, it follows that the denominator of (3.39) can be bounded by

$$\left(1 - \lambda e^{-\mathfrak{C}N}\right) \left(1 + \lambda \left(1 - \lambda e^{-c_{13}N}\right)\right) \leq \frac{1 - \mathbb{E}_{\mu_A} \left[e^{-\frac{\lambda}{T}\tau_A} \mathbb{1}_{\tau_A < \tau_B}\right]}{\mathbb{P}_{\mu_A} \left[\tau_B < \tau_A\right]} \leq 1 + \lambda$$
(3.65)

-

provided that N is chosen large enough. Together with (3.50) the assertion of Lemma 3.14 follows.  $\hfill \Box$ 

# CHAPTER 4

# Metastability and Spectral Theory

In the present chapter we study the relation between the low-lying part of the spectrum of  $-L_N$  and the mean transition times associated to certain metastable sets in  $\mathcal{M}^n$ . The main objective is to show that, under some non-degeneracy condition, to each metastable set corresponds a simple eigenvalue of  $-L_N$  which is equal to the inverse of the mean exit time from this set up to small errors. Let us point out that the methods we are going to use in the sequel are mainly analytic.

The remainder of this chapter is organized as follows. In Section 4.1 we review first the main ideas and methods that were used in the strong recurrent setting [16]. In particular, we explain some of the challenges, we are faced with, and sketch a strategy to solve them. A first observation is a relation between the eigenvalues of the generator  $-L_N$  and the eigenvalues of the generator  $-G^n$  associated to the coarse-grained Markov chain. By employing this observation, in Section 4.2 we prove a rough localization of the small eigenvalues. In the remaining Section 4.3 we focus on improving the previously obtained estimates. Using a posteriori error estimates based on general results due to Kato [74] and Temple [109] combined with properties of metastable sets and pointwise estimates on harmonic functions we derive upper and lower bounds on the small eigenvalues that coincides in the limit when N tends to infinity.

**Contributions in this chapter.** The main contribution of the author is the development of a strategy that allows to reduce the original eigenvalue problem on  $S_N$  to the strong recurrent setting on the coarse-grained level. A crucial step of the presented method is to establish first a rough localization of the eigenvalues. Under the assumption that there are at most as many small eigenvalues as metastable sets, these estimates can be used in a second step to compute precisely the small eigenvalues. Let us point out that this approach is not limited to the particular setting we have chosen. Beyond that, the chapter contains the following novel pieces:

- a lower bound on the eigenvalues of the generator  $-L_N$  in terms of the eigenvalues of the generator  $-G^n$ , see Proposition 4.5;
- an upper bound on the residuum exploiting pointwise estimates on harmonic functions, see Lemma 4.14;
- an analysis of the Rayleigh-Ritz values that appear in the improved upper bound, see 4.13.

# 4.1. Introduction

Investigations in the connection between the dynamical behavior of Markov processes and the existence of small eigenvalues of the corresponding generators dates back at least to the work of Wentzell [112] and Feidlin and Wentzell [51]. In the study of diffusions, they identified  $\lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \ln \lambda(\varepsilon)$  using large deviation methods. Based on variational principles, these estimates could be improved up to a multiplicative error first by Holley, Kusuoka and Stroock [65] for principal eigenvalues and later for the full set of exponentially small eigenvalues by Miclo [86] and Mathieu [84]. By using potential-theoretic ideas that were already suggested in an early work by Wentzell [113], sharp estimates on the small eigenvalues were established by Bovier, Gayrard, Klein and Eckhoff for reversible diffusion processes [18, 45] and for Markov chains on discrete state spaces in the reversible setting [16]. In the context of reversible Markov chains on discrete state spaces, their starting point was the definition of a *set of metastable points*. Under some non-degeneracy conditions they showed that to each metastable point corresponds a simple eigenvalue of the discrete generator associated to the Markov chain. Moreover, each such eigenvalue is equal to the inverse of the mean exit time from the corresponding metastable point up to negligible error terms.

In the previous chapter we have seen, that e.g. in the context of stochastic spin systems at finite temperature, the probability to hit a particular configuration in the state space and the escape probability from a metastable point are of the same order. For this reason, a Markov chain in such a setting can only be  $\rho$ -metastable in the sense of [11, Definition 4.1], if the transition probabilities are exponentially small. For instance, this is the case if we consider a spin system in the low temperature regime. In contrast to that, mean-field systems can be studied as well in the finite temperature regime. The reason behind this fact is that mean-field systems have the property that there exists a macroscopic variable such that the stochastic process on the coarse-grained space induced by this map is still Markovian. Thus, the set of metastable points can be defined on this lower dimensional space. A further advantage of mean-field systems is that the spectrum of the discrete generator,  $-G_N$ , associated to the induced Markov chain on the lower-dimensional space is contained in the spectrum of  $-L_N$ .

As we already mentioned in the preface, our interest in the relation between the metastable behavior and the small eigenvalues of the generator associated to a Markov chain arises from the study of disordered mean-field spin systems at *finite temperature*. In such a situation, we *cannot* exactly reduce the model to a low-dimensional one via lumping techniques.

Before discussing the difficulties, we are faced with when starting from the definition of a set of metastable sets, let us briefly sketch the main ideas of the approach, originally presented in [16], to investigate the small eigenvalues for strongly recurrent, reversible Markov chains on a finite state spaces. For an in-deep presentation of the connection between metastability and small eigenvalues from a potential-theoretic point of view we refer to the lecture notes [9, 11].

Small eigenvalues in the strong recurrent setting: A review of the methods. Let us consider the setting that was introduced in Chapter 1, i.e. let  $\{\sigma(t)\}$  be an irreducible Markov chain on a finite state space,  $S_N$ , that is reversible with respect to a unique invariant measure  $\mu_N$ , and assume additionally that  $\{\sigma(t)\}$  is  $\rho$ -metastable in the sense of [11, Definition 4.1]. This means that, for  $0 < \rho \ll 1$ , there exists a set of *metastable points*(!),  $\mathcal{M}_N \equiv \{\sigma^1, \ldots, \sigma^K\} \subset S_N$ , such that

$$\frac{\max_{\eta \in \mathcal{M}_N} \mathbb{P}_{\eta} \left[ \tau_{\mathcal{M}_N \setminus \{\eta\}} < \tau_{\eta} \right]}{\min_{\eta \in \mathcal{S}_N \setminus \mathcal{M}_N} \mathbb{P}_{\eta} \left[ \tau_{\mathcal{M}_N} < \tau_{\eta} \right]} \leq \rho.$$

It is well known that reversibility implies that the generator,  $L_N$ , is self-adjoint on the weighted space  $L^2(S_N, \mu_N)$ . Hence, the spectrum of  $L_N$  is real and we can order the eigenvalues,  $\lambda_i$ , of  $-L_N$  in increasing order. As an immediate consequence of the Theorem of Perron-Frobenius we have that

$$0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_{|\mathcal{S}_N|} \leq 2.$$

Our main objective is to compute the first K eigenvalues of  $-L_N$ . As a first step, let us consider the eigenvalue problem for  $-L_N$  subject to zero boundary conditions on a given subset  $D \subset S_N$ . In such a situation, the smallest eigenvalue, that we denote by  $\lambda^D \equiv \lambda_1^D$ , is called principal eigenvalue. Notice that, for every set D, the principal eigenvalue has the

#### 4.1. INTRODUCTION

pleasant feature to give rise to an *a priori* estimate of the |D| + 1 smallest eigenvalue of  $-L_N$ . Our starting point is the Theorem of Courant-Fischer. It implies that

$$\lambda_i \leq \lambda_i^D \leq \lambda_{i+|D|}, \quad \forall i = 1, \dots, |\mathcal{S}_N| - |D|.$$
(4.1)

We use this observation in the following way: Let us consider an increasing sequence of sets  $D_1 \subset D_2 \subset \ldots \subset D_K$  such that  $D_1$  and for all  $i \ge 1$  the difference  $D_{i+1} \setminus D_i$  is each a singleton. Then (4.1) implies that

$$\lambda^{D_i} \leq \lambda_{i+1} \qquad \forall i = 1, \dots, K.$$
(4.2)

The Equation (4.2) is one of the main ingredients to characterize completely the K smallest eigenvalues of  $-L_N$ . Note that every increasing sequence of sets  $\{D_i\}$  can be used to derive a lower bound on the corresponding eigenvalue. But, the quality of such lower bounds depends crucially on the chosen sets  $D_i$ . In the iteration procedure that is presented below the idea is to construct an increasing sequence of sets from the set of metastable points.

Since the eigenvalue  $\lambda_{|\mathcal{M}_N|+1}$  is bounded from below by the principal eigenvalue of  $\lambda^{\mathcal{M}_N}$ , a next step in the analysis is to derive bounds on it. Another pleasant feature of principal eigenvalues is that they can be easily characterized via variational principles. Consider a set  $D \subset S_N$ . While an upper bound on  $\lambda^D$  follows from the Rayleigh principle, that reads

$$\lambda^{D} = \inf_{\substack{f \mid D = 0 \\ \|f\|_{\mu_{N}} = 1}} \left\langle f, -L_{N} f \right\rangle_{\mu_{N}},\tag{4.3}$$

a lower bound can be obtained by using a variational formula due to Donsker and Varadhan [40]

$$\lambda^{D} = \inf_{\substack{f|_{D} = 0 \\ \|f\|_{\mu_{N}} = 1}} \sup_{g|_{D^{c}} > 0} \left\langle \frac{f^{2}}{g}, -L_{N} g \right\rangle_{\mu_{N}}.$$
(4.4)

Here, we introduced the notations  $||f||_{2,\mu_N}^2 \equiv \langle f, f \rangle_{\mu_N} \coloneqq \sum_{\sigma \in S_N} \mu_N(\sigma) f(\sigma)^2$ .

Let us now come back to the principal eigenvalue of  $-L_N$ . An upper bound on  $\lambda^{\mathcal{M}_N}$  is obtained by plugging the normalized equilibrium potential  $h_{\eta,\mathcal{M}_N}/\|h_{\eta,\mathcal{M}_N}\|_{2,\mu_N}^2$  for some  $\eta \in \mathcal{S}_N \setminus \mathcal{M}_N$  in the Rayleigh principle and optimizing over all  $\eta$ . For a lower bound, recall that the function  $w_{\mathcal{M}_N}$  equals  $\mathbb{E}_{\sigma}[\tau_{\mathcal{M}_N}]$  on  $\mathcal{M}_N^c$ . In particular, for all  $\sigma \notin \mathcal{M}_N$  it holds that  $(L_N w_{\mathcal{M}_N})(\sigma) = -1$ . Thus, by setting  $g \equiv w_{\mathcal{M}_N}$  the variational principle of Donsker-Varadhan yields that

$$\lambda^{\mathcal{M}_{N}} \geq \left(\max_{\sigma \in \mathcal{S}_{N} \setminus \mathcal{M}_{N}} \mathbb{E}_{\sigma}[\tau_{\mathcal{M}_{N}}]\right)^{-1}.$$
(4.5)

This proves the following

**Lemma 4.1.** Let  $\lambda^{\mathcal{M}_N}$  denote the principal Dirichlet eigenvalue of  $-L_N$  this respect to the set  $\mathcal{M}_N$ . Then

$$\min_{\sigma \in \mathcal{S}_N \setminus \mathcal{M}_N} \frac{\operatorname{cap}(\sigma, \mathcal{M}_N)}{\left\|h_{\sigma, \mathcal{M}_N \setminus \{\sigma\}}\right\|_{1, \mu_N}} \leq \lambda^{\mathcal{M}_N} \leq \min_{\sigma \in \mathcal{S}_N \setminus \mathcal{M}_N} \frac{\operatorname{cap}(\sigma, \mathcal{M}_N)}{\left\|h_{\sigma, \mathcal{M}_N \setminus \{\sigma\}}\right\|_{2, \mu_N}^2}.$$
 (4.6)

Having establish a lower bound for the principal Dirichlet eigenvalue, we turn, in a second step, to a precise characterization of the eigenvalues of  $-L_N$  that are strictly below  $\lambda^{\mathcal{M}_N}$ . Notice that there are at most K such eigenvalues. The strategy is to consider the solution of the Dirichlet problem

$$\begin{cases} -(L_N f^{\lambda})(\sigma) - \lambda f^{\lambda}(\sigma) = 0, & \sigma \in \mathcal{S}_N \setminus \mathcal{M}_N, \\ f^{\lambda}(\sigma) = \phi(\sigma), & \sigma \in \mathcal{M}_N. \end{cases}$$
(4.7)

If  $\lambda < \lambda^{\mathcal{M}_N}$  is an eigenvalue of  $-L_N$  and if we choose  $\phi^{\lambda}$  as the eigenfunction corresponding to this eigenvalue, then  $f^{\lambda}$  equals  $\phi^{\lambda}$  on  $\mathcal{S}_N$ . In order to see this, notice that  $f^{\lambda}(\sigma) - \phi^{\lambda}(\sigma) = 0$ for all  $\sigma \in \mathcal{M}_N$ . Hence, the difference  $f^{\lambda} - \phi^{\lambda}$  solves a Dirichlet problem with respect to the operator  $-L_N - \lambda$  subject to zero boundary conditions. Due to the fact that  $\lambda < \lambda^{\mathcal{M}_N}$ , this Dirichlet problem has a unique solution given by the zero function. This implies that  $f^{\lambda}$  is equal to the eigenfunction  $\phi^{\lambda}$  everywhere.

Notice that by linearity of the Dirichlet problem and due to the properties of the  $\lambda$ -equilibrium potential that was defined in Chapter 1, we can represent  $f^{\lambda}$  as

$$f^{\lambda}(\sigma) = \sum_{\eta \in \mathcal{M}_N} c_{\eta} h^{\lambda}_{\eta, \mathcal{M}_N \setminus \eta}(\sigma), \qquad c_{\eta} \in \mathbb{R}.$$
(4.8)

If for a given  $\lambda < \lambda^{\mathcal{M}_N}$  there exists a non-zero vector  $\{c_\eta : \eta \in \mathcal{M}_N\}$  such that for all  $\sigma \in \mathcal{S}_N - (L_N f^{\lambda})(\sigma) = \lambda f^{\lambda}(\sigma)$ , then  $\lambda$  is an eigenvalue of  $-L_N$ . The existence of such a vector is equivalent to the fact that the determinant of the matrix  $E_{\mathcal{M}_N}^{\lambda} \in \mathbb{R}^{K \times K}$  with elements

$$\left(\mathcal{E}_{\mathcal{M}_{N}}^{\lambda}\right)_{\sigma,\eta} = -\left(L_{N}h_{\eta,\mathcal{M}_{N}\setminus\eta}^{\lambda}\right)(\sigma) - \lambda h_{\eta,\mathcal{M}_{N}\setminus\eta}^{\lambda}(\sigma), \qquad \sigma,\eta\in\mathcal{M}_{N}.$$

$$(4.9)$$

vanishes. The next step in the analysis is to study the matrix  $E_{\mathcal{M}_N}^{\lambda}$ . Although this is in general a quite difficult task, if we assuming additionally the non-degeneracy condition (4.10), we can compute precisely the largest eigenvalues below  $\lambda^{\mathcal{M}_N}$ . This is the statement of the following

**Theorem 4.2** ([11, Theorem 6.8]). Assume that there exists  $\eta \in M_N$  such that for  $0 < \delta \ll 1$ 

$$\delta^{2} \frac{\operatorname{cap}(\eta, \mathcal{M}_{N} \setminus \eta)}{\left\|h_{\eta, \mathcal{M}_{N} \setminus \eta}\right\|_{2, \mu_{N}}^{2}} \geq \max_{\sigma \in \mathcal{M}_{N} \setminus \{\sigma\}} \frac{\operatorname{cap}(\sigma, \mathcal{M}_{N} \setminus \sigma)}{\left\|h_{\sigma, \mathcal{M}_{N} \setminus \sigma}\right\|_{2, \mu_{N}}^{2}}.$$
(4.10)

Then, the largest eigenvalue  $\lambda_n$  of  $-L_N$  below  $\lambda^{\mathcal{M}_N}$  is given by

$$\lambda_{\eta} = \frac{\operatorname{cap}(\eta, \mathcal{M}_{N} \setminus \eta)}{\left\|h_{\eta, \mathcal{M}_{N} \setminus \eta}\right\|_{2, \mu_{N}}^{2}} \left(1 + \mathcal{O}(\rho^{2} + \delta^{2})\right) = \frac{1}{\mathbb{E}_{\eta}\left[\tau_{\mathcal{M}_{N} \setminus \eta}\right]} \left(1 + \mathcal{O}(\rho^{2} + \delta^{2})\right).$$
(4.11)

Based on this theorem and (4.2), we are now in the position to characterize iteratively all eigenvalues of  $-L_N$  below  $\lambda^{\mathcal{M}_N}$ . In order to do so, consider an increasing sequence of metastable sets  $\mathcal{M}_1 \subset \mathcal{M}_2 \subset \ldots \subset \mathcal{M}_K \equiv \mathcal{M}_N$  such that  $\mathcal{M}_1 = \eta^1$  and  $\mathcal{M}_i \setminus \mathcal{M}_{i-1} = \eta^i$  for all  $1 < i \leq K$ . We assume that this increasing sequence can be constructed in such a way that for all  $1 < i \leq K$ ,  $\eta^i$  and  $\mathcal{M}_{i-1}$  satisfy the non-degeneracy condition (4.10), i.e.

$$\delta^{2} \frac{\operatorname{cap}(\eta^{i}, \mathcal{M}_{i-1})}{\|h_{\eta^{i}, \mathcal{M}_{i-1}}\|_{2, \mu_{N}}^{2}} \geq \max_{\sigma \in \mathcal{M}_{i}} \frac{\operatorname{cap}(\sigma, \mathcal{M}_{i} \setminus \sigma)}{\|h_{\sigma, \mathcal{M}_{i} \setminus \sigma}\|_{2, \mu_{N}}^{2}}, \quad \forall i = 2, \dots, K.$$

$$(4.12)$$

Then,  $-L_N$  has exactly K eigenvalues below  $\lambda^{\mathcal{M}_N}$  given by

$$\lambda_{1} = 0, \qquad \lambda_{i} = \frac{\operatorname{cap}(\eta^{i}, \mathcal{M}_{i-1})}{\|h_{\eta^{i}, \mathcal{M}_{i-1}}\|_{2, \mu_{N}}^{2}} \left(1 + \mathcal{O}(\rho^{2} + \delta^{2})\right), \qquad \forall i = 2, \dots, K.$$
(4.13)

whereas the normalized eigenfunction corresponding to  $\lambda_i$  reads

$$\phi^{\lambda_{i}}(\sigma) = \frac{h_{\eta^{i}, \mathcal{M}_{i-1}}(\sigma)}{\left\|h_{\eta^{i}, \mathcal{M}_{i-1}}\right\|_{2, \mu_{N}}^{2}} + \mathcal{O}(\delta + \rho) \sum_{j=1}^{i-1} \frac{h_{\eta^{j}, \mathcal{M}_{j-1}}(\sigma)}{\left\|h_{\eta^{j}, \mathcal{M}_{j-1}}\right\|_{2, \mu_{N}}^{2}}.$$
(4.14)

**Challenges and outline of the strategy.** The starting point for our further investigations in the small eigenvalues of the generator,  $-L_N$ , is the set of metastable sets,  $\mathcal{M}^n$ , that was introduced in Chapter 3. In the sequel, we discuss the difficulties that arise in this setting. In particular, we outline a strategy to overcome some of these problems.

Suppose that  $\mathcal{M}^n$  contains K metastable sets. Let us stress the fact that, typically, the number of configurations in such metastable sets, that we denote by  $|\mathcal{M}^n|$ , is exponentially large in N. By (4.2), the principal eigenvalue,  $\lambda^{\mathcal{M}^n}$ , of  $-L_N$  subject to zero boundary conditions on

#### 4.1. INTRODUCTION

the set of configurations in  $\mathcal{M}^n$  is a lower bound for  $\lambda_{|\mathcal{M}^n|+1}$ . Although we expect that there are only K eigenvalues below  $\lambda^{\mathcal{M}^n}$  that are separated by a gap from the remaining eigenvalues in the spectrum, we do not have a priori such a knowledge at hand. This is the first crucial difference compared to the strong recurrent setting. Since  $K \ll |\mathcal{M}^n|$ , it might be that there are exponentially many small eigenvalues below  $\lambda^{\mathcal{M}^n}$ . Hence, even if we determine for each metastable set  $M \in \mathcal{M}^n$  an eigenvalue that is below  $\lambda^{\mathcal{M}^n}$ , we cannot characterize fully the low-lying spectrum.

# **Open Problem 4.3.** Prove/disprove that there exists $0 < \delta < \mathfrak{C}$ such that $e^{-\delta N} \lambda^{\mathcal{M}^n} < \lambda_{K+1}$ .

Let us assume the existence of such a  $\delta$ . In the sequel, we address the question of characterizing precisely the small eigenvalues. In order to derive necessary conditions on the low-lying spectrum of  $-L_N$ , a possible strategy is to relate the small eigenvalues to a matrix that is to leading order equal to the so-called capacitance matrix. The former is constructed as follows. Fix an eigenvalue  $\lambda < e^{-\delta N} \lambda^{M^n}$  and denote by  $\phi^{\lambda}$  the corresponding eigenfunction. Then, for  $X, Y \in \mathcal{M}^n$ , the entries of the matrix  $E(\lambda, \phi^{\lambda}) \in \mathbb{R}^{K \times K}$  are given by

$$\left(E(\lambda,\phi^{\lambda})\right)_{X,Y} = -\sum_{\sigma\in X} \mu_N(\sigma) \phi^{\lambda}(\sigma) \left(\left(L_N h_{\eta,\mathcal{M}_N\setminus\eta}^{\lambda}\right)(\sigma) + \lambda h_{\eta,\mathcal{M}_N\setminus\eta}^{\lambda}(\sigma)\right).$$
(4.15)

In contrast to (4.9), in the entries of the matrix  $E(\lambda, \phi^{\lambda})$  appear beside the eigenvalue  $\lambda$ , that we want to compute, the eigenfunction  $\phi^{\lambda}$  which we do not know a priori.

Let us point out that this approach was used in the context of reversible diffusions [18, 45]. The reason for its success in that context is based on the following observation. A regularity analysis of the eigenfunction shows, that within balls,  $B_{\varepsilon}(x)$ , of radius  $\varepsilon$  around the metastable point x,  $\phi^{\lambda}$  is either almost constant and, in particular, does not change its sign or its modulus is close to zero. Let us stress the fact that a suitable regularity theory for eigenfunctions of spin systems is still missing. In particular, to establish that  $\phi^{\lambda}$  does not change its sign within a metastable set X is rather challenging. For this reason, we propose a different strategy.

As a first step, we investigate the relation between the spectrum of  $-L_N$  and the spectrum of the generator  $-G^n$  associated to a canonical Markov chain  $\{\varrho^n(t)\}$  on the lower dimensional state space  $\Gamma^n$ . Our goal is to find for each eigenvalue  $\eta \in \operatorname{spec}(-G^n)$  an interval that contains at least one eigenvalue  $\lambda \in \operatorname{spec}(-L_N)$ . For this purpose, we construct an artificial Markov chain  $\{\overline{\sigma}(t)\}$  on  $S_N$  that is exactly lumpable, i.e.  $\varrho^n(\overline{\sigma}(t))$  is a Markov chain on  $\Gamma^n$  that coincides in law with the Markov chain  $\{\varrho^n(t)\}$ . In particular, the spectral properties of the generator associated to  $\{\overline{\sigma}(t)\}$  are used to establish lower bounds on eigenvalues  $\lambda \in \operatorname{spec}(-L_N)$ .

In a second step we improve both the upper and the lower bounds for the low-lying eigenvalues. While the upper bound follows immediately by an analysis of certain Rayleigh-Ritz values, lower bounds, that coincide with the upper bounds asymptotically, are obtained by means of the following

**Theorem 4.4** (Kato and Temple). Let u be an approximate eigenvector of unit length and  $\xi = \langle u, -L_N u \rangle_{\mu_N}$ . Assume that there exists an interval (a, b) that contains  $\xi$  and exactly one eigenvalue  $\lambda$  of  $-L_N$ . Then

$$-\frac{\|(L_N+\xi)u\|_{2,\mu_N}^2}{\xi-a} \le \xi-\lambda \le \frac{\|(L_N+\xi)u\|_{2,\mu_N}^2}{b-\xi}.$$
(4.16)

Proof. See [104, Theorem 3.8].

A key observation, see Lemma 4.14, is that for a suitable approximated eigenfunction, u, the residual  $||(L_N + \xi)u|| \leq e^{2\varepsilon(n)c_1N}\xi$ . Note that the exponential growing factor can be compensated by choosing n appropriately because  $\xi$  is exponentially small in N. Let us stress

 $\Box$ 

the fact that, under the assumption that  $e^{\delta N} \lambda^{\mathcal{M}^n} \leq \lambda_{K+1}$  and a suitable non-degeneracy condition, the first *K* intervals are disjoint and contain exactly one eigenvalue of  $-L_N$ .

#### 4.2. Rough localization of eigenvalues

The main objective in this section is to derive *a priori* estimates on the eigenvalues of  $-L_N$ . Recall that  $\{\boldsymbol{\varrho}^n(t)\}$  is a Markov chain on the coarse-grained space  $\Gamma^n$  which is reversible with respect to  $\mathcal{Q}^n = \mu_N \circ (\boldsymbol{\varrho}^n)^{-1}$ . Its transition probabilities are given by

$$r^n(\boldsymbol{x}, \boldsymbol{y}) = rac{1}{\mathcal{Q}^n(\boldsymbol{x})} \sum_{\sigma \in \mathcal{S}^n[\boldsymbol{x}]} \mu_N(\sigma) \sum_{\eta \in \mathcal{S}^n[\boldsymbol{y}]} p_N(\sigma, \eta)$$

Further, we associate to  $\{\varrho^n(t)\}$  the generator,  $G^n$ , that acts on functions  $g: \Gamma^n \to \mathbb{R}$  as

$$(G^n g)(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \Gamma^n} r^n(\boldsymbol{x}, \boldsymbol{y}) (g(\boldsymbol{y}) - g(\boldsymbol{x})).$$
(4.17)

Let  $\langle \cdot, \cdot \rangle_{Q^n}$  stands for the scalar product that is defined by

$$\langle f,g 
angle_{\mathcal{Q}^n} := \sum_{oldsymbol{x} \in \Gamma^n} \mathcal{Q}^n(oldsymbol{x}) f(oldsymbol{x}) g(oldsymbol{x})$$

and we write  $L^2(\Gamma^n, Q^n)$  for the vector space  $\mathbb{R}^{|\Gamma^n|}$  equipped with the scalar product  $\langle \cdot, \cdot \rangle_{Q^n}$ . Since reversibility implies that the generator,  $G^n$ , is self-adjoint on  $L^2(\Gamma^n, Q^n)$ , the spectrum of  $G^n$  is real. Let us order the eigenvalues increasingly. An immediate consequence of the Perron-Frobenius Theorem is that

$$0 = \eta_1 < \eta_2 \le \eta_3 \le \dots \le \lambda_{|\Gamma^n|} \le 2$$
 (4.18)

Since we regard elements of  $\mathbb{R}^{|\Gamma^n|}$  as functions from  $\Gamma^n$  to  $\mathbb{R}$ , we call eigenvectors of the matrix  $G^n$  eigenfunctions and vice versa.

Let us now consider the following two matrices  $C = (c(\sigma, \boldsymbol{x})), D = (d(\sigma, \boldsymbol{x})) \in \mathbb{R}^{|S_N| \times |\Gamma^n|}$  to which we refer to as *collection matrix* and *distribution matrix*. The elements of these matrices are defined by

$$c(\sigma, \boldsymbol{x}) := \mathbb{1}_{\sigma \in \mathcal{S}^{n}[\boldsymbol{x}]}, \qquad d(\sigma, \boldsymbol{x}) := \frac{\mu_{N}(\sigma)}{\mathcal{Q}^{n}(\boldsymbol{x})} \mathbb{1}_{\sigma \in \mathcal{S}^{n}[\boldsymbol{x}]}.$$
(4.19)

Taking advantage of these matrices, we have that  $G^n = D^T L_N C$ . Moreover, for all functions  $f: S_N \to \mathbb{R}$  and  $g: \Gamma^n \to \mathbb{R}$  it holds that

$$\left\langle g, D^{\mathrm{T}} f \right\rangle_{Q^{n}} = \sum_{\boldsymbol{x} \in \Gamma^{n}} \mathcal{Q}^{n}(\boldsymbol{x}) \ g(\boldsymbol{x}) \sum_{\sigma \in \mathcal{S}_{N}} \frac{\mu_{N}(\sigma)}{\mathcal{Q}^{n}(\boldsymbol{x})} \ \mathbb{1}_{\sigma \in \mathcal{S}^{n}[\boldsymbol{x}]} \ f(\sigma)$$

$$= \sum_{\sigma \in \mathcal{S}_{N}} \mu_{N}(\sigma) \ f(\sigma) \sum_{\boldsymbol{x} \in \Gamma^{n}} \mathbb{1}_{\sigma \in \mathcal{S}^{n}[\boldsymbol{x}]} \ g(\boldsymbol{x}) = \left\langle Cg, f \right\rangle_{\mu_{N}}.$$

$$(4.20)$$

With the notations introduced above, we prove the following

**Proposition 4.5.** Let  $\lambda_i \in \operatorname{spec}(-L_N)$  and  $\eta_i \in \operatorname{spec}(-G^n)$  sorted in increasing order. Then,

$$\lambda_1 = \eta_1 = 0, \qquad \lambda_i \le \eta_i, \qquad \forall i = 2, \dots, |\Gamma^n|.$$
(4.21)

Moreover, for every  $i = 2, ..., |\Gamma^n|$  there exists  $j_i \ge i$  with  $j_i \ne j_{i'}$  for  $i \ne i'$  such that

$$e^{-2\varepsilon(n)N}\eta_i \leq \lambda_{j_i} \leq e^{2\varepsilon(n)N}\eta_i.$$
(4.22)
*Proof.* Let us start with showing the upper bound in (4.21). By exploiting the Theorem of Courant and Fischer we immediately obtain that

$$\eta_{i} = \min_{\substack{\mathcal{V} \subset \mathbb{R}^{|\Gamma^{n}|} \\ \dim \mathcal{V} = i \\ \dim \mathcal{V} = i \\ m = i$$

where we used in the second step that  $\langle g, G^n g \rangle_{Q^n} = \langle C g, L_N C g \rangle_{\mu_N}$ .

Next, we focus on proving the bounds described in (4.22). Let us stress the fact that in later applications we are particularly interested in the lower bound. To start with, let us sketch the strategy behind our proof. The first idea is to consider an additional irreducible Markov chain  $\{\bar{\sigma}(t)\}\)$  on the state space  $S_N$  whose transition probabilities  $\bar{p}_N$  are reversible with respect to a unique invariant measure  $\bar{\mu}_N$ . We denote its generator by  $\bar{L}^n$ . For a given coarsening level n, this Markov chain is constructed in such a way that the induced process  $\{\varrho^n(\bar{\sigma}(t))\}\)$  is a Markov chain. As a consequence, we obtain that the eigenvalues of  $-G^n$  are contained in the spectrum of the generator  $-\bar{L}^n$ . In a second step, we compare the eigenvalues of  $-L_N$  and  $-\bar{L}_N$ , by exploiting the positivity of the terms in the Dirichlet form.

In order to define the Markov chain  $\{\bar{\sigma}(t)\}$ , we choose the transition probabilities  $\bar{p}_N(\sigma, \eta)$ and the probability measure  $\bar{\mu}_N(\sigma)$  for  $\sigma, \eta \in S_N$  in the following way. For  $\boldsymbol{x}, \boldsymbol{y} \in \Gamma^n$  and  $\sigma \in S^n[\boldsymbol{x}], \eta \in S^n[\boldsymbol{y}]$  set

$$\bar{p}_N(\sigma,\eta) = \frac{r^n(\boldsymbol{x},\boldsymbol{y})}{|A_{\boldsymbol{x},\boldsymbol{y}}(\sigma)|} \mathbb{1}_{p_N(\sigma,\eta)\neq 0} \quad \text{and} \quad \bar{\mu}_N(\sigma) = \frac{\mathcal{Q}^n(\boldsymbol{x})}{|\mathcal{S}^n[\boldsymbol{x}]|}, \quad (4.24)$$

where  $A_{x,y}(\sigma)$  for  $\sigma \in S^n[x]$  is a short hand notation for the set  $\{\eta \in S^n[y] \mid d_H(\sigma, \eta) = 1\}$ . As a consequence of the single site dynamics, the cardinality of  $A_{x,y}(\sigma)$  is constant on  $S^n[x]$ . In particular, it holds, for all  $x, y \in \Gamma^n$ , that

$$\left|\mathcal{S}^{n}[\boldsymbol{x}]\right|\left|A_{\boldsymbol{x},\boldsymbol{y}}(\sigma)\right| = \left|\mathcal{S}^{n}[\boldsymbol{y}]\right|\left|A_{\boldsymbol{y},\boldsymbol{x}}(\eta)\right| \quad \forall \sigma \in \mathcal{S}^{n}[\boldsymbol{x}], \ \eta \in \mathcal{S}^{n}[\boldsymbol{y}].$$
(4.25)

As an immediate consequence, the detailed balance condition is satisfied. Further notice that,

$$\sum_{\eta \in \mathcal{S}^{n}[\boldsymbol{y}]} \bar{p}_{N}(\sigma, \eta) = \sum_{\eta \in \mathcal{S}^{n}[\boldsymbol{y}]} \bar{p}_{N}(\sigma', \eta) = r^{n}(\boldsymbol{x}, \boldsymbol{y}), \quad \forall \sigma, \sigma' \in \mathcal{S}^{n}[\boldsymbol{x}].$$
(4.26)

Since (4.26) holds true, employing the classical theorem of Burke and Rosenblatt [21], we can conclude that  $\{\varrho^n(\bar{\sigma}(t))\}$  is a Markov chain.

Moreover, (4.26) implies that the subspace C, spanned by the columns of the matrix C, is invariant under the linear map  $\overline{L}_N$ , i.e.  $\overline{L}_N C \subset C$ . Thus,  $\operatorname{spec}(-G^n) \subset \operatorname{spec}(-\overline{L}_N)$ . In order to see this, notice that  $\overline{L}_N$  is C-invariant if and only if there exists a matrix  $M \in \mathbb{R}^{|\Gamma^n| \times |\Gamma^n|}$  such that  $\overline{L}_N C = CM$ . But, from this equation follows easily that Cg is an eigenfunction of  $\overline{L}_N$  if g is an eigenfunction of M. Since,  $G^n = D^T \overline{L}_N C = D^T C = M$  by construction,  $G^n = M$ . Notice that this relation between the eigenvalues was proven in [82, Lemma 12.8] by similar arguments. Although to each  $\eta \in \operatorname{spec}(-G^n)$  there exists an unique  $\lambda \in \operatorname{spec}(-\overline{L}_N)$ , we do not know to which eigenvalue of  $-L_N$  the eigenvalue  $\eta_i \in \operatorname{spec}(-G^n)$  belongs to. In view of the Courant-Fischer Theorem, we have that for all  $i = 2, \ldots, |\Gamma^n|$  there exists  $j_i \ge i$  such that

$$\eta_1 = \overline{\lambda}_1 = 0, \quad \text{and} \quad \eta_i = \overline{\lambda}_{j_i}, \quad \forall i = 2, \dots, |\Gamma^n|.$$
 (4.27)

We are left with the task to establish a relation between the eigenvalues of  $-L_N$  and  $-\overline{L}_N$ . Suppose that *n* is chosen large enough to ensure that  $\varepsilon(n) \leq 3/5$ . Then, the assumption (2.4) implies that for  $\sigma \in S^n[x]$  and  $\eta \in S^n[y]$  such that  $p_N(\sigma, \eta) > 0$ 

$$e^{-\varepsilon(n)}\frac{p_N(\sigma,\eta)}{\bar{p}_N(\sigma,\eta)} \le e^{\varepsilon(n)}$$
 and  $e^{-\varepsilon(n)(N-1)} \le \frac{\mu_N(\sigma)}{\bar{\mu}_N(\sigma)} \le e^{\varepsilon(n)(N-1)}$ . (4.28)

Recall that the Dirichlet form,  $\mathcal{E}(f)$ , that is defined by  $\mathcal{E}(f) = \langle f, -L_N f \rangle_{\mu_N}$ , can be written as a positive quadratic form. Hence,

$$\left\langle f, -L_N f \right\rangle_{\mu_N} \geq e^{-\varepsilon(n)N} \frac{1}{2} \sum_{\sigma,\eta \in \mathcal{S}_N} \bar{\mu}_N(\sigma) \bar{p}_N(\sigma,\eta) \left( f(\sigma) - f(\eta) \right)^2$$
  
=  $e^{-\varepsilon(n)N} \left\langle f, -\overline{L}_N f \right\rangle_{\mu_N},$  (4.29)

whereas  $\langle f, f \rangle_{\mu_N} \leq e^{\varepsilon(n)N} \langle f, f \rangle_{\overline{\mu}_N}$ . A further application of the Courant-Fischer Theorem, yields

$$\lambda_{j_i} \geq e^{-2\varepsilon(n)N} \min_{\substack{\mathcal{W}_{j_i} \subset \mathbb{R}^{|\mathcal{S}_N|} \\ \dim \mathcal{W}_{j_i} = j_i}} \max_{\substack{f \in \mathcal{W}_i \\ f \neq 0}} \frac{\langle f, -\bar{L}_N f \rangle_{\bar{\mu}_N}}{\langle f, f \rangle_{\bar{\mu}_N}} = e^{-2\varepsilon(n)N} \bar{\lambda}_{j_i}.$$
(4.30)

Hence, by combining (4.27) and (4.30), the lower bound in (4.22) is immediate. Moreover, by interchanging the role of  $L_N$  and  $\overline{L}_N$  in (4.29) we can conclude the proof.

**Remark 4.6.** Notice that the bounds  $\lambda_i \leq \eta_i \leq \lambda_{|S_N| - |\Gamma^n| + i}$  for all  $i = 2, ..., |\Gamma^n|$  have already been established in [60].

**Corollary 4.7.** Suppose that there exists  $K \in \{2, ..., |\Gamma^n|\}$  such that  $\lambda_{K+1} > e^{2\varepsilon(n)N}\eta_K$ . Then,

$$e^{-2\varepsilon(n)N}\eta_i \leq \lambda_i \leq \eta_i, \quad \forall i=2,\ldots,K.$$
 (4.31)

*Proof.* In view (4.22), every interval  $[e^{-2\varepsilon(n)N}\eta_i, e^{2\varepsilon(n)N}\eta_i]$ , for  $i = 2, ..., |\Gamma^n|$ , contains at least one eigenvalue of  $-L_N$ . By inspecting the proof of Proposition 4.5, we see that the assumption  $\lambda_{K+1} > e^{2\varepsilon(n)N}\eta_K$  together with (4.27) implies that  $\bar{\lambda}_K = \eta_K$ . Since  $j_i \ge i$ , it follows that  $\bar{\lambda}_i = \eta_i$  for all i = 1, ..., K and (4.30) yields the assertion.

Let us point out that it is a rather challenging task to prove that the number of eigenvalues below the principal Dirichlet eigenvalue  $\lambda^{\mathcal{M}^n}$  coincide with the number of metastable sets, say K. To our best knowledge, an answer to this problem is so far missing. Notice that for the precise characterization of the small eigenvalues, that we present in the next section, it suffices to know a priori that there exists  $\delta > 0$  such that  $e^{-\delta N} \lambda^{\mathcal{M}^n} \leq \lambda_{K+1}$ . Although differently formulated, we will assume in the sequel the existence of such a separation.

**Lemma 4.8.** Suppose that the set of metastable sets,  $\mathcal{M}^n$ , contains  $K \equiv K(N)$  elements, where K(N) grows at most sub-exponentially. Assume there exists  $0 < \mathfrak{c} < \mathfrak{C}$  such that

$$\min_{X \in \mathcal{S}_N^n \setminus \mathcal{M}^n} \mathbb{P}_{\mu_X} \left[ \tau_{\mathcal{M}^n} < \tau_X \right] \ge e^{-\mathfrak{c}N}, \tag{4.32}$$

where  $S_N^n = \{S^n[x] \subset S_N \mid x \in \Gamma^n\}$ . Suppose that there exists  $0 < \delta < \mathfrak{C} - \mathfrak{c}$  such that

$$\bar{\lambda}_{K+1} \ge e^{-\delta N} \bar{\lambda}^{\mathcal{M}^n}, \tag{4.33}$$

where  $\bar{\lambda}^{\mathcal{M}^n}$  denotes the principal Dirichlet eigenvalue of the generator  $-\overline{L}_N$  subject to zero boundary conditions on  $\mathcal{M}^n$ . If  $n \ll N$  is chosen large enough to ensure that  $2\varepsilon(n) < \mathfrak{C} - \mathfrak{c}$ , then

$$\frac{1}{2KN^{n\cdot q}} e^{(\mathfrak{C}-\mathfrak{c}-\delta-2\varepsilon(n))N} \lambda_K \leq \lambda_{K+1}.$$
(4.34)

*Proof.* Recall that metastable sets are defined as preimages under  $arrho^n$  of points  $m{m}\in m{M}\subset\Gamma^n.$ 

As a first step we claim that  $\bar{\lambda}^{\mathcal{M}^n} = \eta^M$ , where  $\eta^M$  is the principal Dirichlet eigenvalue of  $-G^n$ . Let us denote by  $\overline{L}_{\mathcal{M}^n}$  and  $G^n_M$  the operators subject to zero boundary conditions on  $\mathcal{M}^n$  and M, respectively. We prove the claim above by contradiction. Assume that  $\bar{\lambda}^{\mathcal{M}^n} < \eta^M$ . In view of the discussion above, we know that  $\operatorname{spec}(-G^n_M) \subset \operatorname{spec}(-\overline{L}_{\mathcal{M}^n})$ . Hence, there exists  $\bar{\lambda}_i^{\mathcal{M}^n} \in \operatorname{spec}(-\overline{L}_{\mathcal{M}^n})$  with i > 1 such that  $\bar{\lambda}_i^{\mathcal{M}^n} = \eta^M$ . By the Theorem of Perron-Frobenius we know that the eigenfunctions  $\phi^1$  to  $\bar{\lambda}^{\mathcal{M}^n}$  and  $g^1$  to  $\eta^M$  are strictly positive. Since  $\langle \phi^1, \phi^i \rangle_{\bar{\mu}_N} = 0$ , the eigenfunction to  $\bar{\lambda}_i^{\mathcal{M}^n}$  has to be somewhere negative. Since  $\phi^i$  is constant for all  $\sigma \in S^n[x]$ , see [82, Lemma 12.8] the eigenfunction  $g^1$  is somewhere negative. Contradiction. Using a similar argument if  $\eta^M$  is assumed to be smaller than  $\bar{\lambda}^{\mathcal{M}^n}$  proves the claim.

In view of (4.30) we have that

$$\lambda_{K+1} \geq e^{-2\varepsilon(n)N} \bar{\lambda}_{K+1} \geq e^{-(\delta + 2\varepsilon(n))N} \eta^{M}.$$
(4.35)

Thus, the assertion follows once we have shown an upper bound on  $\lambda_K$  that is strictly smaller than a lower bound for  $e^{-(\delta+2\varepsilon(n))N}\eta^M$ . Combining (4.2) with (4.6) implies that

$$\eta^{\boldsymbol{M}} \geq \min_{\boldsymbol{x}\in\Gamma^{n}\backslash\boldsymbol{M}} \frac{\operatorname{CAP}^{n}(\boldsymbol{x},\boldsymbol{M})}{\left\|h_{\boldsymbol{x},\boldsymbol{M}}\right\|_{1,\mathcal{Q}^{n}}} \geq \frac{1}{N^{n\cdot q}} \left(\min_{\boldsymbol{x}\in\Gamma^{n}\backslash\boldsymbol{M}} \frac{\operatorname{CAP}^{n}(\boldsymbol{x},\boldsymbol{M})}{\mathcal{Q}^{n}(\boldsymbol{x})}\right)^{2},$$
(4.36)

where we used in the second step [11, Lemma 4.8]. Furthermore, from the Dirichlet principle follows immediately that  $CAP^{n}(x, M) \ge cap(X, M^{n})$  where  $X = S^{n}[x]$ . Hence,

$$\eta^{\boldsymbol{M}} \geq \frac{1}{N^{n \cdot q}} \left( \min_{X \in \mathcal{S}_{N}^{n} \setminus \mathcal{M}^{n}} \frac{\operatorname{cap}(X, \mathcal{M}^{n})}{\mu_{N}[X]} \right)^{2} \geq \frac{\mathrm{e}^{-\mathfrak{c}N}}{N^{n \cdot q}} \min_{X \in \mathcal{S}_{N}^{n} \setminus \mathcal{M}^{n}} \mathbb{P}_{\mu_{X}} \big[ \tau_{\mathcal{M}^{n}} < \tau_{X} \big].$$
(4.37)

Let us now derive an upper bound for  $\lambda_K$  that is strictly smaller than the right-hand side of (4.37). From (4.22) we already know that  $\lambda_K < \eta_K$ . Further, by (4.11), it holds that  $\eta_K \leq \max_{\boldsymbol{x} \in \boldsymbol{M}} \operatorname{CAP}^n(\boldsymbol{x}, \boldsymbol{M} \setminus \boldsymbol{x}) / \mathcal{Q}^n(\boldsymbol{x})$ . However, instead of trying to bound from above the mesoscopic capacity  $\operatorname{CAP}^n(\boldsymbol{x}, \boldsymbol{M} \setminus \boldsymbol{x})$  by the microscopic capacity  $\operatorname{cap}(X, \mathcal{M}^n \setminus X)$ , we present in the sequel an alternative proof of an upper bound for  $\lambda_K$ .

Let us consider the *K*-dimensional subspace C of  $\mathbb{R}^{|S^n|}$  that is spanned by normalized equilibrium potentials

$$\mathcal{C} = \operatorname{span}\left\{ h_{X,\mathcal{M}^n \setminus X} / \left\| h_{X,\mathcal{M}^n \setminus X} \right\|_{2,\mu_N}^2 \mid X \in \mathcal{M}^n \right\}.$$

By exploiting the Theorem of Courant and Fischer, we get

$$\lambda_{K} = \min_{\substack{\mathcal{W} \subset \mathbb{R}^{|S^{n}|} \\ \dim \mathcal{W} = K}} \max_{\substack{f \in \mathcal{W} \\ f \neq 0}} \frac{\langle f, -L_{N} f \rangle_{\mu_{N}}}{\langle f, f \rangle_{\mu_{N}}} \le \max_{\substack{f \in \mathcal{C} \\ f \neq 0}} \frac{\langle f, -L_{N} f \rangle_{\mu_{N}}}{\langle f, f \rangle_{\mu_{N}}} = \max_{\substack{z \in \mathbb{R}^{K} \\ z \neq 0}} \frac{\langle z, Kz \rangle}{\langle z, Bz \rangle}, \quad (4.38)$$

where  $K = (K_{X,Y}) \in \mathbb{R}^{K \times K}$  is the so-called *capacitance matrix* that reads

$$\mathcal{K}_{X,Y} = \frac{\left\langle h_{X,\mathcal{M}^n \setminus X}, -L_N h_{Y,\mathcal{M}^n \setminus Y} \right\rangle_{\mu_N}}{\left\| h_{X,\mathcal{M}^n \setminus X} \right\|_{2,\mu_N} \left\| h_{Y,\mathcal{M}^n \setminus Y} \right\|_{2,\mu_N}} \quad \forall X,Y \in \mathcal{M}^n$$
(4.39)

while the matrix  $B = (B_{X,Y}) \in \mathbb{R}^{K \times K}$  is given by

$$B_{X,Y} = \frac{\langle h_{X,\mathcal{M}^n \setminus X}, h_{Y,\mathcal{M}^n \setminus Y} \rangle_{\mu_N}}{\left\| h_{X,\mathcal{M}^n \setminus X} \right\|_{2,\mu_N} \left\| h_{Y,\mathcal{M}^n \setminus Y} \right\|_{2,\mu_N}} \quad \forall X,Y \in \mathcal{M}^n.$$
(4.40)

Since both the matrix K and B are positive definite, we get that for all  $z \in \mathbb{R}^K \setminus \{0\}$ 

$$\frac{\langle z, \mathcal{K}z \rangle}{\|z\|^2} \le \lambda_{\max}(\mathcal{K}) \le \sum_{i=1}^K \lambda_i(\mathcal{K}) = \operatorname{tr}(\mathcal{E}) \le K \max_{X \in \mathcal{M}^n} \mathbb{P}_{\mu_X} \big[ \tau_{\mathcal{M}^n \setminus X} < \tau_X \big], \quad (4.41)$$

where we used in the last step that  $||h_{X,\mathcal{M}^n\setminus X}||_{2,\mu_N}^2 \ge \mu_N[X]$ . Furthermore, if we denote by  $\lambda_{\min}(B)$  the smallest eigenvalue of B then  $\langle z, Bz \rangle \ge \lambda_1(B) ||z||^2$ . By applying Gershgorin's Theorem, we obtain

$$\lambda_{1}(B) \geq \min_{X \in \mathcal{M}^{n}} \left( B_{X,X} - \sum_{Y \in \mathcal{M}^{n} \setminus X} |B_{X,Y}| \right)$$
  
$$\geq \min_{X \in \mathcal{M}^{n}} \left( 1 - \sum_{Y \in \mathcal{M}^{n} \setminus X} \frac{\langle h_{X,\mathcal{M}^{n} \setminus X}, h_{Y,\mathcal{M}^{n} \setminus Y} \rangle_{\mu_{N}}}{\sqrt{\mu_{N}[X] \, \mu_{N}[Y]}} \right).$$
(4.42)

Suppose that  $\mu_N[X] \leq \mu_N[Y]$ . Otherwise, we exchange in the following argument X and Y. In view of Lemma 3.8, we have that  $\langle h_{X,\mathcal{M}^n\setminus X}, h_{Y,\mathcal{M}^n\setminus Y} \rangle_{\mu_N}$  is exponentially small in N compared to  $\mu_N[X]$ . Since the number of metastable sets grows at most sub-exponentially, it holds that  $\lambda_1(B) \geq 1 - \mathcal{O}_N(1)$ . Combining the estimates above and taking into account the properties of metastable sets, we get

$$\lambda_{K} \leq 2 K e^{-\mathfrak{C}N} \min_{X \in \mathcal{S}_{N}^{n} \setminus \mathcal{M}^{n}} \mathbb{P}_{\mu_{X}} [\tau_{\mathcal{M}^{n}} < \tau_{X}]$$
(4.43)

Hence, by combining (4.35), (4.37) and (4.43), we conclude the proof.

**Remark 4.9.** Suppose that we pick from each metastable set  $M_i$  a configuration  $\eta^i \in M_i$ . Then, by (4.2), we know that  $\bar{\lambda}^M \leq \bar{\lambda}_{K+1}$  where  $M = \{\eta^1, \ldots, \eta^K\}$ . Hence, the problem concerning the number of eigenvalues of  $-L_N$  below  $\lambda^{\mathcal{M}^n}$  reduces to compare two principal Dirichlet eigenvalues, namely  $\bar{\lambda}^M$  and  $\bar{\lambda}^{\mathcal{M}^n}$ . This has the following advantages:

- 1) Obviously, by Perron-Frobenius, the eigenfunction,  $\phi_M$ , corresponding to  $\bar{\lambda}^M$  is strictly positive.
- 2) There is a coupling of two versions of the Markov chain  $\{\bar{\sigma}(t)\}$  starting in two different configurations  $\sigma, \eta \in S^n[x]$  such that, for all later times t,  $\varrho^n(\bar{\sigma}(t)) = \varrho^n(\bar{\eta}(t))$ . Moreover, the Hamming distance between  $\bar{\sigma}(t)$  and  $\bar{\eta}(t)$  is non-increasing. Hence, the distribution of the coupling time decays exponentially. This observation may be helpful to control the oscillations of the eigenfunction,  $\phi_M$ .
- 3) It suffices to show that there exist  $0 < \delta < \mathfrak{C} 2\varepsilon(n)$  such that  $e^{-\delta N} \bar{\lambda}^{\mathcal{M}^n} \leq \bar{\lambda}^M$ .

These observations may be helpful to prove (4.2) using only properties of metastable sets.

## 4.3. Characterization of small eigenvalues

In the present section we prove the main result of this chapter. Namely, starting from the rough localization of the spectrum of  $-L_N$ , we improve substantially both the upper and lower bounds for the low-lying part of the spectrum. Moreover, we establish a relation between the low-lying eigenvalues of  $-L_N$  and the metastable exit times associated to certain metastable sets. In the sequel, we construct to each low-lying eigenvalue a pair of approximate eigenvalue/eigenfunction. Since the eigenvalues in the low-lying part of the spectrum are exponentially small in N, the strategy of our proof to derive *relative error bound* that allows to determine *a posteriori* the accuracy of an approximated eigenvalue.

Let us consider a set of metastable sets,  $\mathcal{M}^n$ , as defined in Definition 3.1. Suppose that  $\mathcal{M}^n$  consists of K elements, i.e.  $\mathcal{M}^n = \{M_1, \ldots, M_K\}$  with  $M_i \cap M_j = \emptyset$  for  $i \neq j$ , where K = K(N) may grow at most sub-exponential in N. In view of the discuss in the previous section, we assume in all what follows that there exists  $\delta > 0$  such that  $e^{-\delta N} \lambda^{\mathcal{M}^n} \leq \lambda_{K+1}$ . Further, we take the following *minimal non-degeneracy* condition into account.

**Assumption 4.10.** Let us assume that there exists  $0 < \delta < \mathfrak{C}$  and an unique increasing sequence of sets of sets  $\mathcal{M}_1^n \subset \mathcal{M}_2^n \subset \ldots \subset \mathcal{M}_K^n \equiv \mathcal{M}^n$  such that for all  $i = 2, \ldots, K$ , it holds that



FIGURE 1. Illustration of a landscapes a) and the associated tree structure b). Since the landscape satisfies the minimal non-degeneracy condition, we can associate to each local minimum  $x_i$  a unique saddle point  $z_i$ . Moreover, the height differences between a local minimum and its saddle point are all distinct.

(i) 
$$\mathcal{M}_{i-1}^n = \mathcal{M}_i^n \setminus X_i$$
, where the set  $X_i \in \mathcal{M}_i^n$  satisfies  
$$\max_{M \in \mathcal{M}_{i-1}^n} \mathbb{P}_{\mu_M} \left[ \tau_{\mathcal{M}_i^n \setminus M} < \tau_M \right] < e^{-\delta N} \mathbb{P}_{\mu_{X_i}} \left[ \tau_{\mathcal{M}_{i-1}^n} < \tau_{X_i} \right];$$
(4.44)

(ii) either

$$\mu_N[M] < e^{-\delta N} \mu_N[X_i] \tag{4.45}$$

or

$$\mathbb{P}_{\mu_{X_i}}\left[\tau_{\mathcal{M}_{i-1}^n} < \tau_{X_i}\right] < e^{-\delta N} \mathbb{P}_{\mu_M}\left[\tau_{\mathcal{M}_{i-1}^n} < \tau_M\right]$$
(4.46)

for all  $M \in \mathcal{M}^n \setminus \mathcal{M}_i^n$ .

**Remark 4.11.** The asymptotic ultrametricity of metastable sets induces a tree structure, from which the properties of the long-time behavior of the process can be read off. Notice that, by taking into account the minimal non-degeneracy condition, we obtain a binary tree having the property that the length of the shorter sub-trees below a branching point are all distinct. The construction of this tree goes at follows: At the root split the tree into two sub-trees such that the metastable set  $X_1$  is a leaf in the first and  $X_2$  is a leaf in the second sub-tree. Within each of these two sub-trees we repeat this procedure, see Figure 4.3.

The assumption on  $\lambda_{K+1}$  together with the minimal non-degeneracy condition allows to characterize precisely the *K* smallest eigenvalues of  $-L_N$ .

**Theorem 4.12.** Let us consider an increasing sequence  $\mathcal{M}_1^n \subset \mathcal{M}_2^n \subset \ldots \subset \mathcal{M}_K^n \equiv \mathcal{M}^n$  that satisfies the minimal non-degeneracy condition specified in Assumption 4.10 and assume that there exists  $0 < \mathfrak{c} < \mathfrak{C}$  such that

$$\min_{X \in \mathcal{S}_N^n \setminus \mathcal{M}^n} \mathbb{P}_{\mu_X} \left[ \tau_{\mathcal{M}^n} < \tau_X \right] \ge e^{-\mathfrak{c}N}.$$
(4.47)

Suppose that there exists  $0 < \delta < \mathfrak{C} - \mathfrak{c}$  such that  $e^{-\delta N} \lambda^{\mathcal{M}^n} < \lambda_{K+1}$ . If  $n \ll N$  is large enough to ensure that  $2\varepsilon(n)(c_1+1) < \min\{\mathfrak{C} - \mathfrak{c} - \delta, \delta\}$ , where  $c_1$  is chosen in such a way that  $e^{-c_2N} < \mathbb{P}_{\mu_{X_2}}[\tau_{X_1} < \tau_{X_2}]$ . Then, there exists  $c_{14} > 0$  such that

$$\lambda_{1} = 0, \qquad \lambda_{i} = \frac{\operatorname{cap}(X_{i}, \mathcal{M}_{i-1}^{n})}{\left\|h_{X_{i}, \mathcal{M}_{i-1}^{n}}\right\|_{2, \mu_{N}}^{2}} \left(1 - e^{-c_{14}N}\right), \qquad \forall i = 2, \dots, K.$$
(4.48)

In particular, for all  $i = 2, \ldots, K$ ,

$$\lambda_i = \frac{1}{\mathbb{E}_{\sigma}[\tau_{\mathcal{M}_{i-1}^n}]} \left(1 - e^{-c_{14}N}\right), \qquad \forall \sigma \in X_i.$$
(4.49)

Let us start with improving the upper bound given in Proposition 4.5.

Lemma 4.13. Under the assumptions of Theorem 4.12, it holds that

$$\lambda_{1} = 0, \qquad \lambda_{i} \leq \frac{\operatorname{cap}(X_{i}, \mathcal{M}_{i-1}^{n})}{\left\|h_{X_{i}, \mathcal{M}_{i-1}^{n}}\right\|_{2, \mu_{N}}^{2}} \left(1 + e^{-c_{14}N}\right), \qquad \forall i = 2, \dots, K.$$
(4.50)

*Proof.* Note that the strategy of this proof is similar to the one given in Proposition 4.8. First, consider the vector  $u^1, \ldots, u^K$  with  $u^i \in \mathbb{R}^{|S_N|}$  that are given by

$$u^{1} = 1, \qquad u^{i} = \frac{h_{X_{i},\mathcal{M}_{i-1}^{n}}}{\left\|h_{X_{i},\mathcal{M}_{i-1}^{n}}\right\|_{2,\mu_{N}}^{2}} \quad \text{for } i = 2,\dots,K.$$
 (4.51)

Further, we denote by  $C_i$  the *i*-dimensional subspace  $C_i = \text{span}\{u^1, \ldots, u^i\}$ . Hence, for every  $i = 1, \ldots, K$ , we get

$$\lambda_{i} \leq \min_{\substack{\mathcal{W} \subset \mathbb{R}^{|\mathcal{S}^{n}|} \\ \dim \mathcal{W} = i}} \max_{\substack{f \in \mathcal{W} \\ f \neq 0}} \frac{\langle f, -L_{N} f \rangle_{\mu_{N}}}{\langle f, f \rangle_{\mu_{N}}} \leq \max_{\substack{f \in \mathcal{C}_{i} \\ f \neq 0}} \frac{\langle f, -L_{N} f \rangle_{\mu_{N}}}{\langle f, f \rangle_{\mu_{N}}} = \max_{\substack{z \in \mathbb{R}^{i} \\ z \neq 0}} \frac{\langle z, \mathcal{K}_{i} z \rangle}{\langle z, B_{i} z \rangle}, \quad (4.52)$$

where  $\mathcal{K}_i = (\langle u^k, -L_N u^l \rangle_{\mu_N})_{k,l=1}^i \in \mathbb{R}^{i \times i}$  and  $\mathcal{B}_i = (\langle u^k, u^l \rangle_{\mu_N})_{k,l=1}^i \in \mathbb{R}^{i \times i}$ . But, as a consequence of (4.44), we have that, for any  $z \in \mathbb{R}^i \setminus \{0\}$ ,

$$\frac{\langle z, \mathcal{K}_i z \rangle}{\|z\|^2} \leq \lambda_{\max}(\mathcal{K}_i) \leq \operatorname{tr}(\mathcal{K}_i) \leq \frac{\operatorname{cap}(X_i, \mathcal{M}_{i-1}^n)}{\|h_{X_i, \mathcal{M}_{i-1}^n}\|_{2, \mu_N}^2} \left(1 + K \operatorname{e}^{-\delta N}\right).$$
(4.53)

On the other hand,  $\langle z, B_i z \rangle \geq \lambda_{\min}(B_i) ||z||^2$ . In order to obtain a lower bound for  $\lambda_{\min}(B_i)$ , we exploit again that Gershgorin's Theorem yields

$$\lambda_{\min}(B_{i}) \geq \min_{j=2,...,i} \left( 1 - \frac{\|u^{j}\|_{1,\mu_{N}}}{\|u^{j}\|_{2,\mu_{N}}} - \sum_{\substack{k=2\\k\neq j}}^{i} \frac{\langle u^{j}, u^{k} \rangle}{\|u^{j}\|_{2,\mu_{N}}} \right)$$
  
$$\geq \min_{j=2,...,i} \left( 1 - \frac{\|h_{X_{j},\mathcal{M}_{j-1}^{n}}\|_{1,\mu_{N}}}{\sqrt{\mu_{N}[X_{j}]}} - \sum_{\substack{k=2\\k\neq j}}^{i} \frac{\langle h_{X_{j},\mathcal{M}_{j-1}^{n}}, h_{X_{k},\mathcal{M}_{k-1}^{n}} \rangle_{\mu_{N}}}{\sqrt{\mu_{N}[X_{j}]\mu_{N}[X_{k}]}} \right). \quad (4.54)$$

Thanks to Lemma 3.8 together with the minimal non-degeneracy condition, we get that, for all j = 2, ..., K,  $\mu_N[X_j] < e^{-\delta N} \mu_N[X_1]$  and

$$\begin{aligned} \left\|h_{X_{j},\mathcal{M}_{j-1}^{n}}(\sigma)\right\|_{1,\mu_{N}} &\leq \sum_{M\in\mathcal{M}^{n}}\sum_{\sigma\in\mathcal{V}^{n}(M)}\mu_{N}(\sigma)h_{X_{j},\mathcal{M}_{j-1}^{n}}(\sigma) \\ &\leq N^{n\cdot q}\,\mu_{N}[X_{j}]\left(1+\mathrm{e}^{-(\mathfrak{C}-\varepsilon(n)c_{1})N}+\mathrm{e}^{-(\min\{\delta,\mathfrak{C}-\delta\}-\varepsilon(n)c_{1})N}\right). \end{aligned}$$
(4.55)

Since  $\|h_{X_j,\mathcal{M}_{j-1}^n}\|_{2,\mu_N}^2 \leq \|h_{X_j,\mathcal{M}_{j-1}^n}\|_{1,\mu_N}$ , the estimates above combined with the Cauchy-Schwarz inequality implies that  $\lambda_{\min}(B_i) \geq 1 - N^{n \cdot q} e^{-\delta N/2}$  and the assertion follows immediately.

After having established an upper bound on the small eigenvalues of  $-L_N$  that involves only the capacity between metastable sets and the corresponding equilibrium potential, our next task is to prove lower bounds that coincide in the limit  $N \to \infty$ . In order to apply the Theorem of Kato and Temple, we start with deriving an upper bound on the residual.

66

**Lemma 4.14.** Let  $A = X_i$  and  $B = \mathcal{M}_{i-1}^n$  for  $i = 2, \ldots, K$ . Further, set

$$\xi = \frac{\langle h_{A,B}, -L_N h_{A,B} \rangle_{\mu_N}}{\|h_{A,B}\|_{2,\mu_N}^2}.$$
(4.56)

and suppose that  $c_1 > 1 + \alpha^{-1}$  is chosen in such a way that  $e^{-c_2N} < \mathbb{P}_{\mu_A}[\tau_B < \tau_A]$ . Then,

$$\frac{\|(L_N+\xi)h_{A,B}\|_{2,\mu_N}}{\|h_{A,B}\|_{2,\mu_N}} \le \xi e^{\varepsilon(n)c_1N}.$$
(4.57)

*Proof.* Recall the properties of the equilibrium potential,  $h_{A,B}$ , and the equilibrium measure,  $e_{A,B}$ , i.e.  $h_{A,B}$  is harmonic in  $\sigma \in (A \cup B)^c$  and  $e_{A,B}(\sigma) = -(L_N h_{A,B})(\sigma) = -e_{B,A}(\sigma)$  for  $\sigma \in A \cup B$ . In particular, we have that

$$(L_N + \xi)h_{A,B}(\sigma) = \begin{cases} \xi h_{A,B}(\sigma), & \sigma \notin A \cup B, \\ -e_{A,B}(\sigma) + \xi, & \sigma \in A, \\ e_{B,A}(\sigma), & \sigma \in B. \end{cases}$$
(4.58)

Hence,

$$\begin{split} \left\| \left( L_{N} + \xi \right) h_{A,B} \right\|_{2,\mu_{N}}^{2} \\ &= \sum_{\sigma \in A} \mu_{N}(\sigma) \left( e_{A,B}(\sigma) - \xi \right)^{2} + \xi^{2} \sum_{\sigma \notin A \cup B} \mu_{N}(\sigma) h_{A,B}^{2} + \sum_{\sigma \in B} \mu_{N}(\sigma) e_{B,A}(\sigma)^{2} \\ &= \xi^{2} \| h_{A,B} \|_{2,\mu_{N}}^{2} - 2\xi \operatorname{cap}(A,B) + \sum_{\sigma \in A} \mu_{N}(\sigma) e_{A,B}(\sigma)^{2} + \sum_{\sigma \in B} \mu_{N}(\sigma) e_{B,A}(\sigma)^{2} \\ &= \xi \| h_{A,B} \|_{2,\mu_{N}}^{2} \left( \mathbb{P}_{\nu_{A,B}} \left[ \tau_{B} < \tau_{A} \right] + \mathbb{P}_{\nu_{B,A}} \left[ \tau_{A} < \tau_{B} \right] - \xi \right). \end{split}$$
(4.59)

In view of (2.27), we have that for all  $\sigma, \eta \in A$ 

$$\mathbb{P}_{\sigma}[\tau_B < \tau_A] \leq \mathrm{e}^{\varepsilon(n)c_1 N} \mathbb{P}_{\eta}[\tau_B < \tau_A] + \mathrm{e}^{-c_2 N}.$$
(4.60)

By multiplying both sides with  $\mu_N(\eta)$  and summing over all  $\eta \in A$ , we obtain that

$$\mathbb{P}_{\sigma}\left[\tau_{B} < \tau_{A}\right] \leq e^{\varepsilon(n)c_{1}N} \frac{\operatorname{cap}(A,B)}{\mu_{N}[A]} \left(1 + \frac{e^{-c_{2}N}}{\mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right]}\right) \leq 2e^{\varepsilon(n)c_{1}N} \frac{\operatorname{cap}(A,B)}{\mu_{N}[A]}, \quad (4.61)$$

where we used the assumption  $e^{-c_2N} < \mathbb{P}_{\mu_A}[\tau_B < \tau_A]$  in the last step. A similar computation reveals that

$$\mathbb{P}_{\sigma}\left[\tau_{A} < \tau_{B}\right] \leq e^{\varepsilon(n)c_{1}N} \frac{\operatorname{cap}(A,B)}{\mu_{N}[A]} \left(\frac{\mu_{N}[A]}{\mu_{N}[B]} + \frac{e^{-c_{2}N}}{\mathbb{P}_{\mu_{A}}\left[\tau_{B} < \tau_{A}\right]}\right) \leq 2e^{\varepsilon(n)c_{1}N} \frac{\operatorname{cap}(A,B)}{\mu_{N}[A]}.$$
(4.62)

In the last step we took into account that the assumption  $\mathcal{M}^n(A) \neq \emptyset$  implies that there exists  $M \in \mathcal{M}^n \setminus A$  such that  $\mu_N[A] < \mu_N[M]$ . Note that (4.61) and (4.62) hold uniformly for all  $\sigma \in B$ . Plugging (4.61) and (4.62) into (4.59) and using that  $\xi = \operatorname{cap}(A, B) / \|h_{A,B}\|_{2,\mu_N}^2$ , we get

$$\frac{\left\| (L_N + \xi) h_{A,B} \right\|_{2,\mu_N}^2}{\left\| h_{A,B} \right\|_{2,\mu_N}^2} \leq \xi^2 \left( 4 \, \mathrm{e}^{\varepsilon(n)c_1 N} \, \frac{\| h_{A,B} \|_{2,\mu_N}^2}{\mu_N[A]} - 1 \right) \leq \xi^2 \, \mathrm{e}^{2\varepsilon(n)c_1 N}, \tag{4.63}$$

where we used in the last step that, in view of Lemma 3.8,  $||h_{A,B}||_{2,\mu_N}^2 \leq K N^{n \cdot q} \mu_N[A]$ , while the number of metastable sets in  $\mathcal{M}^n$  grows at most sub-exponential in N. This concludes the proof.

Lemma 4.15. Under the assumptions of Theorem 4.12 it holds that

$$\lambda_{i} \geq \frac{\operatorname{cap}(X_{i}, \mathcal{M}_{i-1}^{n})}{\left\|h_{X_{i}, \mathcal{M}_{i-1}^{n}}\right\|_{2, \mu_{N}}^{2}} \left(1 - e^{-c_{14}N}\right), \quad \forall i = 2, \dots, K.$$
(4.64)

Proof. In order to shorten notation, we set

$$\xi_{i} = \frac{\left\langle h_{X_{i},\mathcal{M}_{i-1}^{n}}, -L_{N} h_{X_{i},\mathcal{M}_{i-1}^{n}} \right\rangle_{\mu_{N}}}{\left\| h_{X_{i},\mathcal{M}_{i-1}^{n}} \right\|_{2,\mu_{N}}^{2}} = \frac{\operatorname{cap}(X_{i},\mathcal{M}_{i-1}^{n})}{\left\| h_{X_{i},\mathcal{M}_{i-1}^{n}} \right\|_{2,\mu_{N}}^{2}}.$$

Due to the fact that the assertion (4.64) is trivial if  $\lambda_i \ge \xi_i$ , let us assume that  $\lambda_i < \xi_i$  for all i = 2, ..., K.

We start with proving a lower bound for  $\lambda_K$ . In order to take advantage of the upper bound in the Theorem of Kato and Temple, we have to find an interval  $(a_K, b_K)$  such that  $(a_K, b_K) \cap \operatorname{spec}(-L_N) = \emptyset$ , see [104, Lemma 3.2]. Since  $\lambda_K < \xi_K$  we set  $a_K = \xi_K$ . Recall that  $\lambda_{K+1} \ge e^{-(\delta+2\varepsilon(n))N}\eta^M$ , where  $\eta^M$  is the principal eigenvalue of  $-G^n$  subject to zero boundary conditions on  $M = \varrho^n(\mathcal{M}^n)$ . In view of (4.37) we have that

$$\eta^{\boldsymbol{M}} \geq \frac{\mathrm{e}^{-\mathfrak{c}N}}{N^{n\cdot q}} \min_{X \in \mathcal{S}_N^n \setminus \mathcal{M}^n} \mathbb{P}_{\mu_X} \left[ \tau_{\mathcal{M}^n} < \tau_X \right] \geq \frac{\mathrm{e}^{(\mathfrak{C}-\mathfrak{c})N}}{N^{n\cdot q}} \xi_K.$$

Now, if we set  $b_K = e^{-(\mathfrak{c}+\delta+2\varepsilon(n))N}/N^{n\cdot q} \min_{X\in S_N^n\setminus \mathcal{M}^n} \mathbb{P}_{\mu_X}[\tau_{\mathcal{M}^n} < \tau_X]$ , the interval  $(a_K, b_K)$  contains no eigenvalue of  $-L_N$ . Hence, the Theorem of Kato and Temple combined with Lemma 4.14 implies that

$$\frac{\xi_K - \lambda_K}{\xi_K} \le e^{2\varepsilon(n)c_1N} \frac{\xi_K}{b_K - \xi_K} \le 2N^{n \cdot q} e^{-(\mathfrak{C} - \mathfrak{c} - \delta - 2\varepsilon(n)(1 + c_1))N},$$
(4.65)

and, for N large enough, (4.65) is equivalent to

$$\lambda_{K} \geq \frac{\operatorname{cap}(X_{K}, \mathcal{M}_{K-1}^{n})}{\|h_{X_{K}, \mathcal{M}_{K-1}^{n}}\|_{2, \mu_{N}}^{2}} \left(1 - e^{-c_{14}N}\right).$$
(4.66)

We prove the lower bound for the remaining eigenvalues  $\lambda_{K-1}, \ldots, \lambda_2$  inductively. Suppose that we have already shown that  $\lambda_{i+1} \ge \xi_{i+1} (1 - e^{-c_{14}N})$ . Under the assumption that  $\lambda_i < \xi_i$ , we set  $a_i = \xi_i$  and  $b_i = \frac{1}{2} \xi_{i+1}$ . As an immediate consequence of the minimal non-degeneracy condition, we have that

$$\mathbb{P}_{\mu_{X_{i+1}}}[\tau_{\mathcal{M}_{i}^{n}} < \tau_{X_{i+1}}] \geq \mathbb{P}_{\mu_{X_{i+1}}}[\tau_{\mathcal{M}_{i-1}^{n}} < \tau_{X_{i+1}}] \geq e^{\delta N} \mathbb{P}_{\mu_{X_{i}}}[\tau_{\mathcal{M}_{i-1}^{n}} < \tau_{X_{i}}].$$
(4.67)

Hence,  $\lambda_{i+1} \geq \frac{1}{2}\xi_{i+1} \geq \frac{1}{2}e^{\delta N}/N^{n \cdot q}\xi_i$  which shows that  $(a_i, b_i) \cap \operatorname{spec}(-L_N) = \emptyset$  for sufficiently large N. A further application of the Theorem of Kato and Temple combined with (4.57) reveals

$$\frac{\xi_i - \lambda_i}{\xi_i} \le e^{2\varepsilon(n)c_1N} \frac{\xi_i}{b_i - \xi_i} \le 4 N^{n \cdot q} e^{-(\delta - 2\varepsilon(n)c_1)N}.$$
(4.68)

This concludes the proof.

*Proof of Theorem 4.12.* From (4.50) and (4.64), (4.48) is immediate. Further, in view of Lemma 3.8, we obtain for any i = 2, ..., K

$$\|h_{X_{i},\mathcal{M}_{i-1}^{n}}\|_{2,\mu_{N}}^{2} = \sum_{\sigma\in\mathcal{S}_{N}} \mu_{N}(\sigma) h_{X_{i},\mathcal{M}_{i-1}^{n}}(\sigma) - \sum_{\sigma\in\mathcal{S}_{N}} \mu_{N}(\sigma) h_{X_{i},\mathcal{M}_{i-1}^{n}}(\sigma) h_{\mathcal{M}_{i-1}^{n},X_{i}}(\sigma)$$

$$\geq \sum_{\sigma\in\mathcal{S}_{N}} \mu_{N}(\sigma) h_{X_{i},\mathcal{M}_{i-1}^{n}}(\sigma) \left(1 - K N^{n \cdot q} e^{-c_{14}N}\right).$$
(4.69)

Since 
$$\|h_{X_i,\mathcal{M}_{i-1}^n}\|_{1,\mu_N} \ge \|h_{X_i,\mathcal{M}_{i-1}^n}\|_{2,\mu_N}^2$$
, the estimate above yields that  

$$\lambda_i = \frac{1}{\mathbb{E}_{\nu_{X_i,\mathcal{M}_{i-1}^n}} [\tau_{\mathcal{M}_{i-1}^n}]} \left(1 - K N^{n \cdot q} e^{-c_{14}N}\right)$$
(4.70)

The pointwise estimates in (4.49) follows from Theorem 2.12.

## CHAPTER 5

# Random field Curie–Weiss–Potts model

The present chapter is devoted to the application of the theory developed in the previous chapters. One particular class of models, we are interested in, are *disordered mean field spin systems*. In general, despite of their lack of physical significance, mean field models serve as a good starting point for understanding the phenomena that occur in disordered systems. As an example, we consider the q-spin Potts model on a complete graph, also known as Curie–Weiss–Potts model, in random fields. From a static point of view, it is one of the simplest and less disordered model.

This chapter has two main objectives. First, we are interested in the metastable behavior of the Curie–Weiss–Potts model in the case when the temperature is finite and the distribution of the random field is continuous. We have already mentioned that the entropy plays a particular role in this kind of models, but an exact reduction to a low-dimensional model via lumping techniques is not possible. However, on a heuristic level it is believed that the dynamical behavior on metastable time scales is well described by a diffusion in a mesoscopic potential landscape. For this reason, we construct a family of mesoscopic variables in such a way that the dynamics induced by them is well approximated by a family of Markov processes which reproduce asymptotically the metastable behavior of the original dynamics. Our aim is to compute precisely the mean exit times from a metastable set and to prove the convergence of normalized metastable exit times to an exponential distribution.

In the Sections 1.6 and 1.7 we have seen that variational principles offer a convenient way to derive upper and lower bounds for capacities and, hence, for mean hitting times. Beyond a particular interest in the Curie–Weiss–Potts model, our second objective is to demonstrate how a suitable test function and a non-negative unit flow can be constructed such that the resulting bounds for the capacities coincide asymptotically.

The remainder of this chapter is organized as follows. In Section 5.1, we start with introducing the model and describing the mesoscopic approximation that allows us to deal with the entropy. Afterwards, we present our results. In order to derive sharp estimates for mean hitting times of metastable sets, a first important ingredient is a precise understanding of the measure that is given by the push forward of the random Gibbs measure under the mesoscopic variables. In Section 5.2, we show that the mesoscopic variables satisfy a sharp large deviation principle. In particular, we study the behavior of the corresponding rate function in a neighborhood of critical points. The core of the present chapter are the Sections 5.3 and 5.4, where we construct upper and lower bounds for capacities. In Section 5.3 we start with constructing a test function that is almost harmonic in a small neighborhood of relevant saddle points of the mesoscopic free energy landscape. In view of the Dirichlet principle, this test function is used to derive upper bounds on capacities. As we will see in Section 5.4, this test function is also used to construct in a first step a mesoscopic unit flow. In a second step we construct for each mesoscopic path a subordinate microscopic unit flow and show certain concentration properties along microscopic paths. By exploiting the Berman-Konsowa principle, we obtain in this way lower bounds on capacities. In the finial Section 5.5, the precise computation of mean hitting times is presented.

**Contributions in this chapter.** In the paper [**6**], Bianchi, Bovier and Ioffe studied the dynamical behavior of the random field Curie–Weiss model. In this model the spin variables can assume only two different values, namely -1 and +1. By adopting their strategy in the construction of upper and lower bounds for capacities, we extend their results to the Curie–Weiss–Potts model. Compared to [**6**], the strategy for controlling the harmonic function by means of an averaged renewal equation combined pointwise estimates on harmonic function, that we presented in Chapter 3, is the main contribution of the author. Beyond that, the chapter contains the following novel bits and pieces:

- a generalization of sharp large deviation estimates to the case when the Hessian of the free energy *F*<sup>n</sup> is degenerate, see Proposition 5.7;
- a characterization of the eigenvalues of  $\text{Hess } F^n$  at critical points of  $F^n$ , see Lemma 5.11;
- a construction of a function that is almost harmonic in a small neighborhood of a saddle point of F<sup>n</sup>, see Lemma 5.20;
- a detailed study of the error propagation along a mesoscopic path, see Proposition 5.31.

## 5.1. Introduction

**5.1.1. The model.** Let  $h = \{h^i\}_{i \in \mathbb{N}}$  be a sequence of independent and identical distributed random variables defined on some probability space, say  $(\Omega, \mathcal{B}, \mathbb{P})$ , which takes values in  $\operatorname{span}\{1\}^{\perp} \subset \mathbb{R}^q$ , where 1 denotes the vector in  $\mathbb{R}^q$  whose components are all 1. For the sake of convenience, we will assume that the distribution,  $\mathbb{P}_h = \mathbb{P} \circ (h^i)^{-1}$  has bounded support. Most of the quantities that we are going to define depend on the realization of h.

The Curie–Weiss–Potts model is a generalization of the *Ising model* to q components on a complete graph, say, on N vertices where N will be a large parameter. Since the ferromagnetic interaction among the spin variables is assumed to be of the same strength, the actual structure of the graph becomes irrelevant and it suffices to consider just a labeling  $\{1, \ldots, N\}$  of the lattice sites. To each site  $i \in \{1, \ldots, N\}$  we associate a spin variable  $\sigma_i$  taking values in  $S_0 := \{1, \ldots, q\}$ , called the set of different colors. In doing so, the state space is given by  $S_N = S_0^N$ . Elements of  $S_N$  are denoted by Greek letters  $\sigma, \eta$  and will be called configurations. Given a realization of h, the random Hamiltonian is defined by

$$H_N(\sigma) := -\frac{1}{N} \sum_{i,j=1}^N \delta(\sigma_i, \sigma_j) - \sum_{i=1}^N \sum_{r \in \mathcal{S}_0} h_r^i \delta(\sigma_i, r), \qquad \sigma \in \mathcal{S}_N,$$
(5.1)

where  $\delta(\sigma_i, \sigma_j)$ , the Kronecker symbol, is equal to 1 when  $\sigma_i = \sigma_j$  and zero otherwise.

On the measurable space  $(S_N, \mathcal{B}(S_N))$ , where  $S_N$  is equipped with the product topology, we define the *finite volume Gibbs measure* of this model as the *random probability measure* 

$$\mu_N(\sigma) := \frac{\exp\left(-\beta H_N(\sigma)\right)}{Z_N} q^{-N}, \qquad (5.2)$$

where  $\beta \ge 0$  is the *inverse temperature* and  $Z_N$  is the normalization constant called the *partition function* which is defined by

$$Z_N := \sum_{\sigma \in \mathcal{S}_N} \exp\left(-\beta H_N(\sigma)\right) q^{-N} \equiv \mathbf{E}_{\sigma[N]} \left[\exp(-\beta H_N)\right].$$
(5.3)

**Remark 5.1.** In the low temperature limit,  $\beta \to \infty$ , the Gibbs measure favors those configurations where the Hamiltonian is minimal. In view of (5.1), the first term, describing the interaction among the spins, is minimal if all spins have the same color. The second term in (5.1) represents the interaction of the spin variables and the external random magnetic field. It is minimal if in a configuration  $\sigma$  each spin  $\sigma_i$  takes exactly the color corresponding to the

#### 5.1. INTRODUCTION

maximal component of the field  $h^i$ . There is a competition between the contribution coming from the interaction among the spins and the external field, respectively.

We will study the dynamics of a discrete-time Markov chain  $\{\sigma(t)\}_{t\in\mathbb{N}_0}$  on the state space  $S_N$  with transition probabilities denoted by  $p_N$ . Its discrete generator,  $L_N$ , acts on functions  $f: S_N \to \mathbb{R}$  as

$$(L_N f)(\sigma) = \sum_{\eta \in \mathcal{S}_N} p_N(\sigma, \eta) (f(\eta) - f(\sigma)).$$
(5.4)

We assume that the process is *irreducible* and *reversible* with respect to Gibbs measure  $\mu_N$ . The chain, we consider, evolves by selecting at each step a vertex  $i \in \{1, ..., N\}$  uniformly at random, proposing a color  $r \in S_0$  uniformly at random among all colors and accepting this proposal according to the probability given by the Metropolis rule. More explicitly, the transition probabilities  $p_N$  are given by

$$p_N(\sigma,\eta) := \frac{1}{Nq} \exp\left(-\beta \left[H_N(\eta) - H_N(\sigma)\right]_+\right) \mathbb{1}_{\mathrm{d}_\mathrm{H}(\sigma,\eta)=1},\tag{5.5}$$

where  $[x]_{+} := \max\{x, 0\}$  and  $d_{H}(\sigma, \eta) = 1$  denotes the Hamming distance between  $\sigma$  and  $\eta$ ,

$$p_N(\sigma,\sigma) := 1 - \sum_{\eta \in S_N} p_N(\sigma,\eta).$$
 (5.6)

**Remark 5.2.** Note that we consider a lazy version of the Metropolis dynamics since we allow that a chosen site *i* can draw randomly the color  $\sigma_i$  as well. The reason behind this choice is that there exist  $\alpha > 0$  such that  $N p_N(\sigma, \sigma^{i,r}) \ge \alpha$  for all  $i \in \{1, ..., N\}$  and  $r \in S_0$ . Here, the configuration  $\sigma^{i,r}$  is obtained from  $\sigma$  by replacing the color  $\sigma_i$  at site *i* by the color *r*.

**5.1.2.** Coarse graining and mesoscopic approximation. Since we are interested in studying the model at finite temperatures, the entropy is crucial in the static and dynamic description. For this reason, we will investigate the model on a coarse grained scale. To start with, we write

$$\mathcal{M}_p := \left\{ \nu = (\nu_1, \dots, \nu_q) \in [0, 1]^q \mid \sum_{k \in \mathcal{S}_0} \nu_k = p \right\} \subset \mathbb{R}^q$$

to denote the simplex in  $\mathbb{R}^q$ , which may be identified with the set of finite measures on  $S_0$ . The relative frequencies of colors appearing in a given configuration  $\sigma$  are recorded by means of the *empirical spin distribution*,  $\varrho_N : S_N \to \Gamma_N \subset \mathcal{M}_1$ ,

$$\sigma \mapsto \varrho_N(\sigma) \coloneqq \frac{1}{N} \sum_{i=1}^N \delta_{\sigma_i}$$
(5.7)

where  $\delta_x$  is the point-mass at  $x \in \mathbb{R}$ . Using that  $\delta(\sigma_i, \sigma_j) = \sum_{r \in S_0} \delta(\sigma_i, r) \, \delta(r, \sigma_j)$  we can rewrite the Hamiltonian (5.1) in terms of the *macroscopic variable*,  $\varrho_N(\sigma)$ , as

$$H_N(\sigma) = -N \left\| \varrho_N(\sigma) \right\|^2 - \sum_{i=1}^N \langle h^i, e^{\sigma_i} \rangle,$$
(5.8)

where  $e^{\sigma_i} \in \mathbb{R}^q$  denotes the coordinate vector pointing in direction  $\sigma_i$  and  $\langle \cdot, \cdot \rangle$  is the usual Euclidean product in  $\mathbb{R}^q$ . The *macroscopic variables*,  $\varrho_N$ , will act as an *order parameter* of this model. We define its distribution under the random Gibbs measure,  $\mu_N$ , by the *induced measure*,  $Q_N := \mu_N \circ \varrho_N^{-1}$ .

**Remark 5.3.** Note that there are further ways using different representations of the Potts spins to rewrite the Hamiltonian.

(1) In the hyper-tetrahedral representation suggested by Wu [115], one can express

$$\delta(\sigma_i, \sigma_j) = \frac{1}{q} + \frac{q-1}{q} \langle v^{\sigma_i}, v^{\sigma_j} \rangle$$

where  $v^r$ , for  $r \in S_0$ , are unit vectors, pointing in one of the q different directions, which span a (q-1)-dimensional hyper-tetrahedron.

(2) In the *polar representation* suggested by Mittag and Stephan [87], one can write

$$\delta(\sigma_i, \sigma_j) = \frac{1}{q} \sum_{r \in S_0} (z_{\sigma_i} \cdot \bar{z}_{\sigma_j})^{r-1}$$

where  $z_r = w^r$  with  $w = e^{2\pi i/q}$  which is the *q*th root of unity.

Let us emphasis that, due to the random field,  $\mathbb{P}$ -a.s., the induced process  $\{\varrho_N(\sigma(t))\}_{t\in\mathbb{N}_0}$  is not Markovian. But, there is a canonical construction of an effective Markov chain  $\{\varrho_N(t)\}_{t\in\mathbb{N}_0}$ on  $\Gamma_N$  that is reversible with respect to  $\mathcal{Q}_N$  having the property that  $\varrho_N(t) = \varrho_N(\sigma(t))$  in law whenever the induced process  $\{\varrho_N(\sigma(t))\}$  is a Markovian. If the Markov chain  $\{\sigma(t)\}$  is started in the reversible distribution  $\mu_N$ , then a classical result of Burke and Rosenblatt [21] implies that  $\{\varrho_N(\sigma(t))\}$  is Markovian if and only if for all  $\sigma, \eta \in \mathcal{S}_N$  with  $\varrho_N(\sigma) = \varrho_N(\eta)$  it holds that  $\sum_{\xi:\varrho_N(\xi)=x} p_N(\sigma,\xi) = \sum_{\xi:\varrho_N(\xi)=x} p_N(\eta,\xi)$  for all  $x \in \Gamma_N$ . However, this condition is only satisfied if all  $h^i$  are the same. Due to the fact that  $\mathbb{P}_h$  is assumed to be continuous, it can not be expected that, on the macroscopic level, the Markov chain  $\{\varrho(t)\}$  and the non-Markovian image chain  $\{\varrho(\sigma(t))\}$  have qualitatively the same long-time behavior than. For this reason, our strategy is to construct a family of *mesoscopic variables* having the property that the corresponding effective Markov chain can be seen as a perturbation of the induced dynamics. This strategy was first used in [6].

Due to the fact that the support of the distribution of  $h^i$  is bounded, given a sequence  $\varepsilon(n) \downarrow 0$  as  $n \uparrow \infty$ , we can find, for any n, a partition,  $\mathcal{H}^n \equiv \{\mathcal{H}^n_1, \ldots, \mathcal{H}^n_{k_n}\}$ , such that

$$\operatorname{supp}(\mathbb{P}_h) = \bigcup_{k=1}^{k_n} \mathcal{H}_k^n, \qquad \mathcal{H}_k^n \cap \mathcal{H}_l^n = \emptyset, \quad \forall k \neq l$$

with diam( $\mathcal{H}_k^n$ )  $\leq \varepsilon(n)$ . Notice that the sequence of partitions,  $\{\mathcal{H}^n\}_{n\in\mathbb{N}}$ , can be constructed in such a way that  $\mathcal{H}^{n+1}$  is a refinement of  $\mathcal{H}^n$ . Hence, each realization of the random field  $\{h^i\}_{i\in\mathbb{N}}$  induces a random partition of the set  $\{1, \ldots, N\}$  into subsets

$$\Lambda_k^n \equiv \Lambda_{k,N}^n := \left\{ i \in \{1, \dots, N\} \mid h^i \in \mathcal{H}_k^n \right\}, \qquad k = 1, \dots, k_n.$$

Let us introduce a family of maps  $\rho^n : S_N \to \Gamma^n \subset \mathbb{R}^{k_n \cdot q}$ ,

$$\sigma \mapsto \boldsymbol{\varrho}^{n}(\sigma) \equiv \boldsymbol{\varrho}_{N}^{n}(\sigma) := \sum_{k=1}^{k_{n}} e^{k} \otimes \frac{1}{N} \sum_{i \in \Lambda_{k}^{n}} \delta_{\sigma_{i}}$$
(5.9)

where  $e^k \in \mathbb{R}^{k_n}$  denotes a coordinate vector in  $\mathbb{R}^{k_n}$ . Each map  $\varrho^n$  represents an averages of *microscopic variables* over blocks of *mesoscopic sizes* which are decreasing in n. Note that the range of  $\varrho^n$  is given by

$$\Gamma^{n} \equiv \Gamma_{N}^{n} := \left| \underset{k=1}{\overset{k_{n}}{\times}} \left\{ \nu \in \frac{1}{N} \mathbb{N}_{0}^{q} \right| \sum_{r \in \mathcal{S}_{0}} \nu_{r} = \pi_{k}^{n} \right\} \subset \left| \underset{k=1}{\overset{k_{n}}{\times}} \mathcal{M}_{\pi_{k}^{n}} \right|.$$
(5.10)

Here, for any  $k, \pi_k^n \equiv \pi_{k,N}^n := |\Lambda_k^n|/N$  is the relative frequencies that  $h^i$  takes values in  $\mathcal{H}_k^n$  for a given realization of the random field. Note that the random variables  $\pi_k$  concentrate exponentially in N around their mean values  $\mathbb{E}[\pi_k] = \mathbb{P}[h^i \in \mathcal{H}_k^n] =: p_k$ .

We adopt the notation to denote vectors  $x \in \mathbb{R}^{k_n \cdot q}$  by bold symbols whereas its 'components'  $x^k$  for  $k = 1, ..., k_n$  are elements in  $\mathbb{R}^q$ . Further, in Kronecker products  $x \otimes y \in \mathbb{R}^{k_n \cdot q}$  the first vector x will always be an element in  $\mathbb{R}^{k_n}$  while the second one y will be in  $\mathbb{R}^q$ .

The crucial feature of mean field models is that the Hamiltonian (5.1) can be rewritten as a function of the mesoscopic variables. In order to do so, let

$$\bar{h}^k := \frac{1}{|\Lambda_k^n|} \sum_{i \in \Lambda_k^n} h^i \quad \text{and} \quad \tilde{h}^i \equiv \tilde{h}^{i,k} := h^i - \bar{h}^k, \quad \forall i \in \Lambda_k^n$$
(5.11)

for each  $k = 1, \ldots, k_n$ . Then

$$H_N(\sigma) = -N E(\boldsymbol{\varrho}^n(\sigma)) - \sum_{k=1}^{k_n} \sum_{i \in \Lambda_k^n} \langle \tilde{h}^i, e^{\sigma_i} \rangle, \qquad (5.12)$$

where the function  $E \colon \mathbb{R}^{k_n \cdot q} \to \mathbb{R}$ , is given by

$$\boldsymbol{x} \mapsto E(\boldsymbol{x}) := \left\| \sum_{k=1}^{k_n} x^k \right\|^2 + \sum_{k=1}^{k_n} \langle \bar{h}^k, x^k \rangle.$$
(5.13)

Notice that for a given configuration  $\sigma$ , the sum of the mesoscopic variables equals the corresponding macroscopic variable, i.e.  $\varrho_N(\sigma) = \sum_{k=1}^{k_n} \varrho^k(\sigma)$ . Let us define the distribution of  $\varrho^n$  under the random Gibbs measure,  $\mu_N$ , through

$$Q^n \equiv Q_N^n := \mu_N \circ (\boldsymbol{\varrho}^n)^{-1}.$$
 (5.14)

Furthermore, we introduce the *mesoscopic free energy*,  $F^n$ , which is defined as  $F^n : \mathbb{R}^{k_n \cdot q} \to \mathbb{R}$ ,

$$\boldsymbol{x} \mapsto F^{n}(\boldsymbol{x}) := - \left\| \sum_{k=1}^{k_{n}} x^{k} \right\|^{2} - \sum_{k=1}^{k_{n}} \langle \bar{h}^{k}, x^{k} \rangle + \frac{1}{\beta} \sum_{k=1}^{k_{n}} \pi_{k} I_{|\Lambda_{k}|}(x^{k}/\pi_{k}).$$
(5.15)

The *entropy* or *action function*,  $I_{|\Lambda_k|} \colon \mathbb{R}^q \to \mathbb{R}$ , that appears in the expression of  $F^n$  is defined as the the Legendre-Fenchel transform

$$I_{|\Lambda_k|}(x) := \sup_{t \in \mathbb{R}^q} \left( \langle x, t \rangle - U_{|\Lambda_k|}(t) \right)$$
(5.16)

of the log–moment generating function  $U_{|\Lambda_k|} \colon \mathbb{R}^q \to \mathbb{R}$ ,

 $\sigma$ 

$$t \mapsto U_{|\Lambda_k|}(t) := \frac{1}{|\Lambda_k|} \ln \mathbf{E}_{\sigma[|\Lambda_k|]}^{\tilde{h}} \Big[ \exp(|\Lambda_k| \langle t, \varrho_{|\Lambda_k|} \rangle) \Big] + \frac{1}{|\Lambda_k|} \ln Z_{|\Lambda_k|}^{\tilde{h}} \\ = \frac{1}{|\Lambda_k|} \sum_{i \in \Lambda_k} \ln \left( \sum_{r \in \mathcal{S}_0} \frac{1}{q} \exp(\beta \tilde{h}_r^i + t_r) \right),$$
(5.17)

where we include the constant in the definition of  $U_{|\Lambda_k|}$  to simplify notation in the later.

The Markov chain on state space  $\Gamma^n$  which is reversible with respect to the measure  $\mathcal{Q}^n$  is denoted by  $\{\varrho(t)\}_{t\in\mathbb{N}_0}$ . Its transition probabilities are given by

$$r^{n}(\boldsymbol{x},\boldsymbol{y}) \equiv r_{N}^{n}(\boldsymbol{x},\boldsymbol{y}) := \frac{1}{\mathcal{Q}^{n}(\boldsymbol{x})} \sum_{\boldsymbol{\sigma} \in \mathcal{S}^{n}[\boldsymbol{x}]} \mu_{N}(\boldsymbol{\sigma}) \sum_{\boldsymbol{\eta} \in \mathcal{S}^{n}[\boldsymbol{y}]} p_{N}(\boldsymbol{\sigma},\boldsymbol{\eta}), \qquad \boldsymbol{x}, \boldsymbol{y} \in \Gamma^{n}, \quad (5.18)$$

where  $S^n[x] := (\varrho^n)^{-1}(x)$  is the set-valued preimage of  $\varrho^n$ . If n is chosen such that for all  $\boldsymbol{x}, \boldsymbol{y} \in \Gamma^n$  with  $r^n(\boldsymbol{x}, \boldsymbol{y}) > 0$  it holds that for  $\beta \varepsilon(n) \leq 1$ , then

$$\max_{\boldsymbol{\in}\mathcal{S}^{n}[\boldsymbol{x}],\,\eta\in\mathcal{S}^{n}[\boldsymbol{y}]}\left|\frac{p_{N}(\sigma,\eta)\left|\{\eta\in\mathcal{S}^{n}[\boldsymbol{y}]\mid d_{\mathrm{H}}(\sigma,\eta)=1\}\right|}{r^{n}(\boldsymbol{x},\boldsymbol{y})}-1\right| \leq 3\beta\varepsilon(n).$$
(5.19)

In other words, the Markov chain  $\{ \rho^n(t) \}$  can be seen as a good approximation of the process  $\{\varrho^n(\sigma(t))\}$ . Finally, we denote by  $L^n$  the discrete generator of the Markov chain  $\{\varrho^n(t)\}$ .

In order to simplify the notation in what follows, we will frequently drop the superscript n, identify  $k_n \equiv n$  and refer to the generic partition  $\Lambda_1, \ldots, \Lambda_n$ .

**5.1.3.** Main results. In the sequel, we assume that the inverse temperature  $\beta$  and the distribution,  $\mathbb{P}_h$ , of the random field are such that there exists more than one local minima of  $F^n$ . Here, the coarsening parameter  $n \ll N$  is arbitrary but fixed. Let  $\mathbf{m} \in \Gamma^n$  be a local minimum and let us denote by  $\mathbf{M} \subset \Gamma^n$  the set of deeper local minima of  $F^n$ , i.e. let  $F^n(\mathbf{m}') < F^n(\mathbf{m})$  for all  $\mathbf{m}' \in \mathbf{M}$ . Further, we assume that there exists a unique relevant saddle point,  $\mathbf{z} \in \Gamma^n$ , of index one that separates  $\mathbf{m}$  and  $\mathbf{M}$ , i.e.  $F^n(\mathbf{z}) = \min_{\gamma} \max_{\mathbf{x} \in \gamma} F^n(\mathbf{x})$  where the minimum is taken over all mesoscopic paths in  $\Gamma^n$  that connects  $\mathbf{m}$  and  $\mathbf{M}$ . To each critical point  $\mathbf{x} \in \Gamma^n$  of  $F^n$  we associate a point  $x \in \Gamma_N$  that is given by  $x = x(\mathbf{x}) = \sum_{k=1}^n x^k$ . In Lemma 5.8 we will show that  $x(\mathbf{x}) \neq y(\mathbf{y})$  if the critical points  $\mathbf{x}$  and  $\mathbf{y}$  are distinct. Further, we set  $A = S^n[\mathbf{m}]$  and  $B = S^n[\mathbf{M}]$ .

The main result of the present chapter is the following

**Theorem 5.4.** With the notations and assumptions introduced above, we have that  $\mathbb{P}_h$ -almost surely, for all but finitely many N,

$$Z_N \operatorname{cap}(A, B) = \frac{\beta |\bar{\gamma}_1|}{2\pi N} \frac{\exp(-\beta N F_N(z))}{\sqrt{\left|\det\left(I_q - 2\beta \left(D(z) - \mathbb{E}_h\left[u^1(2\beta z) \cdot u^1(2\beta z)^{\mathrm{T}}\right]\right)\right)\right|}} \left(1 + \mathcal{O}_N(1)\right),$$
(5.20)

where z = z(z),  $D(z) = \text{diag}(z_1, \ldots, z_q)$  and  $u^i(2\beta z) \in \mathbb{R}^q$  is defined componentwise by

$$u_r^i(t) := \frac{\exp(t_r + \beta h_r^i)}{\sum_{s \in \mathcal{S}_0} \exp(t_s + \beta h_s^i)}, \quad \forall k \in \mathcal{S}_0.$$
(5.21)

Further,  $\bar{\gamma}_1$  is the unique negative solution of the explicit equation that is given in (5.87), and

$$F_N(z) = \|z\|^2 - \frac{1}{\beta N} \sum_{i=1}^N \ln\left(\sum_{r \in S_0} \frac{1}{q} \exp(2\beta z_r + \beta h_r^i)\right).$$
(5.22)

For the mean exit times from A and its distribution in the random field Curie–Weiss-Potts model we prove

**Theorem 5.5.** Let us consider the notation introduced above and in Theorem 5.4. Further, we assume that there exists  $\delta > 0$  such that, for large enough N,  $\delta < F^n(z) - F^n(m)$  and there is no local minimum m' of  $F^n$  with  $F^n(m') \in [F^n(m), F^n(m) + \delta]$ . Then,

(i)  $\mathbb{P}_h$ -almost surely, for all but finitely many values of N, we have that

$$\mathbb{E}_{\nu_{A,B}}\left[\tau_B\right] = \frac{2\pi N}{\beta |\bar{\gamma}_1|} K(\beta, m, z) \exp\left(\beta N \left(F_N(z) - F_N(m)\right) \left(1 + \mathcal{O}_N(1)\right)\right)$$
(5.23)

where

$$K(\beta, m, z) = \sqrt{\frac{\left|\det\left(I_q - 2\beta\left(D(z) - \mathbb{E}_h\left[u^1(2\beta z) \cdot u^1(2\beta z)^{\mathrm{T}}\right]\right)\right)\right|}{\det\left(I_q - 2\beta\left(D(m) - \mathbb{E}_h\left[u^1(2\beta m) \cdot u^1(2\beta m)^{\mathrm{T}}\right]\right)\right)}}.$$
(5.24)

(ii) provided that the coarsening level n, used in the definition of A, is chosen large enough to ensure that  $\varepsilon(n)c_1 < (F_N(z) - F_N(m))/4$  for a sufficiently large  $c_1 > 1 + \alpha^{-1}$ , then

$$\mathbb{E}_{\sigma}[\tau_B] = \mathbb{E}_{\nu_{A,B}}[\tau_B] (1 + \mathcal{O}_N(1)), \quad \forall \sigma \in A.$$
(5.25)

(iii) under the same assumption as in (ii), for all  $t \in \mathbb{R}_+$  and  $\sigma \in A$ 

$$\mathbb{P}_{\sigma}[\tau_B/\mathbb{E}_{\sigma}[\tau_B] > t] \longrightarrow e^{-t}, \quad as \quad N \to \infty.$$
(5.26)

### 5.2. Induced measure and free energy landscape

The first important ingredient to derive sharp estimates for mean hitting times is a precise understanding of the induced measure  $Q^n$ . The infinite-volume equilibrium (Gibbs) measures of the Curie-Weiss model of Ising (q = 2) or Potts ( $q \ge 3$ ) type in the presents of a random field and the particular influence of the disorder on the typical behavior of Gibbs measure in large but finite volumes was analyzed in detail in [1, 79, 70]. In this section, our first aim is to show that the mesoscopic variable,  $\varrho^n$ , satisfies a *sharp large deviation principle* under the Gibbs measure with rate function  $F^n$ . Next, we study the behavior of the induced random measure  $Q^n$  in the neighborhood of critical points of  $F^n$ .

## **5.2.1. Sharp large deviation principle.** Let us begin by writing for $x \in \Gamma^n$

$$Z_{N} \mathcal{Q}^{n}(\boldsymbol{x}) = \exp(\beta N E(\boldsymbol{x})) \prod_{k=1}^{n} \sum_{\sigma \in \mathcal{S}_{|\Lambda_{k}|}} \mathbb{1}_{\varrho^{k}(\sigma) = x^{k}} \exp\left(\beta \sum_{i \in \Lambda_{k}} \langle \tilde{h}^{i}, e^{\sigma_{i}} \rangle\right) q^{-|\Lambda_{k}|}$$
  
=:  $\exp(\beta N E(\boldsymbol{x})) \prod_{k=1}^{n} Z_{|\Lambda_{k}|}^{\tilde{h}} \mathbf{E}_{\sigma[|\Lambda_{k}|]}^{\tilde{h}} \Big[ \mathbb{1}_{\varrho_{|\Lambda_{k}|} = x^{k}/\pi_{k}} \Big],$  (5.27)

where  $Z_{|\Lambda_k|}^h$  is the normalization constant of the  $\tilde{h}$ -tilted measure. It remains to control the probability with respect to the  $\tilde{h}$ -tilted measure of the event,  $\{\varrho_{|\Lambda_k|} = x\}$ , that the macroscopic variable  $\varrho_{|\Lambda_k|}$  takes a given value  $x \in \Gamma_{|\Lambda_k|} \subset \mathcal{M}_1$  in the limit when  $N \to \infty$ .

For this purpose, we begin with recalling some well-known properties of the log-moment generating function,  $U_{|\Lambda_k|}$ , and its Legendre–Fenchel transform,  $I_{|\Lambda_k|}$ , taken from [46, 103]. As a consequence of Hölder's inequality,  $U_{|\Lambda_k|}$  is a convex function on  $\mathbb{R}^q$ . This implies, that  $I_{|\Lambda_k|}$  is a proper convex function and lower semi-continuous. Moreover,  $U_{|\Lambda_k|}$  is infinitely differentiable on  $\mathbb{R}^q$  and the range of its gradient map coincides with the *relative interior* of  $\mathcal{M}_1$  that we denote by  $\operatorname{ri}(\mathcal{M}_1)$ .

We recall that for a convex subset  $C \subseteq \mathbb{R}^q$  the *affine hull* of C,  $\operatorname{aff}(C)$ , is defined to be the intersection of all affine sets which contain C. This concept allows us to define  $\operatorname{ri}(C)$  as the set of all  $x \in C$  for which there exists an  $\varepsilon > 0$  such that  $B_{\varepsilon}(x) \cap \operatorname{aff}(C) \subset C$ . Here  $B_{\varepsilon}(x)$  denotes an open ball in  $\mathbb{R}^q$  around x of radius  $\varepsilon$ .

As an application of [46, Theorem VI.5.7],

$$\operatorname{ri}(\operatorname{dom}(I_{|\Lambda_k|})) = \operatorname{ri}(\mathcal{M}_1) \subseteq \operatorname{dom}(I_{|\Lambda_k|}) \subseteq \mathcal{M}_1,$$

while for all  $x \in \mathbb{R}^q \setminus \mathcal{M}_1$  the entropy  $I_{|\Lambda_k|}(x) = \infty$ . Since  $U_{|\Lambda_k|}$  is smooth,  $I_{|\Lambda_k|}$  is strictly convex on every convex subset of  $\operatorname{ri}(\mathcal{M}_1)$ . Further, notice that for all  $r \in S_0$ 

$$\frac{\partial U_{|\Lambda_k|}}{\partial t_r}(t) = \frac{1}{|\Lambda_k|} \sum_{i \in \Lambda_k} \frac{\exp(t_r + \beta h_r^i)}{\sum_{s \in \mathcal{S}_0} \exp(t_s + \beta \tilde{h}_s^i)} =: \frac{1}{|\Lambda_k|} \sum_{i \in \Lambda_k} u_r^{i,k}(t).$$
(5.28)

Thus, the gradient map  $\nabla U_{|\Lambda_k|}$  is *not* one-to-one on  $\mathbb{R}^q$  because for any  $x \in \operatorname{ri}(\mathcal{M}_1)$  and  $t^* \equiv t^*(x) \in \mathbb{R}^q$  such that  $x = \nabla U_{|\Lambda_k|}(t^*)$ , the one-parameter family  $t^*_{\gamma} := t^* + \gamma 1$  satisfies as well  $x = \nabla U_{|\Lambda_k|}(t^*_{\gamma})$  for all  $\gamma \in \mathbb{R}$ . Nevertheless, since

$$\langle x,t^*\rangle-U_{|\Lambda_k|}(t^*)\ =\ \langle x,t^*_\gamma\rangle-U_{|\Lambda_k|}(t^*_\gamma)\qquad\forall\,\gamma\in\mathbb{R}$$

the entropy is well defined, see also [103, Theorem 26.4], and given by the formula

$$I_{|\Lambda_k|}(x) = \langle x, t^* \rangle - U_{|\Lambda_k|}(t^*).$$
(5.29)

Note that, due to [103, Theorem 23.5], it holds that

$$t^* \in \partial I_{|\Lambda_k|}(x) \qquad \Longleftrightarrow \qquad x = \nabla U_{|\Lambda_k|}(t^*)$$
 (5.30)

where  $\partial I_{|\Lambda_k|}(x)$  denotes the subdifferential of  $I_{|\Lambda_k|}$  at x which is defined through

$$\partial I_{|\Lambda_k|}(x) := \{ t \in \mathbb{R}^q \mid I_{|\Lambda_k|}(z) \ge I_{|\Lambda_k|}(x) + \langle t, z - x \rangle, \ \forall z \in \mathbb{R}^q \}.$$

For obvious reasons,  $\partial I_{|\Lambda_k|}(x)$  is an affine subspace of  $\mathbb{R}^q$  with difference space  $V_0$ , where we decompose  $\mathbb{R}^q$  into a direct sum of

$$V_0 = \text{span}\{1\}$$
 and  $V_1 = V_0^{\perp}$ . (5.31)

Moreover, a simple computation shows, see Lemma 5.6, that the log-moment generating function  $U_{|\Lambda_k|}(t)$  is strict convex in the direction orthogonal the  $V_0$  and the leading principal minor, which we denote by  $[\text{Hess } U_{|\Lambda_k|}(t)]_q$ , is positive definite. This implies that the entropy  $I_{|\Lambda_k|}$ with respect to  $\operatorname{aff}(\mathcal{M}_1)$  is essentially smooth, i.e.

- (a)  $I_{|\Lambda_k|}$  is differentiable throughout  $\operatorname{ri}(\mathcal{M}_1)$  and
- (b) for any sequence  $\{x^i\}$  in  $\operatorname{ri}(\mathcal{M}_1)$  which converge to a boundary point  $x \in \mathcal{M}_1$ ,  $\|\nabla I_{|\Lambda_k|}(x^i)\| \to \infty$ .

**Lemma 5.6.** For  $t \in \mathbb{R}^q$  set  $x = \nabla U_{|\Lambda_k|}(t) \in \mathcal{M}_1$ . Then the matrix  $\operatorname{Hess} U_{|\Lambda_k|}(t)$  has exactly one eigenvalue zero in the direction  $1 \in \mathbb{R}^q$ . Moreover, there exists  $c \equiv c(t) > 0$  such that the eigenvalues corresponding to  $\operatorname{span}\{1\}^{\perp}$ , as well as, the eigenvalues of  $[\operatorname{Hess} U_{|\Lambda_k|}(t)]_q$  are bounded from below by c.

*Proof.* In view of (5.28), the symmetric matrix  $B \equiv \text{Hess } U_{|\Lambda_k|}(t)$  can be written as the difference between a diagonal matrix and a sum of rang-1 matrices, i.e.

$$B = \frac{1}{|\Lambda_k|} \sum_{i \in \Lambda_k} \operatorname{diag}(u_1^{i,k}, \dots, u_q^{i,k}) - u^{i,k} \cdot (u^{i,k})^{\mathrm{T}} = \operatorname{diag}(x_1, \dots, x_q) - \frac{1}{|\Lambda_k|} \sum_{i \in \Lambda_k} u^{i,k} \cdot (u^{i,k})^{\mathrm{T}}$$

where  $u^{i,k} \equiv u^{i,k}(t^*)$ . Since  $U_{|\Lambda_k|}$  is a convex function, the matrix B is positive semi-definite. Moreover, it is easy to check that 1 is an eigenvector to the eigenvalue 0. Let us order the eigenvalues of B such that  $\lambda_1(B) \leq \lambda_2(B) \leq \ldots \leq \lambda_q(B)$ . Using the interlacing property of symmetric matrices, we get

$$0 = \lambda_1(B) \le \lambda_1(A) \le \lambda_2(B) \le \dots \le \lambda_{q-1}(B) \le \lambda_{q-1}(A) \le \lambda_q(B)$$
(5.32)

where  $A \equiv [\text{Hess } U_{|\Lambda_k|}(t)]_{q}$ . By applying Gershgorin's Theorem, we obtain that

$$\operatorname{spec}(A) \subset \bigcup_{r \in \mathcal{S}_0} \left[ A_{rr} - R_r, A_{rr} + R_r \right]$$

where

$$R_r := \sum_{\substack{s=1\\s\neq r}}^{q-1} |A_{rs}| = \frac{1}{|\Lambda_k|} \sum_{i \in \Lambda_k} u_r^{i,k} \sum_{\substack{s=1\\s\neq r}}^{q-1} u_s^{i,k} = x_r - \frac{1}{|\Lambda_k|} \sum_{i \in \Lambda_k} u_r^{i,k} \left( u_q^{i,k} + u_r^{i,k} \right).$$

Since diam( $\mathcal{H}_k$ )  $\leq \varepsilon(n)$  and by (5.28), there exists a  $\delta(t) > 0$  such that the minimal component of x is bounded from below by  $\delta(t)$ . Hence,

$$A_{rr} - R_r = \frac{1}{|\Lambda_k|} \sum_{i \in \Lambda_k} u_r^{i,k} u_q^{i,k} = \frac{1}{|\Lambda_k|^2} \sum_{i,j \in \Lambda_k} u_r^{i,k} u_q^{j,k} \frac{u_q^{i,k}}{u_q^{j,k}} \ge x_r x_q e^{-2\beta\varepsilon(n)} \ge c > 0$$

where we set  $c := \delta(t)^2 e^{-2\beta \varepsilon(n)}$ .

Now, we can state the sharp large deviation estimate for the *h*-tilted expectation.

**Proposition 5.7.** Let  $y \in \operatorname{ri}(\mathcal{M}_1)$  and consider a sequence  $\{x^{|\Lambda_k|}\}$  such that  $x^{|\Lambda_k|} \in \Gamma_{|\Lambda_k|}$  and  $\|x^{|\Lambda_k|} - y\|_2 < \frac{1}{|\Lambda_k|}$ . Further define  $t^{|\Lambda_k|} \in V_1 \subset \mathbb{R}^q$  through  $x^{|\Lambda_k|} = \nabla U_{|\Lambda_k|}(t^{|\Lambda_k|})$  if such a  $t^{|\Lambda_k|}$  exists while otherwise  $\|t^{|\Lambda_k|}\|_2 = \infty$ . Then, for all but finitely many N,

$$Z_{|\Lambda_k|}^{\tilde{h}} \mathbf{E}_{\sigma[|\Lambda_k|]}^{\tilde{h}} \left[ \mathbb{1}_{\varrho_{|\Lambda_k|} = x^{|\Lambda_k|}} \right] = \frac{\exp\left(-|\Lambda_k| I_{|\Lambda_k|}(x^{|\Lambda_k|})\right)}{\sqrt{\left(2\pi|\Lambda_k|\right)^{q-1} \det\left[\operatorname{Hess} U_{|\Lambda_k|}(t^{|\Lambda_k|})\right]_q}} \left(1 + \mathcal{O}_N(1)\right), \quad (5.33)$$

where  $\mathcal{O}_N(1) \to 0$  as  $N \to \infty$ .

*Proof.* Sharp large deviation estimates were obtained in [**26**, **39**] for random vectors and in the Ising model. The proof of (5.33) modifies their strategy to the case, where the Hessian is degenerate.

Recall that,  $\mathbb{P}$ -a.s.,  $|\Lambda_k| = p_k (1 + \mathcal{O}_N(1)) N$ . Now, consider for any  $t \in \mathbb{R}^q$  and  $x \in \Gamma_{|\Lambda_k|}$ 

$$\begin{split} Z_{|\Lambda_k|}^{\tilde{h}} \ \mathbf{E}_{\sigma[|\Lambda_k|]}^{\tilde{h}} \Big[ \mathbbm{1}_{\varrho|\Lambda_k|} = x \Big] \ &= \ \mathrm{e}^{-|\Lambda_k| \left( \langle x,t \rangle - U_{|\Lambda_k|}(t) \right)} \ \mathbf{E}_{\sigma[|\Lambda_k|]}^{\tilde{h}} \Big[ \mathbbm{1}_{\varrho|\Lambda_k|} = x \ \frac{\mathrm{e}^{|\Lambda_k| \langle t,\varrho|\Lambda_k| \rangle}}{\mathbf{E}_{\sigma[|\Lambda_k|]}^{\tilde{h}} \Big[ \mathrm{e}^{|\Lambda_k||} \Big[ \mathrm{e}^{|\Lambda_k| \langle t,\varrho|\Lambda_k| \rangle} \Big] \Big] \\ &= \ \mathrm{e}^{-|\Lambda_k| \left( \langle x,t \rangle - U_{|\Lambda_k|}(t) \right)} \ \mathbf{\widehat{E}}_{\sigma[|\Lambda_k|]}^t \Big[ \mathbbm{1}_{\varrho|\Lambda_k|} = x \Big]. \end{split}$$

Since  $y \in \operatorname{ri}(\mathcal{M}_1)$ , there exists  $\varepsilon > 0$  such that  $B_{\varepsilon}(y) \cap \operatorname{aff}(\mathcal{M}_1) \subset \operatorname{ri}(\mathcal{M}_1)$ . By choosing N large enough, we can ensure that  $||x^{|\Lambda_k|} - y||_2 < \varepsilon/2$ . This implies that  $||t^{|\Lambda_k|}||_2 < \infty$ . Setting  $t = t^{|\Lambda_k|}$  and using (5.29), we obtain that

$$\mathbf{E}_{\sigma[|\Lambda_k|]}^{\tilde{h}} \Big[ \mathbb{1}_{\varrho_{|\Lambda_k|} = x^{|\Lambda_k|}} \Big] = \mathrm{e}^{-|\Lambda_k| I_{|\Lambda_k|}(x^{|\Lambda_k|})} \, \widehat{\mathbf{E}}_{\sigma[|\Lambda_k|]}^{t^{|\Lambda_k|}} \Big[ \mathbb{1}_{\varrho_{|\Lambda_k|} = x^{|\Lambda_k|}} \Big]. \tag{5.34}$$

On the other hand, for any  $x \in \Gamma_{|\Lambda_k|}$ , it holds

$$\mathbb{1}_{\varrho|\Lambda_k|(\sigma)=x} = \prod_{r=1}^{q-1} \mathbb{1}_{\varrho|\Lambda_k|(\sigma)_r=x_r} = \frac{1}{(2\pi)^{q-1}} \int_Q e^{i|\Lambda_k|\langle \varrho|_{\Lambda_k|(\sigma)-x, \binom{z}{0}}\rangle} dz,$$

where  $Q = [-\pi, \pi]^{q-1}$ . Hence,

$$\widehat{\mathbf{E}}_{\sigma[|\Lambda_k|]}^{t^{|\Lambda_k|}} \Big[ \mathbb{1}_{\varrho_{|\Lambda_k|} = x^{|\Lambda_k|}} \Big] = \frac{1}{(2\pi)^{q-1}} \int_Q \exp(|\Lambda_k| G_{|\Lambda_k|}(z)) \,\mathrm{d}z$$
(5.35)

where

$$G_{|\Lambda_k|}(z) = U_{|\Lambda_k|}(t^{|\Lambda_k|} + \mathrm{i} \begin{pmatrix} z \\ 0 \end{pmatrix}) - U_{|\Lambda_k|}(t^{|\Lambda_k|}) - \mathrm{i} \langle x^{|\Lambda_k|}, \begin{pmatrix} z \\ 0 \end{pmatrix}\rangle, \qquad z \in \mathbb{R}^{q-1}.$$

Now, set  $\delta = \pi/4$ . For a given  $\gamma_{|\Lambda_k|} := \sqrt{K_{\varepsilon} \ln |\Lambda_k|/|\Lambda_k|}$ , where  $K_{\varepsilon}$  will be defined below, there exists  $N_{\varepsilon}$  such that  $\gamma_{|\Lambda_k|} < \delta/2$  and  $||x^{|\Lambda_k|} - y||_2 < \varepsilon/2$  for all  $|\Lambda_k| > N_{\varepsilon}$ . Moreover, we rewrite (5.35) as

$$\widehat{\mathbf{E}}_{\sigma[|\Lambda_{k}|]}^{t^{|\Lambda_{k}|}} \Big[ \mathbb{1}_{\varrho_{|\Lambda_{k}|} = x^{|\Lambda_{k}|}} \Big] = \frac{1}{(2\pi)^{q-1}} \int_{Q} e^{|\Lambda_{k}| G_{|\Lambda_{k}|}(z)} \, \mathbb{1}_{||z||_{2} \leq \gamma_{|\Lambda_{k}|}} \, \mathrm{d}z + \frac{1}{(2\pi)^{q-1}} \int_{Q} e^{|\Lambda_{k}| G_{|\Lambda_{k}|}(z)} \, \mathbb{1}_{||z||_{2} > \gamma_{|\Lambda_{k}|}} \, \mathrm{d}z.$$
(5.36)

We start with studying the asymptotics of the first integral. Since  $U_{|\Lambda_k|}$  is holomorphic, the following expansion is valid for all  $||z||_2 < \delta$ 

$$G_{|\Lambda_{k}|}(z) = i \langle \nabla U_{|\Lambda_{k}|}(t^{|\Lambda_{k}|}), {\binom{z}{0}} \rangle - \frac{1}{2} \langle z, A z \rangle + R_{3} \left( t^{|\Lambda_{k}|} + i {\binom{z}{0}} \right) - i \langle x^{|\Lambda_{k}|}, {\binom{z}{0}} \rangle$$
  
$$= -\frac{1}{2} \langle z, A z \rangle + R_{3} \left( t^{|\Lambda_{k}|} + i {\binom{z}{0}} \right), \qquad (5.37)$$

where  $A := \left[\operatorname{Hess} U_{|\Lambda_k|}(t^{|\Lambda_k|})\right]_q$  and  $R_3(t^{|\Lambda_k|} + i\binom{z}{0}) = \sum_{|\alpha| \ge 3} a_{\alpha}^{|\Lambda_k|}(iz)^{\alpha}$ . Notice that the coefficients in the remainder term can be bounded from above by means of Cauchy's theorem,

$$\left|a_{\alpha}^{|\Lambda_{k}|}\right| \leq \frac{1}{\delta^{\alpha}} \sup_{\xi \in \operatorname{cl}(P_{\delta}(t^{|\Lambda_{k}|}))} |U_{|\Lambda_{k}|}(\xi)|.$$
(5.38)

Here, we denote by  $P_{\delta}(t) := \{z \in \mathbb{C}^q \mid ||t - z||_{\infty} < \delta\} \subset \mathbb{C}^q$  the polydisc with center  $t \in \mathbb{R}^q$  and radius  $\delta$ . Thus, we are left with the task to show that, for all  $\xi \in cl(P_{\delta}(t^{|\Lambda_k|}))$ , the supremum of  $|U_{|\Lambda_k|}(\xi)|$  is uniformly bounded from above for all  $|\Lambda_k| > N_{\varepsilon}$ . Since  $\delta = \pi/4$ , it holds for all  $\xi \in \operatorname{cl}(P_{\delta}(t^{|\Lambda_k|}))$  that  $|\Im(U_{|\Lambda_k|}(\xi))| \leq \delta$ . Note that for  $\xi^1 + \mathrm{i}\,\xi^2 \in \mathbb{C}^q$ ,

$$\Re (U_{|\Lambda_k|}(\xi^1 + i\xi^2)) - U_{|\Lambda_k|}(\xi^1) = \frac{1}{2|\Lambda_k|} \sum_{i \in \Lambda_k} \ln \left( \sum_{r,s \in \mathcal{S}_0} u_r^{i,k}(\xi^1) u_s^{i,k}(\xi^1) \cos \left(\xi_r^2 - \xi_s^2\right) \right),$$
(5.39)

which implies that  $|\Re(U_{|\Lambda_k|}(\xi))| \leq U_{|\Lambda_k|}(t^{|\Lambda_k|}) + \delta$  for all  $\xi \in cl(P_{\delta}(t^{|\Lambda_k|}))$ . Recall that  $u^i(t)$  is defined in (5.28). Thus,  $|U_{|\Lambda_k|}(\xi)| \leq |\Re(U_{|\Lambda_k|}(\xi))| + |\Im(U_{|\Lambda_k|}(\xi))| \leq U_{|\Lambda_k|}(t^{|\Lambda_k|}) + 2\delta$ . It remains to control  $U_{|\Lambda_k|}(t^{|\Lambda_k|})$ . However, since

$$x^{|\Lambda_k|} = \nabla U_{|\Lambda_k|}(t^{|\Lambda_k|}) = \frac{1}{|\Lambda_k|} \sum_{i \in \Lambda_k} u^{i,k}(t^{|\Lambda_k|}) \le u^{j,k}(t^{|\Lambda_k|}) e^{2\beta\varepsilon(n)} \quad \forall j \in \Lambda_k.$$

and, by rewriting  $U_{|\Lambda_k|}$ , we obtain

$$U_{|\Lambda_{k}|}(t^{|\Lambda_{k}|}) = -\frac{1}{q} \sum_{r \in S_{0}} \frac{1}{|\Lambda_{k}|} \sum_{j \in \Lambda_{k}} \ln\left(u_{r}^{j,k}(t^{|\Lambda_{k}|})\right)$$
  
$$\leq -\frac{1}{q} \sum_{r \in S_{0}} \ln x_{r}^{|\Lambda_{k}|} + 2\beta\varepsilon(n) \leq \ln\frac{2}{\varepsilon} + 2\beta\varepsilon(n), \qquad (5.40)$$

where we used that  $t^{|\Lambda_k|} \perp 1$ , as well as,  $h^i \perp 1$  for all  $i \in \Lambda_k$  and that by construction the minimal component of  $x^{|\Lambda_k|}$  is at least larger than  $\varepsilon/2$ . Thus, we can conclude that there exists  $M_{\varepsilon} < \infty$  such that, for all  $|\Lambda_k| > N_{\varepsilon}$ ,  $\sup_{\xi \in \operatorname{cl}(P_{\delta}(t^{|\Lambda_k|}))} |U_{|\Lambda_k|}(\xi)| \le M_{\varepsilon}$ . Since  $||z||_2 \le \gamma_{|\Lambda_k|} < \delta/2$  for all  $|\Lambda_k| > N_{\varepsilon}$ , the remainder can be bounded from above by

$$\left| R_3(t_N^* + \mathrm{i} \begin{pmatrix} z \\ 0 \end{pmatrix}) \right| \leq M_{\varepsilon} \frac{\|z\|_2^3}{\delta^3} \sum_{k=0}^{\infty} \left( \frac{\|z\|_2}{\delta} \right)^k \leq 2 M_{\varepsilon} \frac{\|z\|_2^3}{\delta^3}.$$
(5.41)

Hence, (5.37) together with the Lemma 5.6 implies

$$\int_{\mathbb{R}^{q-1}} e^{|\Lambda_k| G_{|\Lambda_k|}(z)} \mathbb{1}_{\|z\|_2 \le \gamma_{|\Lambda_k|}} dz = |\Lambda_k|^{-\frac{q-1}{2}} \int_{\mathbb{R}^{q-1}} e^{-\frac{1}{2} \langle z, A z \rangle} \mathbb{1}_{\|z\|_2 \le \sqrt{K_{\varepsilon} \ln |\Lambda_k|}} dz \left(1 + \mathcal{O}_N(1)\right)$$
$$= \frac{\left(2\pi\right)^{q-1}}{\sqrt{\left(2\pi |\Lambda_k|\right)^{q-1} \det\left[\operatorname{Hess} U_{|\Lambda_k|}(t^{|\Lambda_k|})\right]_q}} \left(1 + \mathcal{O}_N(1)\right).$$
(5.42)

In order to complete the proof, it remains to show that the contribution to (5.33) coming from the second integral in (5.36) is negligible, i.e. that the following *locality condition* holds

$$\left|\Lambda_{k}\right|^{\frac{q-1}{2}} \int_{Q} \left| \mathrm{e}^{\left|\Lambda_{k}\right| G_{\left|\Lambda_{k}\right|}(z)} \right| \, \mathbb{1}_{\left|\left|z\right|\right|_{2} > \gamma_{\left|\Lambda_{k}\right|}} \, \mathrm{d}z = \mathcal{O}_{N}(1).$$

By exploiting the representation (5.39), we obtain that

$$\max_{\substack{z \in Q \\ \|z\|_{2} > \gamma_{|\Lambda_{k}|}}} \Re \left( G_{|\Lambda_{k}|}(z) \right) \leq \frac{1}{2 |\Lambda_{k}|} \sum_{i \in \Lambda_{k}} \ln \left( 1 - \frac{4}{\pi^{2}} u_{q}^{i,k}(t^{|\Lambda_{k}|}) \sum_{r=1}^{q-1} u_{r}^{i,k}(t^{|\Lambda_{k}|}) z_{r}^{2} \right)$$

$$\leq -\frac{2}{\pi^{2}} e^{-2\beta\varepsilon(n)} x_{q}^{|\Lambda_{k}|} \sum_{r=1}^{q-1} x_{r}^{|\Lambda_{k}|} z_{r}^{2}$$

$$\leq -\frac{\varepsilon^{2}}{2\pi^{2}} e^{-2\beta\varepsilon(n)} \gamma_{|\Lambda_{k}|}^{2}.$$
(5.43)

In the first step, we use that  $1 - \cos x \ge 2 x^2 / \pi^2$  for all  $|x| \le \pi$ . In the second step, we took advantage of (5.28) and the fact that  $e^{-2\beta\varepsilon(n)} \le u_r^{i,k}(t)/u_r^{j,k}(t) \le e^{2\beta\varepsilon(n)}$  for all  $i, j \in \Lambda_k$  and  $r \in S_0$ . In the last step we used that by construction the minimal component of  $x^{|\Lambda_k|}$  is at least larger than  $\varepsilon/2$ . Thus, by choosing  $K_{\varepsilon} = \frac{\pi^2}{\varepsilon^2} e^{2\beta\varepsilon(n)} q$ , we get

$$\left|\Lambda_{k}\right|^{\frac{q-1}{2}} \int_{Q} \left| e^{|\Lambda_{k}| G_{|\Lambda_{k}|}(z)} \right| \, \mathbb{1}_{\|z\|_{2} > \gamma_{|\Lambda_{k}|}} \, \mathrm{d}z \, \leq \, (2\pi)^{q-1} \, |\Lambda_{k}|^{-1/2}.$$
(5.44)

Thus, by combing (5.42) and (5.44), we conclude the proof.

For  $\boldsymbol{y} \in \operatorname{ri}(\mathcal{M}_1)^{\otimes n}$  and N large enough, let us consider  $\boldsymbol{x} \in \Gamma^n$  with  $\|\boldsymbol{x}^k/\pi_k - \boldsymbol{y}^k\|_2 < \varepsilon/2$ where  $\varepsilon > 0$  is chosen in such a way that  $B_{\varepsilon}(\boldsymbol{y}^k) \cap \operatorname{aff}(\mathcal{M}_1) \subset \operatorname{ri}(\mathcal{M}_1)$  for all  $k = 1, \ldots, n$ . Then, in view of (5.27) and (5.33), the induced measure  $\mathcal{Q}^n(\boldsymbol{x})$  can be expressed as

$$Z_N \mathcal{Q}^n(\boldsymbol{x}) = \frac{\exp(-\beta N F^n(\boldsymbol{x}))}{\prod\limits_{k=1}^n (2\pi N)^{\frac{q-1}{2}} \sqrt{\det\left[\pi_k \operatorname{Hess} U_{|\Lambda_k|}(t^*(x^k/\pi_k))\right]_q}} \left(1 + \mathcal{O}_N(1)\right).$$
(5.45)

Let us point out that the representation (5.45) of  $Q^n$  is in particular valid for all critical points of the free energy,  $F^n$ , since the essential smoothness of  $I_{|\Lambda_k|}$ , for all k = 1, ..., n, implies that these critical points are in the relative interior of  $X_{k=1}^n \mathcal{M}_{\pi_k} \equiv M$ .

**5.2.2. Free energy landscape near the critical point.** In what follows, we will precisely compute the behavior of the measure  $Q^n$  in the neighborhood of critical points of  $F^n$ . First of all notice that the free energy is essential smooth and hence differentiable throughout  $\operatorname{ri}(M)$ . Thus,  $x \in \operatorname{ri}(M)$  is a critical point if and only if  $dF^n(x) = 0 \in T^*_x M$ . In this particular case, the tangent space  $T_x M$  coincides for all  $x \in M$  with the subspace  $V_1$  where we decompose  $\mathbb{R}^{n\cdot q}$  into a direct sum of

$$\boldsymbol{V}_0 := \operatorname{span}\{e^1 \otimes 1, \dots, e^n \otimes 1\} \quad \text{and} \quad \boldsymbol{V}_1 := \boldsymbol{V}_0^{\perp}.$$
(5.46)

Since *M* is embedded in the Euclidean space  $\mathbb{R}^{n \cdot q}$ , at a critical point  $x \in M$  there exists a  $w \equiv w(x) \in \mathbb{R}^{n \cdot q}$  such that  $0 = dF^n(x)(v) = \langle w, v \rangle = \sum_{k=1}^n \langle w^k, v^k \rangle$  for all  $v \in V_1$ . This implies that  $w^k \in \text{span}\{1\} = V_0$ . Recall that, for all k, the subdifferential of  $I_{|\Lambda_k|}$  is a affine subspace of  $\mathbb{R}^q$  with difference space  $V_0$ . Hence, a critical point is characterized by the solution of the system of equations

span{1} 
$$\ni -2 \sum_{k=1}^{n} x^{k} - \bar{h}^{k} + \frac{1}{\beta} \partial I_{|\Lambda_{k}|}(x^{k}/\pi_{k}), \quad \forall k = 1, \dots, n$$
 (5.47)

Hence, for any  $t^*(x^k/\pi_k) \in \partial I_{|\Lambda_k|}$  and  $a \in \mathbb{R}$ 

$$\frac{1}{\beta}t^*(x^k/\pi_k) = 2\sum_{k=1}^n x^k + \bar{h}^k + \gamma 1, \qquad \forall k = 1, \dots, n.$$
(5.48)

Making again use of the fact that  $\nabla U_{|\Lambda_k|}(t^*(x^k/\pi_k)) = x^k/\pi_k$ , (5.47) is equivalent to

$$x^{k} = \pi_{k} \nabla U_{|\Lambda_{k}|} (2\beta x + \beta \bar{h}^{k}), \quad \forall k = 1, \dots, n \quad \text{with} \quad x = \sum_{k=1}^{n} x^{k}, \quad (5.49)$$

where we set for convenience  $\gamma = 0$  since, by (5.28), the gradient  $\nabla U_{|\Lambda_k|}$  is constant in direction  $1 \in \mathbb{R}^q$ . By summing over k, we see that solutions of the system of equations (5.49) are generated by the solutions of the equation

$$x = \sum_{k=1}^{n} \pi_k \nabla U_{|\Lambda_k|} \left( 2\beta x + \beta \bar{h}^k \right)$$
(5.50)

Let us denote a solution of (5.49) and (5.50) by z and z, respectively. The observation that a critical point z of  $F^n$  is already determined by the solution of the lower dimensional problem (5.50) is not a coincidence.

As we will show below, the structure of the free energy landscape,  $F^n$ , is closely related to a q-1-dimensional landscape which is given to leading order by the rate function,  $F_N : \mathbb{R}^q \to \mathbb{R}$ ,

$$x \mapsto F_N(x) := -\|x\|^2 + \frac{1}{\beta} I_N(x) = -\|x\|^2 + \frac{1}{\beta} \sup_{t \in \mathbb{R}^q} \left( \langle t, x \rangle - U_N(t) \right)$$

of  $\frac{1}{\beta N} \ln \mathcal{Q}_N(x)$  where  $U_N \colon \mathbb{R}^q \to \mathbb{R}$  is the log-moment generating function defined by

$$t \mapsto U_N(t) := \frac{1}{N} \ln \mathbf{E}^h_{\sigma[N]} \Big[ \exp(N \langle t, \varrho_N \rangle) \Big] + \ln Z_N^h = \frac{1}{N} \sum_{i=1}^N \ln \left( \sum_{r \in \mathcal{S}_0} \frac{1}{q} \exp(\beta h_r^i + t_r) \right)$$

As a consequence of the definition of  $U_{|\Lambda_k|}$ , given in (5.17), and  $U_N$ , the equation (5.50) is equivalent to  $x = \nabla U_N(2\beta x)$ , i.e. it has the pleasant feature to be independent of the choice of the coarse graining. In the sequel we call  $F_N$  the macroscopic free energy.

In the following lemma, we collect some fact about the connection between  $F^n$  and  $F_N$ .

**Lemma 5.8.** The functions  $F_N$  and  $F^n$  are related in the following ways.

- (i) If z is a critical point of  $F^n$ , then  $z \equiv z(z) = \sum_{k=1}^n z^k$  is a critical point of  $F_N$ .
- (ii) If z is a critical point of  $F_N$ , then  $z \equiv z(z)$  with components  $z^k = \pi_k \nabla U_{|\Lambda_k|} (2\beta z + \beta \bar{h}^k)$  is a critical point of  $F^n$
- (iii) At a critical point z, it holds that

$$F^{n}(\boldsymbol{z}) = F_{N}(\boldsymbol{z}(\boldsymbol{z})) = \|\boldsymbol{z}(\boldsymbol{z})\|^{2} - \frac{1}{\beta N} \sum_{i=1}^{N} \ln\left(\sum_{r \in \mathcal{S}_{0}} \frac{1}{q} \exp(2\beta z_{r}(\boldsymbol{z}) + \beta h_{r}^{i})\right).$$
(5.51)

(iv) For any  $x \in ri(\mathcal{M}_1)$ ,

$$F_N(x) = \inf_{\substack{\boldsymbol{x} \in \times_{k=1}^n \mathcal{M}_{\pi_k}:\\ \sum_{k=1}^n x^k = x}} F^n(\boldsymbol{x}).$$
(5.52)

*Proof.* By a computation analog to the one presented in the lines (5.47) – (5.49), a critical point of  $F_N$  is determined by the solution of  $x = \nabla U_N(2\beta x)$ . Hence, (i) and (ii) are an immediate consequence of (5.49) and (5.50). In order to prove (iii), notice that (5.48) and (5.29) implies that

$$I_{|\Lambda_k|}(z^k/\pi_k) = \frac{\beta}{\pi_k} \langle 2z + \bar{h}^k, z^k \rangle - U_{|\Lambda_k|} (2\beta z + \beta \bar{h}^k), \qquad \forall k = 1, \dots, n$$

where  $z = \sum_{k=1}^{n} z^{k}$ . By combining this expression with (5.15), we obtain that

$$F^{n}(z) = -\|z\|^{2} - \sum_{k=1}^{n} \langle \bar{h}^{k}, z^{k} \rangle + \sum_{k=1}^{n} \left( \langle 2z + \bar{h}^{k}, z^{k} \rangle - \frac{\pi_{k}}{\beta} U_{|\Lambda_{k}|} (2\beta z + \beta \bar{h}^{k}) \right)$$
  
$$= \|z\|^{2} - \frac{1}{\beta} \sum_{k=1}^{n} \pi_{k} U_{|\Lambda_{k}|} (2\beta z + \beta \bar{h}^{k})$$
  
$$= \|z\|^{2} - \frac{1}{\beta N} \sum_{i=1}^{N} \ln \left( \sum_{r \in \mathcal{S}_{0}} \frac{1}{q} \exp(2\beta z_{r} + \beta h_{r}^{i}) \right).$$
(5.53)

On the other hand, notice that  $z = \nabla U_N(2\beta z)$  implies that  $t^*(z) = 2\beta z + \gamma 1$ . In view of (5.29), we have that  $I_N(z) = 2\beta ||z||^2 - U_N(2\beta z)$  which concludes the proof of (iii). Analog to the considerations leading to (5.49), for a given  $x \in \operatorname{ri}(\mathcal{M}_1)$  the minimum on the set  $\{x \in X_{k=1}^n \mathcal{M}_{\pi_k} \mid \sum_{k=1}^n x^k = x\}$  is attained at x(x) determined by the system of equations

$$x^{k}(x) = \pi_{k} \nabla U_{|\Lambda_{k}|} \Big( \beta \left( 2x + \bar{h}^{k} + \lambda \xi \right) \Big), \qquad \forall k = 1, \dots, n,$$
(5.54)

with  $\xi := \frac{1}{\sqrt{q}} (q e^q - 1) \in \mathbb{R}^q$ . By summing over all k, we obtain that the Lagrange multiplier,  $\lambda \equiv \lambda(x)$ , solves

$$x = \sum_{k=1}^{n} \pi_k \nabla U_{|\Lambda_k|} \Big( \beta \left( 2x + \bar{h}^k + \lambda \xi \right) \Big) = \nabla U_N \big( \beta \left( 2x + \lambda \xi \right) \big).$$
(5.55)

By convex duality (5.30), we have that  $t^*(x^k(x)/\pi_k) = \beta (2x + \bar{h}^k + \lambda \xi)$  for all k = 1, ..., n and  $t^*(x) = \beta (2x + \lambda \xi)$ . Thus, an computation analog to (5.51) reveals that

$$F^{n}(\boldsymbol{x}(x)) = -\|\boldsymbol{x}\|^{2} + \frac{1}{\beta} \left( \langle \beta(2x+\lambda\xi), x \rangle - U_{N}(\beta(2x+\lambda\xi)) \right) = F_{N}(x)$$

which concludes the proof.

**Remark 5.9.** By the strong law of large numbers, the set of critical points of  $F_N$  converge,  $\mathbb{P}_h$ -a.s., to the set of solutions of the equation

$$z_r^* = \mathbb{E}\left[\frac{\exp(2\beta z_r^* + \beta h_r^1)}{\sum_{s \in S_0} \exp(2\beta z_s^* + \beta h_s^1)}\right] \quad \forall r \in S_0,$$
(5.56)

i.e. in some sense the Curie-Weiss-Potts model in a random field is less disordered.

**Remark 5.10.** Let us consider a curve  $\gamma: [0, 1] \to \mathcal{M}_1$  in the macroscopic free energy landscape that connects two minima and passes through the minimal saddle point between them. The corresponding minimal energy curve  $\gamma(t)$  defined by (5.54) and (5.55) has the property that it passes as well through the corresponding minima and the saddle point in the mesoscopic free energy landscape. Suppose there exists an  $\varepsilon > 0$  such that the minimal component of  $\gamma(t)$ , for all  $t \in [0, 1]$ , is larger than  $\varepsilon$ . Then, the assumption on the random field  $\{h^i\}$  implies that there exists two universal constants  $0 < c \le C < \infty$  such that

$$c \pi_k \|\gamma(t)\| \leq \left\| \frac{\mathrm{d}}{\mathrm{d}t} \gamma(t)^k \right\| \leq C \pi_k \|\gamma(t)\|,$$
(5.57)

uniformly in N and in k = 1, ..., n.

Next, we analyze the structure of the critical points. Mind that in a local coordinate system (5.30) reads

$$\mathbb{R}^{q-1} \ni s^* = \nabla (I_{|\Lambda_k|} \circ \psi^{-1})(y) \qquad \Longleftrightarrow \qquad \mathbb{R}^{q-1} \ni y = \nabla (U_{|\Lambda_k|} \circ \psi^{-1})(s^*).$$

Here, we choose as a chart  $\psi : \mathcal{M}_1 \to \mathbb{R}^{q-1}$  the linear map  $\psi(x) = V^T x$  where the columns of the matrix V consist of the vectors of an orthonormal basis of  $V_1$ . Since the gradient map

 $\nabla U_{|\Lambda_k|}$  is invariant with respect to the subspace  $V_0$ , we restricted  $U_{|\Lambda_k|}$  to aff $(\mathcal{M}_1)$ . Thus, we get that for any  $x \in \operatorname{ri}(\mathcal{M}_1)$ 

$$\operatorname{Hess}\left(I_{|\Lambda_k|} \circ \psi^{-1}\right)\left(\psi(x)\right) = \left(V^{\mathrm{T}} \operatorname{Hess} U_{|\Lambda_k|}\left(t^*(x)\right)V\right)^{-1}, \quad \forall k = 1, \dots, n.$$
(5.58)

Hence, the Hessian of the free energy,  $F^n$ , at a critical point  $z \in X_{k=1}^n \mathcal{M}_{\pi_k}$  is given in a local coordinate system,  $\varphi(x) = I_n \otimes V^T$ , by

$$\mathcal{A}^{n}(\boldsymbol{z}) \equiv \operatorname{Hess}\left(F^{n} \circ \varphi^{-1}\right)\left(\varphi(\boldsymbol{z})\right) = 1 \cdot 1^{\mathrm{T}} \otimes (-2I_{q-1}) + \operatorname{diag}(U_{1}^{-1}, \dots, U_{n}^{-1})$$
(5.59)

where  $I_{q-1} \in \mathbb{R}^{q-1 \times q-1}$  denotes the identity matrix and

$$U_k \equiv U_k \left( z^k / \pi_k \right) := \beta \pi_k \, V^{\mathrm{T}} \operatorname{Hess} U_{|\Lambda_k|} \left( 2\beta z + \beta \bar{h}^k \right) V.$$
(5.60)

We are interested in the behavior of  $Q^n$  in a neighborhood of the critical points of  $F^n$ . Suppose that z is a critical point of  $F^n$ . Then, for all  $v \in V_1$  such that  $||v^k|| \leq N^{-1/2+\delta}$  for all k

$$\frac{\mathcal{Q}^{n}(\boldsymbol{z}+\boldsymbol{v})}{\mathcal{Q}^{n}(\boldsymbol{z})} = \exp\left(-\frac{\beta N}{2} \langle \boldsymbol{v}, A^{n}(\boldsymbol{z}) \, \boldsymbol{v} \rangle\right) \left(1 + \mathcal{O}_{N}(1)\right)$$
(5.61)

provided that N large enough to ensure that  $z^k + v^k \in \operatorname{ri}(\mathcal{M}_1)$ . This is an immediate consequence of (5.45) and (5.48). Here, we set  $A^n(z) = (I_n \otimes V) A^n(z)(I_n \otimes V^T)$ .

As a next step, we describe the eigenvalues of  $A^n(z)$  at a critical point z of  $F^n$ . To start with, let us denote by  $\{\xi_1, \ldots, \xi_K\}$  the distinct values that appear in the union of the spectrum of  $U_k(z^k/\pi_k)^{-1}$ . Since  $z \in \operatorname{ri}(X_{k=1}^n \mathcal{M}_{\pi_k})$  and in view of Lemma 5.6 all  $\xi_i$ 's are non-zero. Further, let  $\phi^{k,r}$  the eigenvector corresponding to the eigenvalue  $\lambda_r^k$  of  $U_k(z^k/\pi_k)^{-1}$  and set  $\kappa(\xi_l) := \{(k,r) \mid \lambda_r^k = \xi_l\}$ .

**Lemma 5.11.** Let z be a solution of the equation (5.50). Then  $\gamma$  is an eigenvalue of  $A^n(z)$  if and only if either  $\gamma = \xi_l$  for all  $l \leq K$  which satisfy

$$\kappa(\xi_l)| - \dim\left(\operatorname{span}\{\phi^{k,r} \mid (k,r) \in \kappa(\xi_l)\}\right) =: m_l > 0,$$
(5.62)

where  $m_l$  is the geometric multiplicity of  $\xi_l$  or  $\gamma$  is a solution of the equation

$$\det\left(I_{q-1} - 2\sum_{k=1}^{n} \left(U_k (z^k/\pi_k)^{-1} - \gamma\right)^{-1}\right) = 0.$$
(5.63)

Moreover, (5.63) has exactly one negative simple solution, i.e.  $\gamma$  is a simple root, if and only if, z is a critical point of index 1.

*Proof.* Let  $\xi_l$  be such that  $m_l > 0$ , if such a  $\xi_l$  exists. Then we will construct  $m_l$  orthogonal solutions of

$$\left(\mathcal{A}^{n}(\boldsymbol{z})-\boldsymbol{\gamma}\right)\boldsymbol{v} = -2\left(1\otimes\sum_{k=1}^{n}v^{k}\right) + \operatorname{diag}\left(U_{1}^{-1}-\boldsymbol{\gamma},\ldots,U_{n}^{-1}-\boldsymbol{\gamma}\right)\boldsymbol{v} = 0.$$
(5.64)

with eigenvalue  $\gamma = \xi_l$ . Namely, we set  $0 \neq v = \sum_{(k,r) \in \kappa(\xi_l)} a_r^k (e^k \otimes \phi^{k,r})$  with  $a_r^k \in \mathbb{R}$  which implies that (5.64) is equivalent to  $\sum_{(k,r) \in \kappa(\xi_l)} a_r^k \phi^{k,r} = 0$ . Notice that this equation has  $m_l$  orthogonal solutions. Hence, doing this for every  $\xi_l$  with  $m_l > 0$ , we can construct altogether  $\sum_{l=1}^k m_l = n(q-1) - K$  eigenvectors of  $A^n(z)$ .

Now, let us assume that  $\gamma$  is not an eigenvalue of  $U_k^{-1}$ , i.e.  $\gamma \notin \{\xi_1, \ldots, \xi_K\}$ . To find an eigenvalue of  $A^n(z)$ , an elementary computation shows that

$$0 = \det \left( \mathcal{A}^{n}(\boldsymbol{z}) - \gamma \right) = \prod_{k=1}^{n} \det \left( U_{k}^{-1} - \gamma \right) \cdot \det \left( I_{q-1} - 2 \sum_{k=1}^{n} \left( U_{k}^{-1} - \gamma \right)^{-1} \right), \quad (5.65)$$

Hence, (5.63) is then just the demand that the second term of (5.65) vanishes.

It remains to characterize the solutions of (5.63). Let  $M(\gamma) = I_{q-1} - 2\sum_{k=1}^{n} (U_k^{-1} - \gamma)^{-1}$ . Since the matrices  $U_k$  are positive definite, the matrix  $M(\gamma)$  is regular if  $\gamma \in (-\infty, 0]$ . Clearly,  $M(\gamma)$  converge to the identity matrix as  $\gamma$  tends to  $-\infty$ . On the other hand, for  $\gamma_1, \gamma_2 \in (-\infty, 0]$  with  $\gamma_1 < \gamma_2$  it holds that

$$M(\gamma_1) - M(\gamma_2) = 2(\gamma_2 - \gamma_1) \sum_{k=1}^{n} (U_k - \gamma_1)^{-1} (U_k - \gamma_2)^{-1}$$
(5.66)

i.e. the matrix  $M(\gamma_1) - M(\gamma_2)$  is positive definite. In such a case, we also write  $M(\gamma_1) \succ M(\gamma_2)$ . Let us order the eigenvalues of  $M(\gamma_i)$  in such a way that

$$\lambda_1(\mathcal{M}(\gamma_i)) \leq \lambda_2(\mathcal{M}(\gamma_i)) \leq \ldots \leq \lambda_{q-1}(\mathcal{M}(\gamma_i)) \quad \text{for} \quad i=1,2.$$

By using the Theorem of Courant-Fischer, we obtain that  $M(\gamma_1) \succ M(\gamma_2)$  implies that the eigenvalues satisfy  $\lambda_r(M(\gamma_1)) > \lambda_r(M(\gamma_2))$  for all r. Together with the fact that the eigenvalues depends continuously on the entries,  $\lambda_r(M(\gamma))$ , seen as functions of  $\gamma$ , is strictly monotonic decreasing. Thus,  $M(\gamma)$  has exactly one negative simple eigenvalue if and only if M(0) has exactly one negative eigenvalue. But due to the Theorem of Ostrowski [95], M(0) has exactly one negative eigenvalue if and only if z is a critical point of index 1.

Combining the previous observations, we arrive at the following proposition.

**Proposition 5.12.** Let  $\{z^N\}_{N\in\mathbb{N}}$  be a sequence of critical points of  $F_N$  which converge to a critical point  $z^*$  of the deterministic landscape, i.e.  $z^*$  is the solution of (5.56). For each  $z^N$ , let  $z^N$  be the corresponding critical point of  $F^n$  given by (5.50) and  $\hat{z}^N \in \Gamma^n$  its lattice point approximation. Then, for all but finitely many values of N

$$Z_N \mathcal{Q}^n \left( \hat{\boldsymbol{z}}^N \right) = \frac{\exp\left(-\beta N F_N(\boldsymbol{z})\right) \left(1 + \mathcal{O}_N(1)\right)}{\prod\limits_{k=1}^n \left(2\pi N\right)^{\frac{q-1}{2}} \sqrt{\det\left[\pi_k \operatorname{Hess} U_{|\Lambda_k|} \left(2\beta \boldsymbol{z}^N + \beta \bar{h}^k\right)\right]_q}}.$$
(5.67)

In particular, we have that

$$\prod_{k=1}^{n} \det \left[ \pi_k \operatorname{Hess} U_{|\Lambda_k|} \left( 2\beta z^N + \beta \bar{h}^k \right) \right]_q = \frac{\det \left( I_q - 2\beta \operatorname{Hess} U_N \left( 2\beta z^N \right) \right)}{q^n \beta^{n \cdot (q-1)} \det \left( A^n \left( z^N \right) \right)},$$
(5.68)

Moreover,  $\mathbb{P}_h$ -a.s., for all but finitely many N it holds that

$$\det \left( I_q - 2\beta \operatorname{Hess} U_N(2\beta z^N) \right) = \det \left( I_q - 2\beta \left( \mathcal{D}(z^N) - \mathbb{E}_h[u \cdot u^T] \right) \right) \left( 1 + \mathcal{O}_N(1) \right).$$
(5.69)

Recall that  $D = \text{diag}(z_1^N, \dots, z_q^N)$  and  $u \equiv u^1(2\beta z^N) \in \mathbb{R}^q$  is defined componentwise by (5.21).

*Proof.* By combining (5.45) with (5.48) and (5.51), (5.67) is immediate. Hence, it remains to show the expression for the prefactor (5.68). By an elementary computation, the determinant of the matrix  $A^n(z)$  is of the following form

$$\det\left(\mathcal{A}^{n}(\boldsymbol{z}^{N})\right) = \prod_{k=1}^{n} \det\left(\mathcal{U}_{k}^{-1}\right) \cdot \det\left(\mathcal{I}_{q-1} - 2\sum_{k=1}^{n}\mathcal{U}_{k}\right) = \frac{\det\left(\mathcal{I}_{q} - 2\beta \operatorname{Hess}\mathcal{U}_{N}(2\beta z^{N})\right)}{\prod_{k=1}^{n}\det\left(\mathcal{U}_{k}\right)}.$$
(5.70)

Let  $W \in \mathbb{R}^{q-1 \times q-1}$  be the matrix consisting of the first q-1 rows of V. Then, (5.60) and Lemma 5.6 implies

$$\det \left( U_k \right) = \frac{\det \left( W^{\mathrm{T}} U_k W \right)}{\det \left( W^{\mathrm{T}} W \right)} = q \beta^{q-1} \det \left[ \pi_k \operatorname{Hess} U_{|\Lambda_k|} \left( 2\beta z^N + \beta \bar{h}^k \right) \right]_q$$
(5.71)

where we used that  $\det(W^{T}W) = \det(I_{q-1} - \frac{1}{q}1 \cdot 1^{T}) = 1 - \frac{q-1}{q} = \frac{1}{q}$ . By combining (5.70) and (5.71), we obtain (5.68).

As a consequence of (5.50),  $z^N = \nabla U_N(2\beta z^N)$ . Hence,

$$\operatorname{Hess} U_N(2\beta z^N) = \operatorname{diag} \left( z_1^N, \dots, z_r^N \right) - \frac{1}{N} \sum_{i=1}^N u^i (2\beta z^N) \cdot u^i (2\beta z^N)^{\mathrm{T}}.$$

Moreover, the function  $\mathbb{R}^q \ni t \mapsto \frac{1}{N} \sum_{i=1}^N u^i(t) \cdot u^i(t)^{\mathrm{T}} \in \mathbb{R}^{q \times q}$  converge  $\mathbb{P}_h$ -a.s. to the deterministic function  $\mathbb{E}\left[u(t) \cdot u(t)^{\mathrm{T}}\right]$  uniformly on compact subsets of  $\mathbb{R}^q$ . Since the determinant is a continuous function, we conclude (5.69).

## 5.3. Upper bounds on capacities

In this section we derive upper bounds on capacities. For this purpose, we adapt the strategy that was originally presented in [6]. It relies on the idea to bound the microscopic capacity between two disjoint subsets  $A, B \subset S_N$  in terms of the corresponding capacity of suitable sets in the coarse grained space  $\Gamma^n$ . Proceeding this way, we reduce the problem to the construction of a reasonable mesoscopic test function. First, we specify a function that is *almost harmonic* in a small neighborhood of the relevant saddle point. As a next step, we produce a good test function for the mesoscopic Dirichlet form from it.

**5.3.1.** Mesoscopic capacity and partition of the mesoscopic state space. Let us consider two disjoint subsets  $A, B \subset \Gamma^n$  and set  $A = S^n[A]$  and  $B = S^n[B]$ . As an immediate consequence of the *Dirichlet principle*, we can bound the *microscopic capacity* cap(A, B) from above by the *mesoscopic capacity* CAP<sup>n</sup>(A, B)

$$\operatorname{cap}(A, B) = \inf_{h \in \mathcal{H}_{A,B}} \frac{1}{2} \sum_{\sigma,\eta \in \mathcal{S}_{N}} \mu_{N}(\sigma) p_{N}(\sigma,\eta) \left(h(\sigma) - h(\eta)\right)^{2}$$

$$\leq \inf_{g \in \mathcal{G}_{A,B}} \frac{1}{2} \sum_{\sigma,\eta \in \mathcal{S}_{N}} \mu_{N}(\sigma) p_{N}(\sigma,\sigma') \left(g(\boldsymbol{\varrho}^{n}(\sigma)) - g(\boldsymbol{\varrho}^{n}(\eta))\right)^{2}$$

$$= \inf_{g \in \mathcal{G}_{A,B}} \frac{1}{2} \sum_{\boldsymbol{x},\boldsymbol{y} \in \Gamma^{n}} \mathcal{Q}^{n}(\boldsymbol{x}) r^{n}(\boldsymbol{x},\boldsymbol{y}) \left(g(\boldsymbol{x}) - g(\boldsymbol{y})\right)^{2}$$

$$=: \operatorname{CAP}^{n}(\boldsymbol{A},\boldsymbol{B}),$$

$$(5.72)$$

where

$$\mathcal{H}_{A,B} := \{h: \mathcal{S}_N \to [0,1] \mid h|_A \equiv 1, \ h|_B \equiv 0\},$$
(5.73)

$$\mathcal{G}_{\boldsymbol{A},\boldsymbol{B}} := \left\{ g \colon \Gamma^n \to [0,1] \mid g|_{\boldsymbol{A}} \equiv 1, \ g|_{\boldsymbol{B}} \equiv 0 \right\}.$$
(5.74)

Now, let z be a saddle point of  $F^n$  with index 1, i.e. z is a critical point of  $F^n$  and  $A^n(z)$  has exactly one negative eigenvalue while all other eigenvalues are strictly positive. Further let  $A, B \subset \Gamma^n$  be two subsets that are strictly contained in two different connected components of the level sets  $\{x \in \Gamma^n | F^n(x) < F^n(z)\}$ . Additionally, A, B are chosen in such a way that there exists a path  $\gamma$  from A to B such that  $\max_{x \in \gamma} F^n(x) = F^n(z)$ .

Let us point out that only those points in  $x, y \in \Gamma^n$  contribute to the capacity for which  $r^n(x, y) > 0$ , the modulus of the difference of the harmonic function between x, y is large and the induced measure,  $Q^n(x)$ , is not to small. In [15] it was shown that the harmonic function typically changes from one to zero in a small neighborhood of the saddle point z, see Figure 5.3.1. Hence, for  $\delta_N = c_0 N^{-1/2+\delta}$  with  $c_0 < \infty$  and  $\delta > 0$  let us consider the domain

$$D^{n} \equiv D^{n}(\boldsymbol{z}, \delta_{N}) := \{\boldsymbol{x} \in \Gamma^{n} \mid \|\boldsymbol{z} - \boldsymbol{x}\|_{\infty} \le \delta_{N}\}.$$
(5.75)

86



FIGURE 1. Comparison of the free energy landscape and the harmonic function in the random field Curie-Weiss Potts model for (q = 3) with boundary condition 1 on the shallow local minimum and 0 on the global minimum.

Here,  $D^n(z, \delta_N)$  is a cube in  $\Gamma^n$  centered in z with side length  $2\delta_N$ . Further, for a fixed vector  $v \in \Gamma^n$ , to be defined below, consider the subsets

$$W_{0} := \left\{ \boldsymbol{x} \in \Gamma^{n} \mid |\langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{z} \rangle| < \delta_{N} \right\}$$
  

$$W_{1} := \left\{ \boldsymbol{x} \in \Gamma^{n} \mid \langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{z} \rangle \le -\delta_{N} \right\},$$
  

$$W_{2} := \left\{ \boldsymbol{x} \in \Gamma^{n} \mid \langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{z} \rangle \ge \delta_{N} \right\},$$
  
(5.76)

Our strategy is to compute first an upper bound on the capacity restricted to the set  $D^n$  using an appropriate test function that is almost harmonic in  $D^n$  with boundary condition zero and one, respectively, on the sets  $W_1 \cap D^n$  and  $W_2 \cap D^n$ . Afterwards, we will show that the contribution to the capacity outside  $D^n$  is negligible.

**5.3.2. Distortion by mesoscopic transition probabilities.** In order to construct a test function that is almost harmonic inside  $D^n$ , we have to take into account the Hessian of the free energy,  $F^n$ , at the saddle point, z, as well as the mesoscopic transition probabilities. In what following, we first adopted the construction of a distortion of the form  $B^n(z) = G^T A^n(z)G$  that was originally introduced in [6]. In a second step, we characterize the eigenvalues of  $B^n(z)$  and some aspects of the corresponding eigenvectors.

We begin with considering the block diagonal matrix  $R = \text{diag}(R_1, \ldots, R_n) \in \mathbb{R}^{n \cdot q \times n \cdot q}$ where the matrices  $R_k \in \mathbb{R}^{q \times q}$ , for each  $k = 1, \ldots, n$ , are given by

$$R_k \equiv R_k(\boldsymbol{z}) := \sum_{(r,s)\in\Delta_k} r^n \left( \boldsymbol{z}, \boldsymbol{z} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s} \right) (e^r - e^s) \cdot (e^r - e^s)^{\mathrm{T}}.$$
 (5.77)

Here, z is the saddle point, specified above, and  $\Delta_k \subset S_0^2$ , is the smallest set containing all undirected edges of  $S_0^2$ , e.g.  $\Delta_k = \{(r,s) \in S_0^2 | s < r\}$ . We denote by  $\hat{e}^{k,r} \equiv \frac{1}{N}e^k \otimes e^r$  a coordinate vector in  $\mathbb{R}^{n\cdot q}$  of length  $\frac{1}{N}$ . A characterization of  $\Delta_k$  for  $k = 1, \ldots, n$  will be given in Assumption 5.18. Obviously, for each k, the matrix  $R_k$  is positive semi-definite and has an eigenvalue zero corresponding to the eigenvector  $1 \in \mathbb{R}^q$ . Before we characterize the remaining eigenvalues of  $R_k$ , let us have a closer look at the mesoscopic transition probabilities. As a consequence of the underlying microscopic singlesite dynamics, for any  $x, y \in \Gamma^n$  only those transitions  $r^n(x, y)$  may be different from zero, for which  $y = x - \hat{e}^{k,r} + \hat{e}^{k,s}$  with  $k \in \{1, \ldots, n\}$  and  $r, s \in S_0$ . For a given  $\sigma \in S_N$ , let us denote by

$$\Lambda_{k,r}(\sigma) := \left\{ i \in \Lambda_k \, \middle| \, \sigma_i = r \right\} \tag{5.78}$$

the set of lattice sites *i* in the block  $\Lambda_k$  those spin variable  $\sigma_i$  are equal to the color *r*. For  $x \in \Gamma^n$  set  $x = \sum_{k=1}^n x^k$ . Then, we have

$$r^{n}(\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) = \frac{1}{\mathcal{Q}^{n}(\boldsymbol{x})} \sum_{\sigma \in \mathcal{S}^{n}[\boldsymbol{x}]} \mu_{N}(\sigma) \sum_{i \in \Lambda_{k,r}(\sigma)} p_{N}(\sigma, \sigma^{i,s})$$
$$= \frac{1}{\mathcal{Q}^{n}(\boldsymbol{x})} \sum_{\sigma \in \mathcal{S}^{n}[\boldsymbol{x}]} \mu_{N}(\sigma) \sum_{i \in \Lambda_{k,r}(\sigma)} \frac{\exp\left(-\beta \left[\langle 2x + h^{i}, e^{r} - e^{s} \rangle - \frac{2}{N}\right]_{+}\right)}{Nq},$$

where the configuration  $\sigma^{i,r}$  is obtained from  $\sigma$  by replacing the color  $\sigma_i$  at site *i* through the color  $r \in S_0$ . Notice that for all  $\boldsymbol{x} \in \Gamma^n$  and  $\sigma \in S^n[\boldsymbol{x}]$  the cardinality of the set  $\Lambda_{k,r}(\sigma)$  is equal to  $Nx_r^k$ . In particular,  $|\Lambda_{k,r}(\sigma)|$  assumes the same value for all  $\sigma \in S^n[\boldsymbol{x}]$ . Using that  $h^i = \bar{h}^k + \tilde{h}^i$  and diam  $\mathcal{H}_k \leq \varepsilon(n)$ , we get for  $i \in \Lambda_k$  the estimate

$$r^{n}\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right) = \frac{x_{r}^{k}}{q} \exp\left(-\beta\left[\langle 2x+\bar{h}^{k},e^{r}-e^{s}\rangle-\frac{2}{N}\right]_{+}\right)\left(1+\mathcal{O}(\varepsilon(n))\right).$$
(5.79)

Thus, provided N is large enough, (5.79) implies that there exists a  $c_{15}(n) > 0$  independent of N such that the mesoscopic rates at the saddle point z satisfy  $r^n(z, z - \hat{e}^{k,r} + \hat{e}^{k,s}) \ge c_{15}(n)$ for all k = 1, ..., n and  $r, s \in S_0$ .

Let us know come to a characterization of the eigenvalues in the orthogonal complement,  $V_1$ , to the subspace span{1}.

**Lemma 5.13.** For each k, the eigenvalues of the matrix  $R_k$  with respect to the subspace  $V_1$  are bounded from below by  $c_{15}(n)$ .

*Proof.* The matrix  $R_k$  is symmetric. Hence, by exploiting the interlacing property for symmetric matrices, we can bound the eigenvalues of  $R_k$  with respect to  $V_1$  by the eigenvalues of the leading principle minor, i.e.  $\lambda_s(R_k) \leq \lambda_s([R_k]_q) \leq \lambda_{s+1}(R_k)$  for all  $s = 1, \ldots, q-1$  whereas  $\lambda_1(R_k) = 0$ . By applying Gershgorin's Theorem and computations similar to the ones given in the proof of Lemma 5.6, the eigenvalues of the principle minor  $[R_k]_q$  are bounded from below by

$$\lambda_{s}([R_{k}]_{q}) \geq r^{n}(\boldsymbol{z}, \boldsymbol{z} - \hat{\boldsymbol{e}}^{k,q} + \hat{\boldsymbol{e}}^{k,s}) \mathbb{1}_{(q,s) \in \Delta_{k}} + r^{n}(\boldsymbol{z}, \boldsymbol{z} - \hat{\boldsymbol{e}}^{k,s} + \hat{\boldsymbol{e}}^{k,q}) \mathbb{1}_{(s,q) \in \Delta_{k}} \geq c_{15}(n),$$
(5.80)

for all  $s = 1, \ldots, q - 1$ . This completes the proof.

As an immediate consequence of the lemma above we have that, for each k, the restriction of  $R_k$  to  $V_1$  is regular. Moreover, there exists a positive semi-definite matrix  $G_k$  such that  $R_k = G_k G_k^{\mathrm{T}}$ . The distortion of  $A^n(z)$  by the mesoscopic transition probabilities  $r^n$  is now defined through

$$B^{n}(\boldsymbol{z}) := \operatorname{diag}(G_{1}^{\mathrm{T}}, \dots, G_{n}^{\mathrm{T}}) A^{n}(\boldsymbol{z}) \operatorname{diag}(G_{1}, \dots, G_{n}) = G^{\mathrm{T}} A^{n}(\boldsymbol{z}) G.$$
(5.81)

Let us denote by  $B^n(z) = (I_n \otimes V^T) B^n(z) (I_n \otimes V)$  and  $G = (I_n \otimes V^T) G (I_n \otimes V)$  the corresponding projections to  $V_1$ .

**Remark 5.14.** In comparison with the distortion defined in [6], it turns out that the block diagonal form of the matrix G reflects the coarse graining procedure whereas the particular form of the matrix  $R_k$  is linked to the underlying graph structure of the mesoscopic transition probabilities.

Let  $\{\hat{v}^1, \ldots, \hat{v}^{n \cdot (q-1)}\}$  be an orthonormal basis of eigenvectors to the eigenvalues  $\hat{\gamma}_k$  of  $B^n(z)$  with respect to  $V_1$ . Further define the distorted vectors

$$\boldsymbol{v}^k := G^{-\mathrm{T}} \hat{\boldsymbol{v}}^k$$
 and  $\check{\boldsymbol{v}}^k := G \hat{\boldsymbol{v}}^k = R \boldsymbol{v}^k$ . (5.82)

Here,  $G^{-1}$  denotes the Moore-Penrose pseudo inverse of the matrix G, i.e. it is defined by the property that  $GG^{-1}G = G$ ,  $G^{-1}GG^{-1} = G^{-1}$  and  $GG^{-1}$  is symmetric.

**Remark 5.15.** Notice that the matrix G restricted to the subspace  $V_1$  is regular, and hence the Moore-Penrose pseudo inverse does not cause any further difficulties.

An important fact about these vectors is that

$$\langle \check{\boldsymbol{v}}^l, \boldsymbol{v}^k \rangle = \delta(l,k) \quad \text{and} \quad A^n(\boldsymbol{z}) \,\check{\boldsymbol{v}}^k = \hat{\gamma}_k \, \boldsymbol{v}^k.$$
 (5.83)

This implies the following non-orthogonal decomposition of the quadratic form

$$\langle \boldsymbol{x}, A^n(\boldsymbol{z}) \, \boldsymbol{y} \rangle = \sum_{k=1}^{n \cdot (q-1)} \hat{\gamma}_k \, \langle \boldsymbol{x}, \boldsymbol{v}^k \rangle \, \langle \boldsymbol{v}^k, \boldsymbol{y} \rangle, \qquad \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{V}_1.$$
 (5.84)

In the following lemma we characterize the eigenvalues of  $B^n(z)$ .

**Lemma 5.16.** Let z be a solution of the equation (5.50) and z be given by (5.49). Further, assume that z is a critical point of index 1. Then,  $B^n(z)$  has a unique negative eigenvalue  $\hat{\gamma}_1 \equiv \hat{\gamma}_1(N, n)$  which is given by the negative solution of the equation

$$\det\left(I_{q-1} - 2\sum_{k=1}^{n} \left(U_k (z^k / \pi_k)^{-1} - \gamma R_k^{-1}\right)^{-1}\right) = 0.$$
(5.85)

Moreover, we have that,  $\mathbb{P}_h$ -a.s,

$$\lim_{n \to \infty} \lim_{N \to \infty} \hat{\gamma}_1(N, n) =: \bar{\gamma}_1, \tag{5.86}$$

where  $\bar{\gamma}_1$  is the unique negative solution of the equation

$$\det\left(I_{q-1} - 2\beta \mathbb{E}\left[\left(\left(V^{\mathrm{T}}(\operatorname{diag}(u_{1}, \dots, u_{q}) - u \cdot u^{\mathrm{T}})V\right)^{-1} - \gamma \left(V^{\mathrm{T}}\bar{R}V\right)^{-1}\right]\right) = 0, \quad (5.87)$$

with  $u \equiv u^1(2\beta z)$ , whereas the function  $u^i$  is defined in (5.21) and

$$\bar{R} = \sum_{(r,s)\in\Delta_k} \frac{\exp(2\beta z_r + \beta h_r^1) \exp(-\beta \lfloor \langle 2z + h^1, e^r - e^s \rangle \rfloor_+)}{\beta q \sum_{r'\in\mathcal{S}_0} \exp(2\beta z_{r'} + \beta h_{r'}^1)} (e^r - e^s) \cdot (e^r - e^s)^{\mathrm{T}}.$$
 (5.88)

*Proof.* Since the matrix *G* is positive definite, the Theorem of Ostrowski implies that there exist positive numbers  $\theta_i$  such that  $\lambda_i(B^n(z)) = \theta_i \lambda_i(A^n(z))$ . By combining this fact with Lemma 5.11, we deduce that if *z* is a critical point of index 1,  $B^n(z)$  has a unique negative eigenvalue. Further set  $R_k = G_k G_k^T$ . Then, for  $\gamma \in (\infty, 0]$ ,

$$0 = \det \left( B^{n}(\boldsymbol{z}) - \gamma \right) = \prod_{k=1}^{n} \det \left( R_{k} U_{k}^{-1} - \gamma \right) \cdot \det \left( I_{q-1} - 2 \sum_{k=1}^{n} \left( R_{k} U_{k}^{-1} - \gamma \right)^{-1} R_{k} \right).$$

The fact that the eigenvalues of  $R_k U_k^{-1}$  are positive implies that the unique negative eigenvalue  $\hat{\gamma}_1$  annihilate the last determinant. Hence, we recover (5.85).

It remains to show the convergence property. Inserting the expression for  $U_k$ , given by (5.60), reveals that (5.85) is equivalent to

$$0 = \det \left( I_{q-1} - 2\beta \sum_{k=1}^{n} \pi_k \left( \left( V^{\mathrm{T}} \operatorname{Hess} U_{|\Lambda_k|} \left( 2\beta z + \beta \bar{h}^k \right) V \right)^{-1} - \gamma \left( V^{\mathrm{T}} \frac{1}{\beta \pi_k} R_k V \right)^{-1} \right)^{-1} \right)$$

By substituting (5.49) into (5.79) we get

$$\frac{1}{\beta\pi_k}R_k = \sum_{(r,s)\in\Delta_k} \frac{\exp\left(2\beta z_r + \beta\bar{h}_r^k\right) \exp\left(-\beta\left[\langle 2z+\bar{h}^k, e^r - e^s\rangle - \frac{2}{N}\right]_+\right)}{\beta q \sum_{r'\in\mathcal{S}_0} \exp\left(2\beta z_{r'} + \beta\bar{h}_{r'}^k\right)} \left(e^r - e^s\right) \cdot \left(e^r - e^s\right)^{\mathrm{T}} \left(1 + \mathcal{O}(\varepsilon)\right).$$
(5.89)

On the other hand, a simple computation shows that

 $\operatorname{Hess} U_{|\Lambda_k|} \left( 2\beta z + \beta \bar{h}^k \right)_{r,s}$ 

$$= \left(\frac{\exp(2\beta z_r + \beta \bar{h}_r^k)}{\sum_{r' \in \mathcal{S}_0} \exp(2\beta z_{r'} + \beta \bar{h}_{r'}^k)} \delta(r, s) - \frac{\exp(2\beta z_r + \beta \bar{h}_r^k) \exp(2\beta z_s + \beta \bar{h}_s^k)}{\left(\sum_{r' \in \mathcal{S}_0} \exp(2\beta z_{r'} + \beta \bar{h}_{r'}^k)\right)^2}\right) (1 + \mathcal{O}(\varepsilon)).$$

By exploiting the continuity of the determinant allows to deduce the claimed convergence property.  $\hfill \Box$ 

A consequence of the particular form of the matrix  $A^n(z)$ , see (5.59), is the following

**Lemma 5.17.** Let  $v \equiv v^1$ . Then, there exists a constant  $c_{16} > 0$  such that, independent of n.

$$c_{16} \leq \min_{k} \|v^{k}\| = \max_{k} \|v^{k}\| \leq \frac{1}{c_{16}}.$$
 (5.90)

*Proof.* First of all notice that  $B^n(z)\hat{v}^1 = \hat{\gamma}_1\hat{v}^1$  is equivalent to  $A^n(z)Rv^1 = \hat{\gamma}_1v^1$ . By (5.59) it follows that

$$2\sum_{l=1}^{n} R_{l} V^{\mathrm{T}} v^{l} = \left( U_{k}^{-1} R_{k} - \hat{\gamma}_{1} \right) V^{\mathrm{T}} v^{k}, \qquad \forall k = 1, \dots, n.$$
 (5.91)

Recall that  $V \in \mathbb{R}^{q \times (q-1)}$  is the matrix those columns consist of an orthonormal basis of  $V_1$ . As an immediate consequence of (5.91) we have that for all k, l = 1, ..., n

$$C_{k} v^{k} := \left( U_{k}^{-1} R_{k} - \hat{\gamma}_{1} \right) V^{\mathrm{T}} v^{k} = \left( U_{l}^{-1} R_{l} - \hat{\gamma}_{1} \right) V^{\mathrm{T}} v^{l} = C_{l} v^{l}$$
(5.92)

Hence,  $\hat{v}^l = G_l^T C_l^{-1} C_k v^k$ . On the other hand,  $U_k^{-1} R_k = (V^T \text{Hess } U_{|\Lambda_k|} (2\beta z + \beta \bar{h}^k) V)^{-1} \frac{1}{\beta \pi_k} R_k$ . In view of Lemma 5.6 and (5.89) combined with Lemma 5.13 we conclude that the eigenvalues of the matrix  $C_k$  are bounded away from zero and infinity, uniformly in n, N and  $k = 1, \ldots, n$ . Moreover,

$$1 = \langle \hat{\boldsymbol{v}}, \hat{\boldsymbol{v}} \rangle = \sum_{l=1}^{n} \langle \hat{v}^{l}, \hat{v}^{l} \rangle = \langle v^{k}, C_{k}^{\mathrm{T}} \left( \sum_{l=1}^{n} C_{l}^{-\mathrm{T}} R_{l} C_{l} \right) C_{k} v^{k} \rangle.$$
(5.93)

 $\Box$ 

By exploiting again the Theorem of Ostrowski, the assertion of the lemma follows.

Concerning the construction of the matrices  $R_k$  there is some freedom in choosing the sets  $\Delta_k$ . Let us point out that the eigenvalues of the matrix  $B^n(z)$  as well as the corresponding eigenvectors may depend crucially on the corresponding choice. Further, let us emphasis that the computation of an upper bound on capacities is not sensitive to the particular choice of  $\Delta_k$ , except for the fact that the unique negative eigenvalue  $\hat{\gamma}_1$  appears in the resulting expression. However, in order to construct a non-negative unit flow that leads to a matching lower bound, the choice of  $\Delta_k$  really matters. It turns out that we should construct the matrices  $R_k$  in such a way that the vector  $v^1 \equiv v = \sum_{k=1}^n e^k \otimes v^k$  defined in terms of the eigenvector  $\check{v}^1$ 

corresponding to the unique negative eigenvalue of the matrix  $B^n(z)$  satisfies the property that

either 
$$v_r^k - v_s^k \ge 0$$
 or  $v_r^k - v_s^k \le 0$   $\forall k = 1, \dots, n \text{ and } (r, s) \in \Delta_k.$  (5.94)

One may expect that if we choose the sets  $\Delta_k$  in such a way that the eigenvector to the negative eigenvalue of  $A^n(z)$  satisfies (5.94), this property is preserved for the distorted vector v. However, a proof of this claim is challenging and missing so far. For this reason, we will rely in the sequel on the following

**Assumption 5.18.** Assume that there exists sets  $\Delta_1, \ldots, \Delta_n$  such that the vector  $v^1$  that corresponds to the negative eigenvalue,  $\hat{\gamma}_1$ , of the matrix  $B^n(z)$  satisfy the property

$$v_r^k - v_s^k \ge 0 \qquad \forall k = 1, \dots, n \text{ and } (r, s) \in \Delta_k.$$
(5.95)

**5.3.3.** Approximation of the harmonic function near the saddle point. Now, our aim is to construct a function  $g : \mathbb{R}^{n \cdot q} \to [0, 1]$  which is almost harmonic with respect to the generator  $L^n$ . To start with, consider the function  $f : \mathbb{R} \to [0, 1]$  given by

$$f(s) = \sqrt{\frac{\beta N |\hat{\gamma}_1|}{2\pi}} \int_{-\infty}^{s} \exp\left(-\frac{\beta N |\hat{\gamma}_1|}{2} t^2\right) dt.$$
 (5.96)

Then, we define the function g through

$$g(\boldsymbol{x}) = f(\langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{z} \rangle), \qquad (5.97)$$

with  $v \equiv v^1$  as defined in (5.82). Let us point out that this vector v is also used in the definition of the sets  $W_0, W_1, W_2$ . Due to the choice of  $\delta_N$ , we have that in the limit when  $N \to \infty$  the function g(x) converge exponential fast to 0 for all  $x \in W_1 \cap \partial D^n(z, \delta_N)$  and to 1 for all  $x \in W_2 \cap \partial D^n(z, \delta_N)$ .

For our further computations, the transition rates  $r^n(x, y)$  are in a slightly unpleasant form. Moreover, we would like to replace the measure  $Q^n$  in the neighborhood  $D^n$  of the saddle point z under consideration by the approximation given in (5.61). Based on the fact that the Dirichlet form is monotone in the transition probabilities, in [6] the following comparison result was established

**Lemma 5.19** ([6, Lemma 4.1]). Let  $(r(x, y))_{x,y\in\Gamma}$  and  $(\bar{r}(x, y))_{x,y\in\Gamma}$  be two transition matrices that are reversible with respect to the invariant distributions Q and  $\overline{Q}$ , respectively. Assume that for all  $x, y \in \Gamma$ , there exists  $\delta > 0$  such that

$$\left|\frac{Q(x)}{\overline{Q}(x)} - 1\right| \le \delta \quad \text{and} \quad \left|\frac{r(x,y)}{\overline{r}(x,y)} - 1\right| \le \delta.$$
(5.98)

Then, for any disjoint subsets  $A, B \subset \Gamma$ ,

$$(1-\delta)^2 \leq \frac{\operatorname{CAP}(A,B)}{\overline{\operatorname{CAP}}(A,B)} \leq (1-\delta)^{-2}.$$
 (5.99)

Hence, we are left finding suitable modifications of the transition probabilities  $r^n(x, y)$  and the measure  $Q^n(x)$ , respectively. First, let us define the measure  $\overline{Q}^n$  through

$$\overline{\mathcal{Q}}^{n}(\boldsymbol{x}) := \mathcal{Q}^{n}(\boldsymbol{z}) \exp\left(-\frac{\beta N}{2} \left\langle \boldsymbol{x} - \boldsymbol{z}, A^{n}(\boldsymbol{z})(\boldsymbol{x} - \boldsymbol{z})\right\rangle\right), \qquad \boldsymbol{x} \in \Gamma^{n}.$$
(5.100)

In view of (5.61), there exists a  $0 \le c_{17} < \infty$  such that for all  $x \in D^n$ ,

$$\left|\frac{\mathcal{Q}^{n}(\boldsymbol{x})}{\overline{\mathcal{Q}}^{n}(\boldsymbol{x})}-1\right| \leq c_{17} N \delta_{N}^{3}.$$
(5.101)

By (5.79), the mesoscopic transition probabilities satisfy for all  $x \in D^n$  the bound

$$\frac{x_r^k}{z_r^k} e^{-4\beta\delta_N} e^{-2\beta\varepsilon(n)} \leq \frac{r^n(\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s})}{r^n(\boldsymbol{z}, \boldsymbol{z} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s})} \leq \frac{x_r^k}{z_r^k} e^{4\beta\delta_N} e^{2\beta\varepsilon(n)}.$$

By exploiting the fact that critical points are bounded away from the boundary of  $X_{k=1}^n \mathcal{M}_{\pi_k}$  implies that there exists  $0 \le c_{18} < \infty$  such that for all  $x \in D^n$ ,

$$\left|\frac{r^{n}(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s})}{r^{n}(\boldsymbol{z},\boldsymbol{z}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s})}-1\right| \leq c_{18}\max\{\delta_{N},\varepsilon(n)\}.$$
(5.102)

Hence, we define the following mesoscopic transition rates

$$\bar{r}^n(x, x - \hat{e}^{k,r} + \hat{e}^{k,s}) := r^k_{r,s} \text{ and } \bar{r}^n(x - \hat{e}^{k,r} + \hat{e}^{k,s}, x) := rac{r^k_{r,s} \, \overline{\mathcal{Q}}^n(x)}{\overline{\mathcal{Q}}^n(x - \hat{e}^{k,r} + \hat{e}^{k,s})}$$
(5.103)

with  $r_{r,s}^k \equiv r^n(\mathbf{z}, \mathbf{z} - \hat{\mathbf{e}}^{k,r} + \hat{\mathbf{e}}^{k,s})$  having the property that they are reversible with respect to the measure  $\overline{Q}^n$ . Moreover, we define the operator  $\overline{L}^n$  acting on functions  $g: \Gamma^n \to \mathbb{R}$  as

$$(\bar{L}^{n}g)(\boldsymbol{x}) = \sum_{k=1}^{n} \sum_{r,s=1}^{q} \bar{r}^{n} (\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) (g(\boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) - g(\boldsymbol{x})).$$
(5.104)

The actual advantage of choosing the negative eigenvalue,  $\hat{\gamma}_1$ , of  $B^n(z)$  and the distorted eigenvector  $v^1$  is that we can derive an estimate on  $|(\overline{L}^n g)(x)|$  that is by a factor  $\delta_N^2$  smaller than an arbitrary choice of the parameters  $\hat{\gamma}_1$  and v. We prove this fact in the following lemma.

**Lemma 5.20.** Consider the function g as defined in (5.97). Then, for all  $x \in D^n$ , there exists a constant  $0 \le c < \infty$  such that

$$\left| \left( \overline{L}^n g \right)(\boldsymbol{x}) \right| \leq \sqrt{\frac{\beta |\hat{\gamma}_1|}{2\pi N}} \exp\left( -\frac{\beta N |\hat{\gamma}_1|}{2} \langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{z} \rangle^2 \right) c \, \delta_N^2.$$
(5.105)

*Proof.* To lighten the notation, we assume throughout this proof that the origin of the coordinate system is chosen in such a way that z = 0. Further, we set  $A^n(z) \equiv A$ . Now, for any k = 1, ..., n and  $(r, s) \in \Delta_k$ , let us rewrite the corresponding terms in the generator (5.104) by means of the detailed balance condition as

$$r_{r,s}^{k}\left(g\left(\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right)-g(\boldsymbol{x})\right)\left(1+\frac{\overline{\mathcal{Q}}^{n}\left(\boldsymbol{x}+\hat{\boldsymbol{e}}^{k,r}-\hat{\boldsymbol{e}}^{k,s}\right)}{\overline{\mathcal{Q}}^{n}(\boldsymbol{x})}\cdot\frac{g\left(\boldsymbol{x}+\hat{\boldsymbol{e}}^{k,r}-\hat{\boldsymbol{e}}^{k,s}\right)-g(\boldsymbol{x})}{g\left(\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right)-g(\boldsymbol{x})}\right).$$
(5.106)

Since  $x \in D^n(0, \delta_N)$ , we have that

$$\frac{\overline{\mathcal{Q}}^{n}(\boldsymbol{x}+\hat{\boldsymbol{e}}^{k,r}-\hat{\boldsymbol{e}}^{k,s})}{\overline{\mathcal{Q}}^{n}(\boldsymbol{x})} = \exp\left(-\beta\left\langle \boldsymbol{e}^{k,r}-\boldsymbol{e}^{k,s},A\,\boldsymbol{x}\right\rangle\right) \exp\left(-\frac{\beta}{2N}\left\langle \boldsymbol{e}^{k,r}-\boldsymbol{e}^{k,s},A\left(\boldsymbol{e}^{k,r}-\boldsymbol{e}^{k,s}\right)\right\rangle\right) \\
= \exp\left(-\beta\left\langle \boldsymbol{e}^{k,r}-\boldsymbol{e}^{k,s},A\,\boldsymbol{x}\right\rangle\right) \left(1+\mathcal{O}(1/N)\right).$$
(5.107)

In view of (5.97), let us first consider the difference  $f(a + \frac{1}{N}b) - f(a)$ . A straight forward computation yields

$$\begin{split} f\left(a+\frac{1}{N}b\right) \,-\, f(a) \ &= \ \sqrt{\frac{\beta|\hat{\gamma}_1|}{2\pi N}} \exp\left(-\frac{\beta N|\hat{\gamma}_1|}{2} a^2\right) \int_0^b \exp\left(-\frac{\beta|\hat{\gamma}_1|}{2N} t^2 - \beta|\hat{\gamma}_1| \, a \, t\right) \mathrm{d}t \\ &= \ b \sqrt{\frac{\beta|\hat{\gamma}_1|}{2\pi N}} \exp\left(-\frac{\beta N|\hat{\gamma}_1|}{2} a^2\right) \exp\left(-\frac{\beta|\hat{\gamma}_1|}{2} ab\right) \frac{\sinh\left(\frac{\beta|\hat{\gamma}_1|}{2} ab\right)}{\frac{\beta|\hat{\gamma}_1|}{2} ab} \left(1 + \mathcal{O}(\frac{1}{N})\right). \end{split}$$

Notice that the function  $\sinh(x)/x$  is symmetric. Hence,

$$\frac{g(\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s})-g(\boldsymbol{x})}{g(\boldsymbol{x}+\hat{\boldsymbol{e}}^{k,r}-\hat{\boldsymbol{e}}^{k,s})-g(\boldsymbol{x})} = -\exp\left(\beta\left|\hat{\gamma}_{1}\right|\left\langle\boldsymbol{v},\boldsymbol{x}\right\rangle\left(v_{s}^{k}-v_{r}^{k}\right)\right)\left(1+\mathcal{O}(1/N)\right).$$
(5.108)

Since  $x \in D^n(0, \delta_N)$  and  $\exp(-ax) \sinh(ax)/ax = 1 - ax + \mathcal{O}(x^2)$ , we obtain that to leading order the difference  $g(x - \hat{e}_r^k + \hat{e}_s^k) - g(x)$  is given by

$$g(\boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) - g(\boldsymbol{x})$$

$$= \left(v_s^k - v_r^k\right) \sqrt{\frac{\beta|\hat{\gamma}_1|}{2\pi N}} \exp\left(-\frac{\beta N|\hat{\gamma}_1|}{2} \langle \boldsymbol{v}, \boldsymbol{x} \rangle^2\right) \left(1 - \frac{\beta|\hat{\gamma}_1|}{2} \langle \boldsymbol{v}, \boldsymbol{x} \rangle \left(v_s^k - v_r^k\right) + \mathcal{O}(\delta_N^2)\right). \quad (5.109)$$

Plugging (5.107), (5.108) and (5.109) into (5.106) and using that for all  $x \in D^n(0, \delta_N)$ 

$$1 - \exp\left(\beta \left\langle \hat{\boldsymbol{e}}^{k,s} - \hat{\boldsymbol{e}}^{k,r}, \left(A + |\hat{\gamma}_1| \, \boldsymbol{v} \cdot \boldsymbol{v}^{\mathrm{T}}\right) \boldsymbol{x} \right\rangle\right) = \beta \left\langle \hat{\boldsymbol{e}}^{k,s} - \hat{\boldsymbol{e}}^{k,r}, \left(A + |\hat{\gamma}_1| \, \boldsymbol{v} \cdot \boldsymbol{v}^{\mathrm{T}}\right) \boldsymbol{x} \right\rangle + \mathcal{O}(\delta_N^2),$$

we obtain, by collecting the leading order terms, that there exists  $0 \leq c \leq \infty$  such that

$$\begin{aligned} \left| \left( \overline{L}^{n} g \right)(\boldsymbol{x}) \right| &\leq \sqrt{\frac{\beta |\hat{\gamma}_{1}|}{2\pi N}} \exp\left( -\frac{\beta N |\hat{\gamma}_{1}|}{2} \langle \boldsymbol{v}, \boldsymbol{x} \rangle^{2} \right) \left( c \, \delta_{N}^{2} + \beta \left\langle \boldsymbol{v}, R \left( A + |\hat{\gamma}_{1}| \, \boldsymbol{v} \cdot \boldsymbol{v}^{\mathrm{T}} \right) \boldsymbol{x} \right\rangle \right) \\ &= \sqrt{\frac{\beta |\hat{\gamma}_{1}|}{2\pi N}} \exp\left( -\frac{\beta N |\hat{\gamma}_{1}|}{2} \left\langle \boldsymbol{v}, \boldsymbol{x} \right\rangle^{2} \right) c \, \delta_{N}^{2}. \end{aligned}$$
(5.110)

Here, we used in the first step the definition of R, see (5.77). Let us point out that the second step relies on our choice of v and  $\hat{\gamma}_1$  together with (5.82), (5.83) and the property (5.84). Namely,

$$\langle \boldsymbol{v}, R\left(A + |\hat{\gamma}_{1}| \, \boldsymbol{v} \cdot \boldsymbol{v}^{\mathrm{T}}\right) \boldsymbol{x} \rangle = \sum_{i=1}^{n \cdot (q-1)} \hat{\gamma}_{i} \langle \check{\boldsymbol{v}}^{1}, \boldsymbol{v}^{i} \rangle \langle \boldsymbol{v}^{i}, \boldsymbol{x} \rangle - \hat{\gamma}_{1} \langle \check{\boldsymbol{v}}^{1}, \boldsymbol{v}^{i} \rangle \langle \boldsymbol{v}, \boldsymbol{x} \rangle = 0.$$
 (5.111)

This implies the statement of the lemma.

After having shown that g is a good approximation of the equilibrium potential in a neighborhood of a critical point z, we can now proceed to compute an upper bound for the capacity.

**Proposition 5.21.** For every  $n \in \mathbb{N}$  and N sufficiently large we get

$$\operatorname{cap}(A,B) \leq \mathcal{Q}^{n}(\boldsymbol{z}) \frac{\beta |\hat{\gamma}_{1}|}{2\pi N} q^{-\frac{n}{2}} \prod_{i=1}^{n \cdot (q-1)} \sqrt{\frac{2\pi N}{\beta |\hat{\gamma}_{i}|}} \sqrt{\operatorname{det}\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\right)} \left(1 + \mathcal{O}(\max\{\varepsilon(n), \delta_{N}\})\right).$$
(5.112)

*Proof.* The strategy is to compute first the contribution to the mesoscopic Dirichlet form in the neighborhood  $D^n(z, \delta_N)$  of the relevant saddle point z using the approximate harmonic function g. By exploiting Lemma 5.19, the computations will be done with the modification of the restricted Dirichlet form  $\mathcal{E}_{D_N}^n$ 

$$\overline{\mathcal{E}}_{D_{N}}^{n}(g) := \sum_{\boldsymbol{x}\in D^{n}} \sum_{k=1}^{n} \sum_{(r,s)\in\Delta_{k}} \overline{\mathcal{Q}}^{n}(\boldsymbol{x}) \, \overline{r}^{n}\left(\boldsymbol{x}, \boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right) \left(g\left(\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right) \, - \, g(\boldsymbol{x})\right)^{2}. \tag{5.113}$$

Afterwards, we will show that the contribution to the Dirichlet form outside  $D^n(z, \delta_N)$  is negligible. Again, we choose the origin of the coordinate system in such a way that z = 0, and we set  $A \equiv A^n(z)$ .

First, note that by (5.109)

$$\left(g(\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s})-g(\boldsymbol{x})\right)^{2} = \frac{\beta|\hat{\gamma}_{1}|}{2\pi N} \exp\left(-\beta N|\hat{\gamma}_{1}|\langle \boldsymbol{v},\boldsymbol{x}\rangle^{2}\right) \left(v_{s}^{k}-v_{r}^{k}\right)^{2} \left(1+\mathcal{O}(\delta_{N})\right).$$
(5.114)

Inserting (5.114) together with the definitions of  $\overline{Q}^n$  and  $\overline{r}^n$  into (5.113) gives

$$\overline{\mathcal{E}}_{D_N}^n(g) = \mathcal{Q}^n(0) \frac{\beta |\hat{\gamma}_1|}{2\pi N} \sum_{\boldsymbol{x} \in D^n} \exp\left(-\frac{\beta N}{2} \langle \boldsymbol{x}, \left(A+2 |\hat{\gamma}_1| \, \boldsymbol{v} \cdot \boldsymbol{v}^{\mathrm{T}}\right) \boldsymbol{x} \rangle\right) \left(1 + \mathcal{O}(\delta_N)\right), \quad (5.115)$$

where we take additionally advantage of the fact that due to (5.82) and (5.83)

$$\sum_{k=1}^{n} \sum_{(r,s)\in\Delta_{k}} r_{r,s}^{k} \left( v_{s}^{k} - v_{r}^{k} \right)^{2} = \langle \boldsymbol{v}, R \, \boldsymbol{v} \rangle = \langle \boldsymbol{v}^{1}, \check{\boldsymbol{v}}^{1} \rangle = 1.$$

It remains to evaluate in (5.115) the sum over  $x \in D^n(0, \delta_N)$ . By a standard approximation of the sum by an integral, we get that

$$\sum_{\boldsymbol{x}\in D^{n}} \exp\left(-\frac{\beta N}{2} \langle \boldsymbol{x}, \left(A+2\left|\hat{\gamma}_{1}\right|\boldsymbol{v}\cdot\boldsymbol{v}^{\mathrm{T}}\right)\boldsymbol{x}\rangle\right)\right)$$
$$= \left(\frac{N^{q-1}}{\sqrt{q}}\right)^{n} \int_{\boldsymbol{V}_{1}} \exp\left(-\frac{\beta N}{2} \langle \boldsymbol{x}, \left(A+2\left|\hat{\gamma}_{1}\right|\boldsymbol{v}\cdot\boldsymbol{v}^{\mathrm{T}}\right)\boldsymbol{x}\rangle\right) \mathrm{d}\boldsymbol{x} \left(1+\mathcal{O}\left(\sqrt{\ln N/N}\right)\right).$$
(5.116)

Let us briefly explain the arising of the factor  $\left(N^{q-1}/\sqrt{q}\right)^n$ . In order to replace the sum by an integral we have to take into account the volume of the unit cell associated to each lattice point that is given by the *n*-fold Cartesian product of the parallelepiped

$$\left\{\frac{1}{N}\sum_{s=2}^{q}\lambda_s\left(e^s-e^r\right)\ \middle|\ \lambda_s\in[0,1]\right\}\ \subset\ V_1\ \subset\ \mathbb{R}^q$$

of side length 1/N. Its volume is given by  $\sqrt{\det(I_{q-1} + 1 \cdot 1^T)} / N^{q-1} = \sqrt{q} / N^{q-1}$ . For the purpose of evaluating the integral, we consider the immersion

 $\mathbb{R}^{n \cdot (q-1)} \ni y \mapsto G \cdot \left( \hat{\boldsymbol{v}}^1, \dots, \hat{\boldsymbol{v}}^{n \cdot (q-1)} \right) y \in \mathbb{R}^{n \cdot q},$ 

where  $\{\hat{v}^1, \dots, \hat{v}^{n \cdot (q-1)}\}$  is an orthonormal basis of eigenvectors of  $B^n(z)$ . Then, in view of (5.82), we have

$$\int_{\boldsymbol{V}_{1}} \exp\left(-\frac{\beta N}{2} \langle \boldsymbol{x}, \left(A+2\left|\hat{\gamma}_{1}\right| \boldsymbol{v} \cdot \boldsymbol{v}^{\mathrm{T}}\right) \boldsymbol{x} \rangle\right) \mathrm{d}\boldsymbol{x}$$
$$= \prod_{k=1}^{n \cdot (q-1)} \int_{\mathbb{R}^{n(q-1)}} \exp\left(-\frac{\beta N}{2}\left|\hat{\gamma}_{k}\right| \boldsymbol{y}^{2}\right) \mathrm{d}\boldsymbol{y} \sqrt{\det\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\right)} = \prod_{k=1}^{n \cdot (q-1)} \sqrt{\frac{2\pi}{\beta N\left|\hat{\gamma}_{k}\right|}} \sqrt{\det\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\right)},$$
(5.117)

where we used the non-orthogonal decomposition (5.84) in the second step. Therefore, we obtain the following estimate of the modified Dirichlet form  $\overline{\mathcal{E}}_{D_{N}}^{n}$ 

$$\overline{\mathcal{E}}_{D_N}^n(g) = \mathcal{Q}^n(0) \frac{\beta |\hat{\gamma}_1|}{2\pi N} q^{-\frac{n}{2}} \prod_{k=1}^{n \cdot (q-1)} \sqrt{\frac{2\pi N}{\beta |\hat{\gamma}_k|}} \sqrt{\det\left(\mathcal{G}^{\mathrm{T}}\mathcal{G}\right)} \left(1 + \mathcal{O}(\delta_N)\right).$$
(5.118)

It remains to show that the contribution to the Dirichlet form coming from points x outside of  $D^n(z, \delta_N)$  do not contribute significantly to the capacity. Let us define the test function  $\tilde{g}$  as

$$\tilde{g}(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{x} \in W_1 \\ 1, & \boldsymbol{x} \in W_2 \\ g(\boldsymbol{x}), & \boldsymbol{x} \in W_0 \end{cases}$$
(5.119)

It is worth noting that the only non-zero contribution to the Dirichlet form  $\mathcal{E}^n(\tilde{g})$  comes from  $\overline{W}_0 := W_0 \cup \partial W_0$ . Here, we denote by  $\partial W_0$  the set of all points  $x \notin W_0$  such that there exists



FIGURE 2. Domains for the construction of the test function  $\tilde{g}$ 

 $\boldsymbol{y} \in W_0$  with  $r^n(\boldsymbol{x}, \boldsymbol{y}) > 0$ . Further, let us define the sets

$$W_0^{\text{in}} := W_0 \cap D^n \quad \text{and} \quad W_0^{\text{out}} := W_0 \cap (D^n)^c$$
  
 $\overline{W}_0^{\text{in}} := \overline{W}_0 \cap D^n \quad \text{and} \quad \overline{W}_0^{\text{out}} := W_0^{\text{out}} \cap \partial W_0^{\text{out}}.$ 

We denote by  $\mathcal{E}_{\overline{W}_0^{\text{in}}}^n$  and  $\mathcal{E}_{\overline{W}_0^{\text{out}}}^n$ , respectively, the restriction of the Dirichlet form to the corresponding sets. Then,

$$\mathcal{E}^{n}(\tilde{g}) = \overline{\mathcal{E}}_{\overline{W}_{0}^{\text{in}}}^{n}(\tilde{g}) \left(1 + \mathcal{O}(\max\{\varepsilon(n), \delta_{N}\})\right) + \mathcal{E}_{\overline{W}_{0}^{\text{out}}}^{n}(\tilde{g}) \\
= \left(\overline{\mathcal{E}}_{\overline{W}_{0}^{\text{in}}}^{n}(g) - \left(\overline{\mathcal{E}}_{\overline{W}_{0}^{\text{in}}}^{n}(g) - \overline{\mathcal{E}}_{\overline{W}_{0}^{\text{in}}}^{n}(\tilde{g})\right)\right) \left(1 + \mathcal{O}(\max\{\varepsilon(n), \delta_{N}\})\right) + \mathcal{E}_{\overline{W}_{0}^{\text{out}}}^{n}(\tilde{g}) \quad (5.120)$$

The remaining part of the proof comprises three steps.

STEP 1. Let us consider the term  $\overline{\mathcal{E}}_{W_0^{\text{in}}}^n(g)$ . For a suitable choosen 0 < c < 1 such that  $D^n(\boldsymbol{z}, c \, \delta_N) \subseteq \overline{W}_0^{\text{in}}$  we have that

$$\overline{\mathcal{E}}^n_{D^n(\boldsymbol{z},c\,\delta_N)}(g) \leq \overline{\mathcal{E}}^n_{W^{\rm in}_0}(g) \leq \overline{\mathcal{E}}^n_{D^n(\boldsymbol{z},\delta_N)}(g).$$

Hence, by inspecting the computation above, we obtain immediately that

$$\overline{\mathcal{E}}_{D^{n}(\boldsymbol{z},c\,\delta_{N})}^{n}(g) = \overline{\mathcal{E}}_{D^{n}(\boldsymbol{z},\delta_{N})}^{n}(g) \left(1 + \mathcal{O}(\delta_{N})\right).$$

Thus, it follows that

$$\overline{\mathcal{E}}_{W_0^{\text{in}}}^n(g) = \overline{\mathcal{E}}_{D^n(\boldsymbol{z},\delta_N)}^n(g) \left(1 + \mathcal{O}(\delta_N)\right).$$
(5.121)

STEP 2. Consider now the term  $\overline{\mathcal{E}}_{\overline{W}_0}^n(g) - \overline{\mathcal{E}}_{\overline{W}_0}^n(\tilde{g})$ . Since,  $\tilde{g} \equiv g$  on  $W_0$ , we get

$$\left|\overline{\mathcal{E}}_{\overline{W}_{0}^{\text{in}}}^{n}(g) - \overline{\mathcal{E}}_{\overline{W}_{0}^{\text{in}}}^{n}(\tilde{g})\right| \leq I_{1} + I_{2},$$
(5.122)

say, with

$$I_{1} = \sum_{\boldsymbol{x} \in W_{1} \cap \partial W_{0}^{\text{in}}} \overline{\mathcal{Q}}^{n}(\boldsymbol{x}) \sum_{k=1}^{n} \sum_{(r,s) \in \Delta_{k}} r_{r,s}^{k} \left( \left( g(\boldsymbol{x}) - g(\boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) \right)^{2} + g(\boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s})^{2} \right),$$

$$I_{2} = \sum_{\boldsymbol{x} \in W_{2} \cap \partial W_{0}^{\text{in}}} \overline{\mathcal{Q}}^{n}(\boldsymbol{x}) \sum_{k=1}^{n} \sum_{(r,s) \in \Delta_{k}} r_{r,s}^{k} \left( \left( g(\boldsymbol{x}) - g(\boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) \right)^{2} + \left( 1 - g(\boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) \right)^{2} \right).$$

where we used that, by definition, the function  $\tilde{g}$  has boundary condition zero and one, respectively, on  $W_1$  and  $W_2$ . Due to the symmetry, we focus only on the first term. Notice that, for  $\boldsymbol{x} \in W_1 \cap \partial W_0^{\text{in}}$ , it holds that  $\langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq -\delta_N$  and

$$g(oldsymbol{x}) \ \le \ f(-\delta_N) \ \le \ rac{1}{\sqrt{2\pieta N \left|\hat{\gamma}_1
ight|} \ \delta_N} \ \exp\!\left(-rac{eta N}{2}\!\left|\hat{\gamma}_1
ight| \,\delta_N^2
ight)$$

Inserting this estimate together with (5.109) into the expression for  $I_1$  gives

$$I_{1} \leq \frac{\beta |\hat{\gamma}_{1}|}{2\pi N} \exp\left(-\beta N |\hat{\gamma}_{1}| \,\delta_{N}^{2}\right) \sum_{\boldsymbol{x} \in W_{1} \cap \partial W_{0}^{\text{in}}} \overline{\mathcal{Q}}^{n}(\boldsymbol{x}) \sum_{k=1}^{n} \sum_{(r,s) \in \Delta_{k}} r_{r,s}^{k} \left(\left(v_{s}^{k} - v_{r}^{k}\right)^{2} + c \,\delta_{N}^{-2}\right)\right)$$
$$= \mathcal{Q}^{n}(0) \frac{\beta |\hat{\gamma}_{1}|}{2\pi N} \exp\left(-\beta N |\hat{\gamma}_{1}| \,\delta_{N}^{2}\right) \sum_{\boldsymbol{x} \in W_{1} \cap \partial W_{0}^{\text{in}}} \exp\left(-\frac{\beta N}{2} \langle \boldsymbol{x}, A \, \boldsymbol{x} \rangle\right) c' \,\delta_{N}^{-2}, \tag{5.123}$$

where we used in the second step that

$$\sum_{k=1}^{n} \sum_{(r,s)\in\Delta_{k}} r_{r,s}^{k} \left( v_{s}^{k} - v_{r}^{k} \right)^{2} = \langle \boldsymbol{v}, R \boldsymbol{v} \rangle = \langle \boldsymbol{v}^{1}, \check{\boldsymbol{v}}^{1} \rangle = 1.$$

Let us point out that the constant c' is independent of n and N. The evaluation of the remaining sum over  $x \in W_1 \cap \partial W_0^{\text{in}}$  is similar to the computation we did above except for the fact that the integration runs over a (n(q-1)-1)-dimensional hyperplane W orthogonal to  $V_1$ and shifted by the vector  $\delta_n v$ . Thus, setting d = n(q-1), then for N large enough, we have

$$\sum_{\boldsymbol{x}\in W_{1}\cap\partial W_{0}^{\text{in}}} \exp\left(-\frac{\beta N}{2}\langle\boldsymbol{x},A\,\boldsymbol{x}\rangle\right)$$

$$\leq \left(\frac{N^{q-1}}{\sqrt{q}}\right)^{n} \int_{\boldsymbol{W}} \exp\left(-\frac{\beta N}{2}\langle\boldsymbol{x},A\,\boldsymbol{x}\rangle\right) d\boldsymbol{x}$$

$$= \left(\frac{N^{q-1}}{\sqrt{q}}\right)^{n} \exp\left(-\frac{\beta N}{2}\hat{\gamma}_{1}\,\delta_{N}^{2}\right) \prod_{k=2}^{n(q-1)} \int_{\mathbb{R}^{d-1}} \exp\left(-\frac{\beta N}{2}\left|\hat{\gamma}_{k}\right|\,y^{2}\right) dy\,\sqrt{\det\left[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\right]_{d}}$$

$$\leq \sqrt{N}\left(\frac{N^{q-1}}{\sqrt{q}}\right)^{n} \exp\left(\frac{\beta N}{2}\left|\hat{\gamma}_{1}\right|\,\delta_{N}^{2}\right) \prod_{k=2}^{n(q-1)} \sqrt{\frac{2\pi}{\beta N}\left|\hat{\gamma}_{k}\right|}\,\sqrt{\det\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\right)}.$$
(5.124)

Notice that the second step relies on the non-orthogonal decomposition (5.84). Moreover, by taking advantage of the interlacing property of the eigenvalues of symmetric matrices and the fact that the eigenvalues of the matrix  $R_k$  are bounded from below by  $c_{15}(n) \ge 1/N$ , we bound the determinant of the principle minor  $\det[G^T G]_{n(q-1)}$  by  $N \det(G^T G)$ . This implies that

$$I_1 \leq \mathcal{Q}^n(0) \frac{\beta |\hat{\gamma}_1|}{2\pi N} \prod_{k=2}^{n(q-1)} \sqrt{\frac{2\pi N}{\beta |\hat{\gamma}_k|}} \sqrt{\det \left( \mathcal{G}^{\mathrm{T}} \mathcal{G} \right)} N \exp \left( -\frac{\beta N}{2} |\hat{\gamma}_1| \, \delta_N^2 \right).$$
(5.125)
A similar bound can also be obtained for the term  $I_2$ . Finally, comparing these expression with (5.118) gives

$$\left|\overline{\mathcal{E}}_{\overline{W}_{0}^{\text{in}}}^{n}(g) - \overline{\mathcal{E}}_{\overline{W}_{0}^{\text{in}}}^{n}(\tilde{g})\right| \leq \overline{\mathcal{E}}_{D^{n}}(g) N e^{-\frac{1}{2}\beta |\hat{\gamma}_{1}| N^{2\delta}}.$$
(5.126)

STEP 3. It remains to consider the last Dirichlet form in (5.120). Since by assumption the saddle point is not degenerate and we are interested in an estimate orthogonal to the direction v, it holds for all  $x \in \overline{W}_0^{\text{out}}$  that there exists  $K < \infty$  such that  $F^n(x) \ge F^n(z) + KN^{-1+2\delta}$  for some  $K < \infty$  and N sufficiently large. Hence,

$$\mathcal{E}^{n}_{\overline{W}^{\text{out}}_{0}}(\tilde{g}) \leq \sum_{\boldsymbol{x}\in\overline{W}^{\text{out}}_{0}} \mathcal{Q}^{n}(\boldsymbol{x}) \leq c \, \mathcal{Q}^{n}(0) \, \mathrm{e}^{-\beta K N^{2\delta}} \leq \overline{\mathcal{E}}^{n}_{D_{N}}(g) \, \mathcal{O}_{N}(\delta_{N}).$$
(5.127)

Combining (5.120) with the estimates given in (5.121), (5.126) and (5.127), yields that the Dirichlet form  $\mathcal{E}^n(\tilde{g}) = \overline{\mathcal{E}}_{D_N}^n(g) \left(1 + \mathcal{O}(\max\{\varepsilon(n), \delta_N\})\right)$ . This completes the proof of the upper bound.

In view of Proposition 5.12, we obtain the following upper bound for capacities.

Corollary 5.22. With the notation, introduced above, we obtain

$$Z_N \operatorname{cap}(A, B) \leq \frac{\beta |\bar{\gamma}_1|}{2\pi N} \frac{\exp(-\beta N F_N(z))}{\sqrt{\left|\det\left(I_q - 2\beta \operatorname{Hess} U_N(2\beta z)\right)\right|}} (1 + \mathcal{O}(\delta_N)).$$
(5.128)

*Proof.* First, recall that  $|\hat{\gamma}_i|$  denote the eigenvalues of the matrix  $B^n(z)$ . Hence,

$$\prod_{k=1}^{n \cdot (q-1)} |\hat{\gamma}_k| = \left| \det \left( B^n(\boldsymbol{z}) \right) \right| = \left| \det \left( G^{\mathrm{T}} A^n(\boldsymbol{z}) G \right) \right| = \det \left( G^{\mathrm{T}} G \right) \cdot \left| \det \left( A^n(\boldsymbol{z}) \right) \right|.$$
(5.129)

Substituting the expression (5.67) for the induced measure  $Q^n(z)$  at a critical point together with (5.68) and (5.129) into the upper bound (5.112) yields (5.128) except from the fact that instead of  $\bar{\gamma}_1$  the eigenvalue  $\hat{\gamma}_1$  of the matrix  $B^n(z)$  appears. Notice that the error term  $\mathcal{O}(\max\{\varepsilon(n), \delta_N\})$  becomes independent of n when n is chosen large enough. Hence,  $\hat{\gamma}_1$  is the only n-dependent quantities on the right-hand side of (5.128) while the left-hand side is independent of n. By taking the limit  $n \to \infty$  yields the desired bound.

## 5.4. Lower bounds on capacities

In this section we will demonstrate how the Berman-Konsowa principle can be used to derive lower bounds on capacities. In view of Proposition 1.22, our task is to construct a suitable non-negative unit flow. The actual construction follows the strategy suggested in [6] and is done in two steps. As a first step, we construct a *mesoscopic unit flow* from the approximate harmonic function. In a second step, we construct for each mesoscopic path a *subordinate microscopic unit flow*. By a careful construction of both flows, we can establish a lower bound on the microscopic capacity in terms of the mesoscopic capacity that differs only by a factor of size  $1 + O(\varepsilon(n))$ .

5.4.1. Mesoscopic and microscopic lower bounds: The strategy. Let us consider two minima, m, m', of the mesoscopic free energy  $F^n$ . Further, let us denote by z the lowest saddle point of  $F^n$  between m and m', i.e.  $F^n(\hat{z}) = \min_{\gamma} \max_{x \in \gamma} F^n(x)$  where the minimum is taken over all mesoscopic paths  $\gamma$  on  $\Gamma^n$  between  $\hat{m}$  and  $\hat{m}'$ . Here,  $\hat{x} \in \Gamma^n$  denotes the closest lattice point approximation of  $m \in \bigotimes_{k=1}^n \mathcal{M}_{\pi_k}$ . For convenience, we pretend that  $m, z, m' \in \Gamma^n$ , since the proofs will not be sensitive concerning this correction. As in the previous section, let  $A, B \subset \Gamma^n$  be two subsets which are strictly contained in two different connected components of the level sets  $\{x \in \Gamma^n \mid F^n(x) < F^n(z)\}$  such that  $m \in A$  and  $m' \in B$ .

In view of (1.86), we will construct a non-negative unit flow  $f_{A,B}$  of the form

$$\mathfrak{f}_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathcal{Q}^n(\boldsymbol{x}) r^n(\boldsymbol{x},\boldsymbol{y})}{\mathcal{E}^n(\tilde{g})} \phi_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{x},\boldsymbol{y}), \qquad (5.130)$$

such that with respect to the law,  $\mathbb{P}^{\mathfrak{f}_{A,B}}$ , of the associated Markov chain it holds

$$\mathbb{P}^{\mathbf{f}_{\boldsymbol{A},\boldsymbol{B}}}\left[\sum_{(\boldsymbol{x},\boldsymbol{y})\in\boldsymbol{\gamma}}\phi_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{x},\boldsymbol{y})=1+\mathcal{O}_{N}(1)\right]=1-\mathcal{O}_{N}(1).$$
(5.131)

As an immediate consequence of Proposition 1.22, Equation (5.131) implies that the mesoscopic capacity is bounded from below by

$$\operatorname{CAP}^{n}(\boldsymbol{A},\boldsymbol{B}) \geq \mathbb{E}^{\mathfrak{f}_{\boldsymbol{A},\boldsymbol{B}}}\left[\left(\sum_{(\boldsymbol{x},\boldsymbol{y})\in\boldsymbol{\gamma}}\frac{\mathfrak{f}_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{x},\boldsymbol{y})}{\mathcal{Q}^{n}(\boldsymbol{x})\,r^{n}(\boldsymbol{x},\boldsymbol{y})}\right)^{-1}\right] \geq \mathcal{E}^{n}(\tilde{g})\left(1-\mathcal{O}_{N}(1)\right). \quad (5.132)$$

As demonstrated in [6], for a suitable chosen neighborhood  $\mathfrak{D}^n$  of the saddle point z, the flow  $\mathfrak{f}_{A,B}$  consists of a concatenation of a flow  $\mathfrak{f}_A$  from A to the boundary  $\partial_A \mathfrak{D}^n$ , a flow  $\mathfrak{f}$  through  $\mathfrak{D}^n$  and a flow  $\mathfrak{f}_B$  from  $\partial_B \mathfrak{D}^n$  to B. It turns out that, with a probability close to one, the major contribution to the sum in (5.131) comes from the flow  $\mathfrak{f}$ , i.e.

$$\mathbb{P}^{\mathfrak{f}}\left[\sum_{(\boldsymbol{x},\boldsymbol{y})\in\boldsymbol{\gamma}}\phi(\boldsymbol{x},\boldsymbol{y})=1+\mathcal{O}_{N}(1)\right]=1-\mathcal{O}_{N}(1).$$

The actual construction of  $\mathfrak{f}$ , that is presented in the next subsection, is the most difficult part and relies on the following observation. Although the test function g, which we have constructed in (5.97), does not satisfy Kirchhoff's law, Lemma 5.20 suggests that inside  $\mathfrak{D}^n$  a potential candidate for  $\phi$  seems to be a (small) perturbation of the discrete gradient  $\nabla g$ . In contrast to that, we will show that with a probability close to one the contribution to (5.131) coming from  $\mathfrak{f}_A$  and  $\mathfrak{f}_B$  is indeed negligible.

After having established a mesoscopic lower bound via a suitable mesoscopic flow, our next task is to construct a subordinate *microscopic flow*,  $f_{A,B}$ , from  $A = S^n[A]$  to  $B = S^n[B]$ . In order to do so, let  $X = \{X(t)\}_t$  be the Markov chain on  $\Gamma^n$  with law  $\mathbb{P}^{\mathfrak{f}_{A,B}}$  that starts in A and is stopped on the arrival of B. Given a realization  $\underline{x}$  of this mesoscopic Markov chain, we choose the label in such a way that  $\underline{x} = (x(a_{\underline{x}}), \ldots, x(0), \ldots, x(b_{\underline{x}}))$  with  $x(a_{\underline{x}}) \in A$ ,  $x(b_{\underline{x}}) \in B$  and  $\sum_{k=1}^n x(0)^k = \sum_{k=1}^n z^k$ . Notice that  $a_{\underline{x}} < 0$  and  $b_{\underline{x}} > 0$ . To each path,  $\underline{x}$ , of positive probability associate a subordinate microscopic unit flow  $f^{\underline{x}}$  such that

$$f^{\underline{x}}(\sigma,\eta) > 0 \qquad \Longleftrightarrow \qquad (\underline{\varrho}^n(\sigma),\underline{\varrho}^n(\eta)) \in \underline{x}.$$
 (5.133)

Hence, the total microscopic flow  $f_{A,B}$  can be decomposed as

$$f_{A,B}(\sigma,\eta) = \sum_{\underline{x}} \mathbb{P}^{\mathfrak{f}_{A,B}} \left[ \boldsymbol{X} = \underline{x} \right] f^{\underline{x}}(\sigma,\eta).$$
(5.134)

Due to the fact that  $f^x$  is a unit flow, it holds that

$$\sum_{\sigma \in S^{n}[\boldsymbol{x}]} \sum_{\eta \in S^{n}[\boldsymbol{y}]} f^{\boldsymbol{x}}(\sigma, \eta) = 1, \qquad \forall (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{x}.$$
(5.135)

As an immediate consequence of (5.134) and (5.135), the total microscopic flow has the property

$$\sum_{\sigma \in \mathcal{S}^{n}[\boldsymbol{x}]} \sum_{\eta \in \mathcal{S}^{n}[\boldsymbol{y}]} f_{A,B}(\sigma,\eta) = \sum_{\boldsymbol{x}} \mathbb{P}^{\boldsymbol{f}_{\boldsymbol{A},\boldsymbol{B}}} \left[ \boldsymbol{X} = \boldsymbol{x} \right] \mathbb{1}_{(\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{x}} = \boldsymbol{\mathfrak{f}}_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{x},\boldsymbol{y}).$$
(5.136)

Further, for each subordinate microscopic flow let us associate a microscopic Markov chain  $\Sigma^{\underline{x}} = \{\Sigma^{\underline{x}}(t)\}_t$  on  $\mathcal{S}_N$  with law  $\mathbb{P}^{f^{\underline{x}}}$ . Notice that (5.134) combined with (5.136) give rise to

the following decomposition of unity

$$\mathbb{1}_{\mathfrak{f}_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{\varrho}^{n}(\sigma),\boldsymbol{\varrho}^{n}(\eta))>0} = \sum_{\boldsymbol{x}} \sum_{\boldsymbol{\sigma}\ni(\sigma,\eta)} \frac{\mathbb{P}^{\mathfrak{f}_{\boldsymbol{A},\boldsymbol{B}}}[\boldsymbol{X}=\boldsymbol{x}] \mathbb{P}^{\boldsymbol{f}^{\boldsymbol{x}}}[\boldsymbol{\Sigma}^{\boldsymbol{x}}=\boldsymbol{\sigma}]}{\mathfrak{f}_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{\varrho}^{n}(\sigma),\boldsymbol{\varrho}^{n}(\eta)) f^{\boldsymbol{x}}(\sigma,\eta)}.$$
(5.137)

Hence, by Proposition 1.22, the capacity cap(A, B) is bounded from below by

$$\operatorname{cap}(A,B) \geq \sum_{\underline{x}} \mathbb{P}^{\mathfrak{f}_{A,B}} \left[ \boldsymbol{X} = \underline{x} \right] \mathbb{E}^{f^{\boldsymbol{x}}} \left[ \left( \sum_{t=a_{\boldsymbol{x}}}^{b_{\boldsymbol{x}}-1} \frac{\mathfrak{f}_{A,B} \left( \boldsymbol{x}(t), \boldsymbol{x}(t+1) \right) f^{\boldsymbol{x}} \left( \sigma(t), \sigma(t+1) \right)}{\mu_{N} \left( \sigma(t) \right) p_{N} \left( \sigma(t), \sigma(t+1) \right)} \right)^{-1} \right]$$
  
 
$$\geq \mathcal{E}^{n}(\tilde{g}) \sum_{\underline{x}} \mathbb{P}^{\mathfrak{f}_{A,B}} \left[ \boldsymbol{X} = \underline{x} \right] \left( \sum_{t=a_{\boldsymbol{x}}}^{b_{\boldsymbol{x}}-1} \mathbb{E}^{f^{\boldsymbol{x}}} \left[ \Phi_{t}^{\boldsymbol{x}} \right] \phi_{\boldsymbol{A},B} \left( \boldsymbol{x}(t), \boldsymbol{x}(t+1) \right) \right)^{-1}$$
(5.138)

with

$$\Phi_{t}^{\boldsymbol{x}}(\underline{\sigma}) = \frac{\mathcal{Q}^{n}(\boldsymbol{x}(t)) r^{n}(\boldsymbol{x}(t), \boldsymbol{x}(t+1))}{\mu_{N}(\sigma(t)) p_{N}(\sigma(t), \sigma(t+1))} f^{\boldsymbol{x}}(\sigma(t), \sigma(t+1))$$
(5.139)

where we applied in the last line Jensen's inequality and replaced  $\mathfrak{f}_{A,B}$  by the expression (5.130). Hence, it remains to construct for each macroscopic path  $\underline{x}$  a subordinate unit flow  $f^{\underline{x}}$ .

**Remark 5.23.** Let us consider for a moment the artificial case where all  $\tilde{h}_i \equiv 0$ . In view of (5.79), it holds that  $\sum_{\eta \in S^n[\boldsymbol{x}(t+1)]} p_N(\sigma, \eta) = r^n(\boldsymbol{x}(t), \boldsymbol{x}(t+1))$  for all  $\sigma \in S^n[\boldsymbol{x}(t)]$ . Therefore, we can define

$$f^{\boldsymbol{x}}(\sigma(t), \sigma(t+1)) := \frac{\mu_N(\sigma(t)) p_N(\sigma(t), \sigma(t+1))}{\mathcal{Q}^n(\boldsymbol{x}(t)) r^n(\boldsymbol{x}(t), \boldsymbol{x}(t+1))}$$

which is indeed a unit flow, because Kirchhoff's law is satisfied, i.e.

$$\sum_{\eta \in \mathcal{S}^n[\boldsymbol{x}(t-1)]} f^{\boldsymbol{x}}(\eta, \sigma(t)) = \frac{\mu_N(\sigma(t))}{\mathcal{Q}^n(\boldsymbol{x}(t))} = \frac{1}{|\mathcal{S}^n[\boldsymbol{x}(t)]|} = \sum_{\eta \in \mathcal{S}^n[\boldsymbol{x}(t+1)]} f^{\boldsymbol{x}}(\sigma(t), \eta).$$

( ( ) )

Note, that this choice of a subordinate flow implies that  $\Phi_t^{\underline{x}}(\underline{\sigma}) \equiv 1$  for all  $t = a_{\underline{x}}, \ldots, b_{\underline{x}}$ . Thus it is immediate that  $\operatorname{cap}(A, B) \geq \mathcal{E}^n(\tilde{g})$ .

In view of the assumption on the distribution of the magnetic field and the coarse graining procedure, there is some hope to construct a subordinate flow  $f^x$  such that  $\Phi_t^x(\underline{\sigma})$  is very close to one. Due to the fact that we are interested in proving a lower bound, it is enough to consider a subset of all realizations of the mesoscopic chain. This subset is chosen in such a way that its  $\mathbb{P}^{f_{A,B}}$ -probability tends to one as  $N \uparrow \infty$ , and that we are able to construct a subordinate flow such that uniformly for all those realizations  $\underline{x}$  it holds

$$\sum_{t=a_{\boldsymbol{x}}}^{b_{\boldsymbol{x}}-1} \mathbb{E}^{f^{\boldsymbol{x}}} \left[ \Phi_t^{\boldsymbol{x}} \right] \phi_{\boldsymbol{A},\boldsymbol{B}} \left( \boldsymbol{x}(t), \boldsymbol{x}(t+1) \right) \leq 1 + \mathcal{O}(\epsilon(n)).$$

Clearly, if we are able to find such a subset, we immediately obtain

$$\operatorname{cap}(A, B) \geq \mathcal{E}^n(\tilde{g}) \left(1 - \mathcal{O}(\epsilon(n))\right)$$

**5.4.2. Construction of the mesoscopic flow.** Let us point out that the construction of  $\mathfrak{f}$  and  $\mathfrak{f}_A, \mathfrak{f}_B$  presented in this subsection is an adaptation of the corresponding construction given in [6]. Our first aim is to construct a flow  $\mathfrak{f}$  on  $\Gamma^n$  within the set

$$\mathfrak{D}^{n} \equiv \mathfrak{D}^{n}(\boldsymbol{z},\nu\delta_{N}) := D^{n}(\boldsymbol{z},\delta_{N}) \cap \big\{\boldsymbol{x}\in\Gamma^{n} \mid |\langle\boldsymbol{x}-\boldsymbol{z},\check{\boldsymbol{v}}\rangle| < \nu\delta_{N}\big\},$$
(5.140)



FIGURE 3. Sketch of the set  $D^n(z, \delta_N)$  and the subset  $\mathfrak{D}^n$  that is used in the construction of the flow  $\mathfrak{f}$ . Notice that the set  $\mathfrak{D}_0^n$  is a narrow tube along the direction  $\check{v}^1$  lying inside  $\mathfrak{D}^n$ .

for some small number  $\nu > 0$ . Here,  $\check{v} \equiv \check{v}^1$  is the vector previously defined in (5.82). Note that the boundary of  $\mathfrak{D}^n$ , denoted by  $\partial \mathfrak{D}^n$ , consists of three disjoint parts. Namely,

$$\partial \mathfrak{D}^n = \partial_A \mathfrak{D}^n \cup \partial_B \mathfrak{D}^n \cup \partial_r \mathfrak{D}^n, \qquad (5.141)$$

where  $\partial_A \mathfrak{D}^n := \{x \in \partial \mathfrak{D}^n \mid \langle x - z, \check{v} \rangle \leq -\nu \delta_N \}$  and  $\partial_B \mathfrak{D}^n := \{x \in \partial \mathfrak{D}^n \mid \langle x - z, \check{v} \rangle \geq \nu \delta_N \}$ , while  $\partial \mathfrak{D}_r^n$  denotes all points x at the boundary of  $D^n(z, \delta_N)$  such that  $|\langle x - z, \check{v} \rangle| < \nu \delta_N$ , see Figure 5.4.2. We choose  $\nu$  small enough to guarantee that there exists K > 0 such that

$$F^n(\boldsymbol{x}) \geq F^n(\boldsymbol{z}) + K\delta_N^2, \quad \forall \boldsymbol{x} \in \partial_r \mathfrak{D}^n.$$
 (5.142)

Having in mind the definition of the harmonic flow (1.86) as well as Lemma 5.20, we will construct a non-negative unit flow  $\mathfrak{f}$  of the form

$$\mathfrak{f}(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}) := \frac{c(N)}{\mathcal{E}^n(\tilde{g})} \mathcal{F}[\phi](\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}),$$
(5.143)

for all k = 1, ..., n and  $(r, s) \in \Delta_k$  while  $\mathfrak{f} \equiv 0$  otherwise. Here we introduced the notation

$$\mathcal{F}[\phi]\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right) = \overline{\mathcal{Q}}^{n}(\boldsymbol{x})\,\bar{r}^{n}\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right)\phi\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right). \tag{5.144}$$

Notice that  $c(N) = 1 + o_N(1)$  is a suitable normalization constants.

Hence, our task is to specify a function  $\phi$  of the form  $\phi = \phi_T^0 + u$  where  $\phi_T^0$  is the truncation of a suitable function  $\phi^0$  and u should be a small correction. A natural candidate for  $\phi^0$  would be the discrete gradient of the test function g, see (5.97), because g is almost harmonic as shown in Lemma 5.20. However, due to the fact that we have to add a correction u anyway in order to satisfy Kirchhoff's law, it is more convenient to start with

$$\phi^{0}\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right) := \left(v_{r}^{k}-v_{s}^{k}\right)\sqrt{\frac{\beta|\hat{\gamma}_{1}|}{2\pi N}} \exp\left(-\frac{\beta N|\hat{\gamma}_{1}|}{2}\langle\boldsymbol{x}-\boldsymbol{z},\boldsymbol{v}\rangle^{2}\right)$$
(5.145)

for all k = 1, ..., n and  $(r, s) \in \Delta_k$ . In order to construct the correction u we proceed as follows. First, fix  $0 < \nu_0 \ll \nu$  small enough and define a narrow tube  $\mathfrak{D}_0^n$  along the direction



FIGURE 4. Sketch of the narrow tube  $\mathfrak{D}_0^n$  along the direction  $\check{v}^1$  together with the sets  $\mathfrak{D}_1^n$  and  $\mathfrak{D}_2^n$ .

*v*, i.e.

$$\mathfrak{D}_{0}^{n} \equiv \mathfrak{D}_{0}^{n}(\boldsymbol{z},\nu_{0}\delta_{N}) := \mathfrak{D}^{n} \cap \Big\{ \boldsymbol{x} \in \Gamma^{n} \mid \left\| \boldsymbol{x} - \boldsymbol{z} - \langle \boldsymbol{x} - \boldsymbol{z}, \check{\boldsymbol{v}} \rangle \frac{\check{\boldsymbol{v}}}{\|\check{\boldsymbol{v}}\|_{2}^{2}} \right\|_{2} < \nu_{0}\delta_{N} \Big\}, \quad (5.146)$$

and let C be the cone spanned by the vectors  $\{e^{k,s} - e^{k,r} \mid k = 1, ..., n, (r,s) \in \Delta_k\}$ . Mind that, under the above Assumption 5.18, the vector v lies in the interior of C. Further, we define

$$\mathfrak{D}_{1}^{n} := \left\{ \boldsymbol{x} - \mathcal{C} \mid \boldsymbol{x} \in \partial_{\boldsymbol{B}} \mathfrak{D}_{0}^{n} \right\} \cap \mathfrak{D}^{n}, \qquad \mathfrak{D}_{2}^{n} := \left\{ \boldsymbol{x} + \mathcal{C} \mid \boldsymbol{x} \in \partial_{\boldsymbol{A}} \mathfrak{D}_{1}^{n} \right\} \cap \mathfrak{D}^{n}, \qquad (5.147)$$

where  $\partial_A \mathfrak{D}_1^n$  denotes all points x at the boundary  $\partial \mathfrak{D}^n$  such that  $x - \hat{e}^{k,r} + \hat{e}^{k,s} \in \mathfrak{D}_1^n$  for all  $k = 1, \ldots, n$  and  $(r, s) \in \Delta_k$ . We assume that the constants  $\nu$  and  $\nu_0$  in the definitions of  $\mathfrak{D}^n$  and  $\mathfrak{D}_0^n$  are such that  $\mathfrak{D}_2^n \cap \partial_r \mathfrak{D}^n = \emptyset$ . Moreover, we denote by  $\phi_T^0$  the restriction of  $\phi^0$  to  $\mathfrak{D}_1^n \cup \partial_A \mathfrak{D}_1^n$ 

$$\phi_T^0(\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) = \phi^0(\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) \mathbb{1}_{\boldsymbol{x} \in \mathfrak{D}_1^n \cup \partial_{\boldsymbol{A}} \mathfrak{D}_1^n}.$$
(5.148)

**Remark 5.24.** Let us now explain briefly the idea behind the construction of the sets  $\mathfrak{D}_1^n$  and  $\mathfrak{D}_2^n$ . Consider a point  $x \in \mathfrak{D}_1^n$ . In order to satisfy Kirchhoff's law at x, the incoming flow should equal the outgoing flow, i.e.  $\sum_{y \in \Gamma^n} \mathfrak{f}(y, x) = \sum_{y \in \Gamma^n} \mathfrak{f}(x, y)$ . Recall that for a nonnegative flow  $\mathfrak{f}(x, y) > 0$  implies  $\mathfrak{f}(y, x) = 0$ . By construction of  $\mathfrak{D}_1^n$  and  $\phi_T^0$ , it holds that  $\{y \in \Gamma^n \mid \mathfrak{f}(y, x) > 0\} \subset \mathfrak{D}_1^n$ . Hence, the incoming flow arriving at x can only come from points in  $\mathfrak{D}_1^n$  whereas points in  $\mathfrak{D}_2^n$  cannot contribute to the incoming flow to x. Let us stress the fact that this property holds true for all  $x \in \mathfrak{D}_1^n$ . Hence, the total mass accumulated at  $\partial_B \mathfrak{D}_0^n$  can only come from  $\partial_A \mathfrak{D}_1^n$ . For this reason, we will distribute the initial mass one only among the points in  $\partial_A \mathfrak{D}_1^n$ .

On the other hand, concerning the outgoing flow, a mass transport from  $\mathfrak{D}_1^n$  to  $\mathfrak{D}_2^n \setminus \mathfrak{D}_1^n$  is possible for all points  $x \in \mathfrak{D}_1^n$  and  $y \in \mathfrak{D}_2^n$  such that  $r^n(x, y) > 0$ . Hence, while transporting the initial total mass one from  $\partial_A \mathfrak{D}_1^n$  to  $\partial_B \mathfrak{D}_0^n$  we may loose some mass. The reason behind the truncation of  $\phi^0$  to  $\mathfrak{D}_1^n$  is to minimize these losses. Further, since the total mass can only be transported to  $\partial_B \mathfrak{D}_2^n$  it is enough to construct the correction u on  $\mathfrak{D}_2^n$ . Now, we turn to the construction of the function u. In order to satisfy Kirchhoff's law inside  $\mathfrak{D}^n$ , the function u should satisfy the following recursion

$$\sum_{k=1}^{n} \sum_{(r,s)\in\Delta_{k}} \mathcal{F}[u](\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) = \sum_{k=1}^{n} \sum_{(r,s)\in\Delta_{k}} \mathcal{F}[u](\boldsymbol{x} + \hat{\boldsymbol{e}}^{k,r} - \hat{\boldsymbol{e}}^{k,s}, \boldsymbol{x}) - \mathrm{d}\mathcal{F}[\phi_{T}^{0}](\boldsymbol{x})$$
(5.149)

where

$$d\mathcal{F}[\phi_T^0](\boldsymbol{x}) := \sum_{k=1}^n \sum_{(r,s)\in\Delta_k} \Big(\mathcal{F}[\phi_T^0](\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) - \mathcal{F}[\phi_T^0](\boldsymbol{x} + \hat{\boldsymbol{e}}^{k,r} - \hat{\boldsymbol{e}}^{k,s}, \boldsymbol{x})\Big).$$
(5.150)

Since  $\phi_T^0 \equiv 0$  on  $\mathfrak{D}^n \setminus \mathfrak{D}_1^n$  and due to the construction of  $\mathfrak{D}_2^n$ , we may set  $u \equiv 0$  on  $\mathfrak{D}^n \setminus \mathfrak{D}_2^n$ . In the sequel, we solve the recursion (5.149) explicitly inside  $\mathfrak{D}_1^n$ . For this purpose, we slice  $\mathfrak{D}_1^n$  into layers  $\mathfrak{L}_k$  in the following way. Namely, set  $\mathfrak{L}_0 := \partial_A \mathfrak{D}_1^n$  and for  $j \in \mathbb{N}_0$  define

$$\mathfrak{L}_{j+1} := \left\{ \boldsymbol{x} \in \mathfrak{D}^n \mid \boldsymbol{x} + \hat{\boldsymbol{e}}^{k,r} - \hat{\boldsymbol{e}}^{k,s} \in \bigcup_{i=1}^j \mathfrak{L}_i, \ \forall k = 1, \dots, n, \ (r,s) \in \Delta_k \right\}.$$
(5.151)

Since v lies inside the cone C, there exists  $c_{19}(n) > 0$  and  $T \le c_{19}(n)N\delta_N$  such that

$$\mathfrak{D}_1^n = igcup_{j=0}^T \mathfrak{L}_j.$$

By setting  $u \equiv 0$  on  $\partial_A \mathfrak{D}^n$ , we define recursively, for every  $x \in \mathfrak{L}_{j+1}$ 

$$\mathcal{F}[u]\left(\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}\right) := q_{r,s}^{k} \bigg( \sum_{l=1}^{n} \sum_{(r',s') \in \Delta_{k}} \mathcal{F}[u]\left(\boldsymbol{x} + \hat{\boldsymbol{e}}^{l,r'} - \hat{\boldsymbol{e}}^{l,s'}, \boldsymbol{x}\right) - \mathrm{d}\mathcal{F}[\phi_{T}^{0}](\boldsymbol{x}) \bigg),$$
(5.152)

where the probability distribution  $\{q_{r,s}^k \mid k = 1, ..., n, (r,s) \in \Delta_k\}$  is defined by

$$q_{r,s}^{k} = \frac{r_{r,s}^{k} \left( v_{r}^{k} - v_{s}^{k} \right)}{\sum_{l=1}^{n} \sum_{(t,u) \in \Delta_{k}} r_{t,u}^{l} \left( v_{t}^{l} - v_{u}^{l} \right)}.$$
(5.153)

Obviously, this produces a solution to (5.149) on  $\mathfrak{D}_1^n$  whereas on  $\mathfrak{D}_2^n \setminus \mathfrak{D}_1^n u$  is implicitly given by the solution of (5.149) subject to additional boundary conditions on  $\partial_r \mathfrak{D}_2^n$  and  $\partial_r \mathfrak{D}_1^n$ . Let us emphasis the fact that the positivity of  $q_{r,s}^k$  relies crucially on the Assumption 5.18.

**Proposition 5.25.** The flow f, as constructed above, satisfies the following properties:

(i) For all  $x \in \mathfrak{D}^n$  it holds that

$$d\mathcal{F}[\phi](\boldsymbol{x}) = 0 \quad and \quad \sum_{\boldsymbol{x}\in\partial_{\boldsymbol{A}}\mathfrak{D}^{n}}\sum_{(r,s)\in\Delta_{k}}\mathfrak{f}(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}) = 1. \quad (5.154)$$

(ii) Uniformly for all  $x \in \mathfrak{D}_0^n$  it holds that

$$\phi(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}) = \left(g(\boldsymbol{x}) - g\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right)\right) \left(1+\mathcal{O}_N(1)\right),$$
(5.155)

for all  $k = 1, \ldots, n$  and  $(r, s) \in \Delta_k$ .

(iii) There exists  $\kappa > 0$  such that

$$\max_{\boldsymbol{x}\in\mathfrak{D}^n\setminus\mathfrak{D}_0^n}\max_{k=1,\ldots,n}\max_{(r,s)\in\Delta_k}\mathfrak{f}(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}) \leq N^{-\kappa}.$$
(5.156)

*Proof.* (i) As an immediate consequence of the construction of  $\phi$ , the first statement in (5.154) is satisfied for all  $x \in \mathfrak{D}^n$ , i.e.  $\mathfrak{f}$  is a non-negative, cycle-free flow. It remains to prove that  $\mathfrak{f}$  is a unit flow.

By inspecting the proof of Proposition 5.21 we see that  $\mathcal{E}^n(\tilde{g}) = \overline{\mathcal{E}}_{\mathfrak{D}^n \cup \partial_A \mathfrak{D}^n}^n(g) (1 + \mathcal{O}_N(1))$ . On the other hand, Green's first identify implies that

$$\begin{split} \overline{\mathcal{E}}_{\mathfrak{D}^{n} \cup \partial_{A}\mathfrak{D}^{n}}^{n}(g) &+ \sum_{\boldsymbol{x} \in \mathfrak{D}^{n}} \overline{\mathcal{Q}}^{n}(\boldsymbol{x}) g(\boldsymbol{x}) \left(\overline{L}^{n}g\right)(\boldsymbol{x}) \\ &= -\sum_{\boldsymbol{x} \in \partial_{B}\mathfrak{D}^{n}} \sum_{k=1}^{n} \sum_{(r,s) \in \Delta_{k}(\boldsymbol{x})} \overline{\mathcal{Q}}^{n}(\boldsymbol{x}) g(\boldsymbol{x}) \bar{r}^{n}(\boldsymbol{x}, \boldsymbol{x} + \hat{\boldsymbol{e}}^{k,r} - \hat{\boldsymbol{e}}^{k,s}) \left(g\left(\boldsymbol{x} + \hat{\boldsymbol{e}}^{k,r} - \hat{\boldsymbol{e}}^{k,s}\right) - g(\boldsymbol{x})\right) \\ &- \sum_{\boldsymbol{x} \in \partial_{A}\mathfrak{D}^{n}} \sum_{k=1}^{n} \sum_{(r,s) \in \Delta_{k}(\boldsymbol{x})} \overline{\mathcal{Q}}^{n}(\boldsymbol{x}) g(\boldsymbol{x}) \bar{r}^{n}(\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}) \left(g\left(\boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}\right) - g(\boldsymbol{x})\right), \end{split}$$
(5.157)

where, for  $x \in \partial_A \mathfrak{D}^n$  and  $k \in \{1, ..., n\}$ ,  $\Delta_k(x)$  denotes the subset of all  $(r, s) \in \Delta_k$  such that  $x - \hat{e}^{k,r} + \hat{e}^{k,s} \in \mathfrak{D}^n$  and analog for  $x \in \partial_B \mathfrak{D}^n$ . Recall that the function g converge exponentially fast in N to 0 for all  $x \in \partial_B \mathfrak{D}^n$  and to 1 for all  $x \in \partial_A \mathfrak{D}^n$ . Hence, the first term on the right-hand side of (5.157) is negligible and, in the second term, we can replace g(x) by one. Moreover, a comparison of (5.105) with (5.114) reveals that

$$\left|\sum_{\boldsymbol{x}\in\mathfrak{D}^{n}}\overline{\mathcal{Q}}^{n}(\boldsymbol{x})g(\boldsymbol{x})\left(\overline{L}^{n}g\right)(\boldsymbol{x})\right| = \overline{\mathcal{E}}_{\mathfrak{D}^{n}\cup\partial_{\boldsymbol{A}}\mathfrak{D}^{n}}^{n}(g)\left(1+\mathcal{O}\left(N^{-1/2+2\delta}\right)\right).$$
(5.158)

Hence,

$$\overline{\mathcal{E}}_{\mathfrak{D}^n \cup \partial_{\boldsymbol{A}}\mathfrak{D}^n}^n(g) \left(1 + \mathcal{O}_N(1)\right) = -\sum_{\boldsymbol{x} \in \partial_{\boldsymbol{A}}\mathfrak{D}^n} \sum_{k=1}^n \sum_{(r,s) \in \Delta_k(\boldsymbol{x})} \mathcal{F}[\nabla g] \left(\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}\right).$$
(5.159)

Here, we introduced the discrete gradient,  $\nabla g$ , that is defined by  $(\nabla g)(x, y) := g(y) - g(x)$ . By employing the construction of  $\phi$  we infer that

$$\sum_{\boldsymbol{x}\in\partial_{\boldsymbol{A}}\mathfrak{D}^{n}}\sum_{k=1}^{n}\sum_{(r,s)\in\Delta_{k}(\boldsymbol{x})}\mathcal{F}[\phi](\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}) = \sum_{\boldsymbol{x}\in\partial_{\boldsymbol{A}}\mathfrak{D}_{1}^{n}}\sum_{k=1}^{n}\sum_{(r,s)\in\Delta_{k}(\boldsymbol{x})}\mathcal{F}[\phi^{0}](\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}).$$
(5.160)

Notice that by (5.109)

$$\phi^0\left(\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}\right) = \left(g(\boldsymbol{x}) - g\left(\boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s}\right)\right) \left(1 + \mathcal{O}(\delta_N)\right)$$
(5.161)

uniformly for all  $x \in \mathfrak{D}^n$ . Therefore, we are left with showing that the contribution coming from  $\partial_A \mathfrak{D}^n \setminus \partial_A \mathfrak{D}_1^n$  on the right-hand side of (5.159) is small. By using that

$$-\mathcal{F}[\nabla g](\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}) = \mathcal{F}[\phi^0](\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s})(1+\mathcal{O}(\delta_N))$$

together with the non-orthogonal decomposition (5.84) we have that

$$\frac{\mathcal{F}[\phi^{0}]\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right)}{\mathcal{Q}^{n}(\boldsymbol{z})r_{r,s}^{k}\left(v_{r}^{k}-v_{s}^{k}\right)} = \sqrt{\frac{\beta|\hat{\gamma}_{1}|}{2\pi N}} \exp\left(-\frac{\beta N}{2}\left\langle\boldsymbol{x}-\boldsymbol{z},\left(A^{n}(\boldsymbol{z})+|\hat{\gamma}_{1}|\boldsymbol{v}\cdot\boldsymbol{v}^{\mathrm{T}}\right)(\boldsymbol{x}-\boldsymbol{z})\right\rangle\right) \\
= \sqrt{\frac{\beta|\hat{\gamma}_{1}|}{2\pi N}} \exp\left(-\frac{\beta N}{2}\sum_{i=2}^{n\cdot(q-1)}\hat{\gamma}_{i}\left\langle\boldsymbol{x}-\boldsymbol{z},\boldsymbol{v}\right\rangle^{2}\right). \quad (5.162)$$

Taking advantage of the definition of  $\mathfrak{D}_0^n$  we conclude that there exists c>0 such that

$$\frac{\mathcal{F}[\phi^0](\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s})}{\mathcal{Q}^n(\boldsymbol{z}) r_{r,s}^k \left(v_r^k - v_s^k\right)} \leq \exp\left(-c N^{2\delta}\right)$$
(5.163)

uniformly for all  $x \in \mathfrak{D}^n \setminus \mathfrak{D}_0^n$ , k = 1, ..., n and  $(r, s) \in \Delta_k$ . Hence, by choosing the constant, c(N), in the definition of  $\mathfrak{f}$ , see (5.143), appropriately, (5.154) follows.

(ii) Notice that a computation along the lines of the proof of Lemma 5.20 reveals that there exists  $c_{20} > 0$  such that

$$\frac{\left|\mathrm{d}\mathcal{F}\left[\phi^{0}\right](\boldsymbol{x})\right|}{\mathcal{F}\left[\phi^{0}\right](\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right)} = \left|\left\langle \boldsymbol{v}, R\left(A^{n}(\boldsymbol{z})+|\hat{\gamma}_{1}|\boldsymbol{v}\cdot\boldsymbol{v}^{\mathrm{T}}\right)(\boldsymbol{x}-\boldsymbol{z})\right\rangle + \mathcal{O}(\delta_{N}^{2})\right| \leq c_{20}\,\delta_{N}^{2}$$
(5.164)

uniformly for  $x \in \mathfrak{D}^n$ , k = 1, ..., n and  $(r, s) \in \Delta_k$ . Here, the second step relies on (5.111). In order to establish (5.155) and (5.156) we prove recursively a bound on the correction u. To start with, let  $d_j$  be the smallest constants such that for all  $x \in \mathfrak{L}_j$ 

$$\left|\mathcal{F}[u]\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right)\right| \leq d_j \,\delta_N^2 \,\mathcal{F}\left[\phi^0\right]\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right),\tag{5.165}$$

for all k = 1, ..., n and  $(r, s) \in \Delta_k$ . Then, for any  $x \in \mathfrak{L}_{j+1} \cap \mathfrak{D}_1^n$ , the construction of u given in (5.152) combined with (5.164) implies

$$\frac{\left|\mathcal{F}[u](\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s})\right|}{\mathcal{F}[\phi^{0}](\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s})} \leq q_{r,s}^{k} \left(\sum_{l=1}^{n} \sum_{(r',s')\in\Delta_{k}} \frac{\left|\mathcal{F}[u](\boldsymbol{x} + \hat{\boldsymbol{e}}^{l,r'} - \hat{\boldsymbol{e}}^{l,s'}, \boldsymbol{x})\right|}{\mathcal{F}[\phi^{0}](\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s})} + c_{20} \,\delta_{N}^{2}\right) \\ \leq \delta_{N}^{2} \left(d_{j} \, q_{r,s}^{k} \sum_{l=1}^{n} \sum_{(r',s')\in\Delta_{k}} \frac{\mathcal{F}[\phi^{0}](\boldsymbol{x} + \hat{\boldsymbol{e}}^{l,r} - \hat{\boldsymbol{e}}^{l,s}, \boldsymbol{x})}{\mathcal{F}[\phi^{0}](\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r'} + \hat{\boldsymbol{e}}^{k,s'})} + c_{20}\right). \tag{5.166}$$

By plugging in the definition of  $\phi^0$ , we have that

$$q_{r,s}^{k} \sum_{l=1}^{n} \sum_{(r',s')\in\Delta_{k}} \frac{\mathcal{F}[\phi^{0}](\boldsymbol{x}+\hat{\boldsymbol{e}}^{l,r}-\hat{\boldsymbol{e}}^{l,s},\boldsymbol{x})}{\mathcal{F}[\phi^{0}](\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r'}+\hat{\boldsymbol{e}}^{k,s'})} \\ = 1 + q_{r,s}^{k} \frac{\langle \boldsymbol{v}, R\left(A^{n}(\boldsymbol{z})+|\hat{\gamma}_{1}|\boldsymbol{v}\cdot\boldsymbol{v}^{\mathrm{T}}\right)(\boldsymbol{x}-\boldsymbol{z})\rangle}{r_{r,s}^{k}(v_{r}^{k}-v_{s}^{k})} + \mathcal{O}(\delta_{N}^{2}) = 1 + \mathcal{O}(\delta_{N}^{2})$$
(5.167)

uniformly for all  $x \in \mathfrak{D}_1^n$ , k = 1, ..., n and  $(r, s) \in \Delta_k$ . Here, the first step holds true due to the choice of the probability distribution,  $q_{r,s}^k$ , while in the second step we exploit again (5.111). Thus, the constants  $d_j$  satisfy the following recursive bound

$$d_0 = 0, \qquad d_{j+1} \le d_j \left( 1 + \mathcal{O}(\delta_N^2) \right) + c_{20}.$$
 (5.168)

Since  $T \leq c_{19}(n)N\delta_N$ , we obtain from (5.168) that for all  $1 \leq j \leq T$  and N sufficiently large

$$d_j \leq j c_{20} e^{j \mathcal{O}(\delta_N^2)} \leq T c_{20} e^{T \mathcal{O}(\delta_N^2)} \leq \mathcal{O}(1/\delta_N).$$
(5.169)

As a result, the function u satisfies

$$\left|\mathcal{F}[u]\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right)\right| \leq \mathcal{O}(\delta_N^2) \mathcal{F}[\phi^0]\left(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\right)$$
(5.170)

uniformly for all  $x \in \mathfrak{D}_1^n$ , k = 1, ..., n and  $(r, s) \in \Delta_k$ . In particular, together with (5.161) this concludes the proof of (5.155).

(iii) In view of (5.170), (5.163) and (5.118), (5.156) is satisfied uniformly on  $\mathfrak{D}_1^n \setminus \mathfrak{D}_0^n$ . Moreover, due to our construction of the flow,  $\phi \equiv 0$  on  $\mathfrak{D}^n \setminus \mathfrak{D}_2^n$  and (5.156) is immediately satisfied in the later domain. Thus, it remains to control  $\mathfrak{f}$  on  $\mathfrak{D}_2^n \setminus \mathfrak{D}_1^n$ . Since  $\phi(\boldsymbol{x}) = u(\boldsymbol{x})$  and  $d\mathcal{F}[u](\boldsymbol{x}) = 0$  for all  $\boldsymbol{x} \in \mathfrak{D}_2^n \setminus \mathfrak{D}_1^n$ , in the worst case, the flow  $\mathfrak{f}(\boldsymbol{x}, \boldsymbol{x} - \hat{\boldsymbol{e}}^{k,r} + \hat{\boldsymbol{e}}^{k,s})$  could be equal to the total flow through  $\mathfrak{D}_2^n \setminus \mathfrak{D}_1^n$  that is

$$\sum_{\boldsymbol{x}\in\mathfrak{D}_{1}^{n}}\sum_{k=1}^{n}\sum_{(r,s)\in\Delta_{k}}\mathfrak{f}(\boldsymbol{x},\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s})\ \mathbb{1}_{\boldsymbol{x}-\hat{\boldsymbol{e}}^{k,r}+\hat{\boldsymbol{e}}^{k,s}\in\mathfrak{D}_{2}^{n}}.$$
(5.171)

However, due to (5.170) and (5.163), the latter is of order  $\mathcal{O}(N^{n/2} e^{-cN^{2\delta}})$ . Hence, by combining the estimates on the different domains, we conclude that there exists  $\kappa > 0$  such that (5.156) holds.

**Corollary 5.26.** Let  $\mathfrak{f}$  be the flow constructed above. Then, for the associated Markov chain with the law  $\mathbb{P}^{\mathfrak{f}}$  it holds that

$$\mathbb{P}^{\mathfrak{f}}\left[\sum_{(\boldsymbol{x},\boldsymbol{y})\in\boldsymbol{\gamma}}\phi(\boldsymbol{x},\boldsymbol{y})=1+\mathcal{O}_{N}(1)\right]=1-\mathcal{O}_{N}(1).$$
(5.172)

*Proof.* By (5.154),  $\mathfrak{f}$  is a unit flow that is non-negative by construction. Hence, we can associate a Markov chain  $\mathbf{X} = {\mathbf{X}(t)}$  to it. For some  $0 < \varepsilon \ll 1$ , we denote by  $\partial_{\mathbf{A}} \mathfrak{D}_{0,\varepsilon}^n \equiv \partial_{\mathbf{A}} \mathfrak{D}_0^n (\mathbf{z}, \varepsilon \nu_0 \delta_N)$  the set of all configurations at the boundary  $\partial_{\mathbf{A}} \mathfrak{D}_0^n$  that lies deeply inside the narrow tube. Now, let us consider the set of all realizations of  $\mathbf{X}$  from  $\partial_{\mathbf{A}} \mathfrak{D}_{0,\varepsilon}^n$  to  $\partial_{\mathbf{B}} \mathfrak{D}_0^n$  that stays inside  $\mathfrak{D}_0^n$ , i.e.

$$\mathcal{P}_{\boldsymbol{A},\boldsymbol{B}} := \left\{ \underline{\boldsymbol{x}} \mid \boldsymbol{x}(a_{\underline{\boldsymbol{x}}}) \in \partial_{\boldsymbol{A}} \mathfrak{D}_{0,\varepsilon}^{n}, \ \boldsymbol{x}(b_{\underline{\boldsymbol{x}}}) \in \partial_{\boldsymbol{B}} \mathfrak{D}_{0}^{n}, \ \boldsymbol{x}(t) \in \mathfrak{D}_{0}^{n}, \ \forall \ a_{\underline{\boldsymbol{x}}} < t < b_{\underline{\boldsymbol{x}}} \right\}.$$

In view of (5.155), we have for all  $\gamma \in \mathcal{P}_{A,B}$  that

$$\sum_{\boldsymbol{x},\boldsymbol{y})\in\boldsymbol{\gamma}}\phi(\boldsymbol{x},\boldsymbol{y}) = \left(g(a_{\underline{x}}) - g(b_{\underline{x}})\right)\left(1 + \mathcal{O}_N(1)\right) = 1 + \mathcal{O}_N(1), \quad (5.173)$$

where we used that the function g, as defined in (5.97), converge exponentially fast in N to one for all  $x \in \partial_A \mathfrak{D}^n$  and to zero for all  $x \in \partial_B \mathfrak{D}^n$ .

Hence, it remains to show that the  $\mathbb{P}^{\mathfrak{f}}$ -probability of the event  $\mathcal{P}_{A,B}$  is close to one. Recall that the initial distribution of the Markov chain X is given by  $\sum_{\boldsymbol{y}\in\mathfrak{D}^n}\mathfrak{f}(\boldsymbol{x},\boldsymbol{y})$  for all  $\boldsymbol{x}\in\partial_A\mathfrak{D}^n$  and its transition probabilities are given by  $q^{\mathfrak{f}}(\boldsymbol{x},\boldsymbol{y}) = \mathfrak{f}(\boldsymbol{x},\boldsymbol{y})/\sum_{\boldsymbol{y}'}\mathfrak{f}(\boldsymbol{x},\boldsymbol{y}')$ . Since the flow  $\mathfrak{f}$  is defined in terms of  $\phi^0$  on  $\partial_A\mathfrak{D}^n_0$ , (5.162) implies that the set of all path starting in  $\partial_A\mathfrak{D}^n_{0,\varepsilon}$  is negligible with respect to  $\mathbb{P}^{\mathfrak{f}}$ . Now, set  $\boldsymbol{Y}(t) := \boldsymbol{X}(t) - \check{\boldsymbol{v}} \langle \boldsymbol{X}(t), \check{\boldsymbol{v}} \rangle / \|\check{\boldsymbol{v}}\|_2^2$  and decompose  $\boldsymbol{Y}(t) = \boldsymbol{Y}(0) + \boldsymbol{M}(t) + \boldsymbol{N}(t)$  into a martingale  $\boldsymbol{M}(t)$  and a previsible process  $\boldsymbol{N}(t)$ . As a consequence of (5.170), for all  $t \leq T$ ,  $\|\boldsymbol{N}(t)\| \leq T/N\mathcal{O}(\delta_N) = \delta_N \mathcal{O}(\delta_N)$  on the event  $\{\boldsymbol{X}(t) \in \mathfrak{D}^n_1, \forall 0 \leq t \leq T\}$ . Hence, Doob's maximums inequality for submartingales implies

$$\mathbb{P}\left[\max_{0 \le t \le T} \left\| \boldsymbol{Y}(t) \right\| < \frac{1}{2} \nu_0 \delta_N \right] \ge 1 - \mathbb{P}\left[\max_{0 \le t \le T} \left\| \boldsymbol{M}(t) \right\|^2 \ge c^2 \delta_N^2 \right] \\
\ge 1 - \frac{1}{c^2 \delta_N^{-2}} \mathbb{E}\left[ \left\| \boldsymbol{M}(T) \right\|^2 \right] \\
\ge 1 - \mathcal{O}(N^{-1/2 - \delta}),$$
(5.174)

where  $c < \nu_0(1/2 - \varepsilon) - \mathcal{O}(\delta_N)$  and  $\mathbb{E}[\|\boldsymbol{M}(T)\|^2] \leq T \mathcal{O}(N^{-2})$ . This completes the proof.  $\Box$ 

The remaining part of this subsection is devoted to the construction of the flow  $\mathfrak{f}_A$  from A to  $\partial_A \mathfrak{D}^n$  of the form

$$\mathfrak{f}_{\boldsymbol{A}}(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathcal{Q}^{n}(\boldsymbol{x}) r^{n}(\boldsymbol{x},\boldsymbol{y})}{\mathcal{E}^{n}(\tilde{g})} \phi_{\boldsymbol{A}}(\boldsymbol{x},\boldsymbol{y}) \left(1 + \mathcal{O}_{N}(1)\right)$$
(5.175)

and a of the corresponding flow  $\mathfrak{f}_B$  from  $\partial_B \mathfrak{D}^n$  to B such that the concatenation  $\mathfrak{f}_{A,B} = {\mathfrak{f}_A, \mathfrak{f}, \mathfrak{f}_B}$  satisfies Kirchhoffs law and

$$\mathbb{P}^{\mathfrak{f}_{\boldsymbol{A}}}\left[\sum_{(\boldsymbol{x},\boldsymbol{y})\in\boldsymbol{\gamma}}\phi_{\boldsymbol{A}}(\boldsymbol{x},\boldsymbol{y})=\mathcal{O}_{N}(1)\right]=1-\mathcal{O}_{N}(1)$$
(5.176)

and

$$\mathbb{P}^{\mathfrak{f}_{B}}\left[\sum_{(\boldsymbol{x},\boldsymbol{y})\in\boldsymbol{\gamma}}\phi_{B}(\boldsymbol{x},\boldsymbol{y})=\mathcal{O}_{N}(1)\right]=1-\mathcal{O}_{N}(1).$$
(5.177)



FIGURE 5. Construction of the paths  $\hat{\gamma}^x = \{\hat{\gamma}, \hat{\eta}^x\}$  for each  $x \in \partial_A \mathfrak{D}_0^n$ . While  $\hat{\gamma}$  is defined in terms of the minimal energy curve that ends in some point  $y \in D^n$ ,  $\hat{\eta}^x$  is just the linear interpolation between y and x.

Together with (5.172) these estimates imply that (5.131) holds true. Since the construction of  $\mathfrak{f}_A$  and  $\mathfrak{f}_B$  are completely similar, we will work out the details only for  $\mathfrak{f}_A$ .

The strategy is the following. To each  $x \in \partial_A \mathfrak{D}^n$  associate a nearest neighbor path  $\hat{\gamma}^x = (\hat{\gamma}^x(a^x), \dots, \hat{\gamma}^x(0))$  on  $\Gamma^n$  that starts in A and ends in x, i.e.

$$\hat{\boldsymbol{\gamma}}^{\boldsymbol{x}}(a^{\boldsymbol{x}}) \in \boldsymbol{A}, \quad \hat{\boldsymbol{\gamma}}^{\boldsymbol{x}}(0) = \boldsymbol{x} \quad \text{and} \quad \left\| \hat{\boldsymbol{\gamma}}^{\boldsymbol{x}}(t+1) - \hat{\boldsymbol{\gamma}}^{\boldsymbol{x}}(t) \right\|_{1} = \frac{2}{N}, \quad \forall t = a^{\boldsymbol{x}}, \dots, -1.$$

Then, the flow from A to  $\partial_A \mathfrak{D}^n$  is given by

$$\mathfrak{f}_{oldsymbol{A}}(oldsymbol{y},oldsymbol{y}') \ \coloneqq \ \sum_{oldsymbol{x}\in\partial_{oldsymbol{A}}\mathfrak{D}_0^n} \mathbb{1}_{(oldsymbol{y},oldsymbol{y}')\in\hat{oldsymbol{\gamma}}^{oldsymbol{x}}} \sum_{k=1}^n \sum_{(r,s)\in\Delta_k(oldsymbol{x})} \mathfrak{f}oldsymbol{x},oldsymbol{x}-\hat{oldsymbol{e}}^{k,r}+\hat{oldsymbol{e}}^{k,s}oldsymbol{)}.$$

Notice that the flow  $\mathfrak{f}_{\mathcal{A}}$  constructed above as well as its concatenation  $\{\mathfrak{f}_{\mathcal{A}},\mathfrak{f}\}\$  with the flow  $\mathfrak{f}$  satisfies Kirchhoff's law. The idea, we have in mind to satisfy (5.176), is to choose the paths  $\hat{\gamma}^{\boldsymbol{x}}$  is such a way that  $F^n(\boldsymbol{z}) - F^n(\hat{\gamma}^{\boldsymbol{x}}(t)) > \varepsilon \delta_N^2$  for all  $t = a^{\boldsymbol{x}}, \ldots, 0$  for some  $\varepsilon > 0$  small enough.

In the sequel, we describe the actual construction of the family of paths  $\{\hat{\gamma}^x\}$  for  $x \in \partial_A \mathfrak{D}^n$ . Fix a  $x \in \partial_A \mathfrak{D}^n$ . The path  $\hat{\gamma}^x$  associated to x consists of a concatenation of two paths  $\hat{\gamma}^x = \{\hat{\gamma}, \hat{\eta}^x\}$  where  $\hat{\gamma}$  connects A with some point  $y \in \partial_A D^n$  and  $\hat{\eta}^x$  runs from y to  $x \in \partial_A \mathfrak{D}^n$ , see Figure 5.4.2.

Let  $m \equiv m(\mathbf{m}) \in \mathcal{M}_1$  and  $m' \equiv m'(\mathbf{m}') \in \mathcal{M}_1$  be local minima of the function  $F_N$  that correspond to the minima  $\mathbf{m}, \mathbf{m}'$  of  $F^n$ . In view of Lemma 5.8 it holds that  $m(\mathbf{m}) = \sum_{k=1}^n m^k$ and likewise for m'. In order to construct  $\hat{\gamma}$ , consider the minimal energy curve  $\varphi : [0, 1] \to \mathcal{M}_1$ with respect to  $F_N$  that connects the minimum m and m'. In particular, this curve passes through the saddle point  $z \equiv z(z)$ . Further, the mesoscopic path  $\varphi$  corresponding to  $\varphi$  is uniquely defined by (5.54) and (5.55). We denote its nearest neighbor path approximation in  $\Gamma^n$  by  $\hat{\varphi}$ . Let  $\mathbf{a} \in \mathbf{A}$  be the last point before  $\hat{\varphi}$  leaves the set  $\mathbf{A}$  and  $\mathbf{y} \in D^n$  be the first point where  $\hat{\varphi}$  enters the set  $D^n$ . Then, the path  $\hat{\gamma}$  is given by the portion of the path  $\hat{\varphi}$  from  $\mathbf{a}$  to  $\mathbf{y}$ . On the other hand, for any  $\mathbf{x} \in \partial_{\mathbf{A}} \mathfrak{D}^n$ , let  $\eta^{\mathbf{x}} : [0,1] \to \bigotimes_{k=1}^n \mathcal{M}_{\pi_k}, t \mapsto \mathbf{y} + t(\mathbf{x} - \mathbf{y})$  be the line segment that connects y and x. Then,  $\hat{\eta}^x$  is defined as the nearest neighbor approximation of  $\eta^x$  in  $\Gamma^n$ .

**Lemma 5.27.** Let  $\mathfrak{f}_A$  be the flow constructed above. Then, for the associated Markov chain with law  $\mathbb{P}^{\mathfrak{f}_A}$ , (5.176) is satisfied.

*Proof.* First of all note that, in view of (5.156), the contribution to the flow  $f_A$  coming from paths  $\hat{\gamma}^x$  with  $x \in \partial_A \mathfrak{D}^n \setminus \partial_A \mathfrak{D}^n_0$  is negligible. By plugging the expressions for  $\mathcal{Q}^n(x)$ ,  $r^n(x, y)$  and  $\mathcal{E}^n(\tilde{g})$  given by (5.45), (5.79) and (5.118) into the right-hand side of (5.175), we obtain immediately an upper bound for  $\phi_A(x, y)$ . Namely,

$$\phi_{\boldsymbol{A}}(\boldsymbol{x},\boldsymbol{y}) \leq 2 \frac{\mathcal{E}^{n}(\tilde{g}) \, \boldsymbol{\mathfrak{f}}_{\boldsymbol{A}}(\boldsymbol{x},\boldsymbol{y})}{\mathcal{Q}^{n}(\boldsymbol{x}) \, r^{n}(\boldsymbol{x},\boldsymbol{y})} \leq c \, N^{n(q-1)/2} \, \exp\Big(-\beta N \big(F^{n}(\boldsymbol{z}) - F^{n}(\boldsymbol{x})\big)\Big). \tag{5.178}$$

Thus, it remains to show that there exists  $\varepsilon > 0$  such that  $F^n(z) - F^n(\hat{\gamma}(t)) > \varepsilon \delta_N^2$  for all  $t = a^x, \ldots, 0$ .

As a consequence of the fact that  $F_N$  is essential smooth with respect to  $\mathcal{M}_1$ , there exists a  $\delta > 0$  such that distance between the minimal energy curve and the boundary of  $\mathcal{M}_1$  is at least  $\delta$ . Thus, by construction, the minimal distance between  $\varphi$  and the boundary of  $\bigotimes_{k=1}^n \mathcal{M}_{\pi_k}$  is at least  $\delta \min_k \pi_k$ . In particular,  $\|\nabla F_N(\varphi(t))\|$  is for all  $t \in [0, 1]$  uniformly bounded from above by a constant that is independent of N. Recall that z is the unique minimal saddle point between the sets A and B. By exploiting the boundedness of  $\|\nabla F^n(\varphi(t))\|$  and the fact that the Hausdorff distance between  $\varphi$  and  $\hat{\gamma}$  is at most  $\sqrt{n(q-1)/q}/N$ , we conclude that  $\max_{x \in \hat{\gamma}} F^n(x) \leq F^n(y) + \mathcal{O}(n/N)$ .

Moreover, assuming that the parameter  $\nu$  in the definition of  $\mathfrak{D}^n(\boldsymbol{z},\nu\delta_N)$  is so small that  $\mathfrak{D}^n$  lies deeply inside the set  $D^n(\boldsymbol{z},\delta_N)$ , we can ensure that

$$F^{n}(\boldsymbol{y}) < \min_{\boldsymbol{x} \in \partial_{\boldsymbol{A}} \mathfrak{D}_{0}^{n}} F^{n}(\boldsymbol{x}).$$
(5.179)

Further, by choosing  $\nu_0$  small enough, we can ensure that  $\langle A^n(\boldsymbol{z})(\boldsymbol{x}-\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{z} \rangle > 0$  uniformly for all  $\boldsymbol{x} \in \partial_{\boldsymbol{A}} \mathfrak{D}_0^n$ . Hence,

$$[0,1] \ni t \mapsto \langle A^n(\boldsymbol{z}) (\boldsymbol{y} - \boldsymbol{z} + t(\boldsymbol{x} - \boldsymbol{y})), (\boldsymbol{y} - \boldsymbol{z} + t(\boldsymbol{x} - \boldsymbol{y})) \rangle \leq \langle A^n(\boldsymbol{z}) (\boldsymbol{x} - \boldsymbol{z}), \boldsymbol{x} - \boldsymbol{z} \rangle,$$
(5.180)

i.e. up to negligible corrections  $F^n(\boldsymbol{y} + t(\boldsymbol{x} - \boldsymbol{y})) \leq F^n(\boldsymbol{x})$  uniformly for all  $t \in [0, 1]$  and  $\boldsymbol{x} \in \partial_{\boldsymbol{A}} \mathfrak{D}_0^n$ . Hence,  $\max_{\boldsymbol{x}' \in \hat{\gamma}^{\boldsymbol{x}}} F^n(\boldsymbol{x}') \leq F^n(\boldsymbol{x}) + \mathcal{O}(n/N)$  and there exists a  $\varepsilon > 0$  such that  $F^n(\boldsymbol{z}) - F^n(\boldsymbol{x}) > \varepsilon \delta_N$  uniformly for all  $\boldsymbol{x} \in \partial_{\boldsymbol{A}} \mathfrak{D}_0^n$ .

**5.4.3. Statistics of mesoscopic trajectories.** In order to construct a subordinate flow it suffices to consider a certain subset of all realizations of the mesoscopic chain  $\{X(t)\}$  that starts in A and is stopped at the arrival of B. Recall our convention to choose the label of a realization  $\underline{x} = (x(a_{\underline{x}}), \ldots, x(0), \ldots, x(b_{\underline{x}}))$  such that  $x(a_{\underline{x}}) \in A$ ,  $x(b_{\underline{x}}) \in B$  and  $\sum_{k=1}^{n} x(0)^{k} = \sum_{k=1}^{n} z^{k}$ . For a given path  $\underline{x}$ , we denote by  $\tau_{k}[T] \equiv \tau_{k}[T](\underline{x})$  the number of changed in the k's block compared to x(0) along the path  $\underline{x}$  until time T, i.e. for T > 0 let

$$\tau_k[T] = \frac{N}{2} \sum_{t=0}^{T-1} \left\| x^k(t) - x^k(t-1) \right\|_1$$

while for T < 0 the sum runs from T + 1 to 0.

The set of paths, we are focusing on in the construction of a subordinate flow, are characterized by the following

**Definition 5.28.** A mesoscopic path  $\underline{x} = (x(a_{\underline{x}}), \dots, x(b_{\underline{x}}))$  is called *good* if the following properties are satisfied:

(i)  $\underline{x}$  passes through  $\mathfrak{D}_0^n$ .

(ii) there exists  $C_1 > 0$  such that

$$\max_{k=1,\dots,n} \frac{\tau_k[b_{\underline{x}}]}{|\Lambda_k|} \leq C_1 \quad \text{and} \quad \max_{k=1,\dots,n} \frac{\tau_k[a_{\underline{x}}]}{|\Lambda_k|} \leq C_1.$$
(5.181)

(iii) it holds that

$$\sum_{t=a_{\underline{x}}}^{b_{\underline{x}}-1} \exp\left(\mathcal{O}(\varepsilon(n)) \sum_{k=1}^{n} \frac{\tau_{k}[t]^{2}}{|\Lambda_{k}|}\right) \phi_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{x}(t),\boldsymbol{x}(t+1)) \leq 1 + \mathcal{O}(\varepsilon(n)).$$
(5.182)

The set of all good paths is denoted by  $\mathcal{GP}_{A,B}$ .

**Lemma 5.29** ([6, Proposition 5.4]). Let  $\mathfrak{f}_{A,B}$  be the mesoscopic flow constructed in the previous subsection and denote by  $X = \{X(t)\}_t$  the Markov chain on  $\Gamma^n$  with law  $\mathbb{P}^{\mathfrak{f}_{A,B}}$ . Then,

$$\mathbb{P}^{\mathfrak{f}_{A,B}}\left[\boldsymbol{X}\in\mathcal{GP}_{A,B}\right] = 1 - \mathcal{O}_N(1).$$
(5.183)

*Proof.* Due to the construction of the flow  $\mathfrak{f}$ , the probability of the event that the associated Markov chain does not pass through  $\mathfrak{D}_0^n$  is of order  $\mathcal{O}_N(1)$ . As a further consequence, the step frequencies,  $\tau_k[t]/t$ , are on average proportional to  $\pi_k$ . Therefore, there exist a constant  $C_1$  such that, up to an exponentially small  $\mathbb{P}^{\mathfrak{f}_{A,B}}$ -probability, it holds that

$$\max_{k=1,\dots,n} \frac{\tau_k[b_{\underline{x}}]}{|\Lambda_k|} \leq C_1.$$

Hence, it remains to show that the  $\mathbb{P}^{f_{A,B}}$ -probability of the event (5.182) is close to one. In order to do so, notice that from the construction of the mesoscopic flow, namely the fact that the transition probabilities  $q^{\mathfrak{f}}$  inside  $\mathfrak{D}_0^n$  are given by (5.153) up to small correction, and from the property (5.57) of the minimal energy curve  $\varphi$ , it follows that there exists a  $K < \infty$ , independent of n, such that

$$\mathbb{P}^{\mathfrak{f}_{A,B}}\left[\max_{k=1,\dots,n} \max_{|t|>N^{1/2-\delta}} \frac{\tau_k[t]}{|t|\pi_k} > K\right] = \mathcal{O}_N(1).$$
(5.184)

Now, observe that uniformly for all  $t < N^{1/2-\delta}$  and all paths  $\underline{x}$  it holds that

$$\sum_{k=1}^{n} \frac{\tau_k[t]^2}{|\Lambda_k|} = \mathcal{O}(N^{-2\delta}).$$
(5.185)

On the other hand, for all paths  $\underline{x}$  that satisfy the condition  $\tau_k[t] \leq K\pi_k|t|$  uniformly for all  $|t| > N^{1/2-\delta}$  and k = 1, ..., n, we have

$$\mathcal{O}(\varepsilon(n)) \sum_{k=1}^{n} \frac{\tau_k[t]^2}{|\Lambda_k|} \leq \mathcal{O}(\varepsilon(n)) K^2 \frac{t^2}{N} \sum_{k=1}^{n} \pi_k = \mathcal{O}(\varepsilon(n)) K^2 \frac{t^2}{N}.$$
(5.186)

Moreover, due to the construction of  $\mathfrak{f}_{A,B}$ , there exists a c > 0 such that all those paths  $\underline{x}$  satisfy  $\phi_{A,B}(x(t), x(t+1)) \leq e^{-ct^2/2}$  for all  $|t| > N^{1/2-\delta}$ . Hence, in view of (5.130) we obtain that with  $\mathbb{P}^{\mathfrak{f}_{A,B}}$  close to one

$$\sum_{t=a_{\underline{x}}}^{b_{\underline{x}}-1} \exp\left(\mathcal{O}(\varepsilon(n)) \sum_{k=1}^{n} \frac{\tau_{k}[t]^{2}}{|\Lambda_{k}|}\right) \phi_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{x}(t),\boldsymbol{x}(t+1)) = 1 + \mathcal{O}(\varepsilon(n)).$$
(5.187)

This yields the assertion.

108

**5.4.4. Construction of a subordinate microscopic flow.** Fix a mesoscopic path  $\underline{x} \in \mathcal{GP}_{A,B}$ . Now, our strategy is to construct a directed Markov chain  $\Sigma^{\underline{x}} = \{\Sigma^{\underline{x}}(t)\}_t$  on the set of configurations given by the mesoscopic path, i.e.  $\mathcal{S}^n[\boldsymbol{x}(a_{\underline{x}})] \cup \ldots \cup \mathcal{S}^n[\boldsymbol{x}(b_{\underline{x}})]$ . In view of (1.82), the subordinate flow,  $f^{\underline{x}}$ , is then defined by  $f^{\underline{x}}(\sigma,\eta) = \mathbb{P}^{f^{\underline{x}}}[(\sigma,\eta) \in \Sigma^{\underline{x}}]$ . Actually, since the construction of  $f^{\underline{x}}$  from  $\mathcal{S}^n[\boldsymbol{x}(a_{\underline{x}})]$  to  $\mathcal{S}^n[\boldsymbol{x}(0)]$  and  $\mathcal{S}^n[\boldsymbol{x}(0)]$  to  $\mathcal{S}^n[\boldsymbol{x}(b_{\underline{x}})]$  is completely similar, we will work out the details only for the latter one.

To start with, let us denote by

$$\Theta_{k,t}^{\pm} := \left\{ r \in \mathcal{S}_0 \mid \boldsymbol{x}_r^k(t+1) - \boldsymbol{x}_r^k(t) = \pm \frac{1}{N} \right\}$$
(5.188)

the corresponding color that is created, respectively, annihilated in the k's block from x(t) to x(t+1). Due to the underlying single site dynamic, for each t there exists a unique  $k \in \{1, \ldots, n\}$  such that the sets  $\Theta_{k,t}^+$  and  $\Theta_{k,t}^-$ , respectively, consists of exactly one element while for all  $l \neq k$  the sets  $\Theta_{l,t}^{\pm}$  are empty. For a given configuration  $\sigma$ , recall that we defined by

$$\Lambda_{k,r}(\sigma) := \{ i \in \Lambda_k \mid \sigma_i = r \}, \qquad r \in \mathcal{S}_0$$

the set of lattice site in  $\Lambda_k$  those corresponding spin variables are equal to the color r.

Let us now define a time-inhomogeneous Markov chain  $\{\Sigma^{\boldsymbol{x}}(t)\}$ . Its initial distribution,  $\nu_0^{\boldsymbol{x}}$ on  $\mathcal{S}^n[\boldsymbol{x}(0)]$  is given by the Gibbs measure,  $\mu_N$ , conditioned on the set  $\mathcal{S}^n[\boldsymbol{x}(0)]$ , and we choose the following transition probabilities

$$q_t(\sigma, \sigma^{i,s}) := \begin{cases} \frac{\exp\left(\beta \langle \tilde{h}^i, -e^r + e^s \rangle\right)}{\sum_{j \in \Lambda_{k,r}(\sigma)} \exp\left(\beta \langle \tilde{h}^j, -e^r + e^s \rangle\right)}, & \text{if } \Theta_{k,t}^- = r, \ \Theta_{k,t}^+ = s, \ i \in \Lambda_{k,r}(\sigma), \\ 0, & \text{otherwise.} \end{cases}$$

$$(5.189)$$

Thus, the subordinate flow,  $f^{\mathbf{x}}$ , through the edge  $(\sigma(t), \sigma(t+1))$  where  $\sigma(t) \in S^n[\mathbf{x}(t)]$  and  $\sigma(t+1) \in S^n[\sigma(t+1)]$  is equal to

$$f^{\underline{x}}(\sigma(t), \sigma(t+1)) = \nu_t^{\underline{x}}(\sigma(t)) q_t(\sigma(t), \sigma(t+1)),$$
(5.190)

where  $\nu_t^{\mathbf{x}}$  denotes the marginal distribution of the chain.

**Remark 5.30.** Note that any choice of the transition probabilities  $q_t$  give rise to a subordinate flow. But, in general, the marginal distributions,  $\nu_t^{\underline{x}}$ , of this chain differ from the conditional distribution  $\mu_{S^n[\underline{x}(t)]} \equiv \mu_N[\cdot |S^n[\underline{x}(t)]]$ . Thus, one would like to choose the transition probabilities in such a way that  $\nu_t^{\underline{x}} = \mu_{S^n[\underline{x}(t)]}$  for all  $t = 0, \ldots, b_{\underline{x}}$ . Indeed, in the case where all  $\tilde{h}^i \equiv 0$ , by choosing

$$q_t(\sigma, \sigma^{i,s}) = \begin{cases} 1/|\Lambda_{k,r}(\sigma)|, & \text{if } \Theta_{k,t}^- = r, \ \Theta_{k,t}^+ = s, \ i \in \Lambda_{k,r}(\sigma), \\ 0, & \text{otherwise}, \end{cases}$$

the transported measure  $\nu_t^x$  equals  $\mu_{S^n[\boldsymbol{x}(t)]}$  that is just the uniform measure on  $S^n[\boldsymbol{x}(t)]$ . However, in the general case it is not known under which conditions on size of the distortion of  $\tilde{h}^i$  a transition matrix,  $q_t$ , exists satisfying such property.

Since,

$$q_t(\sigma(t), \sigma(t+1)) = \frac{1 + \mathcal{O}(\varepsilon(n))}{\left|\Lambda_{\Theta_{h,t}^-}(\sigma(t))\right|}, \qquad \frac{p_N(\sigma(t), \sigma(t+1))}{r^n(\boldsymbol{x}(t), \boldsymbol{x}(t+1))} = \frac{1 + \mathcal{O}(\varepsilon(n))}{\left|\Lambda_{\Theta_{h,t}^-}(\sigma(t))\right|}$$
(5.191)

the expression (5.139) can be rewritten as

$$\Phi_t^x(\sigma(t)) = \frac{\nu_t^x(\sigma(t))}{\mu_N[\sigma(t) \mid S^n[\boldsymbol{x}(t)]]} \left(1 + \mathcal{O}(\varepsilon(n))\right) =: \Psi_t(\sigma(t)) \left(1 + \mathcal{O}(\varepsilon(n))\right)$$

where  $\Psi_t$  can be seen as a measure for the deviation of  $\nu_t^x$  from the measure  $\mu_{\mathcal{S}^n[x(t)]}$ . As a consequence of the definition of the Markov chain  $\{\Sigma^x(t)\}, \Psi_0(\sigma) \equiv 1$  for all  $\sigma \in \mathcal{S}^n[x(0)]$ . However, for any t > 0 this is no longer true. Hence, it remains to study the average error propagation along the mesoscopic path x. In particular, our goal is to show that for any good mesoscopic path  $x \in \mathcal{GP}_{A,B}$  it holds that  $\mathbb{E}^{f^x}[\Psi_t] \leq \exp(\mathcal{O}(\varepsilon(n)) t^2/N)$ .

Notice that the Gibbs measure,  $\mu_N$ , conditioned of  $S^n[\boldsymbol{x}(t)]$  is a product measure,

$$\mu_N \left[ \cdot \left| \mathcal{S}^n[\boldsymbol{x}(t)] \right] = \bigotimes_{k=1}^n \mu_t^k(\cdot), \qquad (5.192)$$

where  $\mu_t^k$  is the canonical measure on  $\mathcal{S}_{|\Lambda_k|}:=\,\mathcal{S}_0^{\Lambda_k}$  that is given by

$$\mu_t^k(\sigma) := \frac{1}{Z[x^k(t)]} \exp\left(\beta \sum_{i \in \Lambda_k} \langle \tilde{h}^i, e^{\sigma_i} \rangle\right) \mathbb{1}_{\varrho^k(\sigma) = x^k(t)}, \qquad \sigma \in \mathcal{S}_0^{\Lambda_k}.$$
(5.193)

On the other hand, according to the choice of the transition probabilities (5.189), the Markov chain  $\{\Sigma^{\underline{x}}(t)\}$  splits into a direct product of n Markov chains  $\{\Sigma^{\underline{x},1}(t)\}, \ldots, \{\Sigma^{\underline{x},n}(t)\}$  on the state spaces  $S_{|\Lambda_1|}, \ldots, S_{|\Lambda_n|}$  that evolves independently. Hence, if the Markov chain  $\{\Sigma^{\underline{x}}(t)\}$  has performed T steps, each of the chains  $\{\Sigma^{\underline{x},k}(t)\}$  performs exactly  $\tau_k[T]$  steps. Thus, the corrector,  $\Psi_t$ , is equal to

$$\Psi_t(\sigma(t)) = \prod_{k=1}^n \psi_{\tau_k[t]}^k(\sigma_{\Lambda_k}(t)), \qquad (5.194)$$

where  $\sigma_{\Lambda_k}(t)$  is the projection of  $\sigma(t)$  onto  $S_{|\Lambda_k|}$ . Therefore, it suffices to study the propagation of errors with respect to each  $\{\Sigma^{x,k}(t)\}$ .

**Proposition 5.31.** Fix a  $k \in \{1, ..., n\}$  and let us denote by  $\mathbb{P}_{\mu_0^k}$  the law of the Markov chain  $\{\Sigma^{\mathbf{x},k}(t)\}$  on  $\mathcal{S}_{|\Lambda_k|}$ . Then, there exists  $c < \infty$  such that

$$\mathbb{E}_{\mu_0^k} \left[ \psi_t^k \right] \leq \exp \left( c \,\varepsilon(n) \, \frac{t^2}{|\Lambda_k|} \right), \qquad \forall t = 0, \dots, \tau_k[b_{\underline{x}}]. \tag{5.195}$$

In view of Proposition 5.31, we obtain the following lower bound for capacities.

Corollary 5.32. With the notation introduced above, we obtain

$$\operatorname{cap}(A,B) \geq \frac{\beta|\bar{\gamma}_1|}{2\pi N} \frac{\exp(-\beta N F_N(z))}{\sqrt{\det\left(I_q - 2\beta \operatorname{Hess} U_N(2\beta z)\right)}} \left(1 - \mathcal{O}_N(1)\right).$$
(5.196)

*Proof of Proposition 5.31.* To keep notation simple, we will use the following abbreviations

$$w(\sigma) \equiv \exp\Bigl(\beta \sum_{i \in \Lambda_k} \langle \tilde{h}^i, e^{\sigma_i} \rangle \Bigr), \quad \forall \sigma \in \mathcal{S}_{|\Lambda_k|}, \qquad \text{and} \qquad \tilde{x}(s) \equiv x^k \bigl(\tau_k^{-1}[s]\bigr)$$

where  $\tau_k^{-1}[s] := \min\{t \in \mathbb{N}_0 \mid \tau_k[t] \geq s\}$  and  $x^k(t) \in \mathbb{R}^q$  is the k's block of x(t) for a given path  $\underline{x} = (x(0), \dots, x(b_{\underline{x}}))$ . Further, we define a family of functions  $d_t \equiv d_t^k : S_{|\Lambda_k|} \to \mathbb{R}$ ,

$$\sigma \mapsto d_t(\sigma) := \sum_{i \in \Lambda_{k,r}(\sigma)} \exp\left(\beta \left< \tilde{h}^i, -e^r + e^s \right> \right), \quad \text{for } r = \Theta_{k,t}^-, \ s = \Theta_{k,t}^+.$$
(5.197)

The proof of this proposition comprises three steps.

STEP 1. In the sequel, let  $S[x] := \{ \sigma \in S_{|\Lambda_k|} \mid \varrho^k(\sigma) = x \}$  for  $x \in \mathbb{R}^q$ . We start with showing that for all  $t = 0, \ldots, \tau_k[b_x] - 1$  and  $\eta \in S[\tilde{x}(t)]$ 

$$\psi_t^k(\eta) = \frac{1}{|\mathfrak{P}_t(\eta)|} \sum_{\underline{\sigma} \in \mathfrak{P}_t(\eta)} \prod_{s=0}^{t-1} \frac{\mu_s^k[d_s]}{d_s(\sigma(s))},$$
(5.198)

where  $\mathfrak{P}_t(\eta)$  denotes the set of all paths  $\underline{\sigma} = (\sigma(0), \ldots, \sigma(t-1), \eta)$  from  $\mathcal{S}[\tilde{x}(0)]$  to  $\eta$  of positive  $\mathbb{P}_{\mu_n^k}$ -probability. Let us point out that an analog statement was shown in [6, Proposition 5.1].

To lighten notation, for a fixed t we set  $r = \Theta_{k,t-1}^-$  and  $s = \Theta_{k,t-1}^+$ . Since the marginal distribution,  $\nu_t^k$ , of the Markov chain  $\{\Sigma^{x,k}(t)\}$  satisfies for  $\eta \in S[\tilde{x}(t)]$  the recursion

$$\nu_t^k(\eta) = \sum_{i \in \Lambda_{k,s}(\eta)} \nu_{t-1}^k(\eta^{i,r}) q_{t-1}(\eta^{i,r}, \eta), \qquad (5.199)$$

it follows that

$$\psi_{t}^{k}(\eta) = \sum_{i \in \Lambda_{k,s}(\eta)} \frac{\nu_{t-1}^{k}(\eta^{i,r})}{\mu_{t}^{k}(\eta)} q_{t-1}(\eta^{i,r},\eta) = \sum_{i \in \Lambda_{k,r}(\eta)} \frac{\mu_{t-1}^{k}(\eta^{i,r})}{\mu_{t}^{k}(\eta)} q_{t-1}(\eta^{i,r},\eta) \psi_{t-1}^{k}(\eta^{i,r}).$$
(5.200)

Plugging our choice of the transition probabilities in (5.189), we obtain

$$\frac{\mu_{t-1}^{k}(\eta^{i,r})}{\mu_{t}^{k}(\eta)} q_{t-1}(\eta^{i,r},\eta) = \frac{Z[\tilde{x}(t)]}{Z[\tilde{x}(t-1)]} \left( \sum_{i \in \Lambda_{k,r}(\eta^{i,r})} \exp\left(\beta \left< \tilde{h}^{i}, -e^{r} + e^{s} \right> \right) \right)^{-1}$$
(5.201)

As a consequence of the single site dynamics, we consider, it holds that  $|\Lambda_{k,r}(\sigma)| = |\Lambda_{k,r}|$  for all  $\sigma \in S[\tilde{x}(t)]$ . In particular,  $|\Lambda_{k,r}| = O(N)$ . Notice that this property is implied by the fact that the minimal energy curve,  $\gamma(t)$ , used in the construction of the mesoscopic flow, is bounded away from the boundary of  $\mathcal{M}_1$  uniformly in N. Thus, by considering the ratio of the partition functions we get

$$\frac{Z[\tilde{x}(t)]}{Z[\tilde{x}(t-1)]} = \frac{1}{Z[\tilde{x}(t-1)]} \sum_{\eta \in \mathcal{S}[\tilde{x}(t)]} \frac{1}{|\Lambda_{k,s}|} \sum_{i \in \Lambda_{k,s}(\eta)} w(\eta)$$

$$= \frac{1}{Z[\tilde{x}(t-1)]} \sum_{\eta \in \mathcal{S}[\tilde{x}(t)]} \frac{1}{|\Lambda_{k,s}|} \sum_{i \in \Lambda_{k,s}(\eta)} \exp\left(\beta \langle \tilde{h}^{i}, -e^{r} + e^{s} \rangle\right) w(\eta^{i,r})$$

$$= \frac{1}{Z[\tilde{x}(t-1)]} \sum_{\eta \in \mathcal{S}[\tilde{x}(t-1)]} \frac{1}{|\Lambda_{k,s}|} \sum_{i \in \Lambda_{k,r}(\eta)} \exp\left(\beta \langle \tilde{h}^{i}, -e^{r} + e^{s} \rangle\right) w(\eta)$$

$$= \frac{1}{|\Lambda_{k,s}|} \mu_{t-1}^{k} [d_{t-1}],$$
(5.202)

where  $\mu_t^k[f] = \sum_{\sigma \in S[\tilde{x}(t)]} f(\sigma) \mu_t^k(\sigma)$  denotes the expectation with respect to the measure  $\mu_t^k$ . As a result,

$$\psi_t^k(\eta) = \frac{1}{|\Lambda_{k,s}|} \sum_{i \in \Lambda_{k,s}(\eta)} \frac{\mu_{t-1}^k[d_{t-1}]}{d_{t-1}(\eta^{i,r})} \psi_{t-1}^k(\eta^{i,r}).$$
(5.203)

Iterating the argument above and using that  $\psi_0^k \equiv 1$  yields (5.198).

STEP 2. After having established a different representation for  $\psi_t^k$ , our next task is to compute the average error propagation along the path  $(\tilde{x}(0), \ldots, \tilde{x}(\tau_k[b_x]))$ . Let us remark that the configurations  $\sigma(s)$  and  $\xi(s)$  for two different microscopic paths  $\underline{\sigma}, \xi \in \mathfrak{P}_t(\eta)$  with  $\eta \in S[\tilde{x}(t)]$  differs at most in 2(t-s) coordinates. Therefore, it is easy to see that there exists  $c < \infty$  such that

$$\frac{d_s(\sigma(s))}{d_s(\xi(s))} = 1 - \frac{d_s(\sigma(s)) - d_s(\xi(s))}{d_s(\xi(s))} \le \exp\left(c\,\varepsilon(n)\,\frac{t-s}{|\Lambda_k|}\right).$$
(5.204)

Hence,

$$\mathbb{E}_{\mu_{0}^{k}}\left[\psi_{t}^{k}\right] = \sum_{\eta(0),\dots,\eta(t)} \mu_{0}^{k}(\eta(0)) \prod_{s=1}^{t} q_{s}\left(\eta(s-1),\eta(s)\right) \psi_{t}^{k}\left(\eta(t)\right) \\
= \sum_{\eta(0),\dots,\eta(t)} \mu_{0}^{k}(\eta(0)) \frac{1}{|\mathfrak{P}_{t}(\eta(t))|} \sum_{\underline{\sigma}\in\mathfrak{P}_{t}(\eta(t))} \prod_{s=1}^{t} \frac{\mu_{s-1}^{k}[d_{s-1}]}{d_{s-1}(\sigma(s-1))} q_{s}\left(\eta(s-1),\eta(s)\right) \\
\leq \sum_{\eta(0),\dots,\eta(t-1)} \mu_{0}^{k}(\eta(0)) \frac{\mu_{0}^{k}[d_{0}]}{d_{0}(\eta(0))} \prod_{s=1}^{t-1} \frac{\mu_{s}^{k}[d_{s}]}{d_{s}(\eta(s))} q_{s}\left(\eta(s-1),\eta(s)\right) \exp\left(c\,\varepsilon(n)\,\frac{t^{2}}{|\Lambda_{k}|}\right). \tag{5.205}$$

Note that the constant c may change from line to line throughout the following computations. We can easily bound the ratio  $\mu_s^k \big[ d_s \big] / d_s \big( \eta(s) \big)$  by

$$\frac{\mu_s^k[d_s]}{d_s(\eta(s))} = 1 + \frac{\mu_s^k[d_s] - d_s(\eta(s))}{d_s(\eta(s))} \le \exp\Big(Y_s\big(\eta(s)\big) (1 + c\,\varepsilon(n)\big)\Big), \tag{5.206}$$

where we introduced the random variable

$$Y_{s}(\sigma) := \frac{1}{\left|\Lambda_{k,\Theta_{k,s}^{-}}\right|} \sum_{i \in \Lambda_{k}} a_{i}(s) \left(\mu_{s}^{k} \left[\mathbbm{1}_{i \in \Lambda_{k,\Theta_{k,s}^{-}}}\right] - \mathbbm{1}_{i \in \Lambda_{k,\Theta_{k,s}^{-}}}(\sigma)\right)$$
(5.207)

with  $a_i(s) := \exp\left(\beta \langle \tilde{h}^i, -e^{\Theta_{k,s}^-} + e^{\Theta_{k,s}^+} \rangle\right) - 1$ . Thus, it remains to evaluate the sums in (5.205). For any  $u \in \{1, \ldots, t-1\}$ , we claim that

$$\sum_{\eta(u),\dots,\eta(t-1)} \prod_{s=u}^{t-1} \frac{\mu_s^k[d_s]}{d_s(\eta(s))} q_s(\eta(s-1),\eta(s)) \leq \exp\left(\sum_{s=u}^{t-1} \left(Y_s(\eta(u-1)) + c\,\varepsilon(n)\,\frac{t-u}{|\Lambda_k|}\right)\right)$$
(5.208)

uniformly for all  $\eta(u-1) \in S[\tilde{x}(u-1)]$ . Once we have establish (5.208), it follows that

$$\mathbb{E}_{\mu_0^k} \left[ \psi_t^k \right] \leq \mu_0^k \left[ \exp\left( \sum_{s=0}^{t-1} Y_s \right) \right] \exp\left( c \,\varepsilon(n) \, \frac{t^2}{|\Lambda_k|} \right).$$
(5.209)

The proof of (5.208) relies one the following observation. For any  $u \in \{1, \ldots, t-1\}$  consider an arbitrary chosen but fixed  $\sigma \in S[\tilde{x}(u)]$ . Then, for all  $j \in \Lambda_{k,\Theta_{k,u-1}^-}(\sigma)$  and arbitrary s,

$$Y_{s}(\sigma^{j,\Theta^{+}_{k,u-1}}) = Y_{s}(\sigma) + \frac{1}{|\Lambda_{k,\Theta^{-}_{k,s}}|} \sum_{i \in \Lambda_{k}} a_{i}(s) \mathbb{1}_{i=j} \left( \mathbb{1}_{\Theta^{-}_{k,s}=\Theta^{-}_{k,u-1}} - \mathbb{1}_{\Theta^{-}_{k,s}=\Theta^{+}_{k,u-1}} \right)$$
  
$$\leq Y_{s}(\sigma) + \frac{c \varepsilon(n)}{|\Lambda_{k}|}.$$
(5.210)

Therefore, we obtain by induction that

$$\sum_{\eta(u),\dots,\eta(t-1)} \prod_{s=u}^{t-1} \frac{\mu_s^k[d_s]}{d_s(\eta(s))} q_s(\eta(s-1),\eta(s))$$

$$\leq \sum_{\sigma \in \mathcal{S}[\tilde{x}(u)]} \exp\left(\sum_{s=u}^{t-1} \left(Y_s(\sigma) + c \varepsilon(n) \frac{t-u+1}{|\Lambda_k|}\right)\right) q_u(\eta(u-1),\sigma)$$

$$\leq \exp\left(\sum_{s=u}^{t-1} \left(Y_s(\eta(u-1)) + c \varepsilon(n) \frac{t-u}{|\Lambda_k|}\right)\right),$$

where we take advantage of (5.210) is the second step.

STEP 3. Finally, we are left with the task to bound  $\mu_0^k \left[ \exp\left(\sum_{s=0}^{t-1} Y_s\right) \right]$  from above. Let us consider the log-moment generating function  $h(u) = \ln \mu_0^k \left[ \exp\left(u \sum_{s=0}^{t-1} Y_s\right) \right]$  for  $u \ge 0$ . Since

any function, h, that vanishes at zero, satisfies the bound  $h(t) \leq \frac{t^2}{2} \max_{0 \leq s \leq t} h''(s) + t h'(0)$ , we immediately get that

$$\ln \mu_0^k \Big[ \exp \Big( \sum_{s=0}^{t-1} Y_s \Big) \Big] \leq \max_{0 \le u \le 1} \mathbf{Var}^{t,u} \Big[ \exp \Big( \sum_{s=0}^{t-1} Y_s \Big) \Big] + \sum_{s=0}^{t-1} \mu_0^k \Big[ Y_s \Big],$$
(5.211)

where  $\operatorname{Var}^{t,u}$  is the variance with respect to the tilted measure  $\mu_0^{k,t,u}$  conditioned on  $\mathcal{S}[\tilde{x}(0)]$ ,

$$\mu_{0}^{k,t,u}(\sigma) = \frac{\mu_{0}^{k}(\sigma) \exp\left(u \sum_{s=0}^{t-1} Y_{s}(\sigma)\right)}{\mu_{0}^{k} \left[\exp\left(u \sum_{s=0}^{t-1} Y_{s}\right) \mathbb{1}_{\varrho^{k}(\sigma) = \tilde{x}(0)}\right]}, \quad \forall \sigma \in \mathcal{S}[\tilde{x}(0)].$$
(5.212)

In Lemma 5.34 we derive bounds for the expectation  $\mu_0^k[Y_s]$  and the variance  $\operatorname{Var}^{t,u}[Y_s]$ , respectively. Notice that the proof of this lemma crucially relies on the fact that the distance of the mesoscopic path  $\underline{x}$  from the boundary of  $\times_{k=1}^n \mathcal{M}_{\pi_k}$  is bounded from below by a constant that is independent of N. Consequently, assuming (5.220) we have

$$\sum_{s=0}^{t-1} \mu_0^k \big[ Y_s \big] \leq c_{21} \, \varepsilon(n) \, \sum_{s=0}^{t-1} \frac{s}{|\Lambda_k|} = c_{21} \, \varepsilon(n) \, \frac{t^2}{|\Lambda_k|}.$$

On the other hand, using the Cauchy-Schwartz inequality and assuming (5.221) implies that

$$\mathbf{Var}^{t,u} \Big[ \exp\left(\sum_{s=0}^{t-1} Y_s\right) \Big] \leq \left( \sum_{s=0}^{t-1} \sqrt{\mathbf{Var}^{t,u} [Y_s]} \right)^2 \leq c_{22} \varepsilon(n) \frac{t^2}{|\Lambda_k|}$$

Thus, by combining the estimates above, we finally obtain that there exists  $c < \infty$  such that

$$\ln \mu_0^k \left[ \exp\left(\sum_{s=0}^{t-1} Y_s\right) \right] \le c \varepsilon(n) \frac{t^2}{|\Lambda_k|}.$$
(5.213)

This concludes the proof.

In the remaining part of this subsection we will proof the Lemma 5.34. To start with, notice that the tilted measure,  $\mu_0^{k,t,u}$ , as well as the canonical measure,  $\mu_0^k \equiv \mu_0^{k,0,0}$ , can be written as a product measures,  $\nu^{k,t,u}$ , conditioned on  $S[\tilde{x}(0)]$ , i.e.

$$\mu_0^{k,t,u}(\sigma) = \bigotimes_{i \in \Lambda_k} p_i^{k,t,u} \left[ \sigma \left| \mathcal{S}[\tilde{x}(0)] \right] =: \nu^{k,t,u} \left[ \sigma \left| \mathcal{S}[\tilde{x}(0)] \right] \right],$$
(5.214)

where for all  $r \in S_0$ 

$$p_{i}^{k,t,u}(r) := \frac{\exp\left(\beta \langle \tilde{h}^{i}, e^{r} \rangle + u \sum_{s=0}^{t-1} |\Lambda_{k,\Theta_{k,s}^{-}}|^{-1} a_{i}(s) \mathbb{1}_{r=\Theta_{k,s}^{-}}\right)}{\sum_{r' \in \mathcal{S}_{0}} \exp\left(\beta \langle \tilde{h}^{i}, e^{r'} \rangle + u \sum_{s=0}^{t-1} |\Lambda_{k,\Theta_{k,s}^{-}}|^{-1} a_{i}(s) \mathbb{1}_{r'=\Theta_{k,s}^{-}}\right)}.$$
(5.215)

Let us point out that, as a consequence of the choice of the paths  $\underline{x} \in \mathcal{GP}_{A,B}$ , there exists  $\delta_1 \in (0,1)$  independent of N such that uniformly for all  $i \in \Lambda_k$ ,  $t \in \{0, \ldots, \tau_k[b_{\underline{x}}]\}$  and  $u \in [0,1]$  the probabilities  $p_i^{k,u}$  satisfy

$$\delta_1 \leq p_i^{k,t,u}(r) \leq 1 - \delta_1, \qquad \forall r \in \mathcal{S}_0.$$
(5.216)

A key element in the analysis are sharp large deviation estimates, as well as properties of the involved entropy and corresponding log–moment generating function. Given a subset  $\Lambda \subset \Lambda_k$ , we define  $\rho_{|\Lambda_k \setminus \Lambda|} : S_{|\Lambda_k|} \to \mathcal{M}_{(\Lambda_k - \Lambda)/N}$ ,

$$\sigma \mapsto \varrho_{|\Lambda_k \setminus \Lambda|}(\sigma) := \frac{1}{N} \sum_{i \in \Lambda_k \setminus \Lambda} \delta_{\sigma_i}$$

 $\Box$ 

Note that in later applications we will choose  $\Lambda = \{i\}$  or  $\Lambda = \{i, j\}$ . Further, for any  $x \in \mathbb{R}^q$  consider the entropy

$$I^{t,u}_{|\Lambda_k \setminus \Lambda|}(x) := \sup_{s \in \mathbb{R}^q} \left( \langle s, x \rangle - U^{t,u}_{|\Lambda_k \setminus \Lambda|}(s) \right)$$

which is defined as the Legendre-Fenchel transform of the log-moment generating function  $U^{t,u}_{|\Lambda_k \setminus \Lambda|} \colon \mathbb{R}^q \to \mathbb{R},$ 

$$t \mapsto U^{t,u}_{|\Lambda_k \setminus \Lambda|}(s) := \frac{1}{N} \ln \nu^{k,t,u} \Big[ \exp \left( N \left\langle s, \varrho_{|\Lambda_k \setminus \Lambda|} \right\rangle \right) \Big] = \frac{1}{N} \sum_{i \in \Lambda_k \setminus \Lambda} \ln \left( \sum_{r \in \mathcal{S}_0} e^{s_r} p_i^{k,t,u}(r) \right) \Big]$$

Let us remark that, due to (5.216), a statement analog to Lemma 5.6 can be established. Moreover, by inspecting the proof of [**26**, Theorem 3.1], it can be shown that the error term in (5.33) is actually of order 1/N. Hence, for all  $x \in \operatorname{ri}(\mathcal{M}_{(\Lambda_k - \Lambda)/N})$ ,

$$\nu^{k,t,u} \left[ \mathbb{1}_{\varrho_{|\Lambda_k \setminus \Lambda|} = x} \right] = \frac{\exp\left(-N I_{|\Lambda_k \setminus \Lambda|}^{t,u}(x)\right)}{\sqrt{\left(2\pi N\right)^{q-1} \det\left[\operatorname{Hess} U_{|\Lambda_k \setminus \Lambda|}^{t,u}(t^{\Lambda}(x))\right]_q}} \left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right), \quad (5.217)$$

where  $t^{\Lambda}(x) \in \mathbb{R}^q$  is the solution of  $x = \nabla U^{t,u}_{|\Lambda_k \setminus \Lambda|} (t^{\Lambda}(x))$  such that  $\sum_{r \in S_0} t^{\Lambda}_r(x) = 0$ .

**Lemma 5.33.** Suppose  $y, y' \in \operatorname{ri}(\mathcal{M}_{(\Lambda_k \setminus \Lambda)/N})$ . Then, there exists  $c < \infty$ , independent of N, such that

$$||t^{\Lambda}(y') - t^{\Lambda}(y)|| \le c ||y' - y||.$$
 (5.218)

*Proof.* Due to the fact that  $y, y' \in \operatorname{ri}(\mathcal{M}_{(\Lambda_k - \Lambda)/N})$ , there exists  $\delta \equiv \delta(y, y') > 0$  such that for all  $0 \leq s \leq 1$  the minimal component of sy' - (1 - s)y is at least  $\delta$ . Recall that in a local coordinate system (5.30) reads

$$\psi(t^*) = \nabla \Big( I^{t,u}_{|\Lambda_k \setminus \Lambda|} \circ \psi^{-1} \Big) \big( \psi(x) \Big) \qquad \Longleftrightarrow \qquad \psi(x) = \nabla \Big( U^{t,u}_{|\Lambda_k \setminus \Lambda|} \circ \psi^{-1} \Big) \big( \psi(t^*) \big).$$

Here, we choose as a local chart  $\psi : \mathcal{M}_{(\Lambda_k - \Lambda)/N} \to \mathbb{R}^{q-1}$  the linear map  $\psi(x) = V^T x$  where the columns of the matrix V consists of the vectors of an orthonormal basis of  $V_1$ . Further, we set  $\xi := y' - y \in V_1$ . Then,

$$\begin{aligned} \left\|\psi\left(t^{\Lambda}(y+\xi)\right) - \psi\left(t^{\Lambda}(y)\right)\right\| \\ &= \left\|\nabla\left(I^{t,u}_{|\Lambda_{k}\setminus\Lambda|} \circ \psi^{-1}\right)\left(\psi(y+\xi)\right) - \nabla\left(I^{t,u}_{|\Lambda_{k}\setminus\Lambda|} \circ \psi^{-1}\right)\left(\psi(y)\right)\right\| \\ &\leq \int_{0}^{1} \left\|\operatorname{Hess}\left(I^{t,u}_{|\Lambda_{k}\setminus\Lambda|} \circ \psi^{-1}\right)\left(\psi(y) - s\,\psi(\xi)\right)\,\psi(\xi)\right\|\,\mathrm{d}s \\ &\leq \int_{0}^{1} \left\|\left(V^{\mathrm{T}}\operatorname{Hess}U^{t,u}_{|\Lambda_{k}\setminus\Lambda|}\left(t^{\Lambda}(y-s\,\xi)V\right)^{-1}\psi(\xi)\right\|\,\mathrm{d}s \\ &\leq c\left\|\psi(y'-y)\right\|, \end{aligned}$$
(5.219)

where  $c = \delta^{-2} e^{2\beta \varepsilon(n)}$ . Note that we used (5.58) in the third step and Lemma 5.6 in the last one. This completes the proof.

**Lemma 5.34.** For all  $\underline{x} \in \mathcal{GP}_{A,B}$  and k = 1, ..., n there exists constants  $c_{21}, c_{22} < \infty$  such that for all  $s \in \{0, ..., \tau_k[b_{\underline{x}}]\}$ 

(i) the expectation of  $Y_s$  with respect to  $\mu_0^k$  is bounded by

$$\mu_0^k [Y_s] \leq c_{21} \varepsilon(n) \frac{s}{|\Lambda_k|}, \qquad (5.220)$$

(ii) the variance of  $Y_s$  with respect to  $\mu_0^{k,u}$  is bounded by

$$\mathbf{Var}^{u}[Y_{s}] \leq c_{22} \frac{\varepsilon(n)}{|\Lambda_{k}|}$$
(5.221)

uniformly for all  $u \in [0, 1]$ .

*Proof.* Let us consider an arbitrary but fixed  $k \in \{1, ..., n\}$  and  $s \in \{0, ..., \tau_k[b_x]\}$ . Moreover, we set  $r = \Theta_{k,s}^-$  and abbreviate  $\nu^k \equiv \nu^{k,0,0}$  throughout this proof to lighten notation.

(i) In view of (5.207), it remains to analyze the difference  $|\mu_s^k[\mathbb{1}_{i \in \Lambda_{k,r}}] - \mu_0^k[\mathbb{1}_{i \in \Lambda_{k,r}}]|$  for  $s \ge 1$ . By a straight forward computation we obtain first of all that, for N large enough,

$$\frac{\mu_{s}^{k}\left[\mathbb{1}_{i\in\Lambda_{k,r}}\right] - \mu_{0}^{k}\left[\mathbb{1}_{i\in\Lambda_{k,r}}\right]}{\mu_{s}^{k}\left[\mathbb{1}_{i\in\Lambda_{k,r}}\right]} = \sum_{q\in\mathcal{S}_{0}}\mu_{s}^{k}\left[\mathbb{1}_{i\in\Lambda_{k,q}}\right]\left(1 - \frac{\nu^{k}\left[\mathbb{1}_{\varrho\mid\Lambda_{k}\setminus\{i\}\mid}=\tilde{x}(0)-\hat{e}^{r}\right]\nu^{k}\left[\mathbb{1}_{\varrho\mid\Lambda_{k}\setminus\{i\}\mid}=\tilde{x}(s)-\hat{e}^{q}\right]}{\nu^{k}\left[\mathbb{1}_{\varrho\mid\Lambda_{k}\setminus\{i\}\mid}=\tilde{x}(0)-\hat{e}^{q}\right]\nu^{k}\left[\mathbb{1}_{\varrho\mid\Lambda_{k}\setminus\{i\}\mid}=\tilde{x}(s)-\hat{e}^{r}\right]}\right) = \sum_{q\in\mathcal{S}_{0}}\mu_{s}^{k}\left[\mathbb{1}_{i\in\Lambda_{k,q}}\right]\left(1 - \exp\left(-|\Lambda_{k}|R_{1}\right)\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right)\right)$$
(5.222)

where the remainder,  $R_1$ , is given by

$$R_{1} = I_{|\Lambda_{k}|}^{i} \left( \tilde{x}(s) - \hat{e}^{r} \right) - I_{|\Lambda_{k}|}^{i} \left( \tilde{x}(s) - \hat{e}^{q} \right) - I_{|\Lambda_{k}|}^{i} \left( \tilde{x}(0) - \hat{e}^{r} \right) + I_{|\Lambda_{k}|}^{i} \left( \tilde{x}(0) - \hat{e}^{q} \right).$$
(5.223)

Here, we introduce the notation  $I^i_{|\Lambda_k|} \equiv I^{0,0}_{|\Lambda_k \setminus \{i\}|}$ . Now, for any  $y \in \operatorname{ri}(\mathcal{M}_{(|\Lambda_k|-1)/N})$ , let us define  $t^i(y) \in \mathbb{R}^q$  as a solution of  $y = \nabla U^i_{|\Lambda_k|}(t^i(y))$ . As a immediate consequence of (5.29), we have

$$I^{i}_{|\Lambda_{k}|}(y - \hat{e}^{r}) - I^{i}_{|\Lambda_{k}|}(y - \hat{e}^{q}) = \langle t^{i}(y - \hat{e}^{r}), \hat{e}^{q} - \hat{e}^{r} \rangle + U^{i}_{|\Lambda_{k}|}(t^{i}(y - \hat{e}^{q})) - U^{i}_{|\Lambda_{k}|}(t^{i}(y - \hat{e}^{r})) - \langle t^{i}(y - \hat{e}^{r}) - t^{i}(y - \hat{e}^{q}), y \rangle = \langle t^{i}(y - \hat{e}^{r}), \hat{e}^{q} - \hat{e}^{r} \rangle + \mathcal{O}(1/N^{2}).$$
(5.224)

where we used a Taylor expansion of  $U^i_{|\Lambda_k|}(t^i(y-\hat{e}^q))$  and (5.218) in the second step. Therefore,

$$R_1 = \left\langle t^i(\tilde{x}(s) - \hat{e}^r) - t^i(\tilde{x}(0) - \hat{e}^r), \, \hat{e}^q - \hat{e}^r \right\rangle + \mathcal{O}(1/N^2) \le c \frac{s}{N^2}, \tag{5.225}$$

provided that  $c < \infty$  is chosen appropriately. As a result, there exists  $c_{21} < \infty$  such that

$$\mu_0^k [Y_s] = \mathcal{O}(\varepsilon(n)) \left| \mu_s^k [\mathbb{1}_{i \in \Lambda_{k,r}}] - \mu_0^k [\mathbb{1}_{i \in \Lambda_{k,r}}] \right| \le c_{21} \varepsilon(n) \frac{s}{N}.$$
(5.226)

(ii) To start with, let us rewrite the variance,  $\mathbf{Var}^{t,u}[Y_s]$ , as

$$\mathbf{Var}^{t,u}[Y_s] \leq \frac{c\,\varepsilon(n)}{|\Lambda_{k,r}|} + \frac{1}{\left|\Lambda_{k,r}\right|^2} \sum_{\substack{i,j\in\Lambda_k\\i\neq j}} a_i(s)\,a_j(s)\,\mathbf{Cov}^{t,u}\big[\mathbbm{1}_{i\in\Lambda_{k,r}},\mathbbm{1}_{j\in\Lambda_{k,r}}\big]$$
(5.227)

Hence, it suffices to show that the covariance  $\mathbf{Cov}^{t,u}[\mathbb{1}_{i \in \Lambda_{k,r}}, \mathbb{1}_{j \in \Lambda_{k,r}}]$  is of order 1/N for any distinct  $i, j \in \Lambda_k$ . By a computation similar to the one in (5.222), we obtain

$$\frac{\operatorname{Cov}^{t,u}\left[\mathbbm{1}_{i\in\Lambda_{k,r}},\mathbbm{1}_{j\in\Lambda_{k,r}}\right]}{\mu_{0}^{k,t,u}\left[\mathbbm{1}_{i\in\Lambda_{k,r}},\mathbbm{1}_{j\in\Lambda_{k,r}}\right]} = \sum_{q\in\mathcal{S}_{0}}\mu_{0}^{k,t,u}\left[\mathbbm{1}_{i\in\Lambda_{k,q}}\right] \left(1 - \frac{\nu^{k,u}\left[\mathbbm{1}_{\varrho\mid\Lambda_{k}\setminus\{i\}\mid}=\tilde{x}(0)-\hat{e}^{r}\right]\nu^{k,u}\left[\mathbbm{1}_{\varrho\mid\Lambda_{k}\setminus\{i,j\}\mid}=\tilde{x}(0)-\hat{e}^{r}-\hat{e}^{q}\right]}{\nu^{k,u}\left[\mathbbm{1}_{\varrho\mid\Lambda_{k}\setminus\{i\}\mid}=\tilde{x}(0)-\hat{e}^{q}\right]\nu^{k,u}\left[\mathbbm{1}_{\varrho\mid\Lambda_{k}\setminus\{i,j\}\mid}=\tilde{x}(0)-2\,\hat{e}^{r}\right]}\right)} = \sum_{q\in\mathcal{S}_{0}}\mu_{0}^{k,t,u}\left[\mathbbm{1}_{i\in\Lambda_{k,q}}\right] \left(1 - \exp\left(-\left|\Lambda_{k}\right|R_{2}\right)\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right)\right).$$
(5.228)

By introducing the notation  $I_{|\Lambda_k|}^{i,j} \equiv I_{|\Lambda_k \setminus \{i,j\}|}^{t,u}$ , the remainder can be written as

$$R_{2} = I_{|\Lambda_{k}|}^{i,j} \left( \tilde{x}(0) - 2 \,\hat{e}^{r} \right) - I_{|\Lambda_{k}|}^{i,j} \left( \tilde{x}(0) - \hat{e}^{r} - \hat{e}^{q} \right) - I_{|\Lambda_{k}|}^{i} \left( \tilde{x}(0) - \hat{e}^{r} \right) + I_{|\Lambda_{k}|}^{i} \left( \tilde{x}(0) - \hat{e}^{q} \right)$$
  
$$= \left\langle t^{i,j} \left( \tilde{x}(0) - 2 \,\hat{e}^{r} \right) - t^{i} \left( \tilde{x}(0) - \hat{e}^{r} \right), \, \hat{e}^{q} - \hat{e}^{r} \right\rangle + \mathcal{O} \left( 1/N^{2} \right)$$
  
$$= \mathcal{O} \left( 1/N^{2} \right).$$
(5.229)

In the last step, we took advantage of (5.218) and the fact that  $t^i(\tilde{x}(0) - \hat{e}^r) \in \partial I^{i,j}(y)$  with  $\|\tilde{x}(0) - \hat{e}^r - y\| \leq \sqrt{q}/N$ . Thus, by combining the estimates above, the assertion of (5.221) follows.

## 5.5. Sharp estimates on metastable exit times and exponential distribution

The representation (1.49) of the averaged mean hitting time,  $\mathbb{E}_{\nu_{A,B}}[\tau_B]$ , in terms of the capacity,  $\operatorname{cap}(A, B)$ , and the equilibrium potential,  $h_{A,B}$ , has the advantage that it suffices to compute precisely only  $\operatorname{cap}(A, B)$  and to establish some rough bounds on  $h_{A,B}$ . In the previous sections we have established upper and lower bounds on  $\operatorname{cap}(A, B)$  that coincides in the limit when N tends to infinity. Hence, our main objective now is to show that the equilibrium potential is close to one in the neighborhood,  $\mathcal{U}_{\delta}(A)$ , of the starting set A, whereas outside this neighborhood it is sufficiently small to ensure that the summation of  $\mu_N(\sigma)h_{A,B}(\sigma)$  over all  $\sigma \in \mathcal{S}_N \setminus \mathcal{U}_{\delta}(A)$  compare to  $\mu_N[A]$  is negligible. Notice that in general such a behavior of the equilibrium potential,  $h_{A,B}$ , depends crucially on the choice of the subsets A and B.

To start with, let  $m, m' \in \Gamma^n$  be two local minima of the mesoscopic free energy  $F^n$ . Further, let  $z \in \Gamma^n$  be the lowest saddle point of index one between m from m'. As a consequence of Lemma 5.8, for any n, every mesoscopic critical point m determines uniquely a corresponding macroscopic critical point  $m \equiv m(m) \in \Gamma_N$  and vice versa. Recall that the value of the mesoscopic free energy and macroscopic free energy coincides at critical points, i.e.  $F_N(m) = F^n(m)$ . Let us consider the sets  $A \equiv A^n = S^n[m] \subset S_N$  and  $B \equiv B^n = S^n[m'] \subset S_N$ . In view of (5.67) and (5.128), there exists  $\mathfrak{C}_1 > 0$  such that for all  $n \in \mathbb{N}$  there exists  $N(n) \in \mathbb{N}$  and some  $a_n < \infty$  such that

$$\mathbb{P}_{\mu_A}\left[\tau_B < \tau_A\right] \leq a_n N^{n(q-1)/2-1} \exp\left(-\beta N\left(F_N(z) - F_N(m)\right)\right) \leq e^{-\beta \mathfrak{C}_1 N}, \quad (5.230)$$

for all  $N \ge N(n)$ . By choosing  $\mathfrak{C}_1$  appropriately, we can ensure that the last inequality holds true uniformly for all minima  $\boldsymbol{m}, \boldsymbol{m}'$  of  $F^n$ . For this reason, for every n, the set  $\mathcal{M}^n := \{S^n[\boldsymbol{m}] \mid \boldsymbol{m} \in \Gamma^n \text{ local minima of } F^n\}$  is a suitable candidate for the set of metastable sets. In the following lemma we show that, if n is chosen large enough,  $\mathcal{M}^n$  satisfies (3.1) with  $\mathfrak{C} := \mathfrak{C}_1 - 2\beta\varepsilon(n)$ .

**Lemma 5.35.** For all  $n \in \mathbb{N}$  there exists  $N(n) \in \mathbb{N}$  such that for all  $X = S^n[x] \notin \mathcal{M}^n$  with  $x \in \Gamma^n$ 

$$\mathbb{P}_{\mu_X}\left[\tau_{\mathcal{M}^n} < \tau_X\right] \ge e^{-2\beta\varepsilon(n)N}, \quad \forall N \ge N(n).$$
(5.231)

*Proof.* For every  $x \in \Gamma^n$  there exists a local minimum m of  $F^n$  and a self-avoiding mesoscopic path  $\underline{x}$  from x to m such that the mesoscopic free energy is non-increasing along the path. By exploiting the fact that the Dirichlet form is monotone in the transition probabilities provides an immediate lower bound for  $\mathbb{P}_{\mu_X}[\tau_{\mathcal{M}^n} < \tau_X]$  in terms of the capacity  $\operatorname{cap}^{\underline{x}}(M, X)$  of a chain where all  $p_N(\sigma, \eta)$  are set to zero whenever  $\sigma, \eta \notin S^n[\underline{x}]$ , i.e.

$$\operatorname{cap}(X, \mathcal{M}^n) \geq \inf_{f \in \mathcal{H}_{X, M}} \frac{1}{2} \sum_{\sigma, \eta \in S^n[\boldsymbol{x}]} \mu_N(\sigma) p_N(\sigma, \eta) \left( f(\sigma) - f(\eta) \right)^2 = \operatorname{cap}^{\boldsymbol{x}}(X, A).$$

Here  $A = S^n[m]$  and  $S^n[x]$  denotes the set of configurations that correspond to the mesoscopic path. On the other hand, the Berman-Konsowa principle implies

$$\operatorname{cap}^{\underline{x}}(X,A) \geq \operatorname{CAP}^{n,\underline{x}}(\underline{x},\underline{m}) \left( \sum_{t=0}^{|\underline{x}|-1} \mathbb{E}^{f^{\underline{x}}} [\Phi_t^{\underline{x}}] \left( g_{\underline{x},\underline{m}}(\underline{x}(t)) - g_{\underline{x},\underline{m}}(\underline{x}(t+1)) \right) \right)^{-1}$$

where  $\Phi_t^x$  is defined in (5.139) and  $g_{x,m}$  is the mesoscopic harmonic function along the onedimensional path.

For the subordinate unit flow  $f^{\underline{x}}$  we choose the flow that is induced by a simple directed random walk on  $S^{n}[\underline{x}]$ , i.e. for k = 1, ..., n such that  $||x^{k}(t+1) - x^{k}(t)|| = 2/N$  and  $r = \Theta_{k,t}^{-}$ 

$$f^{\boldsymbol{x}}(\sigma,\eta) = \frac{\mathbb{1}_{\eta \in \Lambda_{k,r}(\sigma)}}{\left|\mathcal{S}^{n}[\boldsymbol{x}(t)]\right| \left|\Lambda_{k,r}(\sigma)\right|} \qquad \forall \sigma \in \mathcal{S}^{n}[\boldsymbol{x}(t)], \ \eta \in \mathcal{S}^{n}[\boldsymbol{x}(t+1)].$$

Since  $Q^n(\boldsymbol{x})/\mu_N(\sigma) \leq e^{\beta \varepsilon(n)N} |\mathcal{S}^n[\boldsymbol{x}]|$  for all  $\sigma \in \mathcal{S}^n[\boldsymbol{x}]$  and  $\boldsymbol{x} \in \Gamma^n$ , our choice of  $f^{\boldsymbol{x}}$  gives rise to  $\Phi_t^{\boldsymbol{x}}(\sigma) \leq e^{\beta \varepsilon(n)N}$  uniformly for all  $\sigma \in \mathcal{S}^n[\boldsymbol{x}(t)]$  and  $t = 0, \ldots, |\boldsymbol{x}| - 1$ . Using that the mesoscopic capacity along a one-dimensional path can be computed explicitly, see e.g. [11]

$$\operatorname{CAP}^{n,\underline{x}}(\boldsymbol{x},\boldsymbol{m}) = \left(\sum_{t=0}^{|\underline{x}|-1} \frac{1}{\mathcal{Q}^n(\boldsymbol{x}(t)) r^n(\boldsymbol{x}(t), \boldsymbol{x}(t+1))}\right)^{-1} \geq c \frac{\mathcal{Q}^n(\boldsymbol{x})}{N\sqrt{N}}$$

where we used that  $r^n(\boldsymbol{x}(t), \boldsymbol{x}(t+1)) \ge c'/N$  and  $F^n(\boldsymbol{x}(t)) - F^n(\boldsymbol{m}) \ge c''(|\boldsymbol{x}| - t)^2/N^2$  in a neighborhood of a local minimum  $\boldsymbol{m}$ . Hence, there exists  $N(n) \in \mathbb{N}$  such that for all  $N \ge N(n)$ 

$$\mathbb{P}_{\mu_X}\left[\tau_{\mathcal{M}^n} < \tau_X\right] \geq \frac{c}{N\sqrt{N}} e^{-\beta\varepsilon(n)N} \geq e^{-2\beta\varepsilon(n)N}.$$

This concludes the proof.

*Proof of Theorem 5.5.* (i) For  $\delta$  specified in Theorem 5.5, we define

$$\mathcal{U}_{\delta} := \left\{ \boldsymbol{x} \in \Gamma^{n} \mid F^{n}(\boldsymbol{x}) < F^{n}(\boldsymbol{m}) + \delta \right\} = \mathcal{U}_{\delta}(\boldsymbol{m}) \cup \bigcup_{\boldsymbol{m}' \in \boldsymbol{M}} \mathcal{U}_{\delta}(\boldsymbol{m}')$$
(5.232)

where  $\mathcal{U}_{\delta}(\boldsymbol{m})$  denotes the connected component of  $\mathcal{U}_{\delta}$  containing  $\boldsymbol{m}$ . Inspecting the proof of Theorem 3.10, we obtain that

$$\mathbb{E}_{\nu_{A,B}}[\tau_B] = \frac{1}{\operatorname{cap}(A,B)} \sum_{\boldsymbol{x} \in \mathcal{U}_{\delta}(\boldsymbol{m})} \mathcal{Q}^n(\boldsymbol{x}) (1 + \mathcal{O}_N(1)).$$
(5.233)

Note that, in view of (5.231), we only have to choose  $c_1 > 1 + \alpha^{-1}$  for  $\alpha^{-1} = \mathcal{O}(q)$  such that  $c_2 > 2\beta\varepsilon(n)$ . It remains to evaluate the right-hand side of (5.233). Analog to the proof of

Proposition 5.21, a standard approximation of the sum by an integral yields

$$\sum_{\boldsymbol{x}\in\mathcal{U}_{\delta}(\boldsymbol{m})} \mathcal{Q}^{n}(\boldsymbol{x}) = \mathcal{Q}^{n}(\boldsymbol{m}) \sum_{\boldsymbol{x}\in\mathcal{U}_{\delta}(\boldsymbol{m})} \exp\left(-\frac{\beta N}{2} \langle \boldsymbol{x}, A^{n}(\boldsymbol{m}) \, \boldsymbol{x} \rangle\right)$$
$$= \mathcal{Q}^{n}(\boldsymbol{m}) \left(\frac{N^{q-1}}{\sqrt{q}}\right)^{n} \int_{V_{1}} \exp\left(-\frac{\beta N}{2} \langle \boldsymbol{x}, A^{n}(\boldsymbol{m}) \, \boldsymbol{x} \rangle\right) \mathrm{d}\boldsymbol{x} \left(1 + \mathcal{O}_{N}(1)\right)$$
$$= \mathcal{Q}^{n}(\boldsymbol{m}) \left(\frac{(2\pi N)^{q-1}}{q \, \beta^{q-1}}\right)^{n/2} \left(\sqrt{\det\left(A^{n}(\boldsymbol{m})\right)}\right)^{-1/2} (1 + \mathcal{O}_{N}(1))$$
(5.234)

Thanks to (5.67) and (5.68), we get that

$$\sum_{\boldsymbol{x}\in\mathcal{U}_{\delta}(\boldsymbol{m})}\mathcal{Q}^{n}(\boldsymbol{x}) = \frac{\exp(-\beta NF_{N}(\boldsymbol{m}))}{Z_{N}\sqrt{\det\left(I_{q}-2\beta\left(D(z)-\mathbb{E}_{h}\left[u^{1}(2\beta z)\cdot u^{1}(2\beta z)^{\mathrm{T}}\right]\right)\right)}} (1+\mathcal{O}_{N}(1)).$$
(5.235)

Thus, by combining (5.233) with (5.234) and the expression for the capacity that is given in (5.20) concludes the proof of (i).

Moreover, (ii) and (iii) follows easily from Corollary 3.12 and Theorem 3.13.  $\hfill \Box$ 

# Bibliography

- 1. J.M.G. Amaro de Matos, A.E. Patrick, and V.A. Zagrebnov, Random infinite-volume Gibbs states for the Curie-Weiss random field Ising model, J. Stat. Phys. 66 (1992), no. 1-2, 139–164.
- S. Arrhenius, Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren, Z. Phys. Chem. 4 (1889), 226–248.
- 3. G. Ben Arous and R. Cerf, Metastability of the tree dimensional Ising model on a torus at very low temperatures, Electron. J. Probab. 1 (1996), no. 10, 1–55.
- 4. K.A. Berman and M.H. Konsowa, *Random-paths and cuts, electrical networks, and reversible Markov-chains*, SIAM J. Discrete Math. **3** (1990), no. 3, 311–319.
- A. Bianchi, A. Bovier, and D. Ioffe, Pointwise estimates and exponential laws in metastable systems via coupling methods, SFB611 Preprint 461 (2009), 1–28.
- \_\_\_\_\_, Sharp asymptotics for metastability in the random field Curie-Weiss model, Elect. J. Probab. 14 (2009), 1541–1603.
- 7. A. Bianchi and A. Gaudillière, Metastable states, quasi-stationary and soft measures, mixing time asymptotics via variational principles, arXiv:1103.1143 (2011), 1–32.
- R.M. Blumenthal and R.K. Getoor, *Markov Processes and Potential Theory*, 1 ed., Pure and Applied Mathematics, vol. 29, Academic Press, 1968.
- 9. A. Bovier, *Metastability and ageing in stochastic dynamics*, Dynamics and Randomness II, Nonlinear Phenomena and Complex Systems, vol. 10, Kluwer Academic Publisher, 2004, pp. 17–80.
- \_\_\_\_\_, Metastability: A potential theoretic approach, Proceedings of the International Congress of Mathematicians, vol. 3, European Mathematical Society Publishing House, 2006, pp. 499–518.
- \_\_\_\_\_, Metastability, Methods of Contemporary Mathematical Statistical Physics, Lecture Notes in Mathematics, vol. 1970, Springer-Verlag, 2009, pp. 177–221.
- Metastability: From mean field models to SPDEs, to appear in Probability in Complex Physical Systems, Springer Proceedings in Mathematics, vol. 11, Springer-Verlag, 2011, pp. 1–16.
- 13. A. Bovier, F. den Hollander, and F.R. Nardi, *Sharp asymptotics for Kawasaki dynamics on a finite box with open boundary*, Probab. Theory Rel. Fields **135** (2006), no. 2, 265–310.
- 14. A. Bovier, F. den Hollander, and C. Spitoni, *Homogenous nucleation for Glauber and Kawasaki dynamics in large volumens at low temperatures*, Ann. Probab. **38** (2010), no. 2, 661–713.
- 15. A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein, *Metastability in stochastic dynamics of disordered mean-field models*, Probab. Theory Rel. Fields **119** (2001), no. 1, 99–161.
- 16. \_\_\_\_\_, Metastability and low lying spectra in reversible Markov chains, Commun. Math. Phys. 228 (2002), no. 2, 219–255.
- 17. \_\_\_\_\_, Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times, J. Europ. Math. Soc. 6 (2004), no. 4, 399–424.
- A. Bovier, V. Gayrard, and M. Klein, Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues, J. Europ. Math. Soc. 7 (2005), no. 1, 69–99.
- A. Bovier and F. Manzo, Metastability in Glauber dynamics in the low-temperature limit: Beyond exponential asymptotics, J. Stat. Phys. 107 (2002), no. 3/4, 757–779.
- S. Brassesco, E. Olivieri, and Vares M.E., Couplings and asymptotic exponentiality of exit times, J. Stat. Phys 93 (1998), no. 1/2, 393–404.
- 21. C.J. Burke and M. Rosenblatt, A Markovian function of a Markov chain, Ann. Math. Statist. 29 (1958), no. 4, 1112–1122.
- D. Capocaccia, M. Cassandro, and E. Olivieri, A study of metastability in the Ising model, Commun. Math. Phys 39 (1974), 185–205.
- 23. M. Cassandro, A. Galves, E. Olivieri, and M.E. Vares, Metastable behaviour of stochastic dynamics A pathwise approach, J. Stat. Phys. 35 (1984), no. 5-6, 603–634.
- M. Cassandro, E. Olivieri, and P. Picco, Small random perturbations of infinite dimensional dynamical systems and nucleation theory, Ann. Inst. H, Poincaré 44 (1986), no. 4, 343–396.

### BIBLIOGRAPHY

- 25. R. Cerf and F. Manzo, Nucleation and growth for the Ising model in d dimensions at very low temperatures, arXiv:1102.1741 (2011), 1-84.
- N.R. Chaganty and J. Sethuraman, Multidimensional large deviation local limit theorems, J. Multivariate Anal. 20 (1986), no. 2, 190–204.
- 27. K.L. Chung, Green, Brown, and Probability & Brownian Motion on the Line, 1 ed., World Scientific Publishing, 2002.
- E.N.M. Cirillo, F.R. Nardi, and C. Spitoni, Metastability for reversible probabilistic cellular automata with selfinteraction, J. Stat. Phys. 132 (2008), no. 3, 431–471.
- 29. E.B. Davies, Metastability and the Ising model, J. Stat. Phys. 27 (1982), no. 4, 657-675.
- 30. \_\_\_\_\_, Metastable states of symmetric Markov semigroups I, Proc. London Math. Soc. 45 (1982), no. 3, 133–150.
- 31. \_\_\_\_\_, Metastable states of symmetric Markov semigroups II, J. London Math. Soc. 26 (1982), no. 2, 541–556.
- P. Dehghanpour and R.H. Schonmann, Metropolis dynamics relaxation via nucleation and growth, Commun. Math. Phys. 188 (1997), 89–119.
- 33. \_\_\_\_\_, A nucleation-and-growth model, Probab. Theory Relat. Field 107 (1997), 123–135.
- F. den Hollander, F.R. Nardi, E. Olivieri, and E. Scoppola, Droplet growth for three-dimensional Kawasaki dynamics, Probab. Theory Relat. Fields 125 (2003), 153–194.
- 35. F. den Hollander, E. Olivieri, and E. Scoppola, Metastability and nucleation for conservative dynamics, J. Math. Phys. 41 (2000), no. 3, 1424–1498.
- 36. P. Deuflhard, W. Huisinga, A. Fischer, and Ch. Schütte, *Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains*, Linear Algebra Appl. **315** (2000), 39–59.
- P. Deuflhard and M. Weber, Robust perron cluster analysis in conformation dynamics, Linear Algebra Appl. 398 (2005), 161–184.
- P.G.L. Dirichlet, Über einen neuen Ausdruck zur Bestimmung der Dichtigkeit einer unendlich d
  ünnen Kugelschale, wenn der Werth des Potentials derselben in jedem Punkt ihrer Oberfl
  äche gegeben ist, K
  önigl. Akad. d. Wiss. 1850 (1852), 99–116.
- R.L. Dobrushin and S.B. Shlosman, Large and moderate deviations in the Ising model, Adv. Soviet Math. 20 (1994), 91–219.
- D. Donsker and S.R.S Varadhan, On the principal eigenvalue of second-order elliptic differential operators, Comm. Pure Appl. Math. 29 (1976), 143–157.
- 41. J.L. Doob, Semimartingales and subharmonic functions, Trans. Amer. Math. Soc. 77 (1954), no. 1, 86–121.
- 42. \_\_\_\_\_, Discrete potential theory and boundaries, J. Math. Mech. 8 (1959), no. 3, 433–458; erratum 993.
- 43. \_\_\_\_\_, Classical Potential Theory and Its Probabilistic Counterpart, 1 ed., Grundlehren der mathematischen Wissenschaften, vol. 262, Springer-Verlag, 1984.
- P.G. Doyle and J.L. Snell, Random Walks and Electric Networks, 1 ed., Carus Mathematical Monographs, vol. 22, Mathematical Association of America, 1984.
- M. Eckhoff, Precise asymptotics of small eigenvalues of reversible diffusions in the metastable regime, Ann. Probab. 33 (2005), no. 1, 244–299.
- R. S. Ellis, Entropy, Large Deviations, and Statistical Mechanics, Grundlehren der mathematischen Wissenschaften, vol. 271, Springer, Berlin, 1985.
- 47. H. Eyring, The activated complex in chemical reations, J. Chem. Phys. 3 (1935), 107–115.
- Martinelli F., Olivieri E., and Scoppola E., Small random perturbations of finite- and infinite-dimensional dynamical systems: Unpredictability of exit times, J. Stat. Phys. 55 (1989), no. 4/5, 477–504.
- 49. W. Feller, An Introduction to Probability Theory and Its Application, 2 ed., vol. 2, John Willey & Sons, 1970.
- L.R. Fontes, P. Mathieu, and P. Picco, On the averaged dynamics of the random field Curie-Weiss model, Ann. Appl. Probab. 10 (2000), no. 4, 1212–1245.
- 51. M.I. Freidlin and A.D. Wentzell, *Random Perturbation of Dynamical Systems*, 2 ed., Grundlehren der mathematischen Wissenschaften, vol. 260, Springer, 1998.
- 52. A. Galves, E. Olivieri, and M.E. Vares, Metastability for a class of dynamical systems subject to small random perturbations, Ann. Probab. 15 (1987), no. 4, 1288–1305.
- 53. A. Gaudillière, Conderse physics applied to Markov chains A brief introduction to potential theory, arXiv:0901.3053 (2008), 1–49.
- A. Gaudillière, F. den Hollander, F.R. Nardi, E. Olivieri, and E. Scoppola, Ideal gas approximation for a twodimensional rarefied gas under Kawasaki dynamics, Stoch. Process. Appl. 119 (2009), 737–774.
- B. Gaveau and L.S. Schulman, Theory of nonequilibrium first-order phase transitions for stochastic dynamics, J. Math. Phys. 39 (1998), no. 3, 1517–1533.
- B. Gentz and N. Berglund, Anomalous behavior of the Kramers rate at bifurcations in classical fild theory, J. Phys. A 42 (2009), no. 5, 052001:1–9.
- H. Grabert, Projection Operator Techniques in Nonequilibrium Systems, Springer Tracts in Modern Physics, vol. 95, Springer-Verlag, 1982.

#### BIBLIOGRAPHY

- H. Grabert, P. Hänggi, and P. Talkner, Microdynamics and nonlinear stochastic processes of gross variables, J. Stat. Phys. 22 (1980), no. 5, 537–552.
- 59. G. Green, An Essay on the Application of Mathematical Analysis to the Theory of Electricity and Magnetism, 1 ed., Nottingham, 1828.
- R. Grone, K.H. Hoffmann, and P. Salamon, An interlacing theorem for reversible Markov chains, J. Phys. A: Math. Theor 41 (2008), no. 212002, 1–7.
- 61. P. Hänggi, P. Talkner, and Borkovec M., *Reaction-rate theory: fifty years after Kramers*, Re. Mod. Phys. **62** (1990), no. 2, 251–341.
- B. Helffer, M. Klein, and F. Nier, Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach, Mat. Contemp. 26 (2004), 1–45.
- 63. B. Helffer and F. Nier, *Hypoelliptic Estimates and Spectral Theory for Fokker-Planck Operators and Witten Laplacians*, 1 ed., Lecture Notes in Mathematics, vol. 1862, Springer-Verlag, 2005.
- 64. Lieberstein H.M., *Theory of Partial Differential Equations*, 1 ed., Mathematics in Science and Engineering, vol. 93, Academic Press, 1972.
- R.A. Holley, S. Kusuoka, and S.W. Stroock, Asymptotics of the spectral gap with applications to the theory of simulated annealing, J. Funct. Anal. 83 (1989), 333–347.
- 66. W. Huisinga, S. Meyn, and Ch. Schütte, *Phase transitions and metastability in Markovian and molecular systems*, Ann. Appl. Probab. **14** (2004), no. 1, 419–458.
- 67. G.A. Hunt, Markoff processes and potentials. I, Illinois J. Math. 1 (1957), no. 1, 44-93.
- 68. \_\_\_\_\_, Markoff processes and potentials. II, Illinois J. Math. 1 (1957), no. 3, 316–369.
- 69. \_\_\_\_\_, Markoff processes and potentials. III, Illinois J. Math. 2 (1958), no. 2, 151–213.
- G. Iacobelli and C. Kuelske, Metastates in finite-type mean-field models: Visibility, invisibility, and random restoration of symmetry, J. Stat. Phys. 140 (2010), no. 1, 27–55.
- 71. J.D. Jackson, Classical Electrodynamics, 3 ed., Wiley, 1999.
- S. Kakutani, Two-dimensional Brownian motion and harmonic functions, Proc. Imp. Acad. Tokyo 20 (1944), no. 10, 706–714.
- 73. I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, 2 ed., Graduate texts in mathematics, vol. 113, Springer-Verlag, 1991.
- 74. T. Kato, On the upper and lower bounds of eigenvalues, J. Phys. Soc. Japan 4 (1949), 334–339.
- O.D. Kellogg, Foundations of Potential Theory, 1 ed., Grundlehren der mathematische Wissenschaften, vol. 31, Springer-Verlag, 1929.
- J.G. Kemeny, J.L. Snell, and A.W. Knapp, *Denumerable Markov Chains*, 2 ed., Graduate texts in Mathematics, vol. 40, Springer-Verlag, 1976.
- 77. Y. Kifer, A discrete-time version of the Wentzell-Friedlin theory, Ann. Probab. 18 (1990), no. 4, 1676–1692.
- H.A. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions, Physica 7 (1940), no. 4, 284–304.
- C. Kulske, Metastates in disordered mean-field models: Random field and Hopfield models, J. Stat. Phys. 88 (1997), no. 5-6, 1257–1293.
- H. Lebesgue, Conditions de régularité, conditions d'irrégularité, conditions d'impossibilité dans le problème de Dirichlet, Comp. Rendue Acad. Sci. 178 (1924), no. 1, 349–254.
- D.A. Levin, M.J. Luczak, and Y. Peres, Glauber dynamics for the mean-field Ising model: Cut-off, critical power law, and metastability, Probab. Theory Rel. Fields 146 (2010), no. 1-2, 223–265.
- D.A. Levin, Y. Peres, and E.L. Wilmer, Markov Chain and Mixing Times, 1 ed., American Mathematical Society, 2008.
- R. Lyons and Y. Peres, Probability on Trees and Networks, Cambridge University Press, 2011, In preparation. Current version available at http://mypage.iu.edu/~rdlyons/.
- P. Mathieu, Spectra, exit times and long time asymptotics in the zero white noise limit, Stoch. and Stoch. Rep. 55 (1995), 1–20.
- P. Mathieu and P. Picco, Metastability and convergence to equilibrium for the random field Curie-Weiss model, J. Stat. Phys. 91 (1998), no. 3/4, 679–732.
- L. Miclo, Comportement de spectres d'opérateurs de schrödinger à basse température, Bull. Sci. Math. 119 (1995), 529–553.
- 87. L. Mittag and M.J. Stephen, *Mean-field theory of the many component Potts model*, J. Phys. A: Math. Nucl. Gen. 7 (1974), no. 9, L109–L112.
- C.St.J.A. Nash-Williams, Random walk and electric currents in networks, Proc. Cambridge Philos. Soc. 55 (1959), no. 2, 181–194.
- E.J. Neves and R.H. Schonmann, Critical droplets and metastability for a Glauber dynamics at very low temperature, Commun. Math. Phys. 137 (1991), 209–230.
- \_\_\_\_\_, Behavior of droplets for a class of Glauber dynamics at very low temperatur, Probab. Theory Relat. Fields 91 (1992), 331–354.

### BIBLIOGRAPHY

- 91. J.R. Noris, Markov Chains, 1 ed., Cambridge University Press, 1997.
- E. Olivieri and E. Scoppola, Markov chains with exponentially small transition probabilities: First exit problem from a general domain. I. The reversible case, J. Stat. Phys. 79 (1995), no. 3/4, 613–647.
- Markov chains with exponentially small transition probabilities: First exit problem from a general domain. II. The general case, J. Stat. Phys 84 (1996), no. 5/6, 987–1041.
- 94. E. Olivieri and M.E. Vares, *Large Deviations and Metastability*, 1 ed., Encyclopedia of Mathematics and its Applications, vol. 100, Cambridge University Press, 2005.
- A.M. Ostrowski, A quantitative formulation of Sylvesters's law of inertia, Proc. Natl. Acad. Sci. U.S.A. 45 (1959), no. 5, 740–744.
- 96. O. Penrose and J.L. Lebowitz, *Rigorous treatment of metastable states in the van der Waals-Maxwell theory*, J. Stat. Phys. **3** (1971), no. 2, 211–236.
- 97. \_\_\_\_\_, Towards a rigorous molecular theory of metastability, Fluctuation Phenomenon, Studies in Statistical Mechanics, vol. 7, North-Holland Publishing Company, 1979, pp. 293–340.
- 98. H. Poincaré, Théorie du Potentiel Newtonien, 1 ed., Carré & Naud, 1899.
- 99. E. Pollak and P. Talkner, Reaction rate theory: What it was, where is it today, and where is it going?, Chaos 15 (2005), no. 026116, 1–11.
- G. Pólya, Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz, Math. Ann. 84 (1921), no. 1-2, 149–160.
- 101. S.C. Port and C.J. Stone, Brownian Motion and Classical Potential Theory, 1 ed., vol. 29, Academic Press, 1978.
- N. Privault, Potential Theory in Classical Probability, Quantum Potential Theory, Lecture Notes in Mathematics, vol. 1954, Springer-Verlag, 2008, pp. 3–59.
- R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series, vol. 28, Princeton University Press, Princeton, NJ, 1972.
- 104. Y. Saad, Numerical Methods for Large Eigenvalue Problems, 1 ed., Classics in Applied Mathematics, vol. 66, SIAM, 2011.
- 105. R.H. Schonmann and S.B. Shlosmann, Wulff droplets and the metastable relaxation of kinetic Ising models, Commun. Math. Phys. **194** (1998), 389–462.
- 106. J. Schwinger, DeRaad L.L.Jr., K.A. Milton, and W. Tsai, *Classical Electrodynamics*, 1 ed., Perseus Books Group, 1998.
- 107. M. Sugiura, Metastable behaviors of diffusion processes with small parameter, J. Math. Soc. Japan 47 (1995), no. 4, 755–788.
- 108. A. Telcs, The Art of Random Walks, 1 ed., Lecture Notes in Mathematics, vol. 1885, Springer-Verlag, 2006.
- G. Temple, The accuracy of Rayleigh's method of calculating the natural frequencies of vibrating systems, Proc. Roy. Soc. London Ser. A 211 (1958), 204–224.
- 110. H. Thorisson, *Coupling, stationarity, and regeneration*, 1 ed., Probability and its Applications, Springer-Verlag, 2000.
- 111. A.D. Ventcel and M.I. Freidlin, On small random perturbations of dynamical systems, Russ. Math. Surveys 25 (1970), 1–55.
- 112. A.D. Venttsel', On the asymptotic of the greatest eigenvalue of a second-order elliptic differential operator with a small parameter in the higher derivatives, Soviet Math. Dokl. **13** (1972), 13–18.
- Formulae for eigenfunctions and eigenmeasures associated with a Markov process, Theory Probab. Appl. 18 (1973), 1–26.
- 114. W. Woess, *Random Walks on Infinite Graphs and Groups*, 1 ed., Cambridge Tracts in Mathematics, vol. 138, Cambridge University Press, 2000.
- 115. F.Y. Wu, The Potts model, Rev. Mod. Phys. 54 (1982), no. 1, 235-268.
- 116. R. Zwanzig, Nonlinear generalized Langevin equations, J. Stat. Phys. 9 (1973), no. 3, 215-220.